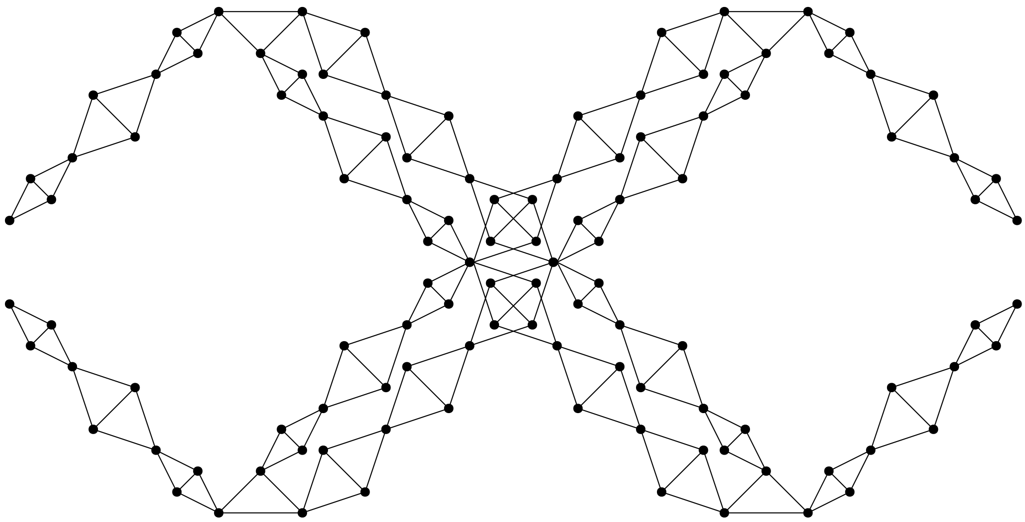
Clique Factors: Extremal and Probabilistic Perspectives



Dissertation zur Erlangung des Grades
eines Doktors der Naturwissenschaften (Dr. rer. nat.)
am Fachbereich Mathematik und Informatik
der Freien Universität Berlin

vorgelegt von

Patrick Morris

Berlin
August 2021

Betreuer und Erstgutachter:
Prof. Tibor Szabó

Zweitgutachter:
Prof. Robert Morris

Tag der Disputation:
11. Oktober 2021

Institut für Mathematik
Freie Universität Berlin

Bend, we don't break, we not the bank
We all we got
Switch whips, relocate
Way out of state
Bada bing, bada bam
It's the puppet master, all of the strings in my hand

Thebe Neruda Kgositsile a.k.a. Earl Sweatshirt

Summary

A K_r*-factor* in a graph G is a collection of vertex-disjoint copies of K_r covering the vertex set of G. In this thesis, we investigate these fundamental objects in three settings that lie at the intersection of extremal and probabilistic combinatorics.

Firstly, we explore *pseudorandom graphs*. An n-vertex graph is said to be (p, β)*-bijumbled* if for any vertex sets $A, B \subseteq V(G)$, we have $e(A, B) = p|A||B| \pm \beta\sqrt{|A||B|}$. We prove that for any $3 \leq r \in \mathbb{N}$ and $c > 0$ there exists an $\varepsilon > 0$ such that any n-vertex (p, β)-bijumbled graph with $n \in r\mathbb{N}$, $\delta(G) \geq cpn$ and $\beta \leq \varepsilon p^{r-1}n$, contains a K_r-factor. This implies a corresponding result for the stronger pseudorandom notion of (n, d, λ)-graphs. For the case of K_3-factors, this result resolves a conjecture of Krivelevich, Sudakov and Szabó from 2004 and it is tight due to a pseudorandom triangle-free construction of Alon. In fact, in this case even more is true: as a corollary to this result, we can conclude that the same condition of $\beta = o(p^2 n)$ actually guarantees that a (p, β)-bijumbled graph G contains every graph on n vertices with maximum degree at most 2.

Secondly, we explore the notion of *robustness* for K_3-factors. For a graph G and $p \in [0, 1]$, we denote by G_p the random sparsification of G obtained by keeping each edge of G independently, with probability p. We show that there exists a $C > 0$ such that if $p \geq C(\log n)^{1/3}n^{-2/3}$ and G is an n-vertex graph with $n \in 3\mathbb{N}$ and $\delta(G) \geq \frac{2n}{3}$, then with high probability G_p contains a K_3-factor. Both the minimum degree condition and the probability condition, up to the choice of C, are tight. Our result can be viewed as a common strengthening of the classical extremal theorem of Corrádi and Hajnal, corresponding to $p = 1$ in our result, and the famous probabilistic theorem of Johansson, Kahn and Vu establishing the threshold for the appearance of K_3-factors (and indeed all K_r-factors) in $G(n, p)$, corresponding to $G = K_n$ in our result. It also implies a first lower bound on the number of K_3-factors in graphs with minimum degree at least $\frac{2n}{3}$, which gets close to the truth.

Lastly, we consider the setting of *randomly perturbed graphs*; a model introduced by Bohman, Frieze and Martin, where one starts with a dense graph and then adds random edges to it. Specifically, given any fixed $0 < \alpha < 1 - \frac{1}{r}$ we determine how many random edges one must add to an n-vertex graph G with $\delta(G) \geq \alpha n$ to ensure that, with high probability, the resulting graph contains a K_r-factor. As one increases α we demonstrate that the number of random edges required 'jumps' at regular intervals, and within these intervals our result is best-possible. This work therefore bridges the gap between the seminal work of Johansson, Kahn and Vu mentioned above, which resolves the purely random case, i.e., $\alpha = 0$, and that of Hajnal and Szemerédi (and Corrádi and Hajnal for $r = 3$) showing that when $\alpha \geq 1 - \frac{1}{r}$ the initial graph already hosts the desired K_r-factor.

Zusammenfassung

Ein K_r*-Faktor* in einem Graphen G ist eine Sammlung von Knoten-disjunkten Kopien von K_r, die die Knotenmenge von G überdecken. Wir untersuchen diese Objekte in drei Kontexten, die an der Schnittstelle zwischen extremaler und probabilistischer Kombinatorik liegen.

Zuerst untersuchen wir *Pseudozufallsgraphen*. Ein Graph heißt (p, β)*-bijumbled*, wenn für beliebige Knotenmengen $A, B \subseteq V(G)$ gilt $e(A, B) = p|A||B| \pm \beta\sqrt{|A||B|}$. Wir beweisen, dass es für jedes $3 \le r \in \mathbb{N}$ und $c > 0$ ein $\varepsilon > 0$ gibt, so dass jeder n-Knoten (p, β)-bijumbled Graph mit $n \in r\mathbb{N}$, $\delta(G) \ge cpn$ und $\beta \le \varepsilon p^{r-1}n$, einen K_r-Faktor enthält. Dies impliziert ein entsprechendes Ergebnis für den stärkeren Pseudozufallsbegriff von (n, d, λ)-Graphen. Im Fall von K_3-Faktoren, löst dieses Ergebnis eine Vermutung von Krivelevich, Sudakov und Szabó aus dem Jahr 2004 und ist durch eine pseudozufällige K_3-freie Konstruktion von Alon bestmöglich. Tatsächlich ist in diesem Fall noch mehr wahr: als Korollar dieses Ergebnisses können wir schließen, dass die gleiche Bedingung von $\beta = o(p^2n)$ garantiert, dass ein (p, β)-bijumbled Graph G jeden Graphen mit maximalem Grad 2 enthält.

Zweitens untersuchen wir den Begriff der *Robustheit* für K_3-Faktoren. Für einen Graphen G und $p \in [0, 1]$ bezeichnen wir mit G_p die zufällige Sparsifizierung von G, die man erhält, indem man jede Kante von G unabhängig von den anderen Kanten mit einer Wahrscheinlichkeit p behält. Wir zeigen, dass, wenn $p = \omega((\log n)^{1/3}n^{-2/3})$ und G ein n-Knoten-Graph mit $n \in 3\mathbb{N}$ und $\delta(G) \ge \frac{2n}{3}$ ist, G_p mit hoher Wahrscheinlichkeit (mhW) einen K_3-Faktor enthält. Sowohl die Bedingung des minimalen Grades als auch die Wahrscheinlichkeitsbedingung sind bestmöglich. Unser Ergebnis ist eine Verstärkung des klassischen extremalen Satzes von Corrádi und Hajnal, entsprechend $p = 1$ in unserem Ergebnis, und des berühmten probabilistischen Satzes von Johansson, Kahn und Vu, der den Schwellenwert für das Auftreten eines K_3-Faktors (und aller K_r-Faktoren) in $G(n, p)$ festlegt, entsprechend $G = K_n$ in unserem Ergebnis. Es impliziert auch eine erste untere Schranke für die Anzahl der K_3-Faktoren in Graphen mit einem minimalen Grad von mindestens $\frac{2n}{3}$, die der Wahrheit nahe kommt.

Schließlich betrachten wir die Situation von *zufällig gestörten Graphen*; ein Modell, bei dem man mit einem dichten Graphen beginnt und dann zufällige Kanten hinzufügt. Wir bestimmen, bei gegebenem $0 < \alpha < 1 - \frac{1}{r}$, wie viele zufällige Kanten man zu einem n-Knoten-Graphen G mit $\delta(G) \ge \alpha n$ hinzufügen muss, um sicherzustellen, dass der resultierende Graph mhW einen K_r-Faktor enthält. Wir zeigen, dass, wenn man α erhöht, die Anzahl der benötigten Zufallskanten in regelmäßigen Abständen “springt", und innerhalb dieser Abstände unser Ergebnis bestmöglich ist. Diese Arbeit schließt somit die Lücke zwischen der oben erwähnten bahnbrechenden Arbeit von Johansson, Kahn und Vu, die den rein zufälligen Fall, d.h. $\alpha = 0$, löst, und der Arbeit von Hajnal und Szemerédi (und Corrádi und Hajnal für $r = 3$), die zeigt, dass der ursprüngliche Graph bereits den gewünschten K_r-Faktor enthält, wenn $\alpha \ge 1 - \frac{1}{r}$ ist.

Acknowledgements

Firstly, I would like to thank Tibor Szabó for his guidance and mentorship, for creating many opportunities for me, and for introducing me to many of the delights of extremal and probabilistic combinatorics, both during the doctorate and before. In this regard, I am also very grateful to Shagnik Das, who taught the course that first sparked my love for combinatorics, and who has been a fantastic source of knowledge and advice ever since. Another extremely influential figure on my mathematical development has been Jie Han, who taught me a lot about clique factors, the topic of this thesis. Indeed, through a series of collaborations with Jie, I have absorbed a tremendous number of techniques and ideas.

I would also like to thank all the further people I have collaborated with on projects during my doctorate: Peter Allen, Yahav Alon, Anurag Bishnoi, Julia Böttcher, Yulin Chang, Jan Corsten, Ewan Davies, David Fabian, Hiệp Hàn, Matthew Jenssen, Charlotte Knierim, Yoshi Kohayakawa, Guilherme Mota, Yury Person, Barnaby Roberts, Wojtek Samotij, Jozef Skokan, Andrew Treglown, Guanghui Wang and Donglei Yang. It has been a pleasure to work with so many people and I have learnt a lot through the process.

Likewise, I thank my mathematical family in Berlin for many interesting conversations and a brilliant working environment. Those who have not featured above are Michaelis Anastos, Simona Boyadzhiyska, Ander Lamaison, Tamás Mészáros, Christoph Spiegel and Raphael Steiner.

I have also had the pleasure of spending extended periods visiting other institutions: IMPA in Rio de Janeiro and USP in São Paulo, Brazil, USACH in Santiago, Chile, MIT in Boston and URI in Rhode Island, USA and LSE in London, UK. I am very grateful to all my hosts for their hospitality, support and enthusiasm during my visits. In particular, the results of Chapter 5 were initiated during my first visit to Brazil and I would like to thank the organisers of the Graphs@IMPA thematic program in the summer of 2018. My second visit to Brazil was also influential and the first breakthrough on the problem covered in Chapter 3 was obtained whilst I was at IMPA. I thank Pedro Araújo and Robert Morris with whom I discussed many aspects of that project. Finally, the work presented in Chapter 4 was initiated and developed during my two trips to LSE and I thank everyone there.

Thanks also go to the Leverhulme Trust, Berlin Mathematical School and the Dahlem Research School, who provided support for this doctorate and without whom it would not have been possible.

Last but certainly not least, I thank my friends and family. I am blessed to have an amazing network of people around me who support me when things get hard and who help me unwind and enjoy other (non-mathematical) aspects of life. In particular, I thank my mum and dad, my sister and Jiselle. Thank you!

Selbstständigkeitserklärung

Gemäß § 7(4) der Promotionsordnung des Fachbereichs Mathematik und Informatik der Freien Universität Berlin versichere ich hiermit, dass ich alle Hilfsmittel und Literatur angegeben und auf dieser Grundlage die Arbeit selbständig verfasst habe. Die vorliegende Arbeit ist frei von Plagiaten. Außerdem versichere ich, dass diese Arbeit zu keinem früheren Promotionsverfahren eingereicht wurde.

Patrick Morris

Unterschrift: ____________________________________

Datum: ____________________________________

Contents

Notation

Here we discuss our notational conventions and general definitions. Many of these are standard but some notation and terminology are more specialised and catered to our needs. A glossary for quick reference is provided at the end of the thesis and lists the notation introduced here as well as elsewhere in the thesis.

Basics: We use $[n]_0$ to denote $[n] \cup \{0\}$. For $0 \leq t \leq n \in \mathbb{N}$, we define $n!_t$ to be the number of ways to select a list of t distinct numbers from $[n]$. That is, $n!_0 := 1$ and for $1 \leq t \leq n$, we have

$$n!_t := \frac{n!}{(n-t)!} = n(n-1)\cdots(n-t+1).$$

We use the notation $x = y \pm z$ to denote that $x \leq y + z$ and $x \geq y - z$ and we say a property holds *with high probability* (whp, for short), if the probability that it holds tends to 1 as some parameter n (usually the number of vertices of a graph) tends to infinity. Given a set A and $k \in \mathbb{N}$ we denote by A^k the set of all ordered k-tuples of elements from A, while $\binom{A}{k}$ denotes the set of all (unordered) k-element subsets of A. For sets X, Y with Y not necessarily contained in X, we use $X \setminus Y$ to denote $X \setminus (X \cap Y)$. We use v_G and e_G to denote the number of vertices and edges in G respectively. We use $\Delta(G)$ (resp. $\delta(G)$) to denote the maximum (resp. minimum) degree of a graph G and $\alpha(G)$ denotes the independence number of G, that is, the size of the largest independent set in G. Throughout we use log to denote the natural (base e) logarithm function. Finally, we drop ceilings and floors unless necessary, so as not to clutter the arguments.

Constants: At times we will define constant hierarchies within proofs, writing statements such as the following: Choose constants

$$0 < c_1 \ll c_2 \ll \ldots \ll c_\ell \ll d.$$

This should be taken to mean that given some constant d (given by the statement we aim to prove), one can choose all the remaining constants (the c_i) from right to left so that all the subsequent constraints are satisfied. That is, there exist increasing functions f_i for $i \in [\ell + 1]$ such that whenever $c_i \leq f_{i+1}(c_{i+1})$ for all $i \in [\ell - 1]$ and $c_\ell \leq f_{\ell+1}(d)$, all constraints on these constants that are in the proof, are satisfied.

Neighbourhoods and degrees: Given a graph G, a vertex $v \in V(G)$ and a set $U \subseteq V(G)$, we define the *neighbourhood* of v in U as $N_G(v; U) := \{u \in U : uv \in E(G)\}$. If $U = V(G)$, we simply write $N_G(v)$ and if G is clear from context we drop the subscript. If two vertices $u_1, u_2 \in$

$V(G)$ are given, then $N_G(u_1, u_2) := N_G(u_1) \cap N_G(u_2)$ denotes the *common neighbourhood* of u_1 and u_2. We will also use this notation for an edge $e = u_1u_2$, taking that $N_G(e) = N_G(u_1, u_2)$. Similarly, if $S \subset V(G)$ is some subset of vertices, $N_G(S) := \cap_{u \in S} N_G(u)$ denotes the common neighbourhood of the vertices in S and if $\underline{u} = (u_1, \ldots, u_\ell)$ is a tuple of vertices (an ordered set), $N_G(\underline{u}) := \cap_{j \in [\ell]} N_G(u_j)$ denotes the common neighbourhood of the set of vertices in $\underline{u}$. The parameters $N_G(u_1, u_2; U)$, $N_G(S; U)$ and $N_G(\underline{u}; U)$ are all defined analogously as the sets of common neighbours that lie in U. We follow the convention that $N_G(\emptyset) = V(G)$. We also define *degrees* $\deg_G(u) = |N_G(u)|$ with $\deg_G(u; U), \deg_G(S), \deg_G(S; U), \deg_G(\underline{u})$ and $\deg_G(\underline{u}; U)$ defined analogously. Again, if the graph G is clear from the context then we drop the subscripts.

Edge subsets as subgraphs: Sometimes, given a graph G and a subset of edges $E' \subseteq E(G)$, we will think of E' as the subgraph $H_{E'} := (V(E'), E')$ of G, where $V(E')$ is the set of vertices that lie in edges in E'. We then use notation like $\delta(E') := \delta(H_{E'})$ and $\deg_{E'}(v) := \deg_{H_{E'}}(v)$. Furthermore, for a vertex set $A \subset V(G)$, $E'[A]$ denotes the edges induced by $H_{E'}$ on X. That is, $E'[A] := \{e \in E' : e \subset A\}$.

Triangles and cliques: For a graph G and $r \in \mathbb{N}, r \geq 2$, we define $K_r(G)$ to be the set of copies of K_r in G. For example, $K_2(G) = E(G)$. When referring to a (copy of a) clique $S \in K_r(G)$, we will sometimes identify the copy with the set of vertices that hosts it. That is, we may think of $S \in K_r(G)$ as a set of r vertices that host a clique in G as well as the copy of the clique itself. Given a set of r-cliques $\Sigma \subseteq K_r(G)$, we use the notation $V(\Sigma)$ to denote all vertices that feature in cliques in Σ, i.e. $V(\Sigma) := \cup_{S \in \Sigma} S$. For $u \in V(G)$ we let $K_r(G, u) \subseteq K_r(G)$ denote the subset of cliques containing u.

Now for a vertex $v \in V(G)$, we let $\mathrm{Tr}_v(G)$ denote the *triangle neighbourhood of* v: the set of edges in $E(G)$ that form a triangle with v in G. That is, $\mathrm{Tr}_v(G) = \{e \in E(G) : v \in N_G(e)\}$. Note that $K_3(G, u) = \{f \cup \{u\} : f \in \mathrm{Tr}_u(G)\}$.

We say that a clique $S \in K_r(G)$ *traverses* vertex subsets $U_1, \ldots, U_r \subseteq V(G)$ if there exists some ordering of S as $S = \{u_1, \ldots, u_r\}$ such that $u_i \in U_i$ for all $i \in [r]$. Note that when the U_i are pairwise disjoint this simplifies to requiring that S contains one vertex from each U_i. However, at times we will deal with not necessarily disjoint sets U_i and so this more delicate definition is needed.

Matchings and factors: For an r-vertex graph H (usually K_r) and an n-vertex graph G, an H-*matching* in G is a collection of vertex-disjoint copies of H in G. The *size* of an H-matching is the number of vertex-disjoint copies of H in the collection. Note that when $H = K_2$ is a single edge, an H-matching is simply a matching and when $H = K_3$, we will also refer to a K_3-matching as a *triangle matching*. If an H-matching covers the vertex set of G (implying that $n \in r\mathbb{N}$), then we refer to the H-matching as an *H-factor* in G. Thus, when $H = K_2$, an H-factor is a perfect matching and when $H = K_3$, we also refer to a K_3-factor as a *triangle factor*. At times, we will refer to an H-matching as a *partial H-factor*. Although these two terms refer to the same objects, we reserve the use of partial factors for when there is an aim for the partial H-factor/H-matching to contribute to a full H-factor. Finally, if a partial H-factor covers the majority of the vertices of G, we will sometimes (informally) refer to it as an *almost H-factor*.

Graph embeddings: Throughout, we will deal exclusively with *ordered* embeddings of graphs, which we also refer to as *labelled* embeddings. Thus when we refer to an embedding of H in G, we implicitly fix an ordering on $V(H)$, say $V(H) := \{h_1, \dots, h_{v_H}\}$ and say that there is an embedding of H onto an (ordered) vertex set $\{v_1, \dots, v_{v_H}\} \subseteq V(G)$ if $v_i v_j \in E(G)$ for all i and j such that $h_i h_j \in E(H)$.

Vertex sets and tuples in tripartite graphs: For a portion of this thesis, we will be concerned with the host graph being a balanced tripartite graph. In such a setting, we will take as convention that the disjoint vertex sets that form the tripartition are labelled V^1, V^2 and V^3 and are each of size n. It will be useful for us to considered ordered tuples of vertices from these vertex sets. We therefore fix $\mathcal{V} := \{\emptyset\} \cup V^1 \cup (V^1 \times V^2) \cup (V^1 \times V^2 \times V^3)$. That is, an element $\underline{u} \in \mathcal{V}$ is a vector of some *length* $0 \leq \ell(\underline{u}) \leq 3$ such that for each $i \leq \ell(\underline{u})$, we have that $\underline{u}$ contains exactly one vertex from V^i.

Vertex sets with elements removed: Given a graph G, a collection of vertices $u_1, \dots, u_\ell \in V(G)$ and a subset of vertices $W \subseteq V(G)$, we use the notation $W_{\hat{u}_1,\dots,\hat{u}_\ell}$ to denote the subset W with the u_i removed. That is,

$$W_{\hat{u}_1,\dots,\hat{u}_\ell} := W \setminus (W \cap \{u_1, \dots, u_\ell\}).$$

Note that we do not impose that the u_i need lie in W. We remark that we add a hat on the removed vertices u_i in this notation to distinguish it from similar notation, which we introduce later, where vertices appear in subscripts without hats, signalling that these vertices are used for certain purposes.

To ease notation, we will sometimes group together some of the collection of vertices we wish to omit, as an ordered tuple. For example, if $\underline{u} = (u_1, \dots, u_\ell) \in \mathcal{V}$ for some $\ell \in [3]_0$ as above, we define $W_{\hat{\underline{u}}} := W_{\hat{u}_1,\dots,\hat{u}_\ell}$.

Partial triangle factors in tripartite graphs: We will be concerned with embedding partial triangle factors in a given host tripartite graph. For $t \in [n]_0$, we therefore define D_t to be the graph on vertex set $[t]\times[3]$, whose edge set consists of the edges $\{\{(s,i),(s,j)\} : s \in [t], i \neq j \in [3]\}$. Thus D_t simply consists of t labelled vertex-disjoint triangles.

Given a graph G on a fixed vertex partition $V^1 \cup V^2 \cup V^3$ as above, we define $\Psi^t(G)$ to be the collection of labelled embeddings of D_t into G, that map $[t] \times \{i\}$ to a subset of V^i for $i \in [3]$. We will be interested in embeddings that fix certain vertices to be isolated. Given $\underline{u} = (u_1, \ldots, u_\ell) \in \mathcal{V}$ of length $\ell \leq 3$ as above and $t \in [n-1]$, we define $\Psi^t_{\underline{\hat{u}}}(G) \subseteq \Psi^t(G)$ to be those $\psi \in \Psi^t(G)$ for which $\psi((s,i)) \neq u_i$ for all $i \in [\ell]$ and $s \in [t]$. That is, we fix the ℓ vertices in $\underline{u}$ to be isolated in the embedding of D_t.

We remark that if $\underline{u} = \emptyset$, then $\Psi^t_{\underline{\hat{u}}}(G) = \Psi^t(G)$ and also note that for an arbitrary $\underline{u} \in \mathcal{V}$ one has that $\Psi^t_{\underline{\hat{u}}}(G) = \Psi^t(G_{\underline{\hat{u}}})$ where $G_{\underline{\hat{u}}}$ is considered as a tripartite graph on partition $V^1_{\underline{\hat{u}}} \cup V^2_{\underline{\hat{u}}} \cup V^3_{\underline{\hat{u}}}$.

Finally, given a vertex $v \in V^1$, we denote by $\Psi^t_v(G) \subseteq \Psi^t(G)$ the set of embeddings $\psi \in \Psi^t(G)$ for which $\psi((1,1)) = v$.

Induced subgraphs: For a graph $G = (V,E)$ and some $U \subseteq V$, we define $G[U]$ to be the subgraph of G *induced* by U, that is $V(G[U]) = U$ and $E(G[U]) = \{e \in E : e \subset U\}$. Similarly, given disjoint subsets $U_1, \ldots, U_k \subset V$, we define $G[U_1, \ldots, U_k]$ to be the k-partite subgraph of G induced by $U_1, \ldots, U_k$, that is $V(G[U]) = U_1 \cup \ldots \cup U_k$ and

$$E(G[U_1, \ldots, U_k]) = \{e \in E : e \subset U_1 \cup \ldots \cup U_k \text{ and } |e \cap U_i| \leq 1 \text{ for all } i \in [k]\}.$$

Given a graph G and a collection of vertices $u_1, \ldots, u_\ell$, we consider the graph induced after removing the u_i, by defining the shorthand $G_{\hat{u}_1,\ldots,\hat{u}_k} := G[V_{\hat{u}_1,\ldots,\hat{u}_k}]$, where $V = V(G)$. For a tuple of vertices $\underline{u}$, the graph $G_{\underline{\hat{u}}}$ is defined analogously.

The notation above will be used for large 'host' graphs G with n vertices. In relation to small graphs (of constant size), we will adopt the following notation. Given a graph F and a vertex subset $W \subset V(F)$, $F \setminus W$ denotes $F[V(F) \setminus W]$ and if $W = \{w\}$ we will drop the set brackets, simply writing $F \setminus w$ to denote $F \setminus \{w\}$.

Hypergraphs: If $\mathcal{H}$ is an r-uniform hypergraph for some $r \in \mathbb{N}$ and $v, u \in V(\mathcal{H})$, $\deg^{\mathcal{H}}(v)$ denotes the number of edges in $\mathcal{H}$ containing v, and $\mathrm{codeg}^{\mathcal{H}}(u,v)$ denotes the number of edges of $\mathcal{H}$ that contain both u and v. If the hypergraph $\mathcal{H}$ is clear from context, we drop the superscripts. If $\mathcal{H}$ is an r-uniform hypergraph with $r \geq 3$ and J is a 2-*uniform* graph on the same vertex set $V(\mathcal{H})$, then $\mathcal{H}_J$ denotes the subhypergraph of $\mathcal{H}$ given by all edges of $\mathcal{H}$ that contain some edge of J.

Graph unions and differences: If $\tilde{G}$ is a graph on the same vertex set as G we write $G \cup \tilde{G}$ to denote the graph on vertex set $V(G)$ with edge set $E(G) \cup E(\tilde{G})$. For graphs $\tilde{G}$ and G on the same vertex set with $\tilde{G}$ a subgraph of G, we let $G \setminus \tilde{G}$ denote the graph on $V(G)$ given by the set of edges that feature in G but not in $\tilde{G}$. If $\mathscr{H}'$ and $\mathscr{H}$ are r-uniform hypergraphs with $\mathscr{H}'$ a subgraph of $\mathscr{H}$, then $\mathscr{H} \setminus \mathscr{H}'$ is defined similarly.

Graph blowups: We write $K^r_{m_1,m_2,\ldots,m_r}$ to denote the complete r-partite graph with parts of size $m_1, \ldots, m_r$. For a graph J on r vertices $\{v_1, \ldots, v_r\}$ and $m_1, \ldots, m_r \in \mathbb{N}$, we define the *blow-up of* J to be the r-partite graph $J_{m_1,\ldots,m_r}$ with vertex set $P_1 \cup P_2 \cup \ldots \cup P_r$, such that $|P_i| = m_i$ and for all $i, j \in [r]$ and $w \in P_i$, $w' \in P_j$ we have $ww' \in E(J_{m_1,\ldots,m_r})$ if and only if $v_i v_j \in E(J)$.

Chapter 1

Introduction

In this thesis we study *clique factors*. We say a graph G contains a *K_r-factor* if there is a collection of vertex-disjoint copies of K_r that completely cover the vertex set of G. When $r = 3$, we often refer to a K_3-factor as a *triangle factor*. As a natural generalisation of a perfect matching in a graph, K_r-factors are a fundamental object in graph theory with a wealth of results studying various aspects and variants. However, unlike perfect matchings, it is not easy to verify whether a graph G contains a K_r-factor or not. Certainly it is necessary that the number of vertices of G must be divisible by r but given this, it was proved by Schaeffer [102] (in the case $r = 3$) and by Hell and Kirkpatrick [92] (in general) that determining if a graph on $n \in r\mathbb{N}$ vertices contains a K_r-factor is an NP-complete problem.

Given that we cannot hope for a nice characterisation of graphs which contain K_r-factors, it is natural to study the relationship between the property of containing a K_r-factor and other graph parameters. In particular, one can investigate notions of *density* and this leads to two fundamental questions. From an *extremal* perspective, we can ask what density condition forces the existence of a K_r-factor. Here it makes sense to use the minimum degree of a graph as the density parameter to avoid superficial examples, such as graphs with isolated vertices. From a *probabilistic* perspective, we can ask what density condition forces the existence of a K_r-factor in *almost all* graphs that satisfy the condition. This corresponds to determining the *threshold* for the appearance of a K_r-factor in $G(n, p)$. These questions are now well-understood, as we will discuss in detail shortly, and they have inspired many further directions of research. Indeed, problems concerning clique factors, as well as the proof methods that have been used to solve them, have been highly influential in extremal and probabilistic combinatorics.

In this thesis, we explore three modern perspectives, each of which merge extremal and probabilistic viewpoints. Firstly, we look at the setting of *pseudorandom* graphs (discussed in Section 1.4), where we ask what conditions on density *and pseudorandomness* force the existence of a K_r-factor. Here, by imposing a pseudorandom condition on a graph we mean (loosely) that we require the graph to emulate a random graph of the same density. Secondly, we look at the *robust* setting (discussed in Section 1.5) which looks at clique factors in *random sparsifications* of dense graphs. That is, in dense graphs G which are guaranteed to contain a K_r-factor (as they are above the extremal minimum degree threshold), we ask what density condition forces the existence of K_r-factors in *almost all subgraphs* of G. Finally, we look at the *randomly perturbed* setting (discussed in Section 1.6) which asks how many random edges need to be added to an arbitrary graph of a given density in order to force the existence of a K_r-factor. In each setting, we are able to provide a complete answer to these questions in certain regimes of parameters, in particular for triangle factors, by giving tight results. The main theorems of this thesis are Theorem I in Section 1.4, Theorem II in Section 1.5 and Theorem III in Section 1.6. Before discussing these results in detail, we first present the relevant context, looking at the extremal setting in Section 1.1, the probabilistic setting in Section 1.2, and discussing the closely related topic of Hamiltonicity in Section 1.3.

1.1 The extremal perspective

The earliest result on clique factors is the well-known theorem of Corrádi and Hajnal [44], who showed that a triangle factor is guaranteed if the host graph is sufficiently dense.

Theorem 1.1.1 (Corrádi–Hajnal [44])**.** *If G is an n-vertex graph with $n \in 3\mathbb{N}$ and $\delta(G) \geq \frac{2n}{3}$, then G contains a triangle factor.*

This theorem was then extended by Hajnal and Szemerédi [78] who determined the minimum degree needed to guarantee a K_r-factor for larger $r \in \mathbb{N}$. In fact, Hajnal and Szemerédi proved a stronger result on so-called *equitable colourings*, solving a conjecture of Erdős, from which the theorem below follows as a corollary. This is discussed further in Section 2.5. See also [105] for a short proof of Theorem 1.1.2.

Theorem 1.1.2 (Hajnal–Szemerédi [78])**.** *If $3 \leq r \in \mathbb{N}$ and G is a graph on $n \in r\mathbb{N}$ vertices with minimum degree $\delta(G) \geq \left(1 - \frac{1}{r}\right)n$, then G contains a K_r-factor.*

The result is tight, as can be seen, for example, by taking G to be a complete graph with a clique of size $\frac{n}{r} + 1$ removed to leave an independent set of vertices, say I. One then has that $\delta(G) = \left(1 - \frac{1}{r}\right)n - 1$ and G does not have a K_r-factor. Indeed, any copy of K_r in a family of vertex-disjoint K_rs can use at most one vertex of I but a K_r-factor should contain $\frac{n}{r} < |I|$ copies of K_r.

This celebrated result thus captures a family of dense graphs that contain K_r-factors, providing a sufficient condition that is computationally easy to verify. However, the condition requires the host graph to be very dense and there are many other graphs containing K_r-factors that are not captured by Theorem 1.1.2. Indeed, as we will see in the next section, *almost all* graphs contain K_r-factors and this remains true when we focus on much sparser graphs. In this thesis, we will explore these ideas further, showing that graphs that fall below the extremal minimum degree threshold given by Theorem 1.1.2 and do *not* contain K_r-factors are *atypical* (Section 1.4) and are *close* to containing a K_r-factor (Section 1.6). Moreover, for triangle factors we will see that the minimum degree threshold given by Theorem 1.1.1 actually guarantees much more than a single triangle factor, showing that graphs with $\delta(G) \geq \frac{2n}{3}$ are *robust* with respect to containing triangle factors (Section 1.5).

Before moving on we mention that Theorem 1.1.1 and Theorem 1.1.2 have inspired many results in extremal graph theory, with generalisations obtained in several directions. Treglown [173] obtained a degree-sequence version of Theorem 1.1.2 and Keevash and Mycroft [104] proved an analogue of the Theorem 1.1.2 in the setting of *r-partite graphs*, whilst there are now several generalisations in the setting of *directed graphs* (see e.g. [51, 52, 172]). Further results are discussed in detail in Chapter 6.

1.2 The probabilistic perspective

Recall that the random graph $G(n, p)$ consists of a vertex set $[n] := \{1, \dots, n\}$ where each edge is present with probability $p = p(n)$, independently of all other choices. We say that a function $p^* = p^*(n)$ is a *threshold* for a (monotone increasing[1]) graph property $\mathcal{P}$ if there exists constants $C, c > 0$ such that:

(i) If $p = p(n) \geq Cp^*$, we have that whp $G(n, p)$ satisfies $\mathcal{P}$.

(ii) If $p = p(n) \leq cp^*$ we have that whp $G(n, p)$ does not satisfy $\mathcal{P}$.

Moreover, we say that p^* is a *sharp threshold* for $\mathcal{P}$ if for all $\varepsilon > 0$, (i) remains valid with the condition $p \geq (1 + \varepsilon)p^*$ and (ii) remains valid with the condition $p \leq (1 - \varepsilon)p^*$. As is usual in random graph theory, we will sometimes refer to *the* threshold for a graph property, despite the fact that the threshold function is not unique (but rather unique up to a constant factor). Thresholds give us a way to study the set of all graphs of a certain density with respect to a certain graph property $\mathcal{P}$. Indeed, standard results (see for example [95, Section 1.4]) show that if $\mathcal{P}$ is monotone increasing with threshold p^* and $1 \leq m \leq \binom{n}{2}$ is such that $m = \omega(p^* n^2)$, then *almost all* graphs with n vertices and m edges satisfy $\mathcal{P}$. That is, a uniformly random graph

[1]A graph property $\mathcal{P}$ is monotone increasing if for two graphs G, G' on n vertices, G being a subgraph of G' and G satisfying $\mathcal{P}$ implies that G' satisfies $\mathcal{P}$.

with m edges will whp satisfy $\mathcal{P}$. Likewise, if $m = o(p^* n^2)$ then almost all graphs with m edges will not satisfy $\mathcal{P}$.

In the early 1990s, the problem of determining the threshold for the property that $G(n,p)$ contains an H-factor[2] attracted the attention of Erdős (see [63]). Indeed, as well as raising the general problem, Erdős particularly focused on the case when $H = K_3$, addressing the expected number $p^*\binom{n}{2}$ of random edges at the threshold p^* for triangle factors in $G(n,p)$ and stating that 'the correct answer will be probably about $n^{4/3}$ edges but perhaps a little more will be needed', and he cautioned that 'the lack of analogs to Tutte's theorem may cause serious trouble' in establishing the threshold. This caution turned out to be well-founded as for a number of years even the case of triangles remained quite stubborn. However, in 2008, spectacular work of Johannson, Kahn and Vu [96] not only resolved the problem for K_3-factors, but the general problem of H-factors for all so-called *strictly balanced* graphs H as well as the analogous problem for hypergraphs. We discuss these variations in more detail in Chapter 6 and state their result only in the case of clique factors here.

Theorem 1.2.1 (Johansson–Kahn–Vu [96])**.** *Let $n \in \mathbb{N}$ be divisible by $r \in \mathbb{N}$ where $r \geq 3$. Then the threshold function for containing a K_r-factor in $G(n,p)$ is*

$$p_r^*(n) := n^{-2/r} (\log n)^{2/(r^2-r)}.$$

The fact that the threshold for containing K_r-factors is at least $p_r^*(n)$ was known before [96] and follows from the fact that $p_r^*(n)$ is the threshold for the property that every vertex is contained in a copy of K_r in $G(n,p)$, as shown originally by Spencer [163] (see also [95, Theorem 3.22 (ii)]). Establishing the upper bound on the threshold was much more challenging and initial progress focused on triangle factors with Ruciński [155] and independently Alon and Yuster [13], giving an upper bound of $n^{-1/2}(\log n)^{1/2}$ on the threshold. Krivelevich [119] then improved this to $n^{-3/5}$ with a proof that is often cited as one of the first instances of the *absorption method* (see Section 2.8). Kim [106] then improved the upper bound further to $n^{-11/18}$ before the problem was finally resolved by Johansson, Kahn and Vu [96] with an involved probabilistic proof relying on the use of *entropy* (see Section 2.3).

Recent results of Kahn [99, 100] in the setting of perfect matchings in random hypergraphs, along with coupling arguments of Riordan [153] and Heckel [90] have established a sharp threshold for K_r-factors in $G(n,p)$. We remark also that a recent breakthrough result of Frankston, Kahn, Narayanan and Park [69], which gives a very general result for thresholds, gives an upper bound on the threshold for containing a K_r-factor in $G(n,p)$ but falls short of the correct threshold, requiring a $\log(n)$ factor rather than $\log(n)^{2/(r^2-r)}$.

[2]Recall from the Notation Section that an H-factor in a graph G is a collection of vertex-disjoint copies of H in G that cover the vertex set of G.

1.3 A detour to Hamiltonicity

Before discussing the results of this thesis, we briefly introduce another fundamental spanning structure in graph theory. A *Hamilton cycle* in a graph G is a cycle covering all the vertices of G and a graph that contains a Hamilton cycle is said to *Hamiltonian*. Hamilton cycles have been thoroughly studied in many contexts and through studying Hamiltonicity, various new phenomena have been discovered, setting trends for further research. Indeed many interesting questions about clique factors (for example those we consider in Sections 1.5 and 1.6) have been inspired by results for Hamiltonicity and establishing analogous results for clique factors often poses a considerable challenge. Here, we collect the seminal results on the subject.

One of the cornerstone theorems of extremal combinatorics is the classical theorem of Dirac [55] establishing the minimum degree threshold for Hamiltonicity.

Theorem 1.3.1 (Dirac [55]). *Any n-vertex graph G with $\delta(G) \geq \frac{n}{2}$ is Hamiltonian.*

This result is tight by considering, for example, a complete bipartite graph with parts of size $\lfloor \frac{n}{2} \rfloor + 1$ and $\lceil \frac{n}{2} \rceil - 1$. In random graphs, the threshold for Hamiltonicity was established by Koršunov [116] and independently Pósa [149].

Theorem 1.3.2 (Koršunov [116], Pósa [149]). *The threshold for Hamiltonicity in $G(n, p)$ is $\frac{\log n}{n}$.*

As with Theorem 1.2.1, the lower bound for the threshold is not difficult. Indeed, if $p \leq c\frac{\log(n)}{n}$ for $c < 1$, then whp $G(n, p)$ contains an isolated vertex and so cannot contain a Hamilton cycle. Proving the upper bound for the threshold was a long-standing open problem in random graph theory. Koršunov [116] in fact established a sharp threshold and Komlós and Szemerédi [114] gave an even more precise threshold result using similar methods. An alternative (simpler) approach to establishing the threshold for Hamiltonicity has only recently been discovered by Frankston, Kahn, Narayanan and Park [69] and follows from their general result on thresholds mentioned above. We refer the reader to the annotated bibliography of Frieze [70] for an in-depth discussion of the many further results on Hamiltonicity in random graphs.

1.4 The pseudorandom perspective

Although Theorem 1.2.1 implies that almost all graphs of density $\omega(p_r^*(n))$ have K_r-factors, we cannot use it to establish if some given graph G contains a K_r-factor. However, it suggests that we can capture much sparser graphs with K_r-factors that are not covered by Theorem 1.1.2, by adding conditions that preclude atypical behaviour.

This naturally leads us to the notion of *pseudorandom graphs*, which are, roughly speaking, graphs which imitate random graphs of the same density. The study of pseudorandom graphs, initiated in the 1980s by Thomason [169, 170], has become a central and vibrant field at the intersection of Combinatorics and Theoretical Computer Science. We refer to the excellent survey of Krivelevich and Sudakov [126] for an introduction to the topic. One way of imposing pseudorandomness is through the spectral notion of the eigenvalue gap. This then leads to the study of (n, d, λ)-graphs G which are d-regular n-vertex graphs with *second eigenvalue* λ. By second eigenvalue, what is actually meant is the second largest eigenvalue in absolute value as follows. Given an n-vertex d-regular graph G, we can look at the eigenvalues of the adjacency matrix A of G which, as A is a symmetric $0/1$-matrix, are real and can be ordered as $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n$. The second eigenvalue is then defined to be $\lambda := \max\{|\lambda_2|, |\lambda_n|\}$. It turns out that this parameter λ controls the pseudorandomness of the graph G, with smaller values of λ giving graphs that have stronger pseudorandom properties. More concretely, the relation is given by the following property of (n, d, λ)-graphs, see e.g. [126, Theorem 2.11], which is known as the *Expander Mixing Lemma* and shows that λ controls the edge distribution between vertex sets. For any vertex subsets A, B of an (n, d, λ)-graph G, one has that

$$\left| e(A, B) - \frac{d}{n}|A||B| \right| \leq \lambda\sqrt{|A||B|}, \tag{1.4.1}$$

where $e(A, B) := |\{uv \in E(G) : u \in A, b \in B\}|$ denotes the number[3] of edges in G with one endpoint in A and the other in B. Note that $\frac{d}{n}$ is the density of the graph G, and hence one would expect to see $\frac{d}{n}|A||B|$ edges between the vertex sets A and B in a random graph G. The pseudorandom parameter λ then controls the discrepancy from this paradigm.

It follows from simple linear algebra, see e.g. [126], that for an (n, d, λ)-graph, one has that $\lambda \leq d$ always and moreover, as long as d is not too close to n, say $d \leq \frac{2n}{3}$, one has that $\lambda = \Omega(\sqrt{d})$. Thus, we think of (n, d, λ)-graphs with $\lambda = \Theta(\sqrt{d})$ as being *optimally pseudorandom*. For example, it is known that random regular graphs are optimally pseudorandom (n, d, λ)-graphs whp [31, 43, 171].

A prominent theme in the study of pseudorandom graphs has been to give conditions on the parameters, n, d and λ that guarantee certain properties of an (n, d, λ)-graph. For example, it follows easily from (1.4.1) that any (n, d, λ)-graph G with $\lambda < \frac{d^2}{n}$ contains a triangle as there is an edge in the neighbourhood of every vertex. In particular, any optimally pseudorandom graph with $d = \omega(n^{2/3})$ must contain a triangle. Moreover, this condition is tight due to a triangle-free construction of an (n, d, λ)-graph due to Alon [7] with $d = \Theta(n^{2/3})$ and $\lambda = \Theta(n^{1/3})$. Alon's construction is optimally pseudorandom and Krivelevich, Sudakov and Szabó [127] generalised it to the whole possible range of densities. That is, for any $d = d(n)$ such that $\Omega(n^{2/3}) = d \leq n$, they gave a sequence of infinitely many n and triangle-free (n', d, λ)-graphs with $n' = \Theta(n)$

[3] Note that edges that lie in $A \cap B$ are counted twice.

and $\lambda = \Theta(\frac{d^2}{n})$. In general, finding optimal conditions for subgraph appearance in (n, d, λ)-graphs seems very hard. Indeed the only tight conditions that are known are those for fixed size odd cycles [11, 126]. With respect to spanning structures, before the work presented here, it was only perfect matchings that had been well understood [32, 37, 126]. Whilst such questions are interesting in their own right, they also have implications in other areas of mathematics. As an example, we mention the beautiful connection given by Alon and Bourgain [8] (see also [2]) who used the existence of certain subgraphs in pseudorandom graphs to prove the existence of additive patterns in large multiplicative subgroups of finite fields.

Here, we answer what has become one of the central problems in this area, by giving a tight condition for an (n, d, λ)-graph to contain a triangle factor.

Theorem I*. *There exists $\varepsilon > 0$ such that any (n, d, λ)-graph with $n \in 3\mathbb{N}$ and $\lambda \le \frac{\varepsilon d^2}{n}$, contains a triangle factor.*

Theorem I* was conjectured by Krivelevich, Sudakov and Szabó [127] in 2004. Focusing solely on optimally pseudorandom graphs, that is, setting $\lambda = \Theta(\sqrt{d})$, Theorem I* gives that any optimally pseudorandom graph with $d = \omega(n^{2/3})$ contains a triangle factor. Comparing this to Theorem 1.1.1, we see that imposing pseudorandomness, which is easy to compute via the second eigenvalue, allows us to capture much sparser graphs which are guaranteed to contain a triangle factor.

Theorem I* (and the more general Theorem I below) conclude a body of work towards the conjecture of Krivelevich, Sudakov and Szabó and the proof of the theorem, discussed in detail in Section 3.1, builds upon the many beautiful ideas of various authors, which have arisen in this study. The first step towards the conjecture was given by Krivelevich, Sudakov and Szabó [127] themselves, who showed that $\lambda \le \frac{\varepsilon d^3}{n^2 \log n}$ for some sufficiently small ε is enough to guarantee a triangle factor. This was improved to $\lambda \le \frac{\varepsilon d^{5/2}}{n^{3/2}}$ by Allen, Böttcher, Hàn, Kohayakawa and Person [3] who also proved that the same condition guarantees the appearance of the square of a Hamilton cycle, a supergraph of a triangle factor (see Section 6.1.2 for more on this). Nenadov [145] then got very close to the conjecture, showing that $\lambda \le \frac{\varepsilon d^2}{n \log n}$ guarantees a triangle factor. Concentrating solely on optimally pseudorandom graphs, these results imply that having degree $d = \omega(n^{4/5}(\log n)^{2/5})$, $\omega(n^{3/4})$ and $\omega((n \log n)^{2/3})$ respectively, guarantees the existence of a triangle factor.

In a different direction, one can fix the condition that $\lambda \le \frac{\varepsilon d^2}{n}$ for some small $\varepsilon > 0$ and prove the existence of other structures giving evidence for a triangle factor. Again, this was initiated by Krivelevich, Sudakov and Szabó [127] who proved that with this condition, one can guarantee the existence of a *fractional triangle factor*. That is, they showed that there is some function w which assigns a weight $w(T) \in [0, 1]$ to each triangle T in a pseudorandom graph G, such that for every vertex $v \in V(G)$, one has that the sum $\sum_{v \in T} w(T)$ of the weights of triangles containing v

is precisely equal to 1. Imposing $\{0,1\}$-weights recovers the notion of a triangle factor and a fractional triangle factor is thus a natural relaxation. Another interesting result of Sudakov, Szabó and Vu [166] showed that when we have $\lambda \leq \frac{\varepsilon d^2}{n}$, we have many triangles and these are well distributed in the (n,d,λ)-graph G. Indeed they proved a Turán-type result showing that any triangle-free subgraph of such a graph G must contain at most half the edges of G. A more recent result due to Han, Kohayakawa and Person [85, 86] shows that $\lambda \leq \frac{\varepsilon d^2}{n}$ guarantees the existence of an *almost triangle factor*; that there are vertex-disjoint triangles covering all but $n^{647/648}$ vertices of such an (n,d,λ)-graph.

We will deduce Theorem I* from a more general theorem (Theorem I below) which deals with K_r-factors for all $r \geq 3$ and works with a larger class of pseudorandom graphs where we do not restrict solely to regular graphs. Indeed, we will work with following notion of *bijumbledness*, whose usage dates back to the original works of Thomason [169, 170], and whose definition captures the key property of edge distribution, given for (n,d,λ)-graphs by (1.4.1).

Definition 1.4.1. Let $n \in \mathbb{N}$, $p = p(n) \in [0,1]$ and $\beta = \beta(n,p) > 0$. An n-vertex graph $G = (V,E)$ is (p,β)-*bijumbled* if for every pair vertex subsets $A, B \subseteq V$, one has that

$$|e(A,B) - p|A||B|| \leq \beta\sqrt{|A||B|}. \tag{1.4.2}$$

Note that, due to (1.4.1), (n,d,λ)-graphs are $\left(\frac{d}{n},\lambda\right)$-bijumbled. As with (n,d,λ)-graphs, we are interested in finding conditions on the parameters n, p and β, that guarantee the existence of certain subgraphs in n-vertex (p,β)-bijumbled graphs. Our main theorem gives conditions for the existence of K_r-factors for all $r \geq 3$ in this setting.

Theorem I. *For every $3 \leq r \in \mathbb{N}$ and $c > 0$ there exists an $\varepsilon > 0$ such that any n-vertex (p,β)-bijumbled graph with $n \in r\mathbb{N}$, $\delta(G) \geq cpn$ and $\beta \leq \varepsilon p^{r-1} n$, contains a K_r-factor.*

We remark that the condition that $\delta(G) \geq cpn$ is natural. Indeed Definition 1.4.1 implies that almost all vertices will have degree at least cpn and some lower bound on minimum degree is necessary to avoid isolated vertices. Theorem I* follows directly from Theorem I and much of the context and past results discussed above have analogous statements when $r \geq 4$ with many authors also working in the more general setting of (p,β)-bijumbled graphs. In particular, for all $r \geq 3$, a condition of $\beta = o(p^{r-1}n)$ guarantees a copy of K_r and before Theorem I the best condition known for ensuring a K_r-factor was $\beta = o\left(\frac{p^{r-1}n}{\log n}\right)$ due to Nenadov [145]. Another result due to Han, Kohayakawa, Person and the author [83] appeared at roughly the same time as that of Nenadov and gave a condition of $\beta = o(p^r n)$ for a K_r-factor, which for $r \geq 4$ gives a stronger result than the previously best known condition of Allen, Böttcher, Hàn, Kohayakawa and Person [3]. Although this condition is weaker than Nenadov's only when the bijumbled graph is very dense, it turns out that the proof methods of both results will be useful in proving Theorem I.

There is one key difference in the picture for the case when $r = 3$ and when $r \geq 4$: the tightness of the condition $\beta = o(p^{r-1}n)$ for *both* the clique and the clique factor when $r \geq 4$ is unknown. We defer a more in depth discussion of this to our concluding remarks (Chapter 6) and conclude this section by again focusing on the most interesting case of triangle factors where we know that Theorem I and Theorem I* are tight due to the construction of Alon [7] (and its generalisation to the whole range of densities by Krivelevich, Sudakov and Szabó [127]) discussed above. Indeed, one of the reasons that the Krivelevich-Sudakov-Szabó conjecture (Theorem I*) has attracted so much attention is that it marks a distinct difference between the behaviour of random graphs and that of (optimally) pseudorandom graphs. In random graphs, we know that triangles appear at density roughly $p = n^{-1}$, whilst for triangle factors the threshold is considerably denser, namely $p = n^{-2/3}(\log n)^{1/3}$ as given by Theorem 1.2.1. On the other hand, there exists triangle-free, optimally pseudorandom graphs with density roughly $n^{-1/3}$, but Theorem I asserts that any pseudorandom graph whose density is a constant factor larger than this is guaranteed to have not only a triangle but a triangle factor. Furthermore, it follows from Theorem I and (the proof of) a result of Han, Kohayakawa, Person and the author [84] that even more is true.

Corollary 1.4.2. *For every $c > 0$ there exists an $\varepsilon > 0$ such that any n-vertex (p, β)-bijumbled graph with $\delta(G) \geq cpn$ and $\beta \leq \varepsilon p^2 n$ is* 2-universal. *That is, given any graph F on at most n vertices, with maximum degree* 2*, G contains a copy of F. In particular, any (n, d, λ)-graph G with $\lambda \leq \frac{\varepsilon d^2}{n}$ is* 2*-universal.*

Our proof of Theorem I incorporates discrete algorithmic techniques, probabilistic methods, fractional relaxations and linear programming duality, and the method of absorption. In Section 3.1 we discuss the proof in detail.

References: The results discussed in this section and proven in Chapter 3, as well as the theory developed in Sections 2.6 and 2.7, represent work of the author [141] which has been submitted for publication. An accompanying conference version [140] of this work deals solely with the setting of Theorem I*.

1.5 The robust perspective

In this section, we restrict the discussion to triangle factors. Although, as we have seen, the premise of Theorem 1.1.1 cannot be weakened, one can ask whether the conclusion can be strengthened. Indeed, as we have seen in Section 1.2 and Section 1.4, the extremal examples that force the minimum degree threshold of Theorem 1.1.1 to be large, are rare and atypical. Therefore it is natural to expect that the minimum degree threshold actually guarantees much more than just a single triangle factor. The aim here is to show that this is indeed the case and we provide a robust version of the Corrádi–Hajnal theorem which shows that any graph G as in

Theorem 1.1.1 is *robust* with the property of having a triangle factor: informally, there are many triangle factors within G and these are well-spread around the graph. Before explaining our results in detail, we take a brief detour to discuss robustness with respect to Hamiltonicity, which has been explored extensively and acts as an indicative example for what type of robustness results one can expect.

The idea that graphs satisfying Dirac's condition (Theorem 1.3.1) are robustly Hamiltonian in some sense, has been around for some time, with various measures of robustness being proposed. For example, Sárközy, Selkow and Szemerédi [156] showed that there exists a constant $c > 0$ such that any n-vertex graph G with $\delta(G) \geq \frac{n}{2}$ contains at least $c^n n! \geq (c^2 n)^n$ Hamilton cycles. This is tight up to the value of c and the authors of [156] conjectured that one can in fact take $c = \frac{1}{2} - o(1)$, which was settled by Cuckler and Kahn [50]. This value of c is best possible, as can be seen by considering $G(n, p)$ with $p = \frac{1}{2} + o(1)$.

Having a large number of Hamilton cycles is compelling evidence for such graphs being robustly Hamiltonian but this property alone does not preclude the possibility that these Hamilton cycles are somehow concentrated on a small part of the graph, for example that many of them share a small subset of edges. Further research has gone into proving stronger notions of robustness, for example showing the existence of many edge-disjoint Hamilton cycles [48, 143, 144] or the existence of a Hamilton cycle when an adversary forbids the use of certain combinations of edges [47, 122, 124]. We refer to the nice survey of Sudakov [165] on the matter for more details.

The notion of robustness we will be interested in here is that *almost all spanning subgraphs of G* contain a Hamilton cycle. To formalise this, for some $p \in [0, 1]$ we define the *random sparsification* of a graph G (with respect to p), denoted G_p, to be the graph obtained by keeping every edge of G independently with probability p. Krivelevich, Lee and Sudakov [122] used random sparsifications to show that graphs satisfying Dirac's condition are robustly Hamiltonian. They proved the following remarkable result.

Theorem 1.5.1 (Krivelevich–Lee–Sudakov [122])**.** *There is a constant $C > 0$ such that for all $n \in \mathbb{N}$ and $p \geq C\frac{\log(n)}{n}$, the following holds. If G is an n-vertex graph with $\delta(G) \geq \frac{n}{2}$, then whp G_p is Hamiltonian.*

Of course, one cannot relax the minimum degree below Dirac's threshold of $\frac{n}{2}$. Moreover, the condition on the probability cannot be significantly relaxed, as if $c < 1$ and $p \leq c\frac{\log n}{n}$ then G_p will whp not be Hamiltonian (see Section 1.3).

Theorem 1.5.1 is a common strengthening of two cornerstone theorems in extremal and probabilistic graph theory, namely Theorem 1.3.1 (corresponding to $p = 1$) and Theorem 1.3.2 (corresponding to $G = K_n$). Another strength of Theorem 1.5.1 is that the robustness given by this notion is relatively strong. Indeed, one can easily infer other notions of robustness from

Theorem 1.5.1. For example, as every Hamilton cycle in a graph G survives in G_p with probability p^n, by considering the expected number of Hamilton cycles in G_p, we can conclude that any graph G with $\delta(G) \geq \frac{n}{2}$ has at least $\left(\frac{cn}{\log n}\right)^n$ Hamilton cycles for some $c > 0$, which is only slightly weaker than the aforementioned results counting Hamilton cycles [50, 156]. One can also obtain many edge-disjoint Hamilton cycles by considering a random partition of the edges of G.

For triangle factors, a robustness version of the Corrádi–Hajnal Theorem follows from the sparse blowup lemma [2, Theorem 1.11]: This general result implies that for $\gamma > 0$ and $p \geq C\left(\frac{\log n}{n}\right)^{1/2}$ any n-vertex graph G with minimum degree $\delta(G) \geq \left(\frac{2}{3}+\gamma\right)n$ satisfies that G_p whp has a triangle factor. Turning this into an exact result in terms of the minimum degree condition requires more work, and moving to smaller probabilities p is substantially harder.

Here we achieve both, showing that graphs G satisfying the properties of the Corrádi–Hajnal Theorem are strongly robust for triangle factors: G_p retains a triangle factor all the way down to the threshold probability p for triangle factors. This is an analogue to Theorem 1.5.1 for triangle factors.

Theorem II. *There is a constant $C > 0$ such that for all $n \in 3\mathbb{N}$ and $p \geq C(\log n)^{1/3} n^{-2/3}$ the following holds. If G is an n-vertex graph with $\delta(G) \geq \frac{2n}{3}$ then whp G_p has a triangle factor.*

As with Theorem 1.5.1, both the minimum degree condition and the condition on the probability are tight and Theorem II provides a common generalisation of Theorem 1.1.1 and Theorem 1.2.1.

Our proof of Theorem II builds on an alternative proof of Theorem 1.2.1 for triangle factors in $G(n, p)$ due to Allen, Böttcher, Davies, Jenssen, Kohayakawa and Roberts [6]. This proof in turn shares some of the key ideas with that of Johansson, Kahn and Vu [96] (as well as [99, 100]), in particular the use of entropy, but follows a different scheme of 'building' our triangle factor one triangle at a time. This scheme provides the opportunity for us to strengthen the proof to deal with incomplete graphs G. We defer a detailed discussion of our proof to Section 4.1.

As a corollary to Theorem II, we can provide a lower bound on the number of triangle factors in every graph G with $\delta(G) \geq \frac{2n}{3}$.

Corollary 1.5.2. *There exists a $c > 0$ such that any graph G with $n \in 3\mathbb{N}$ vertices and $\delta(G) \geq \frac{2n}{3}$ contains at least*

$$\left(\frac{cn}{(\log n)^{1/2}}\right)^{2n/3}$$

triangle factors.

Corollary 1.5.2 follows easily from Theorem II by considering the expected number of triangle factors in G_p and the fact that each triangle factor survives in G_p with probability p^n.

Indeed, for a graph F let $T(F)$ denote the number of triangle factors in F. Theorem II implies $\mathbb{P}\left[T(G_p) \geq 1\right] \geq \frac{1}{2}$ for $p \geq C(\log n)^{1/3}n^{-2/3}$, for G as in Corollary 1.5.2, and for n sufficiently large. Since further $\mathbb{E}\left[T(G_p)\right] = T(G)p^n$ we get

$$\frac{1}{2} \leq \mathbb{P}\left[T(G_p) \geq 1\right] \leq \mathbb{E}\left[T(G_p)\right] = T(G)\left(C\frac{(\log n)^{1/3}}{n^{2/3}}\right)^n,$$

implying Corollary 1.5.2 for c sufficiently small.

To our knowledge, Corollary 1.5.2 is the first of its kind and it gets close to the truth. Indeed, letting $n \in 3\mathbb{N}$ and $H = G(n,q)$ be the binomial random graph with $q = \frac{2}{3} + o(1)$, we have that whp H has minimum degree at least $\frac{2n}{3}$ and the expected number of triangle factors in H is

$$\frac{q^n n!}{(n/3)!6^{n/3}} = \left((1+o(1))\frac{2}{e(\sqrt{3})^3}n\right)^{2n/3}.$$

It is believable that every graph as in Corollary 1.5.2 has at least this many triangle factors. As a first step, removing the $(\log n)^{1/2}$ from the expression in Corollary 1.5.2 poses an interesting open problem.

We conclude this section by discussing several further results that have built on the idea of using random sparsifications to give robustness. Recent results of Johansson [97] and Alon and Krivelevich [14] strengthen Theorem 1.5.1 to establish 'hitting time' results and the existence of families of edge-disjoint Hamilton cycles (whilst requiring a slightly stronger condition that $\delta(G) \geq \left(\frac{1}{2}+\varepsilon\right)n$ for some $\varepsilon > 0$). Frieze and Krivelevich [71] studied Hamilton cycles in random subgraphs of pseudorandom graphs and a recent breakthrough of Condon, Espuny Díaz, Girão, Kühn and Osthus [38, 39] established a tight condition for the Hamiltonicity of random subgraphs of the hypercube. In the setting of graphs satisfying a minimum degree condition, the existence of long paths and cycles [57, 75, 123, 152] have also been extensively studied as well as perfect matchings [74, 77].

References: The results discussed in this section and proven in Chapter 4, as well as the theory developed in Sections 2.2 and 2.3, represent joint work with Peter Allen, Julia Böttcher, Jan Corsten, Ewan Davies, Matthew Jenssen, Barnaby Roberts and Jozef Skokan [5] which is currently being prepared for submission. A preliminary version of this work appeared in [45].

1.6 The randomly perturbed perspective

In the previous section, we discussed graphs G above the minimum degree threshold for clique factors and the effect of taking random sparsifications, looking at subgraphs of the form $G_p = G \cap G(n,p)$. In this section, we will use random edges to *help* with the existence of a clique factor, starting with dense graphs G *below* the extremal threshold and looking at graphs of the form $G \cup G(n,p)$ after adding random edges. We will see that when a graph G is dense and avoids a clique factor, it is, in a sense, *close* to containing a clique factor, in that a small random perturbation of G results in a graph which does contain a clique factor. Moreover, by exploring the optimal amount of random edges needed in the whole range of positive densities, we bridge the gap between the extremal (Section 1.1) and probabilistic (Section 1.2) perspectives for the problem of clique factors.

The idea of studying the effect of random perturbations appeared almost simultaneously in two distinct settings. In Computer Science, Spielman and Teng [164] introduced the notion of smoothed analysis of algorithms. By randomly perturbing an input to an algorithm, they could interpolate between a worst-time case analysis and an average case analysis. Their initial work [164], for which they were awarded the 2009 Fulkerson prize, studied the smoothed analysis of the simplex algorithm. In graph theory, Bohman, Frieze and Martin [23] introduced the randomly perturbed model which, as with smoothed analysis, allows one to understand the interplay between an extremal and probabilistic viewpoint. In their model one starts with a dense graph and then adds m random edges to it. A natural problem in this setting is to determine *how many* random edges are required to ensure that the resulting graph whp contains a given graph F as a spanning subgraph. For example, the main result in [23] states that for every $\tau > 0$, there is a $c = c(\tau)$ such that if we start with an arbitrary n-vertex graph G of minimum degree $\delta(G) \geq \tau n$ and add cn random edges to it, then whp the resulting graph is Hamiltonian. This result characterises how many random edges we require for *every* fixed $\tau > 0$. Indeed, if $\tau \geq \frac{1}{2}$ then Theorem 1.3.1 implies that we do not require any random edges; that is any n-vertex graph G of minimum degree $\delta(G) \geq \tau n$ is already Hamiltonian. On the other hand, if $0 < \tau < \frac{1}{2}$ then the following example implies that we indeed require a linear number of random edges: Let G' be the complete bipartite graph with vertex classes of size $\tau n, (1-\tau)n$. It is easy to see that if one adds fewer than $(1-2\tau)n$ (random) edges to G', the resulting graph is *not* Hamiltonian.

In recent years, a range of results studying similar phenomena in graphs and hypergraphs have been obtained, looking at spanning structures (see Section 6.1 for a detailed discussion of these), as well as other aspects of the model such as Ramsey properties [53, 54, 128, 150]. Much of this work has focused on the range where the minimum degree of the deterministic graph is linear but with respect to some arbitrarily small constant τ. In this range, one thinks of the deterministic graph as 'helping' $G(n,p)$ to get a certain spanning structure and the observed phenomenon is usually a decrease in the probability threshold of a logarithmic factor, as is the case for Hamiltonicity as above. Recently, there has been interest in the other extreme, where one starts with a minimum degree slightly less than the extremal minimum degree threshold for a certain spanning structure and requires a small 'sprinkling' of random edges to guarantee the existence of the spanning structure in the resulting graph, see e.g. [56, 147]. Here, we will study the full range of positive densities.

Balogh, Treglown and Wagner [19] first considered the H-factor problem in the setting of randomly perturbed graphs. Indeed, for every fixed graph H they determined how many random edges one must add to a graph G of linear minimum degree to ensure that whp $G \cup G(n, p)$ contains an H-factor. We only state their result in the case of clique factors.

Theorem 1.6.1 (Balogh–Treglown–Wagner [19])**.** *Let $r \geq 2$. For every $\tau > 0$, there is a constant $C = C(\tau, r) > 0$ such that if $p \geq Cn^{-2/r}$ and G is an n-vertex graph with $n \in r\mathbb{N}$ and minimum degree $\delta(G) \geq \tau n$ then whp $G \cup G(n, p)$ contains a K_r-factor.*

Theorem 1.6.1, unlike Theorem 1.2.1, does not involve a logarithmic term. Thus comparing the randomly perturbed model with the random graph model, we see that starting with a graph of linear minimum degree instead of the empty graph saves a logarithmic factor in terms of how many random edges one needs to ensure the resulting graph whp contains a K_r-factor. Further, Theorem 1.6.1 is best-possible in the sense that given any $0 < \tau < \frac{1}{r}$ and $n \in r\mathbb{N}$, there is a constant $c = c(\tau, r) > 0$ and an n-vertex graph G with minimum degree at least τn so that whp $G \cup G(n, p)$ does not contain a K_r-factor when $p \leq cn^{-2/r}$ (see Section 2.1 in [19] or Section 5.1 of this thesis). However, as suggested in [19], this still leaves open the question of how many random edges one requires if $\tau > \frac{1}{r}$.

Here, we give a sharp answer to this question. Before we state our result we introduce some notation which captures the tightness of our results.

Definition 1.6.2. [Perturbed thresholds for factors] Given some $0 \leq \tau \leq 1$, and a graph H with r vertices, the *perturbed threshold* $p(H, \tau)$ for an H-factor satisfies the following[4]. There exists constants $C = C(H, \tau), c = c(H, \tau) > 0$ such that:

(i) If $p = p(n) \geq Cp(H, \tau)$, then for *any* n-vertex graph G with $n \in r\mathbb{N}$ and $\delta(G) \geq \tau n$, whp $G \cup G(n, p)$ contains an H-factor.

(ii) If $p = p(n) \leq cp(H, \tau)$, then for all $n \in r\mathbb{N}$ there is *some* n-vertex graph G with $\delta(G) \geq \tau n$ such that whp $G \cup G(n, p)$ does not contain an H-factor.

If it is the case that for sufficiently large $n \in r\mathbb{N}$, every n-vertex graph with minimum degree at least τn contains an H-factor we define $p(H, \tau) := 0$.

[4]It is not a priori clear that such a threshold exists, as we require the conclusions to hold whp but only impose a constant factor separation from $p(H, \tau)$. In all the cases we consider, we will show that such a threshold does indeed exist (and determine its value).

Thus, Theorem 1.1.2 implies that $p(K_r, \tau) = 0$ for all $\tau \geq 1 - \frac{1}{r}$ whilst Theorem 1.2.1 precisely states that $p(K_r, 0) = p_r^*(n) = n^{-2/r}(\log n)^{2/(r^2-r)}$. Our main result deals with the intermediate cases (i.e. when $0 < \tau < 1 - \frac{1}{r}$).

Theorem III. *Let $2 \leq k \leq r$ be integers. Then given any $1 - \frac{k}{r} < \tau < 1 - \frac{k-1}{r}$,*

$$p(K_r, \tau) = n^{-2/k}.$$

Thus, Theorem III provides a bridge between the Hajnal–Szemerédi theorem (Theorem 1.1.2) and the Johansson–Kahn–Vu theorem (Theorem 1.2.1). Notice that the value of $p(K_r, \tau)$ demonstrates a 'jumping' phenomenon; given a fixed k the value of $p(K_r, \tau)$ is the *same* for all $\tau \in \left(\frac{r-k}{r}, \frac{r-k+1}{r}\right)$, however if τ is just above this interval the value of $p(K_r, \tau)$ is significantly smaller.

Note in the case when $k = r$, Theorem III is implied by Theorem 1.6.1. Also appearing shortly before this work [87] was a result [147] concerning powers of Hamilton cycles in randomly perturbed graphs which implies the case when $k = 2$ and r is even (we discuss this in more detail in Section 6.1.2). To help provide some intuition for Theorem III, note that $n^{-2/k}$ is the threshold for the property that $G(n, p)$ contains a copy of K_k in every linear sized subset of vertices; this property will be exploited throughout the proof. Our proof uses the absorption method, and in particular the novel 'template absorption method' introduced by Montgomery [136, 137] (see Section 2.8). We also use 'reachability' arguments, introduced by Lo and Markstöm [135], in order to build absorbing structures. We use various probabilistic techniques throughout, such as multi-round exposure, and we use regularity in order to obtain an almost factor. A detailed discussion of our proof is given in Section 5.2.

References: The results discussed in this section and proven in Chapter 5, represent joint work with Jie Han and Andrew Treglown [87].

1.7 Organisation

This thesis centres around three main theorems studying clique factors in different settings. We prove Theorem I in Chapter 3, Theorem II in Chapter 4 and Theorem III in Chapter 5. As mentioned above, our proofs rely on a range of different techniques and methods. In Chapter 2, before proving our main theorems, we introduce these proof methods and build up the relevant theory. After proving our main theorems, in Chapter 6 we then discuss related topics and directions for future research, including several open problems and conjectures.

Further references: Beyond the main results of this thesis, we also present and discuss some further results of the author that were established during the doctorate. The proofs of these are omitted but share some features with those given in the thesis. We mention joint work with Jie Han, Yoshiharu Kohayakawa and Yury Person [83, 84] in Section 1.4 and in particular Corollary 1.4.2 follows from both Theorem I and [84]. In Section 6.2.1, we discuss results relating to factors in hypergraphs and present joint work with Hiệp Hàn and Jie Han [79, 80] in the pseudorandom setting and joint work with Yulin Chang, Jie Han, Yoshiharu Kohayakawa and Guilherme Mota [34] in the randomly perturbed setting. Finally, we present joint work with Jie Han, Guanghui Wang and Donglei Yang [88] on clique factors in the Ramsey–Turán setting, in Section 6.2.2.

Chapter 2

Tools and methods

Our proofs will draw on a variety of different techniques from extremal and probabilistic combinatorics. In this chapter we collect these tools, introducing the key concepts and discussing the relevant theory. In Section 2.1, we list some well-known concentration inequalities. In Section 2.2, we then discuss Szemerédi's famous *Regularity Lemma* (Lemma 2.2.1), an extremely powerful tool in the study of dense graphs. We also list some consequences of the definition of regularity. In Section 2.3, we introduce the entropy function for random variables and discuss some properties of this concept. In Section 2.4 we then introduce the phenomenon of supersaturation in dense graphs. In Section 2.5, we discuss known results that guarantee large H-matchings in dense graphs including a result of Hajnal and Szemerédi [78] for clique matchings and a result of Komlós [110] for general H-matchings. Section 2.6 is devoted to discussing fractional matchings and fractional covers in hypergraphs. Using linear programming duality, we derive simple conditions that guarantee the existence of perfect fractional matchings in hypergraphs. In Section 2.7, we discuss a result of Kostochka and Rödl [117] which uses the semi-random method to show the existence of large "almost perfect" matchings in hypergraphs that satisfy pseudorandom degree conditions. Through a method of random sparsification due to Alon, Frankl, Huang, Rödl, Ruciński and Sudakov [10], we then derive results which guarantee the existence of almost perfect matchings in hypergraphs given the existence of many perfect fractional matchings. Finally, in Section 2.8, we discuss the absorption method and in particular introduce the *template method*, a recent innovation due to Montgomery [136, 137] which has proven to be a powerful tool in absorption arguments.

Most of the results and theory discussed in this chapter were known before the works of this thesis and are given without proof. However in certain places, we deviate from previous literature and sharpen, adjust and simplify proofs to suit our purposes. These results may be of independent interest and useful in future works. In particular, we highlight the following two results. Firstly, in Lemma 2.3.9 we show that random variables that have close to maximal entropy are close to uniform, sharpening a result of a similar flavour due to Johansson, Kahn and Vu [96, Theorem 6.2]. Secondly, in Theorem 2.7.3, we give a new result, giving almost perfect matchings in hypergraphs that are robust with respect to containing perfect fractional matchings.

2.1 Concentration inequalities

We will frequently use concentration inequalities for random variables. The first such inequality, Chernoff's inequality [35] (see also [95, Theorem 2.1, Corollary 2.4 and Theorem 2.8]), deals with the case of binomial random variables.

Theorem 2.1.1 (Chernoff bounds)**.** *Let X be the sum of a set of mutually independent Bernoulli random variables and let $\lambda = \mathbb{E}[X]$. Then for any $0 < \delta < \frac{3}{2}$, we have that*

$$\mathbb{P}[X \geq (1+\delta)\lambda] \leq e^{-\delta^2\lambda/3} \quad \text{and} \quad \mathbb{P}[X \leq (1-\delta)\lambda] \leq e^{-\delta^2\lambda/2}.$$

Furthermore, if $x \geq 7\lambda$, then $\mathbb{P}[X \geq x] \leq e^{-x}$.

We will also be interested in concentration inequalities when dealing with random variables that are not mutually independent. We consider the following general setup. Let Λ be a finite set and let Λ_p be a random subset of Λ such that each element of Λ is included independently with probability p. Let $\mathcal{S}$ be a family of non-empty subsets of Λ and for each $S \in \mathcal{S}$, let I_S be the indicator random variable for the event $S \subseteq \Lambda_p$. Thus each I_S is a Bernoulli random variable $\mathrm{Be}(p^{|S|})$. The following inequality, known as Janson's inequality [94] (see also [95, Theorem 2.14]) provides a bound for the lower tail in this case.

Lemma 2.1.2 (Janson's inequality)**.** *In the setting laid out above, let $X := \sum_{S\in\mathcal{S}} I_S$ and $\lambda := \mathbb{E}(X)$. Let $\Delta_X := \sum_{S\cap T\neq\emptyset} \mathbb{E}[I_S I_T]$, where the sum is over not necessarily distinct ordered pairs $S, T \in \mathcal{S}$. Then for any $0 \leq t \leq \lambda$,*

$$\mathbb{P}[X \leq \lambda - t] \leq \exp\left(-\frac{t^2}{2\Delta_X}\right). \tag{2.1.1}$$

One instance of the setting above, which will be of particular interest, is the appearance of subgraphs in random sparsifications of graphs. Recall that given a graph G and some $p \in [0,1]$, we denote by G_p the random subgraph of G with $V(G_p) = V(G)$ in which every edge of G is present independently with probability p. Given a subgraph $F \subset E(G)$ of G (given by its edge set), we denote by I_F the indicator random variable which is 1 if F is present in G_p and 0 otherwise. Chernoff's inequality (Theorem 2.1.1) can be used to give sharp bounds on random variables of the form $X = \sum_{F\in\mathcal{F}} I_F$, where $\mathcal{F} \subset 2^{E(G)}$ is a collection of edge-disjoint subgraphs of G.

However, when $\mathcal{F}$ consists of not-necessarily edge disjoint subgraphs of G, the situation becomes more complicated and we appeal to Janson's inequality (Lemma 2.1.2) which has the following direct corollary providing a bound for the lower tail in this case.

Lemma 2.1.3 (Janson's inequality for subgraphs)**.** *Let G be a graph and $\mathcal{F} \subset 2^{E(G)}$ be a collection of subgraphs of G and let $p \in [0,1]$. Let $X = \sum_{F \in \mathcal{F}} I_F$, let $\lambda = \mathbb{E}[X]$ and let*

$$\bar{\Delta} = \sum_{(F,F') \in \mathcal{F}^2 \colon F \cap F' \neq \emptyset} \mathbb{E}[I_F I_{F'}] .$$

Then, for every $\varepsilon \in (0,1)$, we have

$$\mathbb{P}[X \leq (1-\varepsilon)\lambda] \leq \exp\left(-\frac{\varepsilon^2 \lambda^2}{2\bar{\Delta}}\right).$$

If we additionally require a bound for the upper tail, we will use the Kim–Vu inequality [107] (see also [12, Theorem 7.8.1]). Let $X = \sum_{F \in \mathcal{F}} I_F$ as above. Given an edge $e \in E(G)$, we write t_e for $I_{\{e\}}$. With this we can write X as a polynomial with variables t_e:

$$X = \sum_{F \in \mathcal{F}} \prod_{e \in F} t_e.$$

Given some $A \subset E(G)$, we obtain X_A from X by deleting all summands corresponding to $F \in \mathcal{F}$ that do not contain A and replacing every t_e with $e \in A$ by 1. That is,

$$X_A = \sum_{F \in \mathcal{F} \colon A \subseteq F} \prod_{e \in F \setminus A} t_e.$$

In other words, X_A is the number of $F \in \mathcal{F}$ that contain A and are present in $G_p \cup A$.

Lemma 2.1.4 (Kim–Vu polynomial concentration)**.** *For every $k \in \mathbb{N}$, there is a constant $c = c(k) > 0$ such that the following is true. Let G be a graph and $\mathcal{F} \subset 2^{E(G)}$ be a collection of subgraphs of G, each with at most k edges. Let $X = \sum_{F \in \mathcal{F}} I_F$ as above and $\lambda \coloneqq \mathbb{E}[X]$. For $i \in [k]$, define $E_i \coloneqq \max\{\mathbb{E}[X_A] : A \subset E(G),\ |A| = i\}$. Further define $E' \coloneqq \max_{i \in [k]} E_i$ and $E = \max\{\lambda, E'\}$. Then, for every $\mu > 1$, we have*

$$\mathbb{P}\left[|X - \lambda| > c(EE')^{1/2}\mu^k\right] \leq ce(G)^{k-1}e^{-\mu}.$$

Finally we will need a basic concentration result for the *hypergeometric distribution*: A random variable X is *hypergeometrically distributed* with parameters $N \in \mathbb{N}$ and $K, t \in [N]_0$ if for all $k \in [K]_0$, $\mathbb{P}[X = k]$ is the probability that when drawing t balls from a set of N balls (K of which are blue and $N - K$ red) without replacement, exactly k are blue. That is,

$$\mathbb{P}[X = k] = \frac{\binom{K}{k}\binom{N-K}{t-k}}{\binom{N}{t}}.$$

We will use the following concentration inequality, which Chvátal [36] deduced from Hoeffding's inequality [93], see also [161].

Lemma 2.1.5. *Let X be hypergeometrically distributed with parameters $N \in \mathbb{N}$, $K \in [N]_0$ and $t \in [N]_0$ and let $\lambda := \mathbb{E}[X] = \frac{tK}{N}$. Then, for all $\varepsilon > 0$, we have*

$$\mathbb{P}[|X - \lambda| > \varepsilon\lambda] \leq 2e^{-2\varepsilon^2(K/N)\lambda}.$$

2.2 Regularity

We will use the famous regularity lemma due to Szemerédi [168] which is an extremely powerful tool in modern extremal combinatorics. The lemma and its consequences appeared in the form we give here, in a survey of Komlós and Simonovits [113], which we also recommend for further details on the subject. First we introduce some necessary terminology. Let G be a graph and let $A, B \subset V(G)$ be disjoint subsets of the vertices of G. For non-empty sets $X \subseteq A$, $Y \subseteq B$, we define the *density of* $G[X,Y]$ to be $d_G(X,Y) := \frac{e_G(X,Y)}{|X||Y|}$. Given $\varepsilon > 0$, we say that a pair (A, B) is *ε-regular* in G if for all sets $X \subseteq A$ and $Y \subseteq B$ with $|X| \geq \varepsilon|A|$ and $|Y| \geq \varepsilon|B|$ we have $|d_G(A,B) - d_G(X,Y)| < \varepsilon$. We say that (A,B) is (ε, d)-regular if (A,B) is ε-regular and $d_G(A,B) = d$.

Furthermore, we say (A,B) is *(ε, d, δ)-super-regular* if (A,B) is (ε, d)-regular and satisfies $\deg_G(v; A) \geq \delta|A|$ for all $v \in B$ and likewise $\deg(v; B) \geq \delta|B|$ for all $v \in A$. We say that (A,B) is (ε, d)-super-regular if it is $(\varepsilon, d, d - \varepsilon)$-super-regular. We say that a k-tuple $(A_1, \ldots, A_k)$ of (pairwise disjoint) subsets of $V(G)$ is (ε, d)-(super-)regular if each of the pairs (A_i, A_j) with $i \neq j \in [k]$ is (ε, d)-(super-)regular. We call a k-partite graph G with parts $A_1, \ldots, A_k$, *(ε, d)-(super-)regular* if $(A_1, \ldots, A_k)$ is an (ε, d)-(super-)regular tuple in G. In the interest of brevity, we use the term *(super-)regular tuple* interchangeably to refer to the tuple of vertex sets $(A_1, \ldots, A_k)$ and also to refer to the (super-)regular k-partite graph $G[A_1, \ldots, A_k]$ that G induces on $A_1 \cup \ldots \cup A_k$. Finally we say that (A,B) is *(ε, d^+)-regular* if it is (ε, d')-regular for some $d' \geq d$. Similarly, we say (A,B) is *$(\varepsilon, d^+, \delta)$-super-regular* if it is $(\varepsilon, d', \delta)$-super-regular for some $d' \geq d$ and we say (A,B) is (ε, d^+)-super-regular if it is $(\varepsilon, d', d - \varepsilon)$-super-regular for some $d' \geq d$. The corresponding definitions are made analogously for regular tuples where we require the densities between all pairs involved to be at least d (and do not require these densities to be equal).

We say that a partition $V(G) = V_0 \cup V_1 \cup \cdots \cup V_t$ is an *ε-regular partition* if $|V_0| \leq \varepsilon |V(G)|$, $|V_1| = \cdots = |V_t|$, and for all but at most εt^2 pairs $(i, j) \in [t] \times [t]$, the pair (V_i, V_j) is ε-regular. We refer to the sets V_i for $i \in [t]$ as *clusters* and also use this term to refer to subsets $V_i' \subset V_i$ for $i \in [t]$. We refer to V_0 as the *exceptional set* and the vertices in V_0 are *exceptional vertices*. Given an ε-regular partition and $d \in [0, 1]$, we say R is the (ε, d)-reduced graph of G (with respect to the partition) if $V(R) = [t]$ and $ij \in E(R)$ if and only if (V_i, V_j) is (ε, d^+)-regular. We will use Szemerédi's Regularity Lemma [168] in the following form which follows easily from e.g. [113, Theorem 1.10].

Lemma 2.2.1 (Regularity Lemma)**.** *For all $0 < \varepsilon \leq 1$ and $m_0 \in \mathbb{N}$ there exists $M_0 \in \mathbb{N}$ such that for every $0 < d < \gamma < 1$, every graph G on $n > M_0$ vertices with minimum degree $\delta(G) \geq \gamma n$ has an ε-regular partition $V_0 \cup V_1 \cup \cdots \cup V_m$ with (ε, d)-reduced graph R on m vertices such that $m_0 \leq m \leq M_0$ and $\delta(R) \geq (\gamma - d - 2\varepsilon)m$.*

We will further make use of the following well-known results about (super-)regular tuples. See, for example, [113, Facts 1.3 and 1.5].

Lemma 2.2.2 (Slicing Lemma)**.** *Let $0 < \varepsilon < \beta, d \leq 1$ and let (V_1, V_2) be an (ε, d)-regular pair. Then any pair (U_1, U_2) with $|U_i| \geq \beta |V_i|$ and $U_i \subseteq V_i$, $i = 1, 2$, is (ε', d')-regular with $\varepsilon' = \max\{\frac{\varepsilon}{\beta}, 2\varepsilon\}$ and some $d' > 0$ such that $|d' - d| \leq \varepsilon$.*

Lemma 2.2.3. *Let $0 < \varepsilon < d \leq 1$ and (V_1, V_2) be an (ε, d)-regular pair and let $X_2 \subseteq V_2$ with $|X_2| \geq \varepsilon |V_2|$. Then all but at most $\varepsilon |V_1|$ vertices $v \in V_1$ satisfy $\deg(v; X_2) \geq (d - \varepsilon)|X_2|$. Likewise, all but at most $\varepsilon |V_1|$ vertices $v \in V_1$ satisfy $\deg(v; X_2) \leq (d + \varepsilon)|X_2|$*

The following lemma can be proven by combining the two previous lemmas.

Lemma 2.2.4. *Let $k \in \mathbb{N}$ and $0 < \varepsilon < d \leq 1$ with $\varepsilon \leq \frac{1}{2k}$. If $Z = (V_1, \ldots, V_k)$ is an (ε, d^+)-regular tuple of disjoint vertex sets of size n, then there are subsets $\tilde{V}_1 \subseteq V_1, \ldots, \tilde{V}_k \subseteq V_k$ with $|\tilde{V}_i| = \lceil (1 - k\varepsilon)n \rceil$ for all $i \in [k]$ so that the k-tuple $\tilde{Z} = (\tilde{V}_1, \ldots, \tilde{V}_k)$ is $(2\varepsilon, (d - \varepsilon)^+, d - k\varepsilon)$-super-regular.*

Our next lemma shows that any sufficiently dense pair is automatically regular. It follows directly from the definition of regularity.

Lemma 2.2.5. *Let $0 < \varepsilon < 1$ and (V_1, V_2) be a pair of vertex sets such that $\deg(v_i; V_{3-i}) \geq (1 - \varepsilon^2) |V_{3-i}|$ for all $i \in [2]$ and $v_i \in V_i$. Then (V_1, V_2) form an $\left(\varepsilon, (1 - \varepsilon^2)^+\right)$-super-regular pair.*

The next lemma is an extremely useful tool, extending the control on the edge count in regular pairs to be able to count the number of embeddings of small subgraphs. For a proof see, for example, [113, Theorem 2.1].

Lemma 2.2.6 (Counting Lemma)**.** *Given $0 < \varepsilon < d \leq 1$, $m \in \mathbb{N}$ and H some fixed graph on r vertices, let G be a graph obtained by replacing every vertex x_i of H with an independent set V_i of size n and every edge of H with an (ε, d^+)-regular pair. If $\varepsilon \leq \frac{d^r}{(2+r)2^r} =: d_0$, then there are at least $(d_0 n)^r$ embeddings of H in G so that each x_i is embedded into the set V_i.*

For triangles, we will need a tighter counting lemma which controls the leading constant for the number of triangles and can be applied in subgraphs of a regular triple. The following lemma can be derived easily from the definition of ε-regularity and we omit the proof here.

Lemma 2.2.7. *Let $0 < \varepsilon < d_{1,2}, d_{1,3}, d_{2,3} \leq 1$ and let Γ be a tripartite graph with parts V^1, V^2, V^3 of size n such that (V^i, V^j) is $(\varepsilon, d_{i,j})$-regular for all $1 \leq i < j \leq 3$. Let $X_i \subseteq V^i$ with $|X_i| \geq \varepsilon n$ for all $i \in [3]$. Then,*

$$|K_3(\Gamma[X_1 \cup X_2 \cup X_3])| = d_{1,2}d_{1,3}d_{2,3}|X_1||X_2||X_3| \pm 10\varepsilon n^3.$$

Finally, the following lemma further allows us to control the exact density of a super-regular pair by deleting edges if necessary. We recall here that we say a pair (A, B) of disjoint vertex sets in (ε, d^+)-super-regular if it is $(\varepsilon, d', d-\varepsilon)$-super-regular for some $d' \geq d$.

Lemma 2.2.8. *For all $0 < \varepsilon < 1$, there is some $n_0 > 0$, such that the following is true for every $n \geq n_0$ and every bipartite graph G with parts V_1, V_2 of size n. Suppose that (V_1, V_2) is (ε^2, d^+)-super-regular for some d such that $4\varepsilon \leq d \leq 1$ and $dn^2 \in \mathbb{N}$. Then there is a spanning subgraph $G' \subseteq G$ so that (V_1, V_2) is $(4\varepsilon, d)$-super-regular in G'.*

Proof. Let $d' \geq d$ be the density of (V_1, V_2). For $i \in [2]$, let $Y_i := \{v \in V_i : \deg(v; V_{3-i}) \leq (d' - \varepsilon^2)n\}$ and observe that by the ε^2-regularity of (V_1, V_2) and Lemma 2.2.3, we have $|Y_i| \leq \varepsilon^2 n$ for both $i \in [2]$. Let $E_Y \subset E(G)$ be the set of edges with at least one vertex in $Y := Y_1 \cup Y_2$ and let $E := E(G) \setminus E_Y$. Let $m := |E_Y| \leq 2\varepsilon^2 n^2$. Let $p := \frac{dn^2 - m}{|E|} = \frac{d \pm 2\varepsilon^2}{d'}$. Let E' be a uniformly random subset of E of size exactly $p|E| \in \mathbb{N}$ and let G' be the spanning subgraph of G with edge set $E' \cup E_Y$. By construction, we have $d_{G'}(V_1, V_2) = d$; we will show that (V_1, V_2) is whp $(4\varepsilon, d, d-\varepsilon)$-super-regular in G'.

Let $A_i \subseteq V_i$ with $A_i \geq 4\varepsilon n$, and let $A_i' = A_i \setminus Y_i$ and $B_i = A_i \setminus A_i'$ for both $i \in [2]$. By ε^2-regularity in G, we have $Z := |E_G(A_1', A_2')| = (d' \pm \varepsilon^2)|A_1'||A_2'|$. Let now $X := |E_{G'}(A_1', A_2')|$. Then X is hypergeometrically distributed with parameters $N = |E|, K = Z, t = p|E|$ and thus $\lambda := \mathbb{E}[X] = pZ = (d \pm 2\varepsilon)|A_1'||A_2'|$. Since $\lambda \geq 8\varepsilon^3 n^2$, it follows from Lemma 2.1.5 that

$$\mathbb{P}[|X - \lambda| > \varepsilon\lambda] \leq 2e^{-2\varepsilon^2(K/N)\lambda} \leq 2e^{-\varepsilon^8 n^2}.$$

In particular, we have $\mathbb{P}[d_{G'}(A_1, A_2) = d \pm 4\varepsilon] \geq 1 - 2e^{-\varepsilon^8 n^2}$. By taking a union bound over all choices of A_1, A_2, we deduce that (V_1, V_2) is 4ε-regular with probability at least $1 - 2e^{2n - \varepsilon^8 n^2}$.

Similarly, we deduce that $\deg_{G'}(v_i; V_{3-i}) \geq (d - \varepsilon)n$ for each $i \in [2]$ and $v_i \in V_i$ with probability at least $1 - 4ne^{-\varepsilon^8 n}$. Note that this is automatically true for all $v \in Y$ as these vertices retain their neighbours from G. Hence, taking another union bound, it follows that (V_1, V_2) is whp $(4\varepsilon, d, d - \varepsilon)$-super-regular in G'. Therefore, for all large enough n, there is a suitable choice for E'. □

2.3 Entropy

In this section we explain basic definitions and properties related to the entropy function, which will play a central rôle in Chapter 4. We will be following the notes of Galvin [72] and all proofs we do not include here can be found or follow immediately from the results there. Throughout this subsection we fix a finite probability space $(\Omega, \mathbb{P})$. Recall also that log denotes the natural logarithm function.

Let $X : \Omega \to S$ be a random variable, and note that we will sometimes use the notation $X(\omega)$, which is an element of S, for the value of X given the outcome $\omega \in \Omega$. Given $x \in S$, we denote $p(x) := \mathbb{P}[X = x]$. We define the *entropy* of X by

$$h(X) := \sum_{x \in S} -p(x) \log p(x).$$

Entropy can be interpreted as a measure of the "uncertainty" of a random variable, or of how much information is "gained" by revealing X. The following lemma shows that the entropy is maximised when X is uniform, corresponding to maximal "uncertainty". Define the *range* of X as the set of values that X takes with positive probability, that is $\mathrm{rg}(X) = \{x \in S : p(x) > 0\}$.

Lemma 2.3.1 (maximal entropy)**.** *For every random variable $X : \Omega \to S$, we have $h(X) \leq \log(|\mathrm{rg}(X)|) \leq \log(|S|)$ with equality if and only if $p(x) = \frac{1}{|S|}$ for all $x \in S$.*

Lemma 2.3.1 provides the key to using entropy in combinatorial arguments. Indeed, the basic method relies on taking a uniformly random object F from some family $\mathcal{F}$ whose cardinality we are interested in estimating. By analysing the entropy of the random variable F, using the tools listed below, we can obtain bounds on the entropy which translate to bounds on the size of $\mathcal{F}$ via Lemma 2.3.1. We now further develop the theory.

Given random variables $X_i : \Omega \to S_i$ for $i \in [n]$, we denote the entropy of the random vector $(X_1, \ldots, X_n)$ by $h(X_1, \ldots, X_n) := h((X_1, \ldots, X_n))$. The entropy function has the following subadditivity property.

Lemma 2.3.2 (subadditivity)**.** *Given random variables $X_i : \Omega \to S_i$, $i \in [n]$, we have*

$$h(X_1, \dots, X_n) \le \sum_{i=1}^{n} h(X_i),$$

with equality if and only if the X_i are mutually independent.

Intuitively, this means that revealing a random vector cannot give us more information than revealing each component separately. We say a random variable $X : \Omega \to S_X$ *determines* another random variable $Y : \Omega \to S_Y$ if the outcome of Y is completely determined by X. For example if X is the outcome of rolling a regular six-sided die and Y is 1 if this outcome is even, and 0 otherwise, then X determines Y. Formally, X *determines* Y if there is a function $f : S_X \to S_Y$ such that $Y(\omega) = f(X(\omega))$ for all $\omega \in \Omega$. If X determines Y, then no additional information is needed to reveal Y once X is revealed. This is formalised in the following lemma.

Lemma 2.3.3 (redundancy)**.** *If $X : \Omega \to S_X$ and $Y : \Omega \to S_Y$ are random variables and X determines Y, then $h(X) = h(X, Y)$.*

If $E \subset \Omega$ is an event with positive probability, we define the *conditional entropy* given the event as

$$h(X|E) := \sum_{x \in S} -p(x|E) \log p(x|E),$$

where $p(x|E) = \mathbb{P}[X = x|E]$. Note that $h(X|E)$ is the entropy of the random variable obtained from X by conditioning on E, so that if Z has distribution $\mathbb{P}[Z = x] = \mathbb{P}[X = x|E]$ then $h(Z) = h(X|E)$. Given two random variables $X : \Omega \to S_X$ and $Y : \Omega \to S_Y$, the conditional entropy of X given Y is defined as

$$h(X|Y) := \mathbb{E}_Y[h(X|Y = y)] = \sum_{y \in S_Y} p(y) h(X|Y = y) \tag{2.3.1}$$

$$= \sum_{\omega \in \Omega} \mathbb{P}[\omega]\, h(X|Y = Y(\omega)), \tag{2.3.2}$$

where $p(y) = \mathbb{P}[Y = y]$. As conditioning on an event or another random variable only gives us more information, we have the following inequalities.

Lemma 2.3.4 (dropping conditioning)**.** *Given random variables $X : \Omega \to S_X$ and $Y : \Omega \to S_Y$, and an event $E \subset \Omega$ we have*

$$h(X|Y) \le h(X) \quad \textit{and} \quad h(X) \ge \mathbb{P}[E]\, h(X|E).$$

Furthermore, if $Y' : \Omega \to S_{Y'}$ is another random variable and Y determines Y', then

$$h(X|Y) \le h(X|Y').$$

The following chain rule strengthens Lemma 2.3.2.

Lemma 2.3.5 (chain rule)**.** *Given random variables $X : \Omega \to S_X$ and $Y : \Omega \to S_Y$, we have*

$$h(X,Y) = h(X) + h(Y|X)$$

and more generally, for random variables $X_i : \Omega \to S_i$, $i \in [n]$, we have

$$h(X_1, \dots, X_n) = \sum_{i=1}^{n} h(X_i | X_1, \dots, X_{i-1}).$$

Lemmas 2.3.1, 2.3.2 and 2.3.5 have the following conditional versions. Given a random variable $X : \Omega \to S_X$ and an event $E \subset \Omega$, we define the conditional range of X given E by $\mathrm{rg}(X|E) = \{x \in S_X : p(x|E) > 0\}$.

Lemma 2.3.6 (maximal conditional entropy)**.** *For every random variable $X : \Omega \to S$ and event $E \subset \Omega$, we have*

$$h(X|E) \le \log\left(|\mathrm{rg}(X|E)|\right).$$

Lemma 2.3.7 (conditional subadditivity)**.** *Given random variables $X_i : \Omega \to S_i$, $i \in [n]$, and $Y : \Omega \to S_Y$, we have*

$$h(X_1, \dots, X_n | Y) \le \sum_{i=1}^{n} h(X_i | Y),$$

with equality if and only if the X_i are mutually independent conditioned on Y.

Lemma 2.3.8 (conditional chain rule)**.** *Given random variables $X_i : \Omega \to S_i$, $i \in [n]$, and $Y : \Omega \to S_Y$, we have*

$$h(X_1, \dots, X_n | Y) = \sum_{i=1}^{n} h(X_i | X_1, \dots, X_{i-1}, Y).$$

The following lemma appears in [5] and will play an essential rôle in the main proof of Chapter 4. It sharpens a similar lemma that appeared in [96]. It states that if a random variable has almost maximal entropy, then it must be close to uniform. This can be seen as a stability result for Lemma 2.3.1.

Lemma 2.3.9 (almost maximal entropy)**.** *For all $\beta > 0$, there is some $\beta' > 0$ such that the following is true for every finite set S and every random variable $X : \Omega \to S$. If $h(X) \ge \log(|S|) - \beta'$, then letting $a := \frac{1}{|S|}$ and $J := \{x \in S : (1-\beta)a \le \mathbb{P}[X = x] \le (1+\beta)a\}$, we have that*

$$|J| \ge (1-\beta)|S| \quad \text{and} \quad \mathbb{P}[X \in J] \ge (1-\beta). \tag{2.3.3}$$

Proof. Let $\beta > 0$ be given and assume that $\beta < \frac{1}{10}$. Fix $\beta' = \frac{\beta^4}{2000}$. Let $X : \Omega \to S$ be a random variable with $h(X) \ge \log(|S|) - \beta'$ and let a and J be as defined in the statement of the lemma.

Further, we define $J^+ = \{y \in S : \mathbb{P}[X = y] > \left(1+\frac{\beta}{4}\right)a\}$ and $J^- = \{y \in S : \mathbb{P}[X = y] < \left(1-\frac{\beta}{4}\right)a\}$. Note that $|J| \geq |S| - (|J^+| + |J^-|)$.

Claim 2.3.10. *We have $|J^+| \leq \frac{\beta}{4}|S|$.*

Proof of Claim: Choose $\eta \leq \frac{\beta}{4}$ so that $\eta|S| = \lfloor \frac{\beta}{4}|S| \rfloor$. Assume for contradiction that $|J^+| > \eta|S|$ and let $\tilde{J}^+ \subset J^+$ be a set of size exactly $\eta|S|$. Define X^+ by

$$\mathbb{P}\left[X^+ = y\right] = \begin{cases} (1+\eta)a & \text{if } y \in \tilde{J}^+ \\ (1-\xi)a & \text{if } y \notin \tilde{J}^+, \end{cases}$$

where $\xi := \frac{\eta^2}{1-\eta}$ is chosen so that $\sum_{y \in S} \mathbb{P}[X^+ = y] = 1$. Now it follows from Karamata's inequality and the fact that $-x\log(x)$ is concave on $[0,1]$, that $h(X^+) \geq h(X)$. We further let $Y = 1$ if $X^+ \in \tilde{J}^+$ and 0 otherwise. We then have that

$$h(X) \leq h(X^+) = h(X^+, Y) = h(X^+|Y=1)\mathbb{P}[Y=1] + h(X^+|Y=0)\mathbb{P}[Y=0] + h(Y),$$

where we used Lemma 2.3.3, the chain rule (Lemma 2.3.5) and the definition of conditional entropy. Note that $\mathbb{P}[Y=1] = \eta(1+\eta)$ and

$$h(Y) = -\eta(1+\eta)\log(\eta(1+\eta)) - (1-\eta(1+\eta))\log(1-(\eta(1+\eta))).$$

Therefore, using also Lemma 2.3.6, we get

$$\begin{aligned} h(X) &\leq \log(\eta|S|)\,\eta(1+\eta) + \log((1-\eta)|S|)\,(1-\eta(1+\eta)) + h(Y) \\ &= \log(|S|) + \log(\eta)\eta(1+\eta) + \log(1-\eta)(1-\eta(1+\eta)) + h(Y) \\ &= \log(|S|) + \eta(1+\eta)\left(\log(\eta) - \log(\eta(1+\eta))\right) \\ &\qquad + (1-\eta(1+\eta))\left(\log(1-\eta) - \log(1-\eta(1+\eta))\right) \\ &= \log(|S|) - \eta(1+\eta)\log(1+\eta) + \left(1-\eta-\eta^2\right)\log\left(\frac{1-\eta}{1-\eta-\eta^2}\right). \end{aligned}$$

Using the approximation $x - \frac{x^2}{2} \leq \log(1+x) \leq x$, which holds for all $x \in (0,1)$, in the forms $\log(1+\eta) \geq \eta\left(1-\frac{\eta}{2}\right)$ and $\log\left(\frac{1-\eta}{1-\eta-\eta^2}\right) = \log\left(1 + \frac{\eta^2}{1-\eta-\eta^2}\right) \leq \frac{\eta^2}{1-\eta-\eta^2}$, we conclude

$$\begin{aligned} h(X) &\leq \log(|S|) - \eta^2(1+\eta)\left(1-\frac{\eta}{2}\right) + (1-\eta-\eta^2)\frac{\eta^2}{1-\eta-\eta^2} \\ &= \log(|S|) - \eta^2 - \frac{\eta^3}{2} + \frac{\eta^4}{2} + \eta^2 \leq \log(|S|) - \frac{\eta^3}{4} < \log(|S|) - \beta', \end{aligned}$$

a contradiction. ∎

Similarly, we can show that $|J^-| \leq \frac{\beta}{4}|S|$ and conclude that $|J| \geq |S| - (|J^+| + |J^-|) \geq (1-\beta)|S|$. Furthermore, by the definition of J^- we have

$$\sum_{y \in J} \mathbb{P}\left[X = y\right] \geq \sum_{y \in S \setminus (J^+ \cup J^-)} \left(1 - \frac{\beta}{4}\right) a \geq \left(1 - \frac{\beta}{2}\right) |S| \left(1 - \frac{\beta}{4}\right) a \geq (1-\beta).$$

This completes the proof. □

2.4 Supersaturation

The following phenomenon was first noticed by Erdős and Simonovits in their seminal paper [64]. It states that if there are many copies of a given small subgraph in some host graph, then we can also find many copies of a blow-up in the host graph (recall the definition of a graph blowup from the Notation Section). It can be proven easily, e.g. by induction.

Lemma 2.4.1. *Let $r, m_1, m_2, \ldots, m_r \in \mathbb{N}$, let J be some graph on r vertices $\{v_1, \ldots, v_r\}$ and $c > 0$. Then there exists $c' = c'(r, m_1, m_2, \ldots, m_r, c) > 0$ such that the following holds. Suppose G is a graph on n vertices with n sufficiently large such that there are subsets $V_1, \ldots, V_r \subseteq V(G)$ and G contains at least cn^r labelled copies of J with $v_i \in V_i$ for $i \in [r]$. Then G contains at least $c'n^{m_1+\ldots+m_r}$ labelled copies of $J_{m_1,m_2,\ldots,m_r}$ with parts $P_1, \ldots, P_r$ such that $P_i \subset V_i$ and $|P_i| = m_i$ for all $i \in [r]$.*

2.5 Matchings and almost factors in dense graphs

The Hajnal–Szemerédi theorem (Theorem 1.1.2) discussed in the introduction is in fact a corollary to a more general theorem from which we can conclude the existence of large K_k-matchings when the minimum degree is slightly less than the extremal threshold for K_k-factors. Indeed Hajnal and Szemerédi [78] proved that any graph with maximum degree Δ has an *equitable colouring* with $\Delta + 1$ colours, that is, a colouring where the colour classes differ in size by at most one. Applying this to the complement of G, which has maximum degree $n - 1 - \delta(G)$, we find a collection of $n - \delta(G)$ vertex-disjoint cliques in G whose sizes differ by at most one and that cover $V(G)$. We will make use of the following general corollary, which we obtain from the fact that when $\delta(G) = \left(\frac{k-1}{k} - x\right)n$ for some $0 \leq x < 1$, then the Hajnal–Szemerédi theorem [78] on equitable colourings provides us with $\left(\frac{1}{k} + x\right)n$ vertex-disjoint cliques. If $0 < x < \frac{1}{k(k-1)}$, some of these cliques, say α, are of size k, and the others are of size $k-1$, hence we have $n = \alpha k + \left(\left(\frac{1}{k} + x\right)n - \alpha\right)(k-1) = \alpha + \frac{n}{k}(1 + kx)(k-1)$. Solving this for α gives the following result.

Theorem 2.5.1 (Hajnal– Szemerédi theorem for K_k-matchings [78]). *Let $n, k \geq 2$ be integers and let $0 \leq x < 1$. Suppose that G is an n-vertex graph with $\delta(G) \geq \left(\frac{k-1}{k} - x\right)n$. Then G contains a K_k-matching of size at least $(1 - (k-1)kx)\lfloor \frac{n}{k} \rfloor$.*

Note that in particular the case $x = 0$ gives the Hajnal–Szemerédi theorem (Theorem 1.1.2) for factors discussed in Section 1.1.

We will also be interested in almost H-factors for general graphs H. Let $\chi(H)$ be the *chromatic number* of a graph H, that is, the minimum number of colours needed to properly colour the vertices of H. Further, let $\chi_{cr}(H) := (\chi(H) - 1)\frac{v_H}{v_H - \sigma(H)}$ where $\sigma(H)$ is the smallest size of a colour class over all colourings of H with $\chi(H)$ colours. The parameter $\chi_{cr}(H)$ is referred to as the *critical chromatic number* of H. The following result of Komlós [110] is a crucial tool in the proof of Theorem III in Chapter 5. It determines the minimum degree threshold for the property of containing an almost H-factor.

Theorem 2.5.2. *For every graph H and every $\alpha > 0$, there exists n_0 such that if G is a graph on $n \geq n_0$ vertices with $\delta(G) \geq \left(1 - \frac{1}{\chi_{cr}(H)}\right)n$, then G contains a partial H-factor that covers all but at most αn vertices of G.*

This was later improved to a constant number of uncovered vertices by Shokoufandeh and Zhao [158], but Komlós' result suffices for our purposes. We will apply Komlós' theorem to find an almost H-factor in a reduced graph R of our (deterministic) graph G from Theorem III; here H will be a carefully chosen auxiliary graph (not K_r!). We discuss this further in Section 5.2.

2.6 Fractional matchings in hypergraphs

Given an r-uniform hypergraph $\mathcal{H}$, a *fractional matching* in $\mathcal{H}$ is a function $f : E(\mathcal{H}) \to \mathbb{R}_{\geq 0}$ such that $\sum_{e : v \in e} f(e) \leq 1$ for all $v \in V(\mathcal{H})$. We say the fractional matching is *perfect* if $\sum_{e : v \in e} f(e) = 1$ for all $v \in V(\mathcal{H})$. The *value* of a fractional matching f is $|f| := \sum_{e \in E(\mathcal{H})} f(e)$. The maximum value $|f|$ over all choices of fractional matching f of $\mathcal{H}$, we call the *fractional matching number of* $\mathcal{H}$, which we denote by $\vartheta^*(\mathcal{H})$.

A *fractional cover* of $\mathcal{H}$ is a function $g : V(\mathcal{H}) \to \mathbb{R}_{\geq 0}$ such that for all $e \in E(\mathcal{H})$, one has $\sum_{v \in e} g(v) \geq 1$. The *value* of a fractional cover g is $|g| := \sum_{v \in V(\mathcal{H})} g(v)$. The *fractional cover number* of $\mathcal{H}$, denoted $\tau^*(\mathcal{H})$ is then the minimum value of a fractional cover g of $\mathcal{H}$.

For an r-uniform hypergraph $\mathcal{H}$, the fractional matching number of $\mathcal{H}$ can be encoded as the optimal solution of a linear program. Taking the dual of this linear program gives another linear program that outputs the fractional cover number as an optimal solution. The strong duality theorem from linear programming thus tells us that $\vartheta^*(\mathcal{H}) = \tau^*(\mathcal{H})$ for any hypergraph $\mathcal{H}$. Using this, one can derive the following simple consequences, see e.g. [118, Proposition 2] or [86, Proposition 2.4].

Proposition 2.6.1. *For any r-uniform hypergraph $\mathcal{H}$ on N vertices, the following hold.*

1. *$\vartheta^*(\mathcal{H}) \leq \frac{N}{r}$ with equality if and only if there exists a perfect fractional matching in $\mathcal{H}$.*
2. *$\vartheta^*(\mathcal{H}) \geq \vartheta(\mathcal{H})$ where $\vartheta(\mathcal{H})$ denotes the size of the largest matching in $\mathcal{H}$.*
3. *If $g : V(\mathcal{H}) \to \mathbb{R}_{\geq 0}$ is a fractional cover and $U \subset V(\mathcal{H})$, then $g' := g|_U : U \to \mathbb{R}_{\geq 0}$ is a fractional cover of $\mathcal{H}[U]$ and hence $|g'| = \sum_{u \in U} g(u) \geq \tau^*(\mathcal{H}[U]) = \vartheta^*(\mathcal{H}[U])$.*

Moreover, it follows from the strong duality theorem for linear programs (and the proof of weak duality) that for an optimal solution, if a variable in the dual program is positive, then its corresponding constraint in the primal program must be tight. These are what are known as the 'complementary slackness conditions'. Using them, one can derive the following further proposition, see e.g. [118, Proposition 2] or [86, Proposition 2.4].

Proposition 2.6.2. *If $g : V(\mathcal{H}) \to \mathbb{R}_{\geq 0}$ is an optimal fractional cover, i.e. $|g| = \tau^*(\mathcal{H})$, then $\vartheta^*(\mathcal{H}) \geq \frac{|W|}{r}$ where $W := \{v \in V(\mathcal{H}) : g(v) > 0\}$.*

We now explore some simple conditions that guarantee the existence of a perfect fractional matching. Given a vertex subset $U \subset V := V(\mathcal{H})$ in a hypergraph $\mathcal{H}$, a *fan focused at U* in $\mathcal{H}$ is a subset $F \subset E(\mathcal{H})$ of edges of $\mathcal{H}$ such that $|e \cap U| = 1$ for all $e \in F$ and $e \cap e' \cap (V \setminus U) = \emptyset$ for all $e \neq e' \in F$. In words, each edge of a fan intersects U in exactly one vertex and outside of U, the edges in a fan are pairwise disjoint. The *size* of a fan is simply the number of edges in the fan. If $U = \{u\}$ is a single vertex, we simply refer to a fan focused at u. The following two lemmas give simple conditions that guarantee a perfect fractional matching. Their proofs follow the method of Krivelevich, Sudakov and Szabó [127], see also [85, 86]. We include the proofs here for completeness. We also remark that both Lemma 2.6.3 and the following Lemma 2.6.4 are provided in a form that eases their application. There is no attempt to optimise the conditions of the statements (for example the upper bound on M in Lemma 2.6.3) and both results hold under looser constraints on the parameters.

Lemma 2.6.3. *Suppose $\mathcal{H}$ is an N-vertex, r-uniform hypergraph and there exists $r \leq M \leq \frac{N}{2r}$ such that given any vertex $v \in V(\mathcal{H})$ and any subset $W \subseteq V(\mathcal{H}) \setminus \{v\}$ of at least M vertices, there exists an edge in $\mathcal{H}$ containing v and $r - 1$ vertices of W. Then $\mathcal{H}$ has a perfect fractional matching.*

Proof. Firstly we note that the condition on $\mathcal{H}$ implies the following two consequences. We have that

(i) for all $v \in V(\mathcal{H})$ there is a fan focused at v in $\mathcal{H}$ of size $2M$, and

(ii) every subset of at least $2M$ vertices of $\mathcal{H}$ induce an edge in $\mathcal{H}$.

Indeed, for each vertex $v \in V(\mathcal{H})$ we can greedily construct such a fan F_v. Whilst F_v has size less than $2M$, taking $W = V(\mathcal{H}) \setminus V(F_v)$, we have that $|W| \geq N - -12M(r-1) \geq M$ (using here that $M \leq \frac{N}{2r}$) and so we can find an edge containing v and $r-1$ vertices of W which extends the fan. Condition (ii) also holds because given a vertex subset W' of at least $2M$ vertices, for any $w \in W'$, we have that there is an edge containing w and $r-1$ vertices of $W' \setminus \{w\}$.

Now suppose for a contradiction that $\mathcal{H}$ does *not* have a perfect fractional matching. Thus, by Proposition 2.6.1 (1), if we take $g : V(\mathcal{H}) \to \mathbb{R}_{\geq 0}$ to be an optimal fractional cover of $\mathcal{H}$ so that $|g| = \vartheta^*(\mathcal{H}) = \tau^*(\mathcal{H})$ we have that $|g| < \frac{N}{r}$. Hence, by Proposition 2.6.2, there exists some vertex w such that $g(w) = 0$. Let F_w be a fan focused at w of size $2M$, whose existence is guaranteed by (i), and define $U \subset V(\mathcal{H})$ to be $U := \bigcup\{e \setminus \{w\} : e \in F_w\}$. So $|U| = 2(r-1)M$ and using that $\sum_{v \in e} g(v) \geq 1$ for each edge $e \in F_w$ and the fact that $g(w) = 0$, we can conclude that $\sum_{u \in U} g(u) \geq 2M$.

Now consider $V' = V(\mathcal{H}) \setminus U$. We have that $|V'| = N - 2(r-1)M$ and

$$\vartheta(\mathcal{H}[V']) \geq \frac{|V'| - 2M}{r} \geq \frac{N}{r} - 2M,$$

as we can greedily build a matching of this size by repeatedly appealing to (ii). Using Proposition 2.6.1 (2) and (3) we have that $\sum_{v \in V'} g(v) \geq \frac{N}{r} - 2M$ which gives a contradiction as

$$|g| = \sum_{u \in U} g(u) + \sum_{v \in V'} g(v) \geq 2M + \left(\frac{N}{r} - 2M\right) = \frac{N}{r}.$$

□

The proof of Lemma 2.6.3 shows that if $\mathcal{H}$ has large fans focused at each vertex (and no large independent sets) then it must have a perfect fractional matching. In fact, it is not necessary that the vertex fans be so large if we have an additional expansion property. This is the content of the following lemma.

Lemma 2.6.4. *Suppose $\mathcal{H}$ is an N-vertex, r-uniform hypergraph and there exists $r \leq M_1 \leq M_2 \leq \frac{N}{2r}$ such that the following hold:*

(i) For all $v \in V(\mathcal{H})$ there is a fan focused at v in $\mathcal{H}$ of size M_1.

(ii) For every subset $W_0 \subset V(G)$ with $|W_0| = M_1$ and every subset $W_1 \subset V(G) \setminus W_0$ with $|W_1| \geq M_2$, there exists an edge of $\mathcal{H}$ with one vertex in W_0 and the other $r-1$ vertices in W_1.

Then $\mathcal{H}$ has a perfect fractional matching.

Proof. We start by noticing that (ii) leads to the following two consequences.

(a) For all $U \subset V(\mathcal{H})$ with $|U| = (r-1)M_1$, fixing $V' := V(\mathcal{H}) \setminus U$ we have that for all $U' \subset V'$ such that $|U'| = M_1$, there is a fan of size $2M_2 - M_1$ focused at U' in $\mathcal{H}[V']$.

(b) Every subset of at least $2M_2$ vertices of $\mathcal{H}$ induce an edge in $\mathcal{H}$.

Indeed, for U' as in (a) we can build the fan $F_{U'}$ focused at U' greedily. Whilst $|F_{U'}| \leq 2M_2 - M_1$ we have that $W := V(G) \setminus (V(F_{U'}) \cup U' \cup U)$ has size at least

$$N - (2M_2 - M_1)(r-1) - M_1 - (r-1)M_1 \geq N - (2r-1)M_2 \geq M_2,$$

as $M_2 \leq \frac{N}{2r}$. Hence we can find an edge using one vertex of U' and $r-1$ vertices of W, which extends the fan $F_{U'}$. The condition (b) also follows easily as taking W' to be a set with $2M_2$ vertices, we have that for any $W'' \subset W'$ with $|W''| = M_1$, there is an edge containing a vertex in W'' and $r-1$ vertices of $W' \setminus W''$ from (ii).

Now we turn to the main proof which is very similar to that of Lemma 2.6.3. We again fix $g : V(\mathcal{H}) \to \mathbb{R}_{\geq 0}$ to be an optimal fractional cover and suppose for a contradiction that $|g| < \frac{N}{r}$. We deduce the existence of a vertex $w \in V(\mathcal{H})$ with $g(w) = 0$ and a fan F_w focused at w of size M_1. Taking $U_1 := \bigcup\{e \setminus \{w\} : e \in F_w\}$, we have that $|U_1| = (r-1)M_1$ and $\sum_{u \in U_1} g(u) \geq M_1$.

Now consider $V' := V(\mathcal{H}) \setminus U_1$. If $\vartheta^*(\mathcal{H}[V']) \geq \frac{N}{r} - M_1$ then we can conclude that $\sum_{v \in V'} g(v) \geq \frac{N}{r} - M_1$ from Proposition 2.6.1 (3) which implies that $|g| \geq \frac{N}{r}$, a contradiction. Hence

$$\vartheta^*(\mathcal{H}[V']) < \frac{N}{r} - M_1 = \frac{N' - M_1}{r}, \tag{2.6.1}$$

where $N' := |V'| = N - (r-1)M_1$. We fix $g' : V' \to \mathbb{R}_{\geq 0}$ to be some optimal fractional cover of $\mathcal{H}[V']$ with $|g'| = \vartheta^*(\mathcal{H}[V'])$. By Proposition 2.6.2, we therefore have that there is some set $U_2 \subset V'$ with $|U_2| = M_1$ and $g'(u') = 0$ for all $u' \in U_2$. By (a) there exists a fan F_{U_2} of size $2M_2 - M_1$ focused at U_2 in $\mathcal{H}[V']$. Taking $Z := \bigcup\{e : e \in F_{U_2}\} \setminus U_2$, we have that $|Z| = (r-1)(2M_2 - M_1)$ and similarly to before, using that for each edge $e \in F_{U_2}$ we have $\sum_{v \in e} g'(v) \geq 1$ and the fact that $g'(u') = 0$ for all $u' \in U_2$, we can conclude that $\sum_{z \in Z} g'(z) \geq |F_{U_2}| = 2M_2 - M_1$.

Finally, we look at $V'' := V' \setminus Z$. We have that $N'' := |V''| = N' - 2(r-1)M_2 + (r-1)M_1$ and using (b) and Proposition 2.6.1 (2), we have that

$$\vartheta^*(\mathcal{H}[V'']) \geq \frac{N'' - 2M_2}{r} = \frac{N' + (r-1)M_1}{r} - 2M_2.$$

Hence, by Proposition 2.6.1 (3), we can conclude that $\sum_{v''\in V''} g'(v'') \geq \frac{N'+(r-1)M_1}{r} - 2M_2$. Combining this with the lower bound on the sum of g' values on Z implies that $|g'| = \vartheta^*(\mathcal{H}[V']) \geq \frac{N'-M_1}{r}$, contradicting (2.6.1). □

2.7 Almost perfect matchings in hypergraphs

It is well-known that hypergraphs that have roughly regular vertex degrees and small codegrees contain large matchings. This is often referred to as Pippenger's Theorem but there are in fact a family of similar results, all following from the "semi-random" or "nibble" method, see e.g. [12, Section 4.7]. Here we use the following explicit version which follows directly from a result of Kostochka and Rödl [117].

Theorem 2.7.1. *For any integers $r \geq 3$ and $K \geq 4$ there exists $\Delta_0 > 0$ such that for all $\Delta \geq \Delta_0$ the following holds. If $\mathcal{H}$ is an r-uniform hypergraph on N vertices such that:*

1. *for all vertices $v \in V(\mathcal{H})$, we have $\deg^{\mathcal{H}}(v) = \Delta\left(1 \pm K\sqrt{\frac{\log\Delta}{\Delta}}\right)$ and*
2. *for all $u \neq v \in V(\mathcal{H})$, we have $\operatorname{codeg}^{\mathcal{H}}(u,v) \leq \Delta^{1/(2r-1)}$,*

then $\mathcal{H}$ has a matching covering all but at most $\Delta^{-1/r}N$ vertices.

Indeed, [117, Theorem 4] states that for all $r \geq 3$, $K_0 \geq 8$ and reals $0 < \delta, \gamma < 1$, there exists a D_0 such that if $\mathcal{H}$ is an r-uniform hypergraph on N vertices with

$$D - K_0\sqrt{D\log D} \leq \deg^{\mathcal{H}}(v) \leq D,$$

for all $v \in V(\mathcal{H})$, where $D \geq D_0$, and $\operatorname{codeg}^{\mathcal{H}}(u,v) \leq C < D^{1-\gamma}$ for all pairs of vertices $u \neq v$, then $\mathcal{H}$ has a matching covering all but at most $O\left(N\left(\frac{C}{D}\right)^{(1-\delta)/(r-1)}\right)$ vertices. In order to derive Theorem 2.7.1 from this we fix $K_0 = 2K$, $\delta = \frac{1}{4r}$ and $\gamma = \frac{2r-2}{2r-1}$. Letting D_0 be the resulting constant given by [117, Theorem 4], we fix $\Delta_0 \geq D_0$ to be some large constant. Hence, our conditions (1) and (2) of Theorem 2.7.1 guarantee that $\mathcal{H}$ satisfies the conditions of [117, Theorem 4] with $D = \Delta + K\sqrt{\Delta\log\Delta}$ and $C = \Delta^{1-\gamma}$. Now note that

$$\frac{C}{D} = (1+o(1))\Delta^{-\gamma} = (1+o(1))\Delta^{-(2r-2)/(2r-1)} = o(\Delta^{-(4r-4))/(4r-1)}).$$

Combined with the fact that $\frac{1-\delta}{r-1} = \frac{4r-1}{4r(r-1)}$ and $\Delta \geq \Delta_0$ is sufficiently large, it follows from [117, Theorem 4] that the number of vertices uncovered by a largest matching is always less than $\Delta^{-1/r}N$, as required.

Clearly, in order to prove that a hypergraph $\mathcal{H}$ has a large matching, it suffices to establish the conditions of Theorem 2.7.1 for a spanning subgraph $\mathcal{H}' \subset \mathcal{H}$. An idea introduced by Alon, Frankl, Huang, Rödl, Ruciński and Sudakov [10] is to find such an $\mathcal{H}'$ as a random subhypergraph of $\mathcal{H}$ and guarantee that the conditions of Theorem 2.7.1 hold for $\mathcal{H}'$ by using perfect fractional matchings to dictate the probability with which we take each edge into $\mathcal{H}'$. This idea was then used in the context of finding almost K_r-factors in pseudorandom graphs by Han, Kohayakawa and Person [85, 86]. We will also adopt this idea and so give the following theorem.

Theorem 2.7.2. *For all $3 \leq r \in \mathbb{N}$ and $0 < \eta < \frac{1}{2}$, there exists an N_0 such that the following holds for all $N \geq N_0$. Suppose $\mathcal{H}$ is an N-vertex, r-uniform hypergraph such that there exists $t := 2N^\eta$ perfect fractional matchings $f_1, \dots, f_t : E(\mathcal{H}) \to \mathbb{R}_{\geq 0}$ in $\mathcal{H}$ with the property that*

$$\sum_{i=1}^{t} \sum_{e \in E(\mathcal{H}) : \{u,v\} \subset e} f_i(e) \leq 2, \tag{2.7.1}$$

for all pairs of vertices $u \neq v \in V(\mathcal{H})$. Then $\mathcal{H}$ has a matching covering all but at most $N^{1-\eta/r}$ vertices.

Proof. We take a random subgraph $\mathcal{H}' \subseteq \mathcal{H}$ be keeping every edge $e \in E(\mathcal{H})$ independently with probability $p_e = \sum_{i=1}^{t} \frac{f_i(e)}{2}$ noting that $p_e \in [0, 1]$ for all $e \in E(\mathcal{H})$ due to (2.7.1). We fix $\Delta := \frac{t}{2} = N^\eta$ and $K := \frac{4}{\eta}$ and claim that $\mathcal{H}'$ satisfies the conditions of Theorem 2.7.1 whp as N tends to infinity.

To check that $\mathcal{H}'$ satisfies the conditions of Theorem 2.7.1, note that for each $v \in V$ we have

$$\mathbb{E}\left[\deg^{\mathcal{H}'}(v)\right] = \sum_{e : v \in e} p_e = \sum_{e : v \in e} \sum_{i=1}^{t} \frac{f_i(e)}{2} = \frac{1}{2} \sum_{i=1}^{t} \left(\sum_{e : v \in e} f_i(e) \right) = \frac{t}{2} = \Delta,$$

using that each f_i is a perfect fractional matching. Applying Theorem 2.1.1 then gives that

$$\begin{aligned} \mathbb{P}\left[\deg^{\mathcal{H}'}(v) \neq \Delta\left(1 \pm K\sqrt{\frac{\log \Delta}{\Delta}}\right)\right] &\leq 2\exp\left(-\frac{K^2 \log \Delta}{3}\right) \\ &\leq 2\exp\left(-\frac{K^2 \eta \log N}{3}\right) \\ &\leq \frac{1}{N^2}, \end{aligned} \tag{2.7.2}$$

for N sufficiently large. Similarly, for $u \neq v \in V(\mathcal{H})$, we have that $\mathbb{E}\left[\operatorname{codeg}^{\mathcal{H}'}(u,v)\right] = \sum_{e : \{u,v\} \subset e} p_e \leq 1$ by (2.7.1) and applying Theorem 2.1.1 gives that

$$\mathbb{P}\left[\operatorname{codeg}^{\mathcal{H}'}(u,v) \geq \Delta^{1/(2r-1)}\right] \leq \exp\left(-\Delta^{1/(2r-1)}\right) \leq \frac{1}{N^3}, \tag{2.7.3}$$

for large N. Hence taking a union bound over all vertices and pairs of vertices and upper bounding the failure probabilities with (2.7.2) and (2.7.3) gives that $\mathcal{H}'$ satisfies the conditions of Theorem 2.7.1 whp. Therefore for N (and hence Δ) sufficiently large, we can fix such an instance of $\mathcal{H}'$ and apply Theorem 2.7.1 which gives the large matching in $\mathcal{H}'$ and hence in $\mathcal{H}$, concluding the proof. □

It will be useful for us to work with the following corollary to Theorem 2.7.2 which gives us a sufficient condition for us to be able to generate the perfect fractional matchings needed in Theorem 2.7.2 via a greedy process. Recall that for a 2-uniform graph J on $V(\mathcal{H})$, $\mathcal{H}_J$ denotes the subhypergraph of $\mathcal{H}$ given by all edges of $\mathcal{H}$ which contain some edge of J.

Theorem 2.7.3. *For all $3 \le r \in \mathbb{N}$ and $0 < \gamma < \frac{1}{2r^2}$, there exists an N_0 such that the following holds for all $N \ge N_0$. Suppose $\mathcal{H}$ is an N-vertex, r-uniform hypergraph such that given any* graph *J on $V(\mathcal{H})$ of maximum degree at most $N^{r^2\gamma}$, we have that $\mathcal{H} \setminus \mathcal{H}_J$ contains a perfect fractional matching. Then $\mathcal{H}$ has a matching covering all but at most $N^{1-\gamma}$ vertices.*

Proof. We will prove this by appealing to Theorem 2.7.2 with $\eta := r\gamma$ and so we set out to find $t := 2N^\eta$ perfect fractional matchings $f_1, \dots, f_t$ such that (2.7.1) holds. We do this algorithmically, finding the f_i one at a time. We begin by defining J_1 to be the empty (2-uniform) graph on $V(\mathcal{H})$ and for $1 \le i \le t$ we do the following. We find a perfect fractional matching f_i in $\mathcal{H} \setminus \mathcal{H}_{J_i}$ and add this to our family of perfect fractional matchings. We then define a graph G_i with vertex set $V(\mathcal{H})$ and a pair of vertices $\rho \in \binom{V(\mathcal{H})}{2}$ forming an edge in G_i if

$$\sum_{\rho \subset e \in E(\mathcal{H})} f_i(e) \ge \frac{N^{-\eta}}{2}.$$

Finally we define $J_{i+1} := J_i \cup G_i$ and move to step $i + 1$.

We claim that this algorithm does not stall and we complete our collection of t perfect fractional matchings. In order to check this, we need to verify that we can find a perfect fractional matching in $\mathcal{H} \setminus \mathcal{H}_{J_j}$ for each $j \in [t]$. This follows because at each step i, we have that for any $v \in V(\mathcal{H})$,

$$\sum_{u \in V(\mathcal{H}) \setminus \{v\}} \left(\sum_{\{u,v\} \subset e \in E(\mathcal{H})} f_i(e) \right) = (r-1) \sum_{v \in e \in E(\mathcal{H})} f_i(e) = r - 1,$$

as f_i is a perfect fractional matching. Hence the number of pairs $\rho \in \binom{V(\mathcal{H})}{2}$ that contain v and form an edge of G_i is at most $2(r-1)N^\eta$. As this holds for all choices of $v \in V(\mathcal{H})$ we have that G_i has maximum degree less than $2(r-1)N^\eta$. Thus for each $j \in [t]$, $J_j := \bigcup_{i=1}^{j-1} G_i$ has maximum degree less than

$$2(r-1)N^\eta (j-1) \le 2(r-1)N^\eta t = 4(r-1)N^{2\eta} \le N^{r\eta} = N^{r^2\gamma},$$

for N sufficiently large. So $\mathcal{H}\setminus\mathcal{H}_{J_i}$ does indeed host a perfect fractional matching by assumption.

Finally we need to check condition (2.7.1) for each pair of vertices $\rho = \{u, v\} \in \binom{V(\mathcal{H})}{2}$. Note that for any pair $\rho \in \binom{V(\mathcal{H})}{2}$ of vertices of $\mathcal{H}$ we have that

$$\sum_{i=1}^{t}\left(\sum_{\rho\subset e\in E(\mathcal{H})} f_i(e)\right) \leq \frac{tN^{-\eta}}{2} = 1,$$

if ρ does not feature as an edge in any of the G_i. On the other hand, we have that if $\rho = \{u, v\} \in E(G_j)$ for some $j \in [t]$, then note that because we forbid the edges of $\mathcal{H}$ containing $\{u, v\}$ from being used again we have that

$$\sum_{i=j+1}^{t}\left(\sum_{\rho\subset e\in E(\mathcal{H})} f_i(e)\right) = 0.$$

Also we have that $\rho \notin E(G_i)$ for $i < j$ as otherwise there could be no weight on (edges containing) ρ in f_j. Hence

$$\sum_{i=1}^{j-1}\left(\sum_{\rho\subset e\in E(\mathcal{H})} f_i(e)\right) \leq \frac{(j-1)N^{-\eta}}{2} \leq 1,$$

and using that

$$\sum_{\rho\subset e\in E(\mathcal{H})} f_j(e) \leq \sum_{u\in e\in E(\mathcal{H})} f_j(e) = 1,$$

gives that (2.7.1) holds for all $\rho \in \binom{V(\mathcal{H})}{2}$ as required. So by Theorem 2.7.2, we have that $\mathcal{H}$ contains a matching covering all but at most $N^{1-\eta/r} = N^{1-\gamma}$ vertices, concluding the proof. □

2.8 The absorption method

We will use the *absorption method*, which has appeared in various guises [60, 119] since the 90s and was widely popularised by Rödl, Ruciński and Szemerédi [154]. The basic idea, which we now sketch, is to split the problem of finding a spanning structure into two tasks: finding some *absorbing structure* and finding an *almost spanning structure* in the host graph G. In more detail, the first task is to prove the existence of an absorbing structure that lies on some *absorbing set* of vertices $A \subset V(G)$ of the n-vertex host graph G. In the context of K_r-factors, this absorbing set will (usually) comprise of a small constant proportion of the vertices of G and the absorbing structure will have a strong absorbing property that implies that for *any* set $L \subset V(G) \setminus A$ of $o(n)$ vertices, if $|L| + |A| \in r\mathbb{N}$, then $G[A \cup L]$ contains a K_r-factor. We then put the absorbing set A to one side and our second task is to find an almost spanning structure which equates to finding an almost K_r-factor in $G[V(G) \setminus A]$. This almost factor will leave some small leftover set $L \subset V(G) \setminus A$ of vertices uncovered, at which point we appeal to the absorbing property of the absorbing structure to complete a K_r-factor.

In recent years, the method has become an extremely important tool for studying the existence of spanning structures in graphs, digraphs and hypergraphs, and novel variations of the method have been developed to overcome certain challenges. Indeed, the method sketched above is quite general (even after restricting the discussion to K_r-factors) and often the challenge in absorbing arguments is to define an absorbing structure in such a way that it has the key absorbing property but can also be found in the setting of interest. Some recent innovations for the use of the method have included the so-called *lattice based absorption* method developed in [81, 82, 103] and the *cascading absorption* method developed in [119, 145], which we will build upon here in our proof of Theorem I (see Section 3.1).

In this section we concentrate on a powerful new approach introduced by Montgomery [136, 137], in his work on spanning trees in random graphs. The general idea is to use the following key notion as an auxiliary graph to define absorbing structures in the host graph of interest.

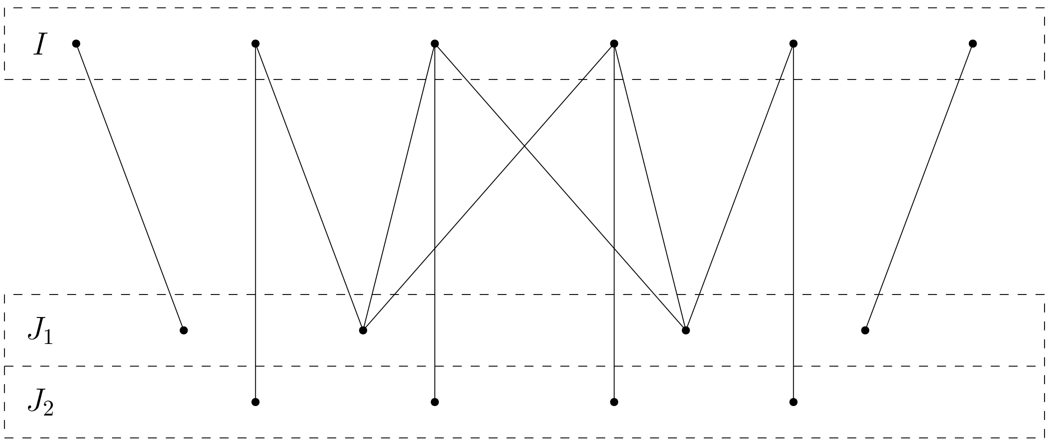


FIGURE 2.1: A template $\mathcal{T}$ of flexibility 2. One can check that the key property is indeed satisfied.

Definition 2.8.1. A *template* $\mathcal{T}$ with *flexibility* $t \in \mathbb{N}$ is a bipartite graph on $7t$ vertices with vertex classes I and $J_1 \cup J_2$, such that $|I| = 3t$, $|J_1| = |J_2| = 2t$, and for any $\bar{J} \subset J_2$, with $|\bar{J}| = t$, the induced graph $\mathcal{T}[V(\mathcal{T}) \setminus \bar{J}]$ has a perfect matching. We call J_2 the *flexible* set of vertices for the template.

See Figure 2.1 for an example of a template. The definition implies that a template is *robust* with respect to having a perfect matching. It is not hard to come up with examples of templates, indeed a complete bipartite graph certainly satisfies the condition. The utility of the notion for defining absorbing structures that are possible to find in the desired host graphs, comes with the fact that *sparse templates* exist. Indeed, Montgomery [136, 137] proved the following using a probabilistic argument.

Theorem 2.8.2. *For all sufficiently large t, there exists a template of flexibility t and maximum degree* 40*.*

Han, Kohayakawa, Person and the author [84] then showed how to derandomise the argument for the existence of templates and find templates with bounded maximum degree efficiently in polynomial time.

We will appeal to the *template absorption* method twice in this thesis, once in our proof of Theorem I (more specifically in proving Proposition 3.1.9) and also in proving the upper bounds of Theorem III. The method has recently found many applications and has played an instrumental rôle in several significant breakthroughs in combinatorics. Indeed, we mention results of Kwan [131], the author [139], and Ferber and Kwan [66] establishing the existence of (families of disjoint) perfect matchings in random Steiner triple systems, the work of Glock, Kühn, Montgomery and Osthus [76] on decompositions of optimally coloured complete graphs into rainbow spanning trees, the proof of Ringel's conjecture by Montgomery, Pokrovskiy and Sudakov [138] and the work of Ferber, Kronenberg and Luh [65] on 2-universality in random graphs.

Chapter 3

Clique factors in pseudorandom graphs

In this chapter we prove Theorem I, which we now restate for convenience, giving a pseudorandom condition for the existence of clique factors in graphs.

Theorem I. *(Restated) For every $3 \leq r \in \mathbb{N}$ and $c > 0$ there exists an $\varepsilon > 0$ such that any n-vertex (p, β)-bijumbled graph with $n \in r\mathbb{N}$, $\delta(G) \geq cpn$ and $\beta \leq \varepsilon p^{r-1} n$, contains a K_r-factor.*

The chapter is organised as follows. In Section 3.1, we discuss the proof in detail, introducing key concepts such as *diamond trees*, *shrinkable orchards* and the *cascading absorption* process. In doing so, we reduce Theorem I to proving two intermediate propositions (Propositions 3.1.8 and 3.1.9) and a lemma (Lemma 3.1.4). In Section 3.2 we then derive some properties of bijumbled graphs. In Section 3.3 we study diamond trees and investigate what kinds of diamond trees we can guarantee in our bijumbled graph. The key result in Section 3.3 is Proposition 3.3.1, which will be crucial at various points in our proof. We then turn, in Section 3.4, to addressing the necessary results for the cascading absorption process which forms a major part of the proof of Theorem I. We prove Lemma 3.1.4 in Section 3.4.1 and discuss Proposition 3.1.8 in Section 3.4.2, reducing it to two intermediate propositions which tackle small and large order shrinkable orchards separately. We go on to prove the existence of shrinkable orchards of small order in Section 3.5 and large order in Section 3.6. Finally we prove Proposition 3.1.9 which provides the final absorption in the proof of Theorem I, in Section 3.7.

3.1 Proof reduction and overview

The proof of Theorem I rests on the shoulders of the previous results [3, 83–86, 127, 145] working towards the conjecture of Krivelevich, Sudakov and Szabó. Indeed it is fair to say that the solution of the conjecture would not have been possible without the insights and ideas of the many authors who tackled this problem. In this section, we discuss these as well as our novel ideas and lay out the key concepts and scheme of the proof. In doing so, we will reduce the theorem to several intermediate results, whose proofs will be the subject of the rest of the chapter.

Our proof, like some of its predecessors [3, 83, 145], works by the method of absorption (see Section 2.8). It turns out that finding an almost K_r-factor in a (p, β)-bijumbled graph G as in Theorem I, is easy. This follows from a simple consequence of Definition 1.4.1 which guarantees that any small linear sized set of vertices contains a copy of K_r, see e.g. Corollary 3.2.5 (2) for a precise statement. Therefore we can greedily choose copies of K_r to be in our K_r-factor and continue this process until we are left with some small leftover set of vertices L, where small here means, of size at most εrn, say. However, at this point we get stuck; we have no way of guaranteeing the existence of a K_r in L and so we do not know how to get a larger set of vertex-disjoint copies of K_r. As usual with absorption, the idea is to put aside an *absorbing set* of vertices which can *absorb* the leftover vertices L into a K_r-factor. That is, before running this greedy process to build a K_r-factor, we find some special set of vertices $X \subset V(G)$ which has the property that for *any* small set of vertices $L \subset V(G) \setminus X$, there is a K_r-factor in $G[X \cup L]$ (provided the trivial divisibility constraint that $r|(|X|+|L|)$). If we can find such an X in G, then we can put it to one side and run the greedy argument to cover almost all the vertices which do not lie in X, with vertex-disjoint copies of K_r. We can then use the absorbing property to *absorb* the leftover vertices L and get a full K_r-factor.

This leaves the challenge of defining some structure in G which has this absorbing property and finding such a structure (on some vertex set X) in G. The building blocks of our absorbing structure will be subgraphs that we call *K_r-diamond trees*. In words, a K_r-diamond tree $\mathcal{D} = (T, R, \Sigma)$ is the graph obtained by taking a tree T and replacing each edge $e \in E(T)$ by a copy of K_{r+1}^- (the $r+1$-vertex clique with a singular edge removed) whose degree $r-1$ vertices are the vertices of e and whose degree r vertices are new and distinct from previous choices, see Figure 3.1 for an example. The following definition formalises this notion.

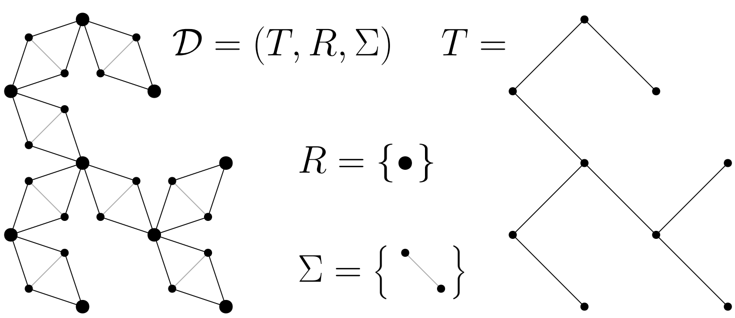


FIGURE 3.1: An example of a K_3-diamond tree $\mathcal{D} = (T, R, \Sigma)$ of order 9 shown on the left. The removable vertices R are the larger vertices of $\mathcal{D}$ and the interior cliques Σ are the edges given in grey. The auxiliary tree T is depicted on the right.

Definition 3.1.1. A *K_r-diamond tree* $\mathcal{D}$ of *order* m in a graph G is a tuple $\mathcal{D} = (T, R, \Sigma)$ where T is an (auxiliary) tree of order m (i.e. with m vertices), $R \subset V(G)$ is a subset of m vertices of G and[1] $\Sigma \subset K_{r-1}(G)$ is a set of $m - 1$ copies of K_{r-1} in G such that the following holds. There are bijective maps $\rho : V(T) \to R$ and $\sigma : E(T) \to \Sigma$ such that:

- Σ is a K_{r-1}-matching in G which is vertex-disjoint from R i.e. $V(S) \cap R = \emptyset$ for all $S \in \Sigma$;
- For all $e = uv \in E(T)$, we have that $V(\sigma(e)) \subseteq N_G(\rho(u)) \cap N_G(\rho(v))$. That is, the $r-1$-clique $\sigma(e) \in K_{r-1}(G)$ can be extended to a copy of K_r in G by adding the vertex $\rho(u)$ and likewise with $\rho(v)$.

We refer to R as the set of *removable vertices* of $\mathcal{D}$ and to Σ as the set of *interior cliques* of $\mathcal{D}$. We define the vertices of $\mathcal{D}$ to be all the removable vertices and the vertices in interior cliques. That is, $V(\mathcal{D}) := (\cup_{S \in \Sigma} V(S)) \cup R$. Finally we define the *leaves* of the diamond tree to be the vertices which are images of leaves in T under ρ.

Note that a K_r-diamond tree of order m has exactly $(m-1)r + 1$ vertices. Krivelevich [119] used K_3-diamond trees in an absorption argument for triangle factors in random graphs which is often cited as one of the first appearances of the absorption method. Nenadov [145] also used this idea in his result that got within a log-factor of Theorem I. The utility of these subgraphs in absorption arguments comes from the following key observation which shows that they can contribute to a K_r-factor in many ways.

Observation 3.1.2. Given a K_r-diamond tree $\mathcal{D} = (T, R, \Sigma)$ in G, we have that for *any* removable vertex $v \in R$, there is a K_r-factor of $G[V(\mathcal{D}) \setminus \{v\}]$. Indeed, consider $u = \rho^{-1}(v)$ in $V(T)$ and the map $\varphi : E(T) \to V(T) \setminus \{u\}$ which maps each edge e of T to the vertex in e which has the larger distance from u in T. Then φ is a bijection and taking the copies of K_r on $\sigma(e) \cup \rho(\varphi(e))$ for each edge $e \in E(T)$ gives the required K_r-factor. See Figure 3.2 for some examples.

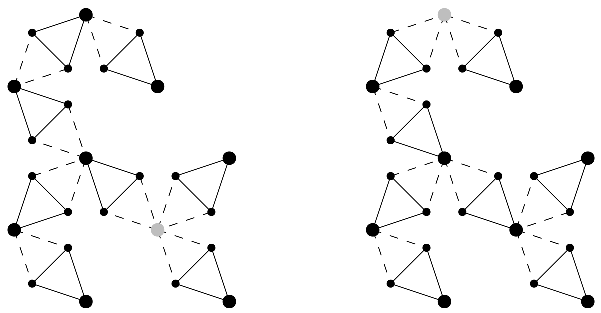


FIGURE 3.2: Some examples of the K_3-factors found after removing a removable vertex from the K_3-diamond tree in Figure 3.1 (see Observation 3.1.2).

[1]Recall that we use the notation $K_{r-1}(G)$ to denote the family of $(r-1)$-cliques in G.

Observation 3.1.2 works for any underlying auxiliary tree T. It turns out that in the (p,β)-bijumbled graphs G we are interested in, one can find K_r-diamond trees of any order up to linear size. Indeed, one can use the argument of Krivelevich [119] to construct these or a different argument due to Nenadov [145]. The method of Nenadov gives diamond trees whose auxiliary tree is a path whilst the argument of Krivelevich gives no control over the underlying auxiliary tree which defines the diamond tree found. As a key part of our argument, we will need to prove the existence of diamond trees which have extra structure, as we discuss shortly.

In order to utilise the absorbing power of diamond trees, we need to group them together in collections. The following definition of an *orchard* captures how we do this.

Definition 3.1.3. We say a collection $\mathcal{O} = \{\mathcal{D}_1, \ldots, \mathcal{D}_k\}$ of pairwise vertex-disjoint K_r-diamond trees in a graph G is a $(k,m)_r$*-orchard* if there are k diamond trees in the collection and each has order at least m and at most $2m$. We refer to k as the *size* of the orchard and m as the *order* of the orchard[2]. We denote by $V(\mathcal{O})$, the vertices featuring in diamond trees in $\mathcal{O}$, that is $V(\mathcal{O}) = \bigcup_{i\in[k]} V(\mathcal{D}_i)$. Finally, if $\mathcal{O}' \subseteq \mathcal{O}$ is a subset of diamond trees in an orchard $\mathcal{O}$, we call $\mathcal{O}'$ a *suborchard* of $\mathcal{O}$.

The term orchard here is supposed to be instructive, indicating that this is a 'neat' collection of diamond trees that all have a similar order and are completely disjoint from one another. As noted in Observation 3.1.2, a K_r-diamond tree can contribute to a K_r-factor in many ways. By grouping together many vertex-disjoint K_r-diamond trees into a $(k,m)_r$-orchard such that $km = \Omega(n)$, we get a structure with a strong absorbing property, as the next lemma shows. We say a $(K,M)_r$-orchard $\mathcal{O}$ *absorbs* a $(k,m)_r$-orchard $\mathcal{R}$ if there is a $((r-1)k, M)_r$-suborchard $\mathcal{O}' \subset \mathcal{O}$, such that there is a K_r-factor in $G[V(\mathcal{R}) \cup V(\mathcal{O}')]$.

Before stating the lemma, we make a few remarks about this definition of absorption of orchards. Firstly, note that the definition gives a K_r-matching which covers all the vertices of $\mathcal{R}$, that is, the vertices of $\mathcal{R}$ are *absorbed* into disjoint copies of K_r. However, we are also careful that we do not destroy the structure of $\mathcal{O}$ in the process of this absorption. Indeed, removing the K_r-matching given by the absorption we are left with a suborchard $\mathcal{O} \setminus \mathcal{O}'$ of $\mathcal{O}$. Finally, we remark that the condition that $\mathcal{O}'$ is a $((r-1)k, M)_r$-orchard stems from how we prove absorption of orchards throughout. Indeed, as we will see, we will always prove that $\mathcal{O}$ absorbs $\mathcal{R}$ by finding, for each diamond tree $\mathcal{B}$ in $\mathcal{R}$, a set of $r-1$ diamond trees $\mathcal{D}_1, \ldots, \mathcal{D}_{r-1} \subset \mathcal{O}$ such that there is a K_r-factor in $G[V(\mathcal{B}) \cup V(\mathcal{D}_1) \cup \ldots \cup V(\mathcal{D}_{r-1})]$. We will ensure that these choices are distinct and thus $\mathcal{O}'$ will be formed by taking all the choices of the $\mathcal{D}_i$ for all $\mathcal{B} \in \mathcal{R}$.

Lemma 3.1.4 (absorption between orchards)**.** *For any $3 \le r \in \mathbb{N}$ and $0 < \zeta, \eta < 1$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p,β)-bijumbled graph G with $\beta \le$*

[2]Note that we abuse notation slightly here. Indeed we refer to *the* order of an orchard although this may not be uniquely defined by the orchard. We take the convention that when we refer to the order of an orchard, we simply fix one of the possible orders arbitrarily, noting that these possible orders differ by a factor of at most 2.

$\varepsilon p^{r-1}n$. Let $\mathcal{O}$ be a $(K,M)_r$-orchard in G such that $KM \geq \zeta n$. Then there exists a set $B \subset V(G)$ such that $|B| \leq \eta p^{2r-4}n$ and $\mathcal{O}$ absorbs any $(k,m)_r$-orchard $\mathcal{R}$ in G with

$$V(\mathcal{R}) \cap (B \cup V(\mathcal{O})) = \emptyset, \qquad k \leq \frac{K}{8r} \qquad \textit{and} \qquad kM \leq mK. \tag{3.1.1}$$

Morally, Lemma 3.1.4 says that *large orchards absorb small orchards*. Here, by large we refer to both the size and the order of the orchards. Indeed the second condition in (3.1.1) shows that the larger orchard has to have a larger size than the smaller orchard. The third condition shows that the ratio between the orders of the orchards is constrained by the ratio of the sizes. That is, the larger $\mathcal{O}$ is compared to $\mathcal{R}$ with respect to their sizes, the smaller $\mathcal{R}$ can be than $\mathcal{O}$ with respect to their orders. The first condition in (3.1.1) simply states that in order for $\mathcal{O}$ to absorb $\mathcal{R}$, we need that $\mathcal{R}$ avoids some small set of bad vertices B. This will be easy to implement in applications.

Lemma 3.1.4 will be proven in Section 3.4.1. It provides us with an absorption property between two distinct orchards. We will also need an absorption property within orchards themselves, showing that we can find a large suborchard which hosts a K_r-factor in G. Given Observation 3.1.2, in order to find K_r-factors on suborchards it suffices to find copies of K_r which traverse sets of removable vertices. We therefore make the following definition.

Definition 3.1.5. Given a $(k,m)_r$-orchard $\mathcal{O} = \{\mathcal{D}_1, \dots, \mathcal{D}_k\}$ in a graph G, the *K_r-hypergraph* generated by $\mathcal{O}$, denoted $\mathcal{H} = \mathcal{H}(\mathcal{O})$, is the r-uniform hypergraph with vertex set $V(\mathcal{H}) = \mathcal{O}$ and with $\{\mathcal{D}_{i_1}, \dots, \mathcal{D}_{i_r}\}$ for distinct $i_1, \dots, i_r \in [k]$ forming a hyperedge in $\mathcal{H}$ if and only if there is a copy of K_r traversing[3] the sets $R_{i_1}, \dots, R_{i_r}$ in G, where R_{i_j} is the set of removable vertices of $\mathcal{D}_{i_j}$ for all j.

Appealing to Observation 3.1.2 then gives the following, as finding copy of K_r traversing r sets of removable vertices removes exactly one vertex from each set.

Observation 3.1.6. If $\mathcal{O}$ is an orchard of K_r-diamond trees in a graph G and $\mathcal{H}(\mathcal{O})$ contains a perfect matching, then $G[V(\mathcal{O})]$ contains a K_r-factor.

We will be particularly interested in orchards which contain almost K_r-factors in a robust way. This gives us the notion of a shrinkable orchard.

Definition 3.1.7. Given $0 < \gamma < 1$, we say a $(k,m)_r$-orchard in a graph G is *γ-shrinkable* if there exists a suborchard $\mathcal{Q} \subset \mathcal{O}$ of size at least γk such for *any* suborchard $\mathcal{Q}' \subseteq \mathcal{Q}$, we have that there is a matching in $\mathcal{H} := \mathcal{H}(\mathcal{O} \setminus \mathcal{Q}')$ covering all but $k^{1-\gamma}$ of the vertices of $\mathcal{H}$.

[3]Here and throughout, when we say a copy of K_r *traverses* r disjoint sets of vertices if it contains one vertex from each set.

Our first key proposition gives the existence of shrinkable orchards. It will be discussed in Section 3.4.2 and proven in Sections 3.5 and 3.6.

Proposition 3.1.8 (existence of shrinkable orchards)**.** *For any $3 \leq r \in \mathbb{N}$ and $0 < \alpha, \gamma < \frac{1}{2^{12r}}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$ and any vertex subset $U \subseteq V(G)$ with $|U| \geq \frac{n}{2}$. For any $m \in \mathbb{N}$ with $1 \leq m \leq n^{7/8}$ there exists a γ-shrinkable $(k, m)_r$-orchard $\mathcal{O}$ in $G[U]$ with $k \in \mathbb{N}$ such that $\alpha n \leq km \leq 2\alpha n$.*

Given Lemma 3.1.4 and Proposition 3.1.8, we know that we can find orchards which contain almost K_r-factors and that large orchards can absorb smaller orchards. This suggests the following approach for giving an absorbing structure which can absorb leftover vertices in our (p, β)-bijumbled graph G. Find a sequence of vertex-disjoint shrinkable orchards, each on a linear number of vertices. Each orchard in the sequence will have a larger order than that of the previous orchard and the first orchard in the sequence will be composed of linearly many K_r-diamond trees of constant size. We can then run a *cascading absorption* through the sequence of orchards. That is, given some small leftover set of vertices L (which is itself a $(|L|, 1)$-orchard), we use the first orchard in the sequence to absorb L. We then use that the first orchard is shrinkable and so we can cover most of what remains of the first orchard with vertex-disjoint copies of K_r. There will be some K_r-diamond trees of the first orchard left at the end of this and for these we appeal to Lemma 3.1.4 to absorb this small suborchard using the second orchard. Then again, the second orchard is shrinkable and so the remainder of the second orchard can be almost fully covered with vertex-disjoint K_rs, leaving some small leftover suborchard uncovered. We then repeat to absorb this leftover with the third orchard and continue in this fashion. In this way we cascade the absorption through the orchards and each time we do this, we increase the order of the orchard which we need to absorb.

This approach is promising but we need to cut this process off at some point and find a *full* K_r-factor on the vertices which have not already been covered by vertex-disjoint copies of K_r. The next proposition states that once the orchard has a large enough order, we can find a structure that can *fully absorb* any leftover.

Proposition 3.1.9 (the final absorption)**.** *For any $3 \leq r \in \mathbb{N}$ and $0 < \alpha, \eta < \frac{1}{2^{3r}}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$ and any vertex subset $W \subseteq V(G)$ with $|W| \geq \frac{n}{2}$.*

There exist vertex subsets $A, B \subset V$ such that $A \subset W$, $|A| \leq \alpha n$, $|B| \leq \eta p^{2r-4} n$ and for any $(k, m)_r$-orchard $\mathcal{R}$ whose vertices lie in $V(G) \setminus (A \cup B)$, with $|A| + |V(\mathcal{R})| \in r\mathbb{N}$, $k \leq \alpha^2 n^{1/8}$ and $m \geq n^{7/8}$, we have that $G[A \cup V(\mathcal{R})]$ has a K_r-factor.

We are now in a position to prove Theorem I, using only Lemma 3.1.4, Propositions 3.1.8 and 3.1.9, some simple properties of (p, β)-bijumbled graphs and Chernoff's Theorem (Theorem 2.1.1).

Proof of Theorem I. For convenient reference throughout the proof, let us fix our constants

$$\gamma := \frac{c}{2^{24r}}, \quad \lambda := \gamma^2, \quad \alpha := \lambda^2, \quad \zeta = \alpha^2, \quad \eta := \zeta^2 \quad \text{and} \quad t := \frac{7}{8\lambda}. \tag{3.1.2}$$

We further fix $n_0 = \frac{1}{\eta^{100}}$ and $\varepsilon > 0$ to be much smaller than $\frac{1}{n_0}$ and sufficiently small enough to apply Lemma 3.1.4 and Propositions 3.1.8 and 3.1.9 with the parameters above. We also use some simple consequences of Definition 1.4.1 which imply that, by choosing $\varepsilon > 0$ sufficiently small, we guarantee that any vertex subset of size ζpn contains a copy of K_{r-1} whilst any vertex set of size ζn contains a copy of K_r, see e.g. Corollary 3.2.5. Moreover we note that if $\delta(G) \geq \left(1 - \frac{1}{r}\right)n$ then it follows from Theorem 1.1.2 that G has a K_r-factor and so we can assume that $\delta(G) < \left(1 - \frac{1}{r}\right)n$. For such n-vertex (p, β)-bijumbled graphs G with $\beta \leq \varepsilon p^{r-1} n$, a well-known fact (see Fact 3.2.1) implies that by choosing $\varepsilon > 0$ to be sufficiently small, we can assume that $n \geq n_0$ in what follows as otherwise no n-vertex (p, β)-bijumbled graphs with $\beta \leq \varepsilon p^{r-1} n$ exist and the theorem is vacuously true. Finally, another well-known fact (see Fact 3.2.2) implies that (p, β)-bijumbled graphs cannot be too sparse. In particular, with our condition on β, by choosing $\varepsilon > 0$ to be sufficiently small, we can also assume that $p \geq n^{-1/3}$ in what follows.

Before finding our K_r-factor in G we need to do some preparation. We begin by setting aside a randomly chosen subset of vertices $Y \subset V(G)$. We let each vertex be in Y with probability α. It follows from Chernoff's Theorem (see Theorem 2.1.1) and a union bound that whp, we have that $|Y| \leq 2\alpha n$ and $\deg(v; Y) \geq \frac{c\alpha pn}{2}$ for all $v \in V(G)$. Indeed, this follows because $\mathbb{E}[\deg(v; Y)] \geq c\alpha pn = \Omega(n^{2/3})$ for each $v \in V(G)$. Therefore, as n is large, we can fix such an instance of Y. We will use the vertices of Y to find copies of K_r containing 'bad' vertices later in the argument.

Next, we apply Proposition 3.1.9 to obtain vertex sets $A \subset V(G) \setminus Y$ and B such that $|A| \leq \alpha n$, $|B| \leq \eta p^{2r-4} n$ and we have the following key absorption property. For any $(k, m)_r$-orchard $\mathcal{R}$ whose vertices lie in $V(G) \setminus (A \cup B)$, we have that if $|A| + |V(\mathcal{R})| \in r\mathbb{N}$, $k \leq \zeta n^{1/8}$ and $m \geq n^{7/8}$ then $G[A \cup V(\mathcal{R})]$ has a K_r-factor. That is, A can absorb orchards whose order is sufficiently large.

As sketched above, the idea is now to provide constantly many (namely, $t + 1$) vertex-disjoint shrinkable orchards $\mathcal{O}_0, \mathcal{O}_1, \ldots \mathcal{O}_t$, each on a linear number of vertices and whose vertices are disjoint from A. The order of these orchards will increase slightly (namely, by a factor of n^{λ}) at each step in the sequence. Due to our definition of t (3.1.2), we can have that $\mathcal{O}_0$ has $\Omega(n)$ orchards of constant order while $\mathcal{O}_t$ has orchards of order $\Omega(n^{7/8})$. The point is that we will be able to repeatedly apply Lemma 3.1.4 and the fact that each orchard is shrinkable to create a cascading absorption through the shrinkable orchards. Indeed $\mathcal{O}_0$ will be able to absorb leftover vertices and each $\mathcal{O}_i$ will be able to absorb any leftover K_r-diamond trees in $\mathcal{O}_{i-1}$, after using that $\mathcal{O}_{i-1}$ is shrinkable to cover almost all of the vertices of $\mathcal{O}_{i-1}$ with disjoint copies of K_r. Once this absorption reaches $\mathcal{O}_t$, we will be able to use A to absorb the leftover K_r-diamond trees in $\mathcal{O}_t$ and complete a K_r-factor. In fact, when absorbing between orchards we do not use *all* of $\mathcal{O}_i$ to absorb leftover diamond trees in $\mathcal{O}_{i-1}$ but rather a suborchard $\mathcal{Q}_i \subset \mathcal{O}_i$ which contains a γ proportion of the K_r-diamond trees in $\mathcal{O}_i$. Indeed this $\mathcal{Q}_i$ is provided by the fact that $\mathcal{O}_i$ is shrinkable (see Definition 3.1.7) and guarantees that removing diamond trees from $\mathcal{Q}_i$ will not prevent us from covering almost all of what remains of $\mathcal{O}_i$ with vertex-disjoint copies of K_r.

In detail, we collect what we require in the following claim.

Claim 3.1.10. *There exists vertex-disjoint orchards $\mathcal{O}_0, \mathcal{O}_1, \dots, \mathcal{O}_t$ in G such that following properties hold.*

(i) *For all $0 \le i \le t$, we have that $V(\mathcal{O}_i) \cap (A \cup B \cup Y) = \emptyset$.*

(ii) *For each $0 \le i \le t$, fixing $m_i := n^{i\lambda}$, we have that $\mathcal{O}_i$ is a $(k_i, m_i)_r$-orchard for some k_i such that $\alpha n \le k_i m_i \le 2\alpha n$.*

(iii) *Each $\mathcal{O}_i$ is γ-shrinkable with respect to some suborchard $\mathcal{Q}_i \subset \mathcal{O}_i$ such that*

$$k_i^* := |\mathcal{Q}_i| \ge \gamma k_i.$$

(iv) *For $1 \le i \le t$, given any suborchard $\mathcal{P} \subset \mathcal{O}_{i-1}$ such that $|\mathcal{P}| \le k_{i-1}^{1-\gamma}$, we have that $\mathcal{Q}_i$ absorbs $\mathcal{P}$.*

Before verifying the claim, let us see how we can derive the theorem using the claim. So suppose we have found such orchards $\mathcal{O}_0, \dots, \mathcal{O}_t$ and fix

$$X := A \bigcup \left(\bigcup_{i=0}^{t} V(\mathcal{O}_i) \right).$$

Furthermore, note that as $k_0^* m_0 = k_0^* \ge \gamma\alpha n \ge \zeta n$, by Lemma 3.1.4, there exists some set $B_0 \subset V(G)$ such that $|B_0| \le \eta p^{2r-4} n$ and $\mathcal{Q}_0$ absorbs any $(k, 1)$-orchard[4] $\mathcal{R}$ such that

$$k \le \zeta n \le \frac{\gamma\alpha n}{8r} \le \frac{k_0^*}{8r}, \tag{3.1.3}$$

[4]Note that a $(k, m)_r$-orchard with $m = 1$ is simply a set of vertices. Each K_r-diamond tree in the orchard has order 1 and so is a single isolated vertex.

and $V(\mathcal{R}) \cap (B_0 \cup V(\mathcal{Q}_0)) = \emptyset$. Indeed the condition on k comes from (3.1.1), using that $m_0 = 1$ and our lower bound on k_0^*. Fix $Z := B_0 \setminus X$ and note that $z := |Z| \leq \eta p^{2r-4} n$ as Z is a subset of B_0. Note also that $X \cap Y = \emptyset$ due to part (i) of Claim 3.1.10 and how we defined A.

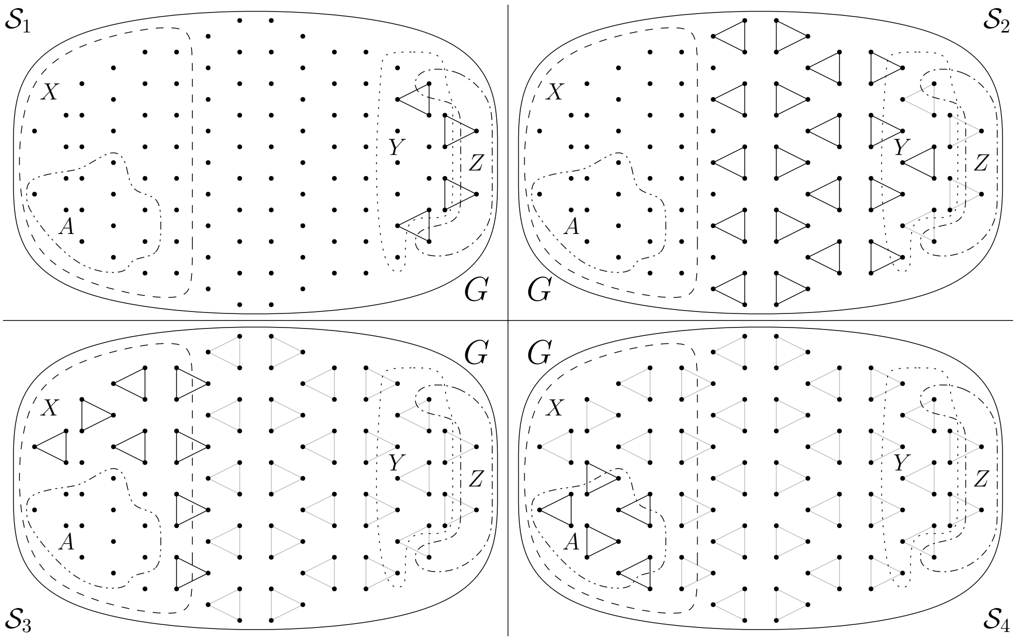


FIGURE 3.3: A schematic to demonstrate the triangles found (and the vertex sets they cover) by our four-phase algorithm that finds a K_3-factor in G.

We are now ready to find our K_r-factor $\mathcal{S}$ which we do algorithmically in four phases. See Figure 3.3 for a visual guide to the cliques found in each phase. So let us initiate with $\mathcal{S}_1 = \emptyset$. In the first phase we find copies of K_r containing the vertices in Z, using some vertices in Y. So let us order the vertices of Z arbitrarily as $Z := \{b_1, \ldots, b_z\}$ and fix $Y_1 := Y \setminus Z$. Now for $1 \leq j \leq z$, we find an r-clique S_j containing b_j and $r-1$ vertices of Y_j. We add S_j to $\mathcal{S}_1$, fix $Y_{j+1} := Y_j \setminus V(S_j)$ and move to step $j+1$. To see that we can always find such a clique, note that for each $j \in [z]$ we have that

$$\deg(b_j; Y_j) \geq \deg(b_j; Y) - |Z| - r(j-1) \geq \frac{c\alpha pn}{2} - r\eta p^{2r-4} n \geq \zeta pn,$$

recalling our key property of Y and using the definitions of our constants (3.1.2). A simple consequence of (1.4.2) (see e.g. Corollary 3.2.5 (1) (i)) implies that there is a copy of K_{r-1} in $N(b_j) \cap Y_j$ and so this forms an r-clique S_j with b_j. In this way, we see that we succeed at every step j and at the end of the first phase we have a set of vertex-disjoint r-cliques $\mathcal{S}_1$ in G of size z, such that every vertex in Z is contained in a clique in $\mathcal{S}_1$.

In the second phase we find the majority of the K_r-factor which we do greedily. We initiate with $\mathcal{S}_2 = \emptyset$ and $W = V(G) \setminus (X \cup V(\mathcal{S}_1))$. Now whilst $|W| \geq \zeta n$, we can find an r-clique S in W. Again, this is a simple consequence of (1.4.2), see Corollary 3.2.5 (2). We add S to $\mathcal{S}_2$

and delete its vertices from W. Therefore at the end of the second phase, we are left with some vertex set $L \subset V(G) \setminus X$ such that $|L| \le \zeta n$ and $\mathcal{S}_1 \cup \mathcal{S}_2$ form a K_r-factor in $G[V(G) \setminus (X \cup L)]$.

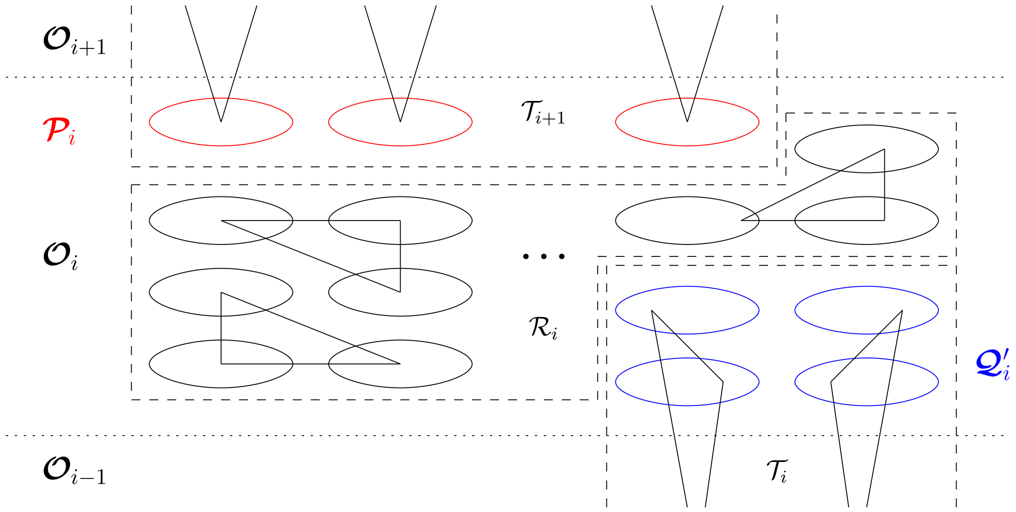


FIGURE 3.4: A closer look at phase 3 of the algorithm in the case $r = 3$.

In our third phase, we will find vertex-disjoint r-cliques $\mathcal{S}_3$ which cover L and use almost all the vertices of $X \setminus A$. We begin by fixing $\ell := |L|$ and noting that L is an $(\ell, 1)$-orchard which we relabel as $\mathcal{P}_{-1}$. Now we run the following procedure for $0 \le i \le t$ (see Figure 3.4). We first absorb $\mathcal{P}_{i-1}$ using $\mathcal{Q}_i$. That is, we find a suborchard $\mathcal{Q}_i' \subset \mathcal{Q}_i$ such that there is a K_r-factor $\mathcal{T}_i$ in $G[V(\mathcal{P}_{i-1}) \cup V(\mathcal{Q}_i')]$. We add the r-cliques in $\mathcal{T}_i$ to $\mathcal{S}_3$. Then, using that $\mathcal{O}_i$ is γ-shrinkable (Claim 3.1.10 (iii)), we can define some $\mathcal{P}_i \subset \mathcal{O}_i \setminus \mathcal{Q}_i'$ such that $|\mathcal{P}_i| \le k_i^{1-\gamma}$ and there is a perfect matching in the K_r-hypergraph $\mathscr{H}(\mathcal{O}_i \setminus (\mathcal{Q}_i' \cup \mathcal{P}_i))$. By Observation 3.1.6, this perfect matching gives a K_r-factor $\mathcal{R}_i$ in $G[V(\mathcal{O}_i \setminus (\mathcal{Q}_i' \cup \mathcal{P}_i))]$. We add $\mathcal{R}_i$ to $\mathcal{S}_3$ and move to step $i+1$ or finish if $i = t$. Note that in order to find $\mathcal{T}_i$ and $\mathcal{Q}_i'$ in each step $i \ge 1$, we appeal to Claim 3.1.10 (iv), whilst when $i = 0$, the existence of $\mathcal{T}_0$ and $\mathcal{Q}_0'$ is guaranteed by the fact that L is an $(\ell, 1)$-orchard with $\ell \le \zeta n$ as in (3.1.3) and L is disjoint from Z and hence B_0.

Let $\boldsymbol{\mathcal{R}} := \mathcal{P}_t \subset \mathcal{O}_t$. We have that $\mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3$ is a K_r-factor in $G[V(G) \setminus (A \cup V(\boldsymbol{\mathcal{R}}))]$. Hence as $r|n$, we must have that $r|(|A| + |V(\boldsymbol{\mathcal{R}})|)$. Moreover, $\boldsymbol{\mathcal{R}}$ is a $(k, m)_r$-orchard with $k \le k_t^{1-\gamma} \le (2\alpha n^{1/8})^{1-\gamma} < \alpha^2 n^{1/8}$ and $m = n^{7/8}$. Finally, note that $V(\boldsymbol{\mathcal{R}}) \cap B = \emptyset$ due to property (i) of Claim 3.1.10. Therefore, by the key property of the absorbing vertex set A in Proposition 3.1.9, we have that there is a K_r-factor $\mathcal{S}_4$ in $G[A \cup V(\boldsymbol{\mathcal{R}})]$. It follows that $\mathcal{S} := \mathcal{S}_1 \cup \mathcal{S}_2 \cup \mathcal{S}_3 \cup \mathcal{S}_4$ is a K_r-factor in G, completing the proof.

It remains to establish Claim 3.1.10 and find the shrinkable orchards as stated. We will do this algorithmically in decreasing order. The reason for this is that in order for (iv) to hold we will appeal to Lemma 3.1.4 and therefore there will be some set of bad vertices B_i which we want $\mathcal{O}_{i-1}$ to avoid. In fact, we will ensure that $\mathcal{O}_{i-1}$ avoids B_j for all $i \le j \le t$. This is not necessary but eases our definitions (as we do not have to reintroduce vertices into the pool U_i of available vertices); the important condition in what follows is that $\mathcal{O}_{i-1}$ avoids B_i for all i.

We start by fixing $U_{t+1} := V(G) \setminus (A \cup B \cup Y)$. Now for $t \geq i \geq 0$ in descending order, we apply Proposition 3.1.8 to find a γ-shrinkable $(k_i, m_i)_r$-orchard $\mathcal{O}_i$ such that $\alpha n \leq k_i m_i \leq 2\alpha n$ and $V(\mathcal{O}_i) \subset U_{i+1}$. We then define U_i as follows. As $\mathcal{O}_i$ is γ-shrinkable, it defines some suborchard $\mathcal{Q}_i \subset \mathcal{O}_i$ as in condition (iii) of the claim. Now as $k_i^* m_i \geq \gamma\alpha n \geq \zeta n$, it follows from Lemma 3.1.4 that there exists some $B_i \subset V(G)$ with $|B_i| \leq \eta p^{2r-4} n$ such that if k and m satisfy $k \leq \frac{k_i^*}{8r}$ and $km_i \leq mk_i^*$ and $\mathcal{R}$ is a $(k, m)_r$-orchard with $V(\mathcal{R}) \subset V(G) \setminus (B_i \cup V(\mathcal{Q}_i))$ then $\mathcal{Q}_i$ absorbs $\mathcal{R}$. We fix $U_i := U_{i+1} \setminus (V(\mathcal{O}_i) \cup B_i)$ and move onto the next index $i - 1$.

Let us first check that the process succeeds in finding the shrinkable orchards $\mathcal{O}_t, \ldots, \mathcal{O}_0$ at each step. Note that we start with $|U_{t+1}| \geq n - 3\alpha n - \eta p^{2r-4} n \geq n - 4\alpha n$. Moreover at each step i, we remove at most $\eta p^{2r-4} n \leq \alpha n$ vertices which lie in B_i and at most $4r\alpha n$ vertices from U_{i+1} which lie in the orchard $\mathcal{O}_i$. Indeed the orchard is composed of k_i vertex-disjoint K_r-diamond trees of order at most $2m_i$, the number of vertices in each diamond tree is less than r times its order and $k_i m_i \leq 2\alpha n$. Hence for all $t \geq i \geq 0$, we have that

$$|U_i| \geq n - (t+2) \cdot 5r\alpha n \geq \frac{n}{2},$$

using that $t \leq \frac{1}{\lambda}$, $\alpha = \lambda^2$ and the definition of λ (see (3.1.2)) here. Hence Proposition 3.1.8 gives the existence of $\mathcal{O}_i$ at each step and verifies part (iii) of the claim. Note that the conditions (i) and (ii) also hold simply from how we defined the $\mathcal{O}_i$ and the fact that we found them in the sets U_i, each of which is a subset of U_{t+1}.

Thus it remains to verify the absorption property between orchards, namely (iv). For each $1 \leq i \leq t$, we chose $\mathcal{O}_{i-1}$ to have vertices in U_i and hence $V(\mathcal{O}_{i-1}) \cap (B_i \cup V(\mathcal{Q}_i)) = \emptyset$. Therefore we have by Lemma 3.1.4 that $\mathcal{Q}_i$ absorbs any suborchard $\mathcal{P} \subset \mathcal{O}_{i-1}$, with $|\mathcal{P}| \leq k_{i-1}^{1-\gamma}$ if $k_{i-1}^{1-\gamma} \leq \frac{k_i^*}{8r}$ and $k_{i-1}^{1-\gamma} m_i \leq k_i^* m_{i-1}$.

Now as $m_i = n^\lambda m_{i-1}$ and $n^{-\lambda} \leq \frac{1}{8r}$ for $n \geq n_0$, it suffices to show that $k_{i-1}^{1-\gamma} \leq k_i^* n^{-\lambda}$. To see this, note that due to the fact that $\alpha n \leq k_{i-1} m_{i-1}, k_i m_i \leq 2\alpha n$, we have

$$k_{i-1} \leq \frac{2\alpha n}{m_{i-1}} = \frac{2\alpha n^{1+\lambda}}{m_i} \leq 2k_i n^\lambda \leq \frac{2k_i^* n^\lambda}{\gamma}$$

and using this as a lower bound for k_i^*, it suffices to show that

$$k_{i-1}^\gamma \geq \frac{2n^{2\lambda}}{\gamma}.$$

This is certainly true as $k_{i-1} \geq k_t \geq \alpha n^{1/8} > n^{4\lambda/\gamma}$, recalling that $\frac{4\lambda}{\gamma} = 4\gamma$ from (3.1.2). This shows that (iv) holds for all i and concludes the proof of the claim and hence the proof. □

We remark that this proof scheme builds on that of Nenadov [145] (which in turn is influenced by that of Krivelevich [119]) who proved that $\beta \leq \frac{\varepsilon p^2 n}{\log n}$ suffices for a triangle factor in an n-vertex (p, β)-bijumbled graph. Indeed, Nenadov also uses a result akin to Lemma 3.1.4 albeit between orchards whose orders only differ by a constant factor. His absorbing structure then contains a sequence of $\Theta(\log n)$ orchards whose order increases by a constant factor along the sequence. Therefore the last orchard in the sequence contains constantly many diamond trees of large order (of order $\Theta(\frac{n}{\log n})$). These can be fully absorbed because any three large sets host a transversal triangle and so transversal triangles between removable sets can be greedily found, completing a triangle factor in the last step. Similarly, the $(k, m)_3$-orchards used in his argument are not imposed to be shrinkable but can be seen to host a triangle factor on all but $o(k)$ of the diamond trees by again applying a greedy approach of finding transversal triangles. The necessity of the $\log n$ in the condition of Nenadov is thus due to needing $\Theta(\log n)$ orchards in the absorbing structure and thus requiring slightly stronger properties of the (p, β)-bijumbled graph, for example the existence of triangles on sets of $\Omega(\frac{n}{\log n})$ vertices.

The key challenge in this chapter is then to prove Propositions 3.1.8 and 3.1.9. Both results rely heavily on a technique we develop to provide the existence of K_r-diamond trees in which we have some control over the set of removable vertices. This control is rather weak; we cannot guarantee that any fixed vertices appear as removable vertices but we can give some flexibility over the choice of removable vertices. See Proposition 3.3.1 for the technical statement of what we prove.

In order to prove Proposition 3.1.8, we build on the approach of Han, Kohayakawa and Person [85, 86]. Indeed, their result showing the existence of an almost K_r-factor (covering all but some $n^{1-\varepsilon'}$ vertices) in (n, d, λ)-graphs can be seen as a step towards proving the existence of shrinkable orchards of order 1. The approach involves showing the existence of an almost-perfect matching in a subhypergraph $\mathcal{H}'$ of the K_r-hypergraph generated by $V(G)$. In order to do this, one needs to carefully choose $\mathcal{H}'$ and this is done by finding many fractional K_r-factors in G which do not put too much weight on (copies of K_r containing) any given edge. Therefore, the methods of Krivelevich, Sudakov and Szabó [127], who proved the existence of singular fractional K_r-factors, become pertinent. They use the power of linear programming duality to prove that certain expansion properties guarantee the existence of fractional factors. In our setting, it turns out that we need several distinct arguments to prove the existence of shrinkable orchards of different orders. We follow the scheme of using fractional factors (in fact, fractional perfect matchings in K_r-hypergraphs, see Section 2.6) but need to adapt the method for different applications and we rely crucially on probabilistic methods to actually prove the existence of orchards which satisfy the necessary expansion properties.

It can be seen that Proposition 3.1.8 alone (for all orders of orchards) would lead via the same proof scheme to a condition of $\beta \leq \frac{\varepsilon p^{r-1} n}{\log \log n}$. In order to close the gap and achieve Theorem I, Proposition 3.1.9 is necessary. To prove this, we appeal to the *template absorption method* (see Section 2.8) to define an absorbing structure. This method was previously used by Han, Kohayakawa, Person and the author [83] to find clique factors in pseudorandom graphs and we used this approach again in our result on 2-universality [84]. Here we combine this idea with the absorbing power of orchards and prove Proposition 3.1.9 with a three-stage algorithm which finds the absorbing structure necessary.

3.2 Properties of bijumbled graphs

Here we collect some properties of bijumbled graphs. These range from simple consequences of Definition 1.4.1 to more involved statements catered to our purposes. We begin by showing that we can assume that the graphs we consider have an arbitrarily large number of vertices.

Fact 3.2.1. *Given any $3 \leq r \in \mathbb{N}$ and $n_0 \in \mathbb{N}$, there exists $\varepsilon > 0$ such that any n-vertex (p, β)-bijumbled graph G with $n \in r\mathbb{N}$, $\delta(G) < \left(1 - \frac{1}{r}\right)n$ and $\beta \leq \varepsilon p^{r-1}n$ must have $n \geq n_0$.*

Proof. Let $\varepsilon > 0$ such that $\varepsilon < \frac{1}{2n_0 r}$. Suppose for a contradiction that there exists an n-vertex (p, β)-bijumbled graph with $\delta(G) < \left(1 - \frac{1}{r}\right)n$, $\beta \leq \varepsilon p^{r-1}n$ and $n < n_0$. Then due to the upper bound on the minimum degree of G, there exists a vertex $u \in V(G)$ and a set $W \in V(G)\backslash\{u\}$ such that $|W| = \frac{n}{r}$ and $\deg_G(u; W) = 0$. However, from Definition 1.4.1, we have that

$$e(\{u\}, W) \geq p|W| - \varepsilon p^{r-1}n\sqrt{\frac{n}{r}} \geq \frac{pn}{r}\left(1 - \varepsilon\sqrt{nr}\right) \geq \frac{pn}{2r} > 0,$$

a contradiction. □

Fact 3.2.1 shows that by choosing $\varepsilon > 0$ sufficiently small, we guarantee that any bijumbled graph G we are interested in either has a large number of vertices or has $\delta(G) \geq \left(1 - \frac{1}{r}\right)n$, in which case Theorem 1.1.2 implies the existence of a K_r-factor and we are done. We will use this at various points in our argument and simply state that we choose $\varepsilon > 0$ sufficiently small to force n to be sufficiently large.

The following well-known fact states that bijumbled graphs cannot to be too sparse.

Fact 3.2.2. *For any $3 \leq r \in \mathbb{N}$ and any $C > 0$, there exists an $\varepsilon > 0$ such that any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1}n$ has $p \geq Cn^{-1/(2r-3)} \geq Cn^{-1/3}$.*

Proof. Let $\varepsilon > 0$ be such that $\varepsilon^2 \leq \frac{1}{32C^{2r-3}}$ and small enough that we can assume that

(i) $n \geq 9$;

(ii) $p \leq \frac{1}{16}$.

Indeed, from Fact 3.2.1, we can choose ε so that (i) holds and $Cn^{-1/(2r-3)} < \frac{1}{16}$ and so we are done if we are not in case (ii).

We will also restrict to the case that

(iii) $p \geq \frac{1}{2n}$.

To see that we can do this, suppose for a contradiction that there exists a (p, β)-bijumbled graph $G = (V, E)$ with $pn < \frac{1}{2}$. We appeal to Definition 1.4.1 and upper bound $2e(G) = e(V, V)$ by $pn^2 + \varepsilon p^{r-1} n^2 < n - 1$. Hence there must be some vertex $u \in V$ which is isolated in G. But then defining $W := V \setminus \{u\}$, the lower bound of Definition 1.4.1 gives that $e(\{u\}, W) \geq p(n-1) - \varepsilon p^{r-1} n\sqrt{n-1} \geq pn\left(\frac{1}{2} - \varepsilon pn\right) > 0$, a contradiction.

We now turn to proving the statement in full generality. Our aim is to construct large (disjoint) vertex subsets U and W such that $e(U, W) = 0$. We do this in the following greedy fashion. We initiate a process by setting $U = \emptyset$ and $W = V(G)$. Now whilst $|W| \geq \frac{3n}{4}$, there exists some $u \in W$ with $\deg(u; W) \leq 2p|W| \leq 2pn$. Indeed this follows from Definition 1.4.1 as

$$\sum_{w \in W} \deg(w; W) = e(W, W) \leq p|W|^2 + \varepsilon p^{r-1}|W| \leq 2p|W|^2.$$

We then choose such a u, delete it from W and add it to U, and also remove $N(u; W)$ from W.

Let U and W be the resulting sets after this process terminates. It is clear that $e(U, W) = 0$ as we have removed all the neighbours of each vertex $u \in U$ from W during the process. We also claim that $|W| \geq \frac{n}{2}$ and $U \geq \frac{1}{16p}$. Indeed, the last step removed at most $1 + 2pn$ vertices from W. Due to our assumptions (i) and (ii), we have that $1 + 2pn < \frac{n}{4}$ and so as W had size greater than $\frac{3n}{4}$ before this step, we indeed have that $|W| \geq \frac{n}{2}$ as the process terminates. To see the lower bound on the size of U, note that if this was not the case, then

$$|V(G) \setminus W| = |\cup_{u \in U} (\{u\} \cup N_G(u))| \leq \sum_{u \in U} |\{u\} \cup N_G(u)| \leq |U|(1 + 2pn) \leq \frac{1}{16p} + \frac{n}{8} \leq \frac{n}{4},$$

using assumption (iii) in the last inequality. This implies then that $|W| \geq \frac{3n}{4}$, a contradiction as the process terminated.

Thus $|W| \geq \frac{n}{2}$, $|U| \geq \frac{1}{16p}$ and from Definition 1.4.1, we have that

$$0 = e(U, W) \geq p|U||W| - \varepsilon p^{r-1} n\sqrt{|U||W|},$$

implying that $p^{2r-3} \geq \frac{1}{32\varepsilon^2 n}$. Given our upper bound on ε, this implies that $p \geq Cn^{-1/(2r-3)}$ as required. □

Our first lemma shows that few vertices have degree much smaller or much larger than expected to a given set.

Lemma 3.2.3. *For any $3 \leq r \in \mathbb{N}$ and $\eta > 0$ there exists an $\varepsilon > 0$ such that if G is an n-vertex (p, β)-bijumbled graph with $\beta \leq \varepsilon p^{r-1} n$ then for $W \subseteq V(G)$ we have that:*

(i) The number of vertices $v \in V(G)$ such that $\deg(v; W) < \frac{p}{2}|W|$, is less than

$$\frac{\eta p^{2r-4} n^2}{|W|}.$$

(ii) For any q such that $2p \leq q \leq 1$, the number of vertices $v \in V(G)$ such that $\deg(v; W) > q|W|$, is less than

$$\frac{\eta p^{2r-2} n^2}{q^2 |W|}.$$

Proof. Fix $\varepsilon > 0$ such that such that $4\varepsilon^2 < \eta$. We prove only (ii), the proof of (i) is both similar and simpler. We set B to be the set of 'bad' vertices i.e. vertices v such that $\deg(v; W) > q|W|$. Thus we have that

$$q|B||W| < e(B, W) \leq p|B||W| + \varepsilon p^{r-1} n \sqrt{|B||W|},$$

using the definition of B and (1.4.2). Rearranging gives that

$$|B| < \frac{\varepsilon^2 p^{2r-2} n^2}{(q-p)^2 |W|},$$

and using that $p \leq \frac{q}{2}$ gives the desired conclusion using our choice of ε. □

Next, we state some further consequences of Definition 1.4.1, showing that we can find cliques traversing large enough subsets of vertices. The following lemma is very general and will be used at various points in our argument. Due to its generality, there are some technical features. Whilst these are all necessary for certain parts of our argument, we do not need all of these at once. In fact for easy reference, we list the consequences of Lemma 3.2.4 that we will use, in Corollary 3.2.5. This may also serve to digest the statement of Lemma 3.2.4, seeing how it is applied in practice. We also refer the reader to the Notation Section for relevant definitions, for example recalling that for a clique S and vertex subset U of a graph, $\deg(S; U)$ denotes the number of common neighbours of (the vertices in) S, that lie in U.

Lemma 3.2.4. *For any $3 \leq r \in \mathbb{N}$ and $0 < \tau < \frac{1}{2^{2r}}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$. Suppose that there are integers $x_i \in \mathbb{N}$, $i \in [r+1]$ such that $x_1 \geq \ldots \geq x_{r+1} \geq 0$ and for some $r^* \in [r]$, one has that*

$$x_i + x_{i+1} + 2i \leq 2r - 2, \tag{3.2.1}$$

for all $1 \leq i \leq r^*$. *Define* $y := \max\{x_{i+1} + i : i \in [r^*]\}$. *Then for any collection of subsets* $U_i \subseteq V(G)$ *such that* $|U_i| \geq \tau p^{x_i} n$ *for all* $i \in [r+1]$, *and any subgraph* $\tilde{G}$ *of* G *with maximum degree less than* $\tau^2 p^y n$, *defining* $G' := G \setminus \tilde{G}$, *we have that there exists a clique* $S \in K_{r^*}(G')$ *traversing* $U_1, \ldots, U_{r^*}$ *such that*

$$\deg_{G'}(S; U_j) \geq \tau p^{r^*} |U_j|,$$

for $r^* + 1 \leq j \leq r + 1$.

Proof. Fix $\varepsilon > 0$ small enough to apply Lemma 3.2.3 (i) with $\eta := \frac{\tau^2}{2^{4r} r}$. Further, fix y and $\tilde{G}$ as in the statement, setting $G' := G \setminus \tilde{G}$. We will prove inductively that for $i = 1, \ldots, r^*$, there exists an i-clique $S_i \in K_i(G')$ traversing $U_1, \ldots, U_i$ such that $\deg_{G'}(S_i; U_j) \geq \left(\frac{p}{4}\right)^i |U_j|$ for all j with $i+1 \leq j \leq r+1$. Note that S_{r^*} is the desired copy of K_{r^*} in the statement, using that $\tau \leq \frac{1}{4^{r^*}}$ here.

So fix some $i \in [r^*]$. If $i \geq 2$, by induction we deduce the existence of S_{i-1} as claimed and for $i \leq j \leq r+1$, define $W_j \subseteq U_j$ so that $W_j := N_{G'}(S_i; U_j)$. If $i = 1$, we simply set $W_j := U_j$ for all j. We thus have that

$$|W_j| \geq \left(\frac{p}{4}\right)^{i-1} |U_j| \geq \tau 4^{1-i} p^{x_j + i - 1} n, \tag{3.2.2}$$

for $i \leq j \leq r+1$. Now we appeal to Lemma 3.2.3 (i) and conclude that for each j with $i+1 \leq j \leq r+1$, there is some set $B_j \subset V(G)$ such that $\deg_G(v; W_j) \geq \frac{p}{2}|W_j|$ for all $v \in V(G) \setminus B_j$ and

$$|B_j| \leq \frac{\eta p^{2r-4} n^2}{|W_j|} \leq \frac{\eta 4^{i-1} p^{2r-3-i-x_j} n}{\tau} \leq \frac{\tau p^{2r-3-i-x_{i+1}} n}{4^i r} \leq \frac{\tau p^{x_i+i-1} n}{4^i r} \leq \frac{|W_i|}{2r}. \tag{3.2.3}$$

Here, we used (3.2.2) in the second inequality, the definition of η and the fact the $x_j \leq x_{i+1}$ in the third, (3.2.1) in the fourth and (3.2.2) once again in the final inequality. We can thus conclude from (3.2.3) that there exists a vertex $w_i \in W_i$ such that $w_i \notin B_j$ for all $i+1 \leq j \leq r+1$. We claim that choosing $S_i = S_{i-1} \cup \{w_i\}$ completes the inductive step. Indeed $S_i \in K_i(G')$ as w_i was chosen from the common neighbourhood of S_{i-1} in G'. Also, fixing some $i+1 \leq j \leq r-1$, we have that $N_G(w_i)$ intersects $W_j = N_{G'}(S_{i-1}; U_j)$ in at least $\frac{p}{2}|W_j|$ vertices. Furthermore, at most

$$\tau^2 p^y n \leq \tau^2 p^{x_{i+1}+i} n \leq \frac{\tau}{2^{2r}} p^{x_j + i} n \leq \frac{p}{4}|W_j|$$

edges adjacent to w_i lie in $\tilde{G}$, using the definition of y, the upper bound on τ, the fact that $x_j \leq x_{i+1}$ and (3.2.2). Therefore we can conclude that for all $i+1 \leq j \leq r+1$, we have $\deg_{G'}(S_i; U_j) \geq \deg_{G'}(S_i; W_j) \geq \frac{p}{4}|W_j| \geq \left(\frac{p}{4}\right)^i |U_j|$, as required. This completes the induction and the proof. □

We now collect some easy consequences of Lemma 3.2.4 for reference later in the proof.

Corollary 3.2.5. *For any $3 \leq r \in \mathbb{N}$ and $0 < \tau < \frac{1}{2^{2r}}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$. We have that:*

1. *For any subgraph $\tilde{G}$ of G with maximum degree less than $\tau^2 p^{r-1} n$ we have the following.*
   - *(i) For any vertex subsets $U_1, \ldots, U_{r-1} \subseteq V(G)$ such that $|U_i| \geq \tau p n$ for $i \in [r-1]$, there exists an $(r-1)$-clique $S \in K_{r-1}(G \setminus \tilde{G})$, traversing the U_i.*
   - *(ii) For any $U_1, \ldots, U_r \subseteq V(G)$ such that $|U_1| \geq \tau p^{2r-4} n$ and $|U_i| \geq \tau n$ for $2 \leq i \leq r$, there exists an r-clique $S \in K_r(G \setminus \tilde{G})$, traversing the U_i.*
2. *For any $U_1, \ldots, U_r \subseteq V(G)$ such that $|U_1| \geq \tau p^{r-1} n$, $|U_i| \geq \tau p n$ for $2 \leq i \leq r-2$ and $|U_{r-1}|, |U_r| \geq \tau n$, there exists an r-clique $S \in K_r(G)$, traversing the U_i.*
3. *For any $W_0, W_1, W_2 \subseteq V(G)$ such that $|W_0|, |W_1|, |W_2| \geq \tau n$, there exists an $S \in K_{r-1}(G[W_0])$ such that* $\deg(S; W_i) \geq \tau^2 p^{r-1} n$ *for $j = 1, 2$.*

Proof. Fix $\varepsilon > 0$ small enough to apply Lemma 3.2.4. This is predominantly a case of plugging in the values and checking the conditions of Lemma 3.2.4. For part (1), we let $G' = G \setminus \tilde{G}$. Then for (1)(i), we take $r^* = r-2$, $x_i = 1$ for $1 \leq i \leq r+1$ and $y = r-1$. We thus have that for $i \in [r^*]$, $x_i + x_{i+1} + 2i = 2 + 2i \leq 2r - 2$ and $x_{i+1} + i = 1 + i \leq r - 1 = y$. Therefore taking U_i for $1 \leq i \leq r-1$ with $|U_i| \geq \tau p n$ (and defining $U_{r+1} = U_r = U_{r-1}$), Lemma 3.2.4 gives us an $(r-2)$-clique $S' \in K_{r-1}(G')$ traversing $U_1, \ldots, U_{r-2}$ such that $\deg_{G'}(S'; U_{r-1}) \geq \tau^2 p^{r-1} n > 0$ (here Fact 3.2.2 shows positivity). Therefore choosing any vertex $v \in N_{G'}(S'; U_{r-1})$ and fixing $S = S' \cup \{v\}$ gives the required clique.

The other cases are similar. For part (1) (ii), we fix $r^* = r-1$, $x_1 = 2r-4$, $x_i = 0$ for $2 \leq i \leq r+1$ and $y = r-1$. Again, it easily checked that the conditions on the x_i are all satisfied and so applying Lemma 3.2.4 (fixing $U_{r+1} = U_r$) gives an $(r-1)$-clique S' in G' traversing $U_1, \ldots, U_{r-1}$ such that S' has a nonempty G'-neighbourhood in U_r. Therefore adding any vertex in this neighbourhood to S' gives the required r-clique $S \in K_r(G')$.

For part (2), we fix $r^* = r-1$, $x_1 = r-1$, $x_i = 1$ for all i such that $2 \leq i \leq r-2$ and $x_{r-1} = x_r = x_{r+1} = 0$. We also let $\tilde{G}$ be the empty graph and so $G = G'$. Now note that for $r = 3$, we have $x_1 = 2$ and $x_2 = 0$ and so $x_1 + x_2 + 2 = 4 = 2r - 2$, whilst for $r \geq 4$, we have $x_1 + x_2 + 2 = r + 2 \leq 2r - 2$. The conditions (3.2.1) for $2 \leq i \leq r^* = r - 1$ can be similarly checked. Therefore Lemma 3.2.4 gives an $(r-1)$-clique $S' \in K_{r-1}(G)$ traversing $U_1, \ldots, U_{r-1}$ such that $N_G(S'; U_r) \neq \emptyset$ and so as above, we extend S' to the required r-clique S.

Finally, for part (3) we fix $r^* = r-1$, $x_i = 0$ for all $1 \leq i \leq r$ and define our sets as $U_i = W_0$ for $i \in [r-1]$ and $U_r = W_1$, $U_{r+1} = W_2$. Applying Lemma 3.2.4 then directly gives the required $(r-1)$-clique $S \in K_{r-1}(G[W_0])$ (again here $\tilde{G}$ is taken to be empty). □

3.3 Diamond trees

Recall the definition of diamond trees from Section 3.1, namely Definition 3.1.1. In this section we prove the existence of diamond trees in our bijumbled graphs. The main aim is to prove the following proposition which gives us some flexibility over which vertices feature as removable vertices of our diamond tree. This will turn out to be very valuable at various points in our proof.

Proposition 3.3.1. *For any $3 \le r \in \mathbb{N}$ and $0 < \tau < \frac{1}{2^{2r}}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p,β)-bijumbled graph G with $\beta \le \varepsilon p^{r-1}n$. For any $2 \le z \le \alpha n$ and any pair of disjoint vertex subsets $U, W \subset V(G)$ such that $|U|, |W| \ge 4\alpha r n$, there exists disjoint vertex subsets $X, Y \subset U$ such that the following hold:*

1. $|X| + |Y| = z$;
2. $|X| \le \max\left\{1, \frac{2z}{\delta}\right\}$ *with* $\delta = \alpha^2 p^{r-1} n$;
3. *for* any *subset $Y' \subset Y$, there exists a K_r-diamond tree $\mathcal{D} = (T, R, \Sigma)$ such that $R = X \cup Y'$ and $\Sigma \subset K_{r-1}(G[W])$ is a K_{r-1}-matching in W.*

Let us pause to digest the proposition. Firstly, note that by choosing $Y' = Y$ in (3) and varying z, we can guarantee the existence of K_r-diamond trees of any order up to linear in our bijumbled graph G. However, the proposition is much more powerful than just this. The vertex set Y and property (3) allow us *flexibility* in which vertices appear in the removable set of vertices of the diamond tree we take from the proposition. We can start with z much larger than the desired order of the diamond tree we want and then remove unwanted vertices from Y to end up with some Y' that we include in the removable vertices of the diamond tree. The point is that by starting with a larger z (and hence larger $|Y|$), we can deduce stronger properties about the vertices in Y, allowing us to then ensure properties of the set of removable vertices R that we would otherwise have no hope in guaranteeing. There is a catch, as we are forced to include the set X in any diamond tree we produce, but note that due to property (2), the size of X is negligible compared to the size of Y. Indeed due to Fact 3.2.2, we have that δ is polynomial in n (of order at least $\Omega(n^{(r-1)/(2r-3)})$, to be precise). Thus we can choose Y' to be much smaller than Y and still have the vertices in Y' contribute a significant subset (at least half, say) of the removable vertices of the diamond tree we obtain. We delay applications of Proposition 3.3.1 to later in the proof but refer the reader to Lemmas 3.5.6, 3.6.2 and 3.7.4 for a flavour of the consequences of the proposition.

The rest of this section is concerned with proving Proposition 3.3.1. The idea behind the proof is simple, we look to find a large (order z) K_r-diamond tree in G with the property that many of the removable vertices are leaves (the set Y). This allows us to pick and choose which leaves (the set Y') we include in our desired diamond tree, as we can simply remove the other leaves and their corresponding interior cliques[5], see Figure 3.5 for an example. In order to find diamond trees with many leaves, we introduce the notion of a scattered diamond tree and deduce the existence of such diamond trees in a suitably pseudorandom graph.

3.3.1 Scattered diamond trees

One way to ensure a large set of leaves in a tree is to impose a minimum degree on all non-leaf vertices. This leads to the following definition.

Definition 3.3.2. We say a tree T (of order at least 2) is δ-*scattered* if every vertex in $V(T)$ which is not a leaf in T has degree at least δ. As a convention we will also say that a tree of order 1 (a single vertex) is δ-scattered for all δ. We say a diamond tree is $\mathcal{D} = (T, R, \Sigma)$ is δ-scattered if its underlying auxiliary tree T is δ-scattered.

See Figure 3.5 for an example of a scattered K_3-diamond tree. The following simple lemma shows that most of the vertices in a scattered tree (and hence most of the removable vertices in a scattered diamond tree) are leaves.

Lemma 3.3.3. *Let $\delta \geq 2$ and suppose that T is a δ-scattered tree of order $m \geq 3$. Then defining $X \subset V(T)$ to be the vertices[6] which have degree greater than 1 in T, we have that*

$$|X| \leq \frac{m-2}{\delta-1}.$$

Proof. By the definition of δ-scattered trees, we have that every vertex in X has degree at least δ. We define $x := |X|$. Note that $T[X]$ is a connected subtree of T. Indeed, the interior vertices of a path between any two vertices of T must lie in X (as they have degree at least 2). Hence $T[X]$ has exactly $x - 1$ edges and we can estimate the number of edges in T as follows:

$$e(T) = m - 1 = \sum_{v \in X} \deg(v) - e(T[X]) \geq x\delta - (x-1).$$

Rearranging, one obtains that $x \leq \frac{m-2}{\delta-1}$, as required. □

We will show that we can find large scattered diamond trees in our bijumbled graph. To begin with, we focus on diamond trees for which the auxiliary tree is a star, which we call *diamond stars*. The next lemma shows that we can find large diamond stars in a suitably pseudorandom graph.

[5]That is, for each unwanted leaf $v \in R$ in the diamond tree $\mathcal{D} = (T, R, \Sigma)$, we remove the interior clique in Σ which corresponds to the edge adjacent to the (preimage of) v in the defining tree T, as well as the leaf v itself.

[6]That is, X is the set of vertices of T which are not leaves.

Lemma 3.3.4. *For any $3 \leq r \in \mathbb{N}$ and $0 < \alpha < \frac{1}{2^{2r}}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$, fixing $\delta := \alpha^2 p^{r-1} n$. Let $U_0, U_1, U_2 \subseteq V(G)$ be disjoint vertex subsets such that $|U_i| \geq \alpha n$ for $i = 0, 1$ and $|U_2| \geq \alpha r n$. Then there exists a K_r-diamond tree $\mathcal{D}^* = (T^*, R^*, \Sigma^*)$ in G such that T^* is a star of order $1 + \delta$ centred at x, say, with*[7] *$\rho^*(x) \in U_0$, $R^* \setminus \{\rho^*(x)\} \subset U_1$ and $\Sigma^* \subset K_{r-1}(G[U_2])$ is a K_{r-1}-matching in U_2.*

Proof. Fix $\varepsilon > 0$ small enough to apply Corollary 3.2.5 (3) with $\tau = \alpha$. Shrink U_0 (if necessary) to be a set of exactly αn vertices. We claim that there is a matching $M \subset K_{r-1}(G[U_2])$ of $(r-1)$-cliques such that $|M| = \alpha n$ and each clique $S \in M$ has $\deg(S; U_i) \geq \delta$, for $i = 0, 1$. Indeed, we can find M greedily by applying Corollary 3.2.5 (3) (with $W_i = U_{2-i}$ for $i = 0, 1, 2$) repeatedly, adding an $(r-1)$-clique S to M and removing its vertices from U_2 after each application. While $|M| \leq \alpha n$, we have that $|U_2| \geq \alpha n$ and so we are indeed in a position to apply Corollary 3.2.5 (3) throughout the process.

Now once we have found M, for each $S \in M$ and for $i = 0, 1$, let $N_i(S) := N(S; U_i)$, that is, the set of vertices in U_i which form a K_r with S. By construction we have that $|N_0(S)| \geq \delta$ for each S in M and so

$$|\{(v, S) \in U_0 \times M : v \in N_0(S)\}| \geq |M|\delta = \alpha n \delta.$$

Hence, as $|U_0|$ has size αn (as we imposed at the start of the proof), by averaging, there exists a vertex $v_0 \in U_0$ and a subset Σ^* of δ cliques in M such that v_0 is in $N_0(S)$ for all $S \in \Sigma^*$. We can now construct our diamond star greedily, with v_0 as the image of the large degree vertex. Sequentially, for each clique S in Σ^*, choose a vertex u in $N_1(S)$ which has not been previously chosen and add the copy of K_{r+1}^- on S, v_0 and u to the diamond star (adding u to R^*). As $N_1(S) \geq \delta$ for all $S \in \Sigma^* \subseteq M$, there is always an option for u and so this process succeeds in building the required diamond star. □

Our next lemma builds on the scheme of Krivelevich [119] to construct large diamond trees. We adapt his proof to guarantee that the diamond tree obtained is scattered.

Lemma 3.3.5. *For any $3 \leq r \in \mathbb{N}$ and $0 < \alpha < \frac{1}{2^{2r}}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$, fixing $\delta := \alpha^2 p^{r-1} n$. For any $2 \leq z \leq \alpha n$ and any pair of disjoint vertex subsets $U, W \subset V(G)$ such that $|U|, |W| \geq 4\alpha r n$, there exists a δ-scattered K_r-diamond tree $\mathcal{D}_{sc} = (T_{sc}, R_{sc}, \Sigma_{sc})$ of order m such that $z \leq m \leq z + \delta$, $R_{sc} \subset U$ and $\Sigma_{sc} \subset K_{r-1}(G[W])$ is a K_{r-1}-matching in $G[W]$.*

[7] Here $\rho^* : V(T^*) \to R^*$ is the associated bijection in the definition of $\mathcal{D}^*$.

Proof. Our proof is algorithmic and works by building a diamond tree forest, that is, a set of pairwise vertex-disjoint diamond trees. At each step of the algorithm, we will add to one of the trees in our forest, boosting the degree of a vertex in the underlying auxiliary tree by δ, using Lemma 3.3.4. By discarding trees when the sum of the orders of the trees gets too large, we will show that one of the trees in our forest will eventually obtain the desired order after finitely many steps of the algorithm. The details follow.

Initiate the process by fixing $U_0 \subset U$ to be an arbitrary subset of αn vertices, $W_0 = \emptyset \subset W$ to be empty and $\mathcal{D}_1, \dots, \mathcal{D}_\ell$ with $\ell = \alpha n$, to be the diamond trees which are defined to be the single vertices in U_0. That is, for $i \in [\ell]$, the K_r-diamond tree $\mathcal{D}_i = (T_i, R_i, \Sigma_i)$ corresponds to an auxiliary tree T_i which is just a single vertex and thus R_i is also a single vertex and Σ_i is empty. In general, at each step of the process we will have a family $\mathcal{D}_1, \dots, \mathcal{D}_\ell$ (for some $\ell \in \mathbb{N}$) of vertex-disjoint K_r-diamond trees such that for each i, the diamond tree $\mathcal{D}_i = (T_i, R_i, \Sigma_i)$ is δ-scattered, has $R_i \subset U_0$ and $\Sigma_i \subset G[W_0]$. Furthermore, we will have that $U_0 = \bigcup_{i \in [\ell]} R_i$ and $W_0 = \bigcup_{i \in \ell} V(\Sigma_i) \subset W$ and maintain throughout that $\alpha n \leq |U_0| \leq 2\alpha n$ and $|W_0| \leq 2(r-1)\alpha n$

Now at each step, given such a set U_0 and family $\mathcal{D}_1, \dots, \mathcal{D}_\ell$, we apply Lemma 3.3.4 with $U_1 = U \setminus U_0$ and $U_2 = W \setminus W_0$, noting that the conditions on the size of U and W in the statement of the lemma and the imposed conditions on the size of U_0 and W_0 throughout the process indeed allow Lemma 3.3.4 to be applied. Thus, we find a K_r-diamond star $\mathcal{D}^* = (T^*, R^*, \Sigma^*)$ of order $\delta + 1$ with centre $v_0 \in U_0$, $R^* \setminus \{v_0\} \subset U \setminus U_0$ and $\Sigma^* \subset K_{r-1}(G[U_2])$ a K_{r-1}-matching. As U_0 is the union of the removable vertices of the family of diamond trees, we have that there is some $i_0 \in [\ell]$ such that $v_0 \in R_{i_0}$. We then update $\mathcal{D}_{i_0}$ by adjoining the diamond star to the tree at v_0, we add all the vertices of R^* to U_0 and all the vertices of the $(r-1)$-cliques in Σ^*, to W_0. Now if there is a K_r-diamond tree among the (new) family $\mathcal{D}_1, \dots, \mathcal{D}_\ell$ which has order at least z, we take such a diamond tree as $\mathcal{D}_{sc}$ and finish the process. If not, then we look at the size of U_0. If $|U_0| < 2\alpha n$, we continue to the next step. If $|U_0| \geq 2\alpha n$, then we sequentially discard arbitrary K_r-diamond trees $\mathcal{D}_j = (T_j, R_j, \Sigma_j)$ from the family. That is, we choose a $\mathcal{D}_j$ in the family, delete R_j from U_0 and delete the vertices that belong to $(r-1)$-cliques in Σ_j from W_0. We continue discarding diamond trees until $|U_0| \leq 2\alpha n$. Note that as $|R_j| < z \leq \alpha n$ for all j, the updated U_0 at the end of this discarding process will have size at least αn as required. We then move to the next step.

All the diamond trees in our family are δ-scattered throughout the process and also W_0, as the set of vertices featuring in interior cliques of a family of K_r-diamond trees whose orders add up to less than $2\alpha n$, has size less than $2(r-1)\alpha n$ throughout. It is also clear that as the order of any diamond tree in our collection grows by at most δ in each step, the order of the diamond tree which is found by the algorithm will be at most $z + \delta$. It only remains to check that the algorithm terminates but this is guaranteed because the number of diamond trees is decreasing throughout the process. Indeed, we never add new diamond trees to the family and every $\frac{\alpha n}{\delta}$ steps we have to discard at least one diamond tree from the family. If the algorithm does not terminate after finding an appropriate $\mathcal{D}_{sc}$, then eventually we will be left with just one diamond tree $\mathcal{D}_1$ in the family, but at this point the order of $\mathcal{D}_1$ would be at least $\alpha n \geq z$, contradicting that the algorithm is still running. $\square$

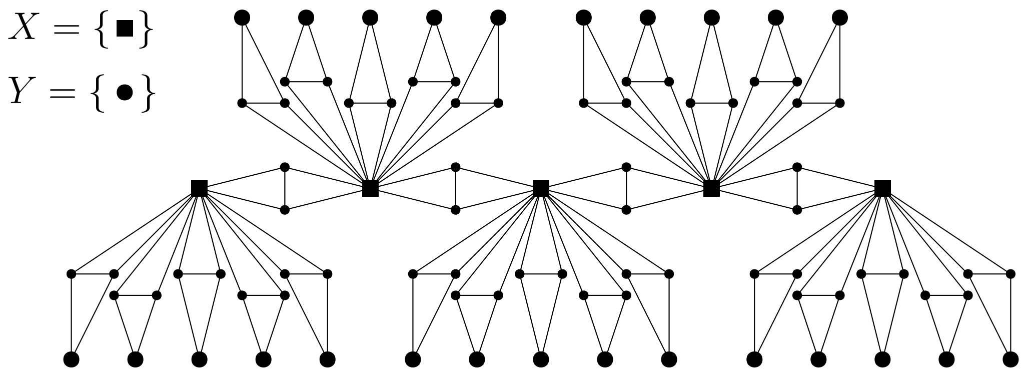


FIGURE 3.5: A 6-scattered K_3-diamond tree.

Using Lemmas 3.3.4 and 3.3.5 we can now deduce Proposition 3.3.1.

Proof of Proposition 3.3.1. Fix $\varepsilon > 0$ small enough to apply Lemmas 3.3.4 and 3.3.5 and small enough to force n to be sufficiently large in what follows. Let us first deal with the case when $z \leq \delta := \alpha^2 p^{r-1} n$. Here, we arbitrarily partition U into U_0 and U_1 of size at least αn, fix $U_2 = W$ and apply Lemma 3.3.4 to get a K_r-diamond star $\mathcal{D}^* = (T^*, R^*, \Sigma^*)$ of order $1 + \delta$ with $R^* \subset U$ and $\Sigma^* \subset K_{r-1}(G[W])$ a K_{r-1}-matching in W. Let $x \in R^*$ be the only non-leaf vertex in R^* and define $X = \{x\}$. Further, let $Y \subset R^* \setminus X$ be an arbitrary subset of $z - 1$ vertices. Now taking $\rho^* : V(T^*) \to R^*$ and $\sigma^* : E(T^*) \to \Sigma^*$ to be the defining bijective maps for $\mathcal{D}^*$, note that for any $Y' \subset Y$, the set of vertices $\{\rho^{*-1}(v) : v \in Y' \cup X\} \subset V(T^*)$ span a sub-tree (or rather a sub-star) of T^*, say T. Therefore, taking $\mathcal{D} = (T, X \cup Y', \Sigma)$ where $\Sigma := \{\sigma^*(e) : e \in E(T)\}$ defines a K_r-diamond tree with removable vertices $Y' \cup X$. Therefore (1), (2) and (3) of the proposition are all satisfied.

When $\delta < z \leq \alpha n$, the proof is similar. We apply Lemma 3.3.5 to get a δ-scattered K_r-diamond tree $\mathcal{D}_{sc} = (T_{sc}, R_{sc}, \Sigma_{sc})$ as given by the lemma and define $X \subset R_{sc}$ to be the non-leaves of $\mathcal{D}_{sc}$. See Figure 3.5 for an example. In order to bound $|X|$ and prove property (2), we appeal to Lemma 3.3.3 which gives that

$$|X| \leq \frac{|R_{sc}| - 2}{\delta - 1} \leq \frac{z + \delta - 2}{\delta - 1} \leq \frac{2z}{\delta},$$

using that $z \geq \delta$ in the final inequality.

We note that for n large (using Fact 3.2.2) we have that $\delta \geq 4$ (indeed δ is $\Omega(n^{(r-1)/(2r-3)})$), implying that $|X| \leq \frac{z}{2}$. We fix $Y \subset R_{sc} \setminus X$ to be an arbitrary subset of size $z - |X|$ and claim that the conditions (1), (2) and (3) of the proposition are all satisfied. Indeed it remains only to prove (3) and this follows similarly to above, by taking sub-diamond trees of $\mathcal{D}_{sc}$. In detail, fix some $Y' \subset Y$ and let $R = Y' \cup X$. Then if $\rho_{sc} : V(T_{sc}) \to R_{sc}$ and $\sigma_{sc} : E(T_{sc}) \to \Sigma_{sc}$ are the defining bijective maps for $\mathcal{D}_{sc}$, we have that the set of vertices $\{\rho_{sc}^{-1}(v) : v \in R\}$ spans a subtree $T \subset T_{sc}$. Indeed we simply deleted leaves from T_{sc}, namely $\rho_{sc}^{-1}(x)$ for $x \in R_{sc} \setminus Y'$. Taking $\Sigma = \{\sigma_{sc}(e) : e \in E(T)\}$, we have that $\mathcal{D} = (T, R, \Sigma)$ is the desired diamond tree. □

3.4 Cascading absorption through orchards

In this section we discuss orchards in our (p, β)-bijumbled graphs. We begin in Section 3.4.1 by proving Lemma 3.1.4 which details conditions for when one orchard absorbs another. In Section 3.4.2, we then discuss the existence of shrinkable orchards, addressing Proposition 3.1.8 which tells us that we can find shrinkable orchards of all desired orders in the graphs we are interested in. The proof of Proposition 3.1.8 requires many ideas and two distinct approaches. Therefore, we defer the majority of the work to later sections and simply reduce the proposition here, splitting it into two 'subpropositions' which will be tackled separately. Recall that Lemma 3.1.4 and Proposition 3.1.8 were the two ingredients we needed to prove the cascading absorption through constantly many orchards in the proof of Theorem I.

3.4.1 Absorbing orchards

Recall the definition (Definition 3.1.3) of an orchard and that we say a $(K, M)_r$-orchard $\mathcal{O}$ *absorbs* a $(k, m)_r$-orchard $\mathcal{R}$ if there is a $((r-1)k, M)_r$-suborchard $\mathcal{O}' \subset \mathcal{O}$, such that there is a K_r-factor in $G[V(\mathcal{R}) \cup V(\mathcal{O}')]$.

In this section we prove Lemma 3.1.4, which is a generalisation of [145, Lemma 3.5]. The lemma gives some sufficient conditions for an orchard to be able to absorb another orchard. Our proof scheme follows that of [84] which gives a polynomial time two-phase algorithm for finding the necessary K_r-factor. The algorithm is a simple greedy algorithm and works by absorbing each diamond tree $\mathcal{B}$ in the small orchard $\mathcal{R}$, one at a time. In more detail, for each diamond tree $\mathcal{B}$ in $\mathcal{R}$, we find $r-1$ diamond trees $\mathcal{D}_1, \ldots, \mathcal{D}_{r-1} \in \mathcal{O}$ such that there is a copy of K_r traversing the sets of removable vertices of $\mathcal{B}$ and the diamond trees $\mathcal{D}_1, \ldots, \mathcal{D}_{r-1}$. This implies that there is a K_r-factor in $G[V(\mathcal{B}) \cup V(\mathcal{D}_1) \cup \ldots \cup V(\mathcal{D}_{r-1})]$ (see Observation 3.1.2) and so we can add the $\mathcal{D}_i$ to the suborchard $\mathcal{O}'$, forbid them from being used again, and move to the next diamond tree $\mathcal{B}' \in \mathcal{R}$. Note that typically, we expect to succeed with this process. Indeed, the set of removable vertices of diamond trees in $\mathcal{O}$ is linear in size (and remains linear even after forbidding diamond trees $\mathcal{D} \in \mathcal{O}$ used for previous $\mathcal{B} \in \mathcal{R}$) and so a typical vertex has $\Omega(pn)$ neighbours among this set of removable vertices. Hence, appealing to Corollary 3.2.5 (1) (i) which states that sets of size $\Omega(pn)$ host copies of K_{r-1}, we can expect to find a copy of K_{r-1} in the neighbourhood of a typical removable vertex of $\mathcal{B} \in \mathcal{R}$ which lies on the removable vertices of diamond trees in $\mathcal{O}$. As long as this copy of K_{r-1} traverses sets of removable vertices of distinct diamond trees in $\mathcal{O}$, we will succeed. With a few extra ideas and a bit of preprocessing (for example partitioning $\mathcal{O}$ into $r-1$ suborchards at the start), this intuition holds true and we can successfully greedily start to build $\mathcal{O}'$. In fact, if kM is small compared to pn, we can fully form $\mathcal{O}'$ in this way and no second phase is necessary. However, if kM is large compared to pn we may run into trouble as with this greedy approach, it may be the case that the neighbourhood of a removable vertex v of a diamond tree $\mathcal{B} \in \mathcal{R}$ has too small a size by the time we come to considering $\mathcal{B}$. Indeed, as we run this greedy process, we forbid the diamond trees (and their removable vertices) which we add to $\mathcal{O}'$, from being used again. This could result in v having much fewer than pn neighbours in the removable vertices of diamond trees in (the remainder of) $\mathcal{O}$ and so we have no guarantee of finding a copy of K_{r-1} in this neighbourhood. We resolve this issue by running a two-phase algorithm and reserving half of $\mathcal{O}$ for the second phase. The key point is that if a diamond tree $\mathcal{B}$ fails in the first round then it must be the case that *all* of the removable vertices of $\mathcal{B}$ have small neighbourhoods amongst the removable vertices of diamond trees in $\mathcal{O}$. Given that throughout the process, many diamond trees in $\mathcal{O}$ will remain available to use, pseudorandomness (more precisely, Corollary 3.2.5 (1) (ii)) tells us that the number of vertices that do not have large enough neighbourhoods, is relatively small. Hence, as each diamond tree $\mathcal{B} \in \mathcal{R}$ which failed in the first phase, has a set of removable vertices which are atypical in this way, we can upper bound the number of diamond trees in $\mathcal{R}$ that fail in the first round. This upper bound will then be used to show that in the second phase, we are successful with each diamond tree, as throughout the second round, the number of removable vertices being forbidden (due to being used to absorb other diamond trees in $\mathcal{R}$) will be negligible and so the neighbourhoods of vertices amongst the removable vertices of diamond trees in the half of $\mathcal{O}$ reserved for this second phase, will remain large.

Proof of Lemma 3.1.4. We fix constants $\tau, \eta' < \frac{\eta\zeta^2}{2^{3r}r^2}$ and choose $\varepsilon > 0$ small enough to apply Lemma 3.2.3 with $\eta_{3.2.3} = \eta'$ and Corollary 3.2.5 with τ as above. Let $\mathcal{O} = \{\mathcal{D}_1, \ldots, \mathcal{D}_K\}$ be the $(K, M)_r$-orchard with each $\mathcal{D}_i = (T_i, R_i, \Sigma_i)$ being a K_r-diamond tree of order between M and $2M$. We start by arbitrarily partitioning $\mathcal{O}$ into $2(r-1)$ suborchards of size as equal as possible so that $\mathcal{O} = \bigcup_{j=1}^{2(r-1)} \mathcal{O}_j$ and each $\mathcal{O}_j$ is a $(K_j, M)_r$-orchard with $K_j = \frac{K}{2(r-1)} \pm 1 \geq \frac{K}{2r}$. For $j \in [2(r-1)]$, we let

$$Y_j := \bigcup_{i:\mathcal{D}_i \in \mathcal{O}_j} R_i$$

be the set of removable vertices of the diamond trees which feature in the j^{th} suborchard. Note that $|Y_j| \geq K_j M \geq \frac{KM}{2r} \geq \frac{\zeta n}{2r}$ for each $j \in [2(r-1)]$. We define B to be the set of vertices $v \in V \setminus V(\mathcal{O})$ such that for some $j \in [2(r-1)]$, $\deg(v; Y_j) < \frac{p}{2}|Y_j|$. By Lemma 3.2.3 (i), we have that

$$|B| < \frac{2(r-1)\eta' p^{2r-4} n^2}{\min_j |Y_j|} \leq \eta p^{2r-4} n,$$

due to our lower bound on the size of the $|Y_j|$ and our upper bound on η'.

Now as in the statement of the lemma, consider a $(k,m)_r$-orchard $\mathcal{R} = \{\mathcal{B}_1, \ldots, \mathcal{B}_k\}$ of diamond trees whose vertices lie in $V \setminus (B \cup V(\mathcal{O}))$. For $i' \in [k]$, let $Q_{i'}$ be the set of removable vertices of the diamond tree $\mathcal{B}_{i'}$. We will show that for each $i' \in [k]$, there exists distinct indices $i_1 = i_1(i'), \ldots, i_{r-1} = i_{r-1}(i') \in [K]$ such that there is copy of K_r which traverses the sets $Q_{i'}$ and $R_{i_1}, \ldots, R_{i_{r-1}}$, where R_{i_1} is the set of removable vertices of $\mathcal{D}_{i_1}$ and likewise for $i_2, \ldots, i_{r-1}$. Now, from Observation 3.1.2, we have that for such an r-tuple $\mathcal{B}_{i'}, \mathcal{D}_{i_1}, \ldots, \mathcal{D}_{i_{r-1}}$, there is a K_r-factor in $G[V(\mathcal{B}_{i'}) \cup V(\mathcal{D}_{i_1}) \cup \ldots \cup V(\mathcal{D}_{i_{r-1}})]$. We will prove that one can choose such indices $i_1, \ldots, i_{r-1}$ for each $i' \in [k]$ in such a way that no $i \in [K]$ is chosen more than once. That is, for $i' \neq j' \in [k]$, the sets $\{i_1(i'), \ldots, i_{r-1}(i')\}$ and $\{i_1(j'), \ldots, i_{r-1}(j')\}$ are disjoint. Therefore our suborchard $\mathcal{O}' \subset \mathcal{O}$ can simply be defined to be the union of all the choices of $\mathcal{D}_{i_j(i')}$ for $i' \in [k]$ and $j \in [r-1]$.

We now show how to find the indices $i_1(i'), \ldots, i_{r-1}(i')$ for each $i' \in [k]$. We will achieve this via the following simple algorithm. We initiate the first round of the algorithm with $\mathcal{O}' = \emptyset$, $I = [k]$, $\mathcal{P}_j = \mathcal{O}_j$ and $Z_j = Y_j$ for $1 \leq j \leq r-1$. Note that the $\mathcal{O}_j$ for $r \leq j \leq 2(r-1)$ do not feature in these definitions. This is because we will not use any diamond trees that lie $\cup_{j=r}^{2(r-1)} \mathcal{O}_j$ in this first round. Now the algorithm runs as follows. For $i' = 1, \ldots, k$, we check if there exists some set $\{\mathcal{D}_{i_j} \in \mathcal{P}_j : j \in [r-1]\}$ such that there is a K_r traversing $Q_{i'}$ and the sets of removable vertices $R_{i_1}, \ldots, R_{i_{r-1}}$. If this is the case then we delete $\mathcal{D}_{i_j}$ from $\mathcal{P}_j$ and add it to $\mathcal{O}'$ for $j \in [r-1]$ and we also delete R_{i_j} from Z_j for all $j \in [r-1]$. Furthermore, we delete i' from I and move to the next index $i'+1$ (or finish this round if $i' = k$). If it is not the case that such diamond trees exist in the orchards $\mathcal{P}_j$, then we simply leave i' as a member of I and move on to the next index.

At the end of the first round, we have some set I of indices remaining. We define $t := |I|$ at this point. We will now use diamond trees in the orchards $\mathcal{O}_j$ with $r \leq j \leq 2(r-1)$ to absorb these remaining diamond trees $\mathcal{B}_{i'}$ with $i' \in I$. Thus we reset the process, setting $\mathcal{P}_j = \mathcal{O}_{j+r-1}$ and $Z_j = Y_{j+r-1}$ for all $j \in [r-1]$. We then follow the same simple process in the second round as we did in the first, running through the (remaining) $i' \in I$ in order and trying to find an appropriate set of diamond trees $\{\mathcal{D}_{i_j} \in \mathcal{P}_j : j \in [r-1]\}$ at each step. We claim that in this second round, we can find such a set for every $i' \in I$ and so by the end of the second round, the set I is empty and $\mathcal{O}'$ is such that $G[V(\mathcal{R}) \cup V(\mathcal{O}')]$ hosts a K_r-factor.

In order to prove this, our analysis splits into two cases. First consider when $kM < \frac{\zeta pn}{16r}$. In this case, we in fact have that the second round is not even necessary as all indices succeed in the first round. Indeed, note that every time we are successful for an index i', we delete at most $2M$

vertices from each of the Z_j. Therefore, at any instance in the first round of the process, we have that any vertex v which is not in B has

$$\deg(v; Z_j) \geq \frac{p|Y_j|}{2} - 2kM \geq \frac{\zeta pn}{4r} - 2kM \geq \frac{\zeta pn}{8r} \geq \tau pn,$$

for all $j \in [r-1]$, using our lower bound on the $|Y_j|$ and our upper bound on kM. But then, by Corollary 3.2.5 (1) (i) (applied in this instance with $\tilde{G}$ being the empty graph and $G' = G$), there exists a copy of K_{r-1} traversing the sets $N(v; Z_j)$ for $1 \leq j \leq r-1$. When v is any vertex in the removable set of vertices $Q_{i'}$ for some diamond tree $\mathcal{B}_{i'}$ in the process, this gives a copy of K_r traversing $Q_{i'}$ and some sets of removable vertices R_{i_j} for diamond trees $\mathcal{D}_{i_j} \in \mathcal{P}_j$, $j \in [r-1]$, as desired. In this way, we see that the process succeeds in every step of the first round to find a suitable $\{i_j(i') : j \in [r]\}$ for each $i' \in [k]$ and I is empty (i.e. $t = 0$) at the end of the round. Note that we used here that the vertices of $Q_{i'}$ are not in B.

When $\frac{\zeta pn}{16r} \leq kM \leq mK$, the second round may be necessary and we start with estimating t, the size of I after the first round. Now note that at the end of the first round, *before* we reassign the sets Z_j to removable vertices in diamond trees in $\mathcal{O}_{j+r-1}$ for $j \in [r-1]$, if we take $Q = \bigcup_{i' \in I} Q_{i'}$, we have that there is no K_r traversing Q and the sets $Z_1, \ldots, Z_{r-1}$. Indeed, otherwise there would be an $i' \in I$ and a vertex $v \in Q_{i'} \subseteq Q$ which is contained in a K_r with a set of vertices $\{v_{i_j} \in Z_j : j \in [r-1]\}$. This contradicts that the index i' failed to find a suitable set of i_j in the first round. Thus, at the end of the first round, there is no K_r traversing Q, and the Z_j, $j \in [r-1]$. Moreover we have that

$$|Z_j| \geq \frac{KM}{2r} - 2kM \geq \frac{KM}{4r} \geq \frac{\zeta n}{4r},$$

using the upper bound on k from (3.1.1) and the fact that at most $2M$ vertices are deleted from Z_j every time we are successful with an index $i' \in I$. Thus, we can conclude from Corollary 3.2.5 (1) (ii) that at the end of the first round, $tm < |Q| < \tau p^{2r-4} n$. Therefore

$$t < \frac{\tau p^{2r-4} n}{m} \leq \frac{\tau p^2 n}{m} \leq \frac{16\tau rpK}{\zeta} \leq \frac{16\tau rpn}{\zeta M} \leq \frac{\zeta pn}{16rM},$$

where we used here our lower and upper bounds on kM to give an upper bound on $\frac{pn}{m}$ in the third inequality, the fact that $KM \leq n$ in the fourth inequality and our upper bound on α in the final inequality.

We now turn to analyse the second round. Using our upper bound on t, we can upper bound the number of vertices deleted in each Z_j throughout the second round, and using this we have that for any vertex v not in B, any $j \in [r-1]$ and at any point in the second round,

$$\deg(v; Z_j) \geq \frac{p|Y_{j+r-1}|}{2} - 2tM \geq \frac{\zeta pn}{4r} - \frac{\zeta pn}{8r} \geq \frac{\zeta pn}{8r}.$$

Thus we can repeat the argument used for the case when kM was small, seeing that at every step in the second round we are successful in finding an appropriate set of i_j for $j \in [r-1]$ for each $i' \in I$. This completes the proof. □

3.4.2 Shrinkable orchards

Here we are concerned with the existence of shrinkable orchards in pseudorandom graphs and verifying Proposition 3.1.8. We encourage the reader to remind themselves of Definitions 3.1.5 and 3.1.7 as well as Observation 3.1.6. In order to prove Proposition 3.1.8, we will appeal to the methods of Sections 2.6 and 2.7. We will use Theorem 2.7.3 to reduce the problem to establishing the existence of perfect fractional matchings in the appropriate K_r-hypergraphs and we will then employ Lemmas 2.6.3 and 2.6.4 to find these perfect fractional matchings. In order that our hypergraph has the desired properties to apply these lemmas we need to choose the diamond trees which define our orchard carefully.

It turns out that different arguments are needed for finding shrinkable orchards of different orders. In Section 3.5 we show how to find shrinkable orchards of small order, establishing the following intermediate proposition.

Proposition 3.4.1 (existence of shrinkable orchards of small order)**.** *For any $3 \leq r \in \mathbb{N}$ and $0 < \tau, \gamma < \frac{1}{2^{3r}}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$ and any vertex subset $U \subseteq V(G)$ with $|U| \geq \frac{n}{2}$. For any $m \in \mathbb{N}$ with*

$$1 \leq m \leq \min\{p^{r-2} n^{1-2r^3\gamma}, n^{7/8}\},$$

there exists a γ-shrinkable $(k, m)_r$-orchard $\mathcal{O}$ in $G[U]$ with $k \in \mathbb{N}$ such that $\alpha n \leq km \leq 2\alpha n$.

In Section 3.6 we then address shrinkable orchards with large order, which results in the following.

Proposition 3.4.2 (existence of shrinkable orchards of large order)**.** *For any $3 \leq r \in \mathbb{N}$ and $0 < \alpha, \gamma < \frac{1}{2^{12r}}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$ and any vertex subset $U \subseteq V(G)$ with $|U| \geq \frac{n}{2}$. For any $m \in \mathbb{N}$ with*

$$p^{r-1} n \leq m \leq n^{7/8},$$

there exists a γ-shrinkable $(k, m)_r$-orchard $\mathcal{O}$ in $G[U]$ with $k \in \mathbb{N}$ such that $\alpha n \leq km \leq 2\alpha n$.

The proof of Proposition 3.1.8 is basically immediate from Propositions 3.4.1 and 3.4.2 but we spell it out nonetheless.

Proof of Proposition 3.1.8. We split into a case analysis based on the density p of our graph G. First consider when $p \geq n^{-1/(10r)}$. Then we claim that $p^{r-2}n^{1-2r^3\gamma} \geq n^{7/8}$ and so the desired γ-shrinkable orchard of all orders up to $n^{7/8}$ can derived from Proposition 3.4.1. Indeed we have that $p^{r-2}n^{1-2r^3\gamma} \geq n^{1-\frac{r-2}{10r}-2r^3\gamma}$ and

$$1 - \frac{r-2}{10r} - 2r^3\gamma > 1 - \frac{1}{10} - \frac{1}{40} = \frac{7}{8},$$

due to our upper bound on γ (and lower bound on r).

When $p < n^{-1/(10r)}$, we have that $p \leq n^{-2r^3\gamma}$ again due to our upper bound on γ. Hence we can apply Proposition 3.4.1 to find γ-shrinkable orchards of orders $m \leq n^{7/8}$ such that $m < p^{r-1}n \leq p^{r-2}n^{1-2r^3\gamma}$ and apply Proposition 3.4.2 to find γ-shrinkable orchards with orders m such that $p^{r-1}n \leq m \leq n^{7/8}$. This settles all cases and so gives the proposition. $\square$

In both cases, a simpler argument works for the extreme cases, that is, when the order is small in Proposition 3.4.1 or when the order is large in Proposition 3.4.2. Extra ideas are then needed to push the approaches, extending the ranges of the two propositions so that they meet and cover all desired orders. In more detail, an easier form of Proposition 3.4.1 can cover orders which get close to $p^{r-1}n$ (see Proposition 3.5.5). Again the separation required depends on γ, explicitly $m \leq p^{r-1}n^{1-r^3\gamma}$. This is already enough to cover all desired orchard orders when p is large. On the other hand, a basic form of the argument for large order orchards gives shrinkable orchards of order at least $p^{r-1}n$ when p is large and of order at least p^{1-r} when p is smaller (see Proposition 3.6.3). Interestingly, Fact 3.2.2 implies exactly that $p^{1-r} = \Omega(p^{r-2}n)$ always and so proves that when p is small (close to the lower bound of $\Omega(n^{-1/(2r-3)})$) and our bijumbled graph is sparse, both the simpler arguments for small orders and large orders *as well as* their extensions are needed. Indeed using the simpler version, Proposition 3.5.5, for small orders and the full power of Proposition 3.4.2 leaves a small gap in the orders and so does using Proposition 3.4.1 in conjunction with the easier Proposition 3.6.3. In order to help the reader through the next two sections, in both cases we begin by presenting the weaker versions of the statements we need. This then lays the foundation for the full proofs and allows us to discuss the more technical aspects needed to push the ranges for which we can prove the existence of shrinkable orchards.

3.5 Shrinkable orchards of small order

Our first argument for proving the existence of shrinkable orchards works provided the order of the orchard is not too large, establishing Proposition 3.4.1. Before embarking on this we have to go through several steps. Firstly, in Section 3.5.1, we generalise the theory of shrinkable orchards built up in Section 3.1, allowing slightly more flexibility for our consequent proofs. In Section 3.5.2, we then use the theory of perfect fractional matchings (see Sections 2.6 and 2.7) to give conditions that guarantee an orchard is shrinkable. In Section 3.5.3, we show how this immediately implies the existence of shrinkable orchards of small order. However this falls short of Proposition 3.4.1 and in the rest of this section we push the ideas to extend the range of orders we can cover, showing how to cleverly choose diamond trees of our orchard in Section 3.5.4 which allows us to prove the full Proposition 3.4.1 in Section 3.5.5.

3.5.1 From orchards to systems

We begin by generalising our definitions slightly, allowing us to work not just with orchards but also set systems.

Definition 3.5.1. Given a graph G we say a set of pairwise disjoint subsets $\mathbf{\Lambda} \subset 2^{V(G)}$ is a (k,m)*-system* if $m \leq |Q| \leq 2m$ for each $Q \in \mathbf{\Lambda}$ and $|\mathbf{\Lambda}| = k$. That is, a (k,m)-system is just a set family of k disjoint vertex sets of size between m and $2m$.

Now given a (k,m)-system $\mathbf{\Lambda}$ in a graph G, the K_r*-hypergraph* generated by $\mathbf{\Lambda}$, denoted $\mathscr{H} = \mathscr{H}(\mathbf{\Lambda};r)$ is the r-uniform hypergraph with vertex set $V(\mathscr{H}) = \mathbf{\Lambda}$ and with $\{Q_{i_1},\dots,Q_{i_r}\} \in \binom{\mathbf{\Lambda}}{r}$ forming a hyperedge in $\mathscr{H}$ if and only if there is a copy of K_r traversing the sets $Q_{i_1},\dots,Q_{i_r}$ in G.

Finally for $0 < \gamma < 1$, we say a (k,m)-system $\mathbf{\Lambda}$ in a graph G is γ*-shrinkable* (with respect to r) if there exists a subsystem $\mathbf{\Gamma} \subset \mathbf{\Lambda}$ of size at least γk such that for *any* subsystem $\mathbf{\Gamma}' \subseteq \mathbf{\Gamma}$, we have that there is a matching in $\mathscr{H} := \mathscr{H}(\mathbf{\Lambda} \setminus \mathbf{\Gamma}';r)$ covering all but $k^{1-\gamma}$ of the vertices of $\mathscr{H}$.

Note that given a $(k,m)_r$-orchard $\mathcal{O}$ we can define a (k,m)-system $\mathbf{\Lambda}$ as the sets of removable vertices of diamond trees in $\mathcal{O}$. That is, $\mathbf{\Lambda} := \{R_{\mathcal{D}} : \mathcal{D} \in \mathcal{O}\}$. Then the K_r-hypergraphs generated by $\mathcal{O}$ and $\mathbf{\Lambda}$ coincide i.e. $\mathscr{H}(\mathbf{\Lambda};r) = \mathscr{H}(\mathcal{O})$, and $\mathcal{O}$ is γ-shrinkable if and only if $\mathbf{\Lambda}$ is a γ-shrinkable. However Definition 3.5.1 allows us slightly more flexibility, giving us the ability to focus on subsets of removable vertices. The next observation highlights this and although the result is trivial, it will be important for our proofs.

Observation 3.5.2. Suppose $r \geq 3$, $0 < \gamma < 1$ and $\mathcal{O} = \{\mathcal{D}_1,\dots,\mathcal{D}_k\}$ is a $(k,m)_r$ orchard in a graph G with R_i being the set of removable vertices of $\mathcal{D}_i$ for $i \in [k]$. Then if $\mathbf{\Lambda} = \{Q_1,\dots,Q_k\}$ is some (k,m')-system (for some m') such that $Q_i \subseteq R_i$ for $i \in [k]$ and $\mathbf{\Lambda}$ is γ-shrinkable (with respect to r), then $\mathcal{O}$ is also γ-shrinkable.

It will become clear why such a relaxation is useful for us and thus why we make this switch to working with set systems.

3.5.2 Sufficient conditions for shrinkability

We now explore conditions on set systems which guarantee shrinkability. We begin by giving some local conditions on a set system which guarantee that it is shrinkable given that it lies in the pseudorandom graphs we are interested in.

Lemma 3.5.3. *For any $3 \leq r \in \mathbb{N}$ and $0 < \alpha, \gamma < \frac{1}{2^r r^2}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$. Suppose $\mathbf{\Lambda} \subset 2^{V(G)}$ is a (k, m)-system such that $m \leq n^{7/8}$, $km \geq \alpha n$, $pk \geq n^{\gamma}$ and:*

1. *There exists a subsystem $\mathbf{\Gamma} \subset \mathbf{\Lambda}$ such that $|\mathbf{\Gamma}| = \gamma k$ and the following holds with $Y := \cup\{P : P \in \mathbf{\Lambda} \setminus \mathbf{\Gamma}\}$. For every $Q \in \mathbf{\Lambda}$, there exists a vertex $v \in Q$ such that*
$$\deg_G(v; Y) \geq \alpha p k m;$$
2. *For any $u \in \cup_{P \in \mathbf{\Lambda}} P$ and $Q \in \mathbf{\Lambda}$, we have that $\deg_G(u; Q) \leq p^{r-1} n^{1-r^3\gamma}$.*

Then $\mathbf{\Lambda}$ is γ-shrinkable with respect to r.

Let us make a few remarks before proving the lemma. Firstly, note that condition (1), despite the slight technicality necessary to avoid dependence on sets in $\mathbf{\Gamma}$, is a natural condition. Indeed, we are requiring that at least one vertex in each set is well connected to the other sets and has a constant fraction of the degree that we would expect on average. Condition (2) is perhaps more mysterious as it is unclear why having an *upper* bound on the degree of a vertex to another set in the system is advantageous. The point is that this guarantees that each of the vertices has a neighbourhood that is well-spread across the other sets of the set system, without being too concentrated on any one other set. Within the proof this necessity manifests itself as we appeal to Theorem 2.7.3 which tells us that a hypergraph $\mathcal{H}$ contains a large matching if whenever we remove a small collection of edges of $\mathcal{H}$ (that contain edges of some 2-uniform graph J which satisfies a maximum degree condition), the resulting hypergraph has a perfect fractional matching. So here we will need that when we disallow edges between certain pairs of sets from being used (dictated by the graph J), we do not significantly alter the graph G in which we work and the graph is still dense enough to prove the existence of a perfect fractional matching in the relevant K_r-hypergraph. Indeed, we will appeal to Lemma 2.6.4 to prove the existence of our perfect fractional matchings. That lemma requires us to prove the existence of large fans at each vertex of the hypergraph and the existence of edges in many different subsets of vertices. These conditions boil down to finding appropriate transversal copies of K_r in the underlying graph G. It turns out that we can find these copies of K_r given our conditions in the graph and the minimum degree condition (1) of the statement of Lemma 3.5.3. In fact, we can find the appropriate copies of K_r even when we forbid the edges of some sparse subgraph $\tilde{G}$ of G from being used (this is

Corollary 3.2.5 (1)). The condition (2) will guarantee that when looking for perfect fractional matchings in sub-hypergraphs (dictated by some choice of J) of the K_r-hypergraph, the forbidden edges of the *graph* G (also dictated by J) form a sparse subgraph $\tilde{G}$ and so we can still find the necessary copies of K_r. The details follow in the proof.

Proof of Lemma 3.5.3. Fix $\varepsilon > 0$ small enough to apply Corollary 3.2.5 with $\tau < \frac{\alpha^3}{2r}$ and small enough to force n to be sufficiently large in what follows. Fixing $\mathbf{\Gamma} \subset \mathbf{\Lambda}$ as in condition (1), we will show that for any $\mathbf{\Gamma}' \subset \mathbf{\Gamma}$, the K_r-hypergraph $\mathscr{H} = \mathscr{H}(\mathbf{\Lambda} \setminus \mathbf{\Gamma}'; r)$ has a matching covering all but $k^{1-\gamma}$ vertices of $\mathscr{H}$. So fix such a $\mathbf{\Gamma}'$, let $\mathbf{\Lambda}^* := \mathbf{\Lambda} \setminus \mathbf{\Gamma}'$ and let $\mathscr{H} := \mathscr{H}(\mathbf{\Lambda}^*; r)$.

In order to show the existence of a large matching in $\mathscr{H}$, we appeal to Theorem 2.7.3. So let us fix $N = |V(\mathscr{H})|$ and note that as $N \geq (1-\gamma)k$ and $k \geq \alpha n^{1/8}$ due to our conditions on k and m, we can assume that N is sufficiently large in what follows. Now fix some 2-uniform graph J on $V(\mathscr{H})$ of maximum degree at most $N^{r^2\gamma}$. If we can show that $\mathscr{H} \setminus \mathscr{H}_J$ contains a perfect fractional matching, then we are done by Theorem 2.7.3 as, because J was arbitrary, the theorem guarantees a matching covering all but at most $N^{1-\gamma} \leq k^{1-\gamma}$ vertices of $\mathscr{H}$.

In order to study $\mathscr{H} \setminus \mathscr{H}_J$, we look at the forbidden edges of G which J imposes. That is, we define

$$\tilde{G}_J := \bigcup_{\{Q_1,Q_2\}\in E(J)} G\left[Q_1, Q_2\right] \bigcup_{Q\in\mathbf{\Lambda}^*} G[Q]$$

where we recall that $G\left[Q_1, Q_2\right]$ denotes the set of all edges in G between the sets Q_1 and Q_2 and $G[Q]$ denotes all the edges induced by G in the set Q (which are also not used when considering $\mathscr{H} \setminus \mathscr{H}_J$). We then have that for any vertex $v \in V(G)$, $\deg_{\tilde{G}_J}(v) = 0$ if $v \notin \cup_{P\in\mathbf{\Lambda}^*} P$ and if $v \in Q \in \mathbf{\Lambda}^*$, then

$$\deg_{\tilde{G}_J}(v) \leq \sum_{P\in N_J(Q)\cup\{Q\}} \deg_G(v; P) \leq (N^{r^2\gamma}+1)p^{r-1}n^{1-r^3\gamma} \leq \tau^2 p^{r-1}n^{1-\gamma}, \tag{3.5.1}$$

using (2), the upper bound on the degrees in J and the fact that $N \leq n$.

Now defining $G'_J := G \setminus \tilde{G}_J$, we have that $\mathscr{H} \setminus \mathscr{H}_J$ is precisely the hypergraph obtained by viewing $\mathbf{\Lambda}^*$ as an (N, m)-system in G'_J and taking the K_r-hypergraph $\mathscr{H}^* = \mathscr{H}(\mathbf{\Lambda}^*; r)$ in G'_J. Indeed, as there are no edges of G'_J between two sets, Q_1 and Q_2 say, which form an edge in J, there can be no edge of $\mathscr{H}$ in $\mathscr{H}^*$ which contains both Q_1 and Q_2. We therefore switch from now on to considering $\mathscr{H}^*$ as the K_r-hypergraph generated by $\mathbf{\Lambda}^*$ in G'_J.

In order to prove the existence of a perfect fractional matching in $\mathscr{H}^*$, we will appeal to Lemma 2.6.4, fixing $M_1 = \alpha^2 pk$ and $M_2 = \alpha k$. Note that due to our lower bound on pk, we certainly have that $M_1 \geq r$. We thus need to check that conditions (i) and (ii) of that lemma hold. For (i), fix some $Q \in \mathbf{\Lambda}^*$. From (1) we have that there exists some vertex v in Q such that $\deg_G(v; W) \geq \alpha pkm$ where $W := \cup_{P \in \mathbf{\Lambda}^*} P$ and so taking $U := N_{G'_J}(v; W)$ we have that

$$|U| \geq \alpha pkm - \deg_{\tilde{G}_J}(v) \geq \frac{\alpha pkm}{2},$$

using (3.5.1). Moreover, due to (2), we can split U into disjoint sets $U_1, \ldots, U_{r-1}$ such that $|U_i| \geq \frac{\alpha pkm}{2r}$ for each i and we have that for any $P \in \mathbf{\Lambda}^*$, there exists an $i \in [r-1]$ such that $P \cap U \subset U_i$. That is, we simply partition U into $r-1$ roughly equal size parts such that vertices which lie in the same P end up in the same part. Condition (2) of the lemma guarantees that $U \cap P$ is small enough for each $P \in \mathbf{\Lambda}^*$ and so we can do this partition in such a way that each of the U_i are roughly equal in size. We will now repeatedly find $(r-1)$-cliques in G'_J traversing the U_i and build a fan F_Q of size M_1 in $\mathscr{H}^*$ focused at Q. We start with F_Q being empty and each time we find a copy $S = \{u_1, \ldots, u_{r-1}\} \in K_{r-1}(G'_J)$ in G'_J with $u_i \in U_i$ for $i \in [r-1]$ we have that there exists some sets $P_1, \ldots, P_{r-1} \in \mathbf{\Lambda}^*$ such that $u_i \in P_i$. We add the hyperedge between $P_1, \ldots, P_{r-1}$ and Q to the fan F_Q and delete any vertices in P_i from U_i for $i \in [r-1]$. We repeat this process and note that we are successful in every step until F_Q has size M_1. Indeed this follows from Corollary 3.2.5 (1) (i) as while $|F_Q| < M_1$, we have deleted at most[8] $M_1 2m \leq \frac{\alpha pkm}{4r}$ vertices from each U_i and so $|U_i| \geq \frac{\alpha pkm}{4r} \geq \tau pn$, using our upper bound on τ and our lower bound on km.

We now turn to verifying condition (ii) of Lemma 2.6.4. We will show that given any r-tuple of disjoint subsystems $\mathbf{\Gamma}_1, \mathbf{\Gamma}_2, \ldots, \mathbf{\Gamma}_r \subset \mathbf{\Lambda}^*$ such that $|\mathbf{\Gamma}_1| = M_1$ and $|\mathbf{\Gamma}_i| \geq \frac{\alpha k}{r}$ for $2 \leq i \leq r$, there exists a hyperedge of $\mathscr{H}^*$ with one endpoint in each of the $\mathbf{\Gamma}_i$. Indeed, this follows from Corollary 3.2.5 (1) (ii) as taking $U_i := \bigcup_{P \in \mathbf{\Gamma}_i} P$ for $i \in [r]$, we have that there exists an r-clique $S \in K_r(G'_J)$ traversing the U_i which in turn gives the hyperedge. The condition (ii) of Lemma 2.6.4 then clearly follows as any subsystem of $\mathbf{\Gamma}^*$ of size $M_2 = \alpha k$ can be split into $(r-1)$ subsystems of size at least $\frac{\alpha k}{r}$. Note that in both applications of Corollary 3.2.5 (1) above we used (3.5.1) to show that we could find cliques that avoid using edges of $\tilde{G}_J$. The lemma now follows from Lemma 2.6.4. □

As previously noted, condition (1) of Lemma 3.5.3 is somewhat weak and just requires that each set in the set system contains a vertex that acts typically. We now show how we can 'clean up' a set system; losing sets which do not have a typical vertex in order to recover condition (1). This allows us to focus on finding systems which satisfy condition (2) of Lemma 3.5.3.

Lemma 3.5.4. *For any $3 \leq r \in \mathbb{N}$ and $0 < \alpha, \gamma < \frac{1}{2^r r^2}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$. Suppose $\mathbf{\Lambda}' \subset 2^{V(G)}$ is a $((1+\gamma)k, m)$-system such that $m \leq n^{7/8}$, $\alpha n \leq km \leq 2\alpha n$, $pk \geq n^{\gamma}$ and for any $u \in \cup_{P \in \mathbf{\Lambda}'} P$ and $Q \in \mathbf{\Lambda}'$, we have that $\deg_G(u; Q) \leq p^{r-1} n^{1-r^3\gamma}$. Then there exists a (k, m)-system $\mathbf{\Lambda} \subset \mathbf{\Lambda}'$ which is γ-shrinkable with respect to r.*

[8]Here we use that every set in $\mathbf{\Lambda}$ has size at most $2m$.

Proof. We fix $\varepsilon > 0$ small enough to apply Lemma 3.5.3 and to apply Lemma 3.2.3 with $\eta < \frac{\gamma\alpha^2}{2}$. The method of the proof is simple; we aim to apply Lemma 3.5.3 and so obtain $\mathbf{\Lambda}$ from $\mathbf{\Lambda}'$ by losing the sets which violate condition (1) of that lemma. By Lemma 3.2.3 (i), there are few vertices which have small degree ($\leq \frac{p}{2}|Y|$) to any set Y which is large enough and so we can expect that we do not lose many sets when transitioning from $\mathbf{\Lambda}'$ to $\mathbf{\Lambda}$. One complication is that the definition of Y in condition (1) of Lemma 3.5.3 *depends* on the sets in the system and so we cannot guarantee that a set satisfying (1) continues to satisfy the condition once other sets have been removed. In order to handle this, we delete sets in the system one by one, creating a process which will terminate with a system which has the desired minimum degree condition. The details now follow.

We begin by fixing $\mathbf{\Gamma}_0 \subset \mathbf{\Lambda}'$ to be some arbitrary subsystem of size $(1-\gamma)k$ and we initiate the process by setting $\mathbf{\Theta} = \mathbf{\Lambda}'$ and setting a 'bin' system $\mathbf{\Phi}$ which we initiate as being empty, that is, we set $\mathbf{\Phi} = \emptyset$. Throughout the process we also define W so that

$$W := \bigcup_{Q \in \mathbf{\Gamma}_0 \cap \mathbf{\Theta}} Q$$

is the subset of vertices that lie in (sets that belong to) the current system $\mathbf{\Theta}$ as well as the system $\mathbf{\Gamma}_0$. Now the process runs as follows. If there is a set P in $\mathbf{\Theta}$ such that $\deg(v; W) < \alpha pkm$ for all $v \in P$, then we delete P from $\mathbf{\Theta}$ and add it to $\mathbf{\Phi}$. Hence if $P \in \mathbf{\Gamma}_0$ then we also remove P from W. We claim that this process terminates with $|\mathbf{\Phi}| \leq \gamma k$. Indeed if this were not the case then consider the process at the point where $|\mathbf{\Phi}| = \gamma k$. At this point we have that

$$|W| = \left| \bigcup_{Q \in \mathbf{\Gamma}_0 \setminus \mathbf{\Phi}} Q \right| \geq (|\mathbf{\Gamma}_0| - |\mathbf{\Phi}|)\, m = (k - 2\gamma k) m \geq \frac{km}{2} \geq \frac{\alpha n}{2}.$$

Now Lemma 3.2.3 (i) implies that at most

$$\frac{\eta p^{2r-4} n^2}{|W|} \leq \frac{2\eta p^{2r-4} n}{\alpha} < \gamma \alpha n \leq \gamma k m = |\mathbf{\Phi}| m,$$

vertices can have degree less than $\frac{p}{2}|W|$ to W. This leads to a contradiction. Indeed, it follows from how $\mathbf{\Phi}$ is defined that at this point in the process, $|\cup_{P \in \mathbf{\Phi}} P| \geq |\mathbf{\Phi}| m$ and for all $v \in \cup_{P \in \mathbf{\Phi}} P$, we have $\deg(v; W) < \alpha pkm < \frac{p}{2}|W|$. Indeed, if a vertex $v \in P \in \mathbf{\Phi}$ had a larger degree to W then P would not have been added to $\mathbf{\Phi}$ in the process.

Hence when the process terminates we have that $|\mathbf{\Phi}| \leq \gamma k$ and we have that $|\mathbf{\Theta}| = |\mathbf{\Lambda}' \setminus \mathbf{\Phi}| \geq k$. We fix $\mathbf{\Lambda} \subseteq \mathbf{\Theta}$ of size k so that $\mathbf{\Gamma}_0 \cap \mathbf{\Theta} \subseteq \mathbf{\Lambda}$. We also fix $\mathbf{\Gamma} \subseteq \mathbf{\Lambda} \setminus (\mathbf{\Gamma}_0 \cap \mathbf{\Theta})$ of size γk (which is possible as $|\mathbf{\Gamma}_0 \cap \mathbf{\Theta}| \leq (1-\gamma)k$). We claim that $\mathbf{\Lambda}$ is γ-shrinkable with respect to r. Indeed, this follows directly from Lemma 3.5.3 noting that condition (1) is satisfied in $\mathbf{\Lambda}$ with respect to $\mathbf{\Gamma}$ due to how we constructed $\mathbf{\Lambda}$. □

3.5.3 The existence of shrinkable orchards of small order

We are now ready to prove the existence of shrinkable orchards by appealing to Lemma 3.5.4. Indeed, we simply need to find orchards which satisfy the maximum degree condition given there. This condition is immediate when the order of the orchards which we aim for is sufficiently small, leading to the following easy consequence.

Proposition 3.5.5. *For any $3 \leq r \in \mathbb{N}$ and $0 < \alpha, \gamma < \frac{1}{2^{3r}}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p,β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$ and any vertex subset $U \subseteq V(G)$ with $|U| \geq \frac{n}{2}$. For any $m \in \mathbb{N}$ with*

$$1 \leq m \leq \min\{p^{r-1} n^{1-r^3\gamma}, n^{7/8}\},$$

there exists a γ-shrinkable $(k,m)_r$-orchard $\mathcal{O}$ in $G[U]$ with k such that $\alpha n \leq km \leq 2\alpha n$.

Proof. We fix $\varepsilon > 0$ small enough to apply Proposition 3.3.1 and Lemma 3.5.4 with α, γ as defined here and $k \in \mathbb{N}$ such that $\alpha n \leq km \leq 2\alpha n$. Note that due to our upper bound on m, we certainly have that $pk \geq n^\gamma$. We begin by finding a $((1+\gamma)k, m)_r$-orchard $\mathcal{O}'$ in $G[U]$. This can be done by repeated applications of Proposition 3.3.1. Indeed we initiate a process by fixing $U' = U$ and $\mathcal{O}' = \emptyset$ and at each step we find some K_r-diamond tree $\mathcal{D}$ of order m in U', add it to $\mathcal{O}'$ and delete its vertices from U'. We claim that we can do this until $\mathcal{O}'$ has size $(1+\gamma)k$. Indeed this follows because at any point in the process, $|V(\mathcal{O}')| \leq (1+\gamma)kmr = (1+\gamma)\alpha rn \leq \frac{n}{4}$ due to our upper bounds on α and γ. Therefore, throughout the process, we have that $|U'| \geq \frac{n}{4}$ and so it can be split into two disjoint sets of size at least $4\alpha rn$. Applying Proposition 3.3.1 with $z = m$ (and taking $Y' = Y$ in (3)) gives us the existence of the diamond tree at each step of this process.

Now defining $\mathbf{\Lambda}' = \{R_{\mathcal{D}} : \mathcal{D} \in \mathcal{O}'\}$, to be the $((1+\gamma)k, m)$-system generated by taking the sets of removable vertices of diamond trees that lie in $\mathcal{O}'$, we have that $\mathbf{\Lambda}'$ satisfies the hypothesis of Lemma 3.5.4 due to the fact that $m \leq p^{r-1} n^{1-r^3\gamma}$. Hence Lemma 3.5.4 implies the existence of a subsystem $\mathbf{\Lambda} \subset \mathbf{\Lambda}'$ of size k which is γ-shrinkable with respect to r. Finally taking $\mathcal{O} := \{\mathcal{D} \in \mathcal{O}' : R_{\mathcal{D}} \in \mathbf{\Lambda}\}$, we have that $\mathcal{O}$ is the required γ-shrinkable $(k,m)_r$-orchard by Observation 3.5.2. □

For dense graphs (that is, when p is large), Proposition 3.5.5 is already enough to establish Proposition 3.1.8. On the other hand, for sparse graphs Proposition 3.5.5 can only be used for orchards of very small order and becomes redundant as the order m approaches $p^{r-1}n$. However, in deriving Proposition 3.5.5, we were quite naïve in our application of Lemma 3.5.4, using the order of a diamond tree as an upper bound on the degrees of vertices to the removable set of vertices of the diamond tree. For a set Q we expect a typical vertex $v \in V(G)$ to have $\deg(v;Q) \le p|Q|$ and so we can hope that Lemma 3.5.4 can be applied to imply the existence of shrinkable orchards whose orders approach $p^{r-2}n$, gaining an extra power of p over Proposition 3.5.5. This is the content of the rest of this section.

3.5.4 Controlling degrees to removable sets of vertices

A reasonable approach when trying to apply Lemma 3.5.4 to deduce the existence of larger order shrinkable orchards is to start with a larger (in size) orchard than we desire and crop diamond trees which fail the bounded degree condition. This approach is reminiscent of how we derived Lemma 3.5.4 from Lemma 3.5.3, where we greedily lost diamond trees which violated condition (1) of Lemma 3.5.3. In this case though, our condition is harder to satisfy. Indeed, we require that *all* vertices in a set in our system satisfy the degree condition and not just a single vertex. In order to achieve this, we will need to appeal to (the full power of) Proposition 3.3.1 to choose our diamond trees. As Proposition 3.3.1 does not give full control over the set of vertices which end up as the set of removable vertices, we have to settle with being able to conclude our desired upper bound on the degrees of vertices to a *subset* of the removable vertices. The detailed statement is as follows.

Lemma 3.5.6. *For any $3 \le r \in \mathbb{N}$ and $0 < \gamma, \eta < \frac{1}{r^2}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p,β)-bijumbled graph G with $\beta \le \varepsilon p^{r-1}n$ and any vertex subset $U \subseteq V(G)$ with $|U| \ge \frac{n}{4}$. For any $m \in \mathbb{N}$ with $p^{r-1}n^{1-\gamma} \le m \le p^{r-2}n^{1-2\gamma}$, there exists a K_r-diamond tree $\mathcal{D} = (T, R, \Sigma)$, of order at most $2m$ such that $V(\mathcal{D}) \subset U$ and there exists a subset $Q \subseteq R$ of removable vertices such that $|Q| = m$ and all but at most ηm vertices $v \in V(G)$ have $\deg(v;Q) \le p^{r-1}n^{1-\gamma}$.*

Proof. Fix $\varepsilon > 0$ small enough to apply Proposition 3.3.1 with $\alpha = \frac{1}{2^{2r}r}$, small enough to apply Lemma 3.2.3 with $\eta_{3.2.3} = \eta' < \frac{\alpha^2\eta}{2^6}$ and small enough to force n to be sufficiently large in what follows. We begin by splitting U into disjoint subsets U' and W' arbitrarily so that $|U'|, |W'| \ge \frac{n}{8} = 4\alpha r n$, noting that this is possible due to our definition of α.

Now fix some m with $p^{r-1}n^{1-\gamma} \le m \le p^{r-2}n^{1-2\gamma}$ and define $q := p^{r-1}n^{1-\gamma}m^{-1}$. Note that $8p \le pn^{\gamma} \le q \le 1$ due to our conditions on m. As we aim to find a set Q of size m, the condition that $\deg(v;Q) > p^{r-1}n^{1-\gamma}$ is equivalent to having that $\deg(v;Q) > q|Q|$. As discussed above, given a diamond tree of the correct order and a subset Q of m removable vertices, we can appeal to Lemma 3.2.3 (ii) to bound the number of vertices which have high degree to Q. However, the bound is not strong enough for our purposes so we instead appeal to the full power of Proposition 3.3.1. The idea is to take Y to be much bigger than m. Therefore applying Lemma 3.2.3 (ii) with respect to Y gives a stronger upper bound on the number of vertices which have large degree (at least $q|Y|$, say) to Y. If we then take Q to be a *random subset* of Y, then we expect the density of the neighbourhood of a vertex in Q to have roughly the same density as the neighbourhood of that vertex in Y. Hence, we can bound the number of vertices which have large degree to Q by 'carrying over' the bound on the number of vertices which had large degree to Y. The details follow.

First we fix $\delta := \alpha^2 p^{r-1} n$ and $z := \min\left\{\alpha n, \frac{\delta m}{2}\right\}$. Note that for n large, due to Fact 3.2.2, δ will also be large. Now apply Proposition 3.3.1 to obtain disjoint subsets $X, Y \subset U'$ as in the statement of Proposition 3.3.1. Note that $|X| \leq \frac{2z}{\delta} \leq m$ and $|Y| \geq z - |X| \geq \frac{z}{2}$ for n sufficiently large. Fix a subset $Z \subset Y$ of size $\frac{z}{2}$ and let $B \subset V(G)$ be the set of vertices $v \in V(G)$ such that $\deg(v; Z) > \frac{q}{4}|Z|$. We claim that $|B|$ has size at most ηm. Indeed, noting that $\frac{q}{4} \geq 2p$, Lemma 3.2.3 (ii) gives that

$$|B| \leq \frac{2^4 \eta' p^{2r-2} n^2}{q^2 |Z|} = \frac{2^5 \eta' n^{2\gamma} m^2}{z} \leq \begin{cases} \eta n^{2\gamma-1} p^{1-r} m & \text{if } z = \frac{\delta m}{2}, \\ \eta n^{2\gamma-1} m^2 & \text{if } z = \alpha n. \end{cases} \tag{3.5.2}$$

In the case that $z = \frac{\delta m}{2}$, the estimate in (3.5.2) is less than ηm for large n due to the condition that $\gamma < \frac{1}{r^2}$ and the fact that $p \geq n^{-1/(2r-3)}$ (Fact 3.2.2). In the case that $z = \alpha n$, the estimate in (3.5.2) is less than ηm due to the fact that $m \leq p^{r-2} n^{1-2\gamma} \leq n^{1-2\gamma}$.

For each $v \notin B$, we have that $\deg(v; Z) \leq \frac{q}{4}|Z|$ and so we let $N_v \subset Z$ be a subset of exactly $\frac{q}{4}|Z|$ vertices in Z such that N_v contains all the neighbours of v which lie in Z. Now consider a random subset $Q_1 \subset Z$ where we keep each vertex independently with probability $p' = \frac{4m}{z}$, noting that $0 \leq p' \leq 1$ for large enough n. Clearly $\mathbb{E}\big[|Q_1|\big] = p'|Z| = 2m$ and for each $v \in V(G) \setminus B$, have that $\mathbb{E}\big[|Q_1 \cap N_v|\big] = p'|N_v| = \frac{qm}{2}$. We get concentration for these random variables from Theorem 2.1.1 which is strong enough to do a union bound and conclude that whp as n (and hence m and qm) tend to infinity, we have that $|Q_1| \geq m$ and $|Q_1 \cap N_v| \leq qm$ for all $v \in V(G) \setminus B$. Therefore, for sufficiently large n, we can fix such an instance of Q_1 and take Q to be a subset of Q_1 such that $|Q| = m$. Therefore for all vertices $v \in V(G) \setminus B$, we have that

$$\deg(v; Q) \leq |Q \cap N_v| \leq |Q_1 \cap N_v| \leq qm = p^{r-1} n^{1-\gamma}.$$

We have that $|B| \leq \eta m$ from above and we use the conclusion of Proposition 3.3.1 to give a K_r-diamond tree $\mathcal{D} = (T, R, \Sigma)$ with removable vertices $R := X \cup Q \subset U' \subset U$ and Σ a K_{r-1}-matching in $W' \subset U$. We thus have that $V(\mathcal{D}) \subset U$ as required and the order of $\mathcal{D}$ is $|Q| + |X| \leq 2m$. □

3.5.5 The existence of shrinkable orchards of larger order

Lemma 3.5.6 gives us the key to being able to push the methods above (which culminated in Lemma 3.5.4) to be able to handle orchards with larger order. We remark that the flexibility given by dealing with (k, m)-systems and Observation 3.5.2 is necessary in order to handle this extension. Indeed, this is due to Lemma 3.5.6 only giving control over the degree to a subset of the removable vertices of the diamond tree generated.

Proof of Proposition 3.4.1. By Proposition 3.5.5 we can focus on the case that

$$p^{r-1}n^{1-r^3\gamma} \leq m \leq \min\{p^{r-2}n^{1-2r^3\gamma}, n^{7/8}\}.$$

We fix $\varepsilon > 0$ small enough to apply Lemma 3.5.4 with $\alpha_{3.5.4} = \alpha' := \frac{\alpha}{4}$ and $\gamma_{3.5.4} = \gamma$ and small enough to apply Lemma 3.5.6 with $\gamma_{3.5.6} = \gamma' := r^3\gamma$ and $\eta_{3.5.6} = \eta < \frac{\alpha}{8}$. Finally we fix some $k \in \mathbb{N}$ such that $\alpha n \leq km \leq 2\alpha n$. By repeatedly applying Lemma 3.5.6, we find a $(2k, m)$-orchard $\mathcal{O}_0$ with $V(\mathcal{O}_0) \subset U$ and each $\mathcal{D} = (T, R, \Sigma) \in \mathcal{O}_0$ has the property that there exists some distinguished subset $Q_{\mathcal{D}} \subset R$ of removable vertices such that $|Q_{\mathcal{D}}| = m$ and all but at most ηm vertices v in $V(G)$ have that $\deg(v; Q_{\mathcal{D}}) \leq p^{r-1}n^{1-\gamma'}$. Indeed, we can find $\mathcal{O}_0$ by sequentially choosing diamond trees and deleting their vertices from U, using that $|V(\mathcal{O}_0)| \leq 4r\alpha n$ at all times in this process and so $|U \setminus V(\mathcal{O}_0)| \geq \frac{n}{4}$ and we can apply Lemma 3.5.6.

Now we will crop our orchard $\mathcal{O}_0$ to arrive at an orchard for which we can apply Lemma 3.5.4 to subsets of removable vertices. Similarly to the proof of Lemma 3.5.4, we do this by a process of 'cleaning up'; losing diamond trees in the orchard which have lots of removable vertices which are atypical. So let $B_1 \subset V(G)$ be the set of vertices $v \in V(G)$ such that $\deg(v; Q_{\mathcal{D}}) > p^{r-1}n^{1-\gamma'}$ for some $\mathcal{D} \in \mathcal{O}_0$ as above. It follows that $|B_1| \leq \eta m \cdot 2k \leq \frac{\alpha n}{4}$. Next we delete $\mathcal{D}'$ from $\mathcal{O}_0$ if $|B_1 \cap Q_{\mathcal{D}'}| \geq \frac{m}{2}$. Due to our upper bound on $|B_1|$, we delete at most $\frac{k}{2}$ of the diamond trees $\mathcal{D}'$ from $\mathcal{O}_0$. Let the resulting suborchard be $\mathcal{O}_1 \subseteq \mathcal{O}_0$ and for each diamond tree $\mathcal{D} = (T, R, \Sigma) \in \mathcal{O}_1$ define a distinguished subset $S_{\mathcal{D}} \subset Q_{\mathcal{D}} \subseteq R$ of removable vertices such that $|S_{\mathcal{D}}| = \frac{m}{2}$ and

$$\deg(v; S_{\mathcal{D}}) \leq p^{r-1}n^{1-\gamma'} = p^{r-1}n^{1-r^3\gamma} \text{ for all } \mathcal{D} \in \mathcal{O}_1 \text{ and all } v \in \bigcup_{\mathcal{D}' \in \mathcal{O}_1} S_{\mathcal{D}'}. \tag{3.5.3}$$

Let $\mathcal{O}_2$ be an arbitrary suborchard of $\mathcal{O}_1$ so that $|\mathcal{O}_2| = (1+\gamma)k$. Moreover, let $\mathbf{\Lambda}' = \{S_{\mathcal{D}} : \mathcal{D} \in \mathcal{O}_2\}$ be the $((1+\gamma)k, \frac{m}{4})$-system defined by the distinguished subsets of removable vertices for the K_r-diamond trees in $\mathcal{O}_2$. Now due to (3.5.3), we have that Lemma 3.5.4 gives the existence of some γ-shrinkable (with respect to r) subsystem $\mathbf{\Lambda} \subset \mathbf{\Lambda}'$. Taking $\mathcal{O} := \{\mathcal{D} \in \mathcal{O}_2 : S_{\mathcal{D}} \in \mathbf{\Lambda}\}$ thus gives a γ-shrinkable $(k, m)_r$-orchard as required, appealing to Observation 3.5.2. □

3.6 Shrinkable orchards of large order

In this section, we establish the existence of shrinkable orchards with large order, proving Proposition 3.4.2. Our approach is to find an orchard such that the K_r-hypergraph $\mathscr{H}$ generated by the orchard is very dense. Again we will appeal to Theorem 2.7.3 which tells us that a hypergraph $\mathscr{H}$ has a large matching if whenever we remove a small collection of edges (dictated by some 2-uniform graph J on $V(\mathscr{H})$), there exists a perfect fractional matching in the remaining hypergraph. In order to find these perfect fractional matchings in subhypergraphs of the K_r-hypergraph $\mathscr{H}$, we will appeal to Lemma 2.6.3, which gives the existence of a perfect fractional matching given that any vertex of the hypergraph is contained in many edges. More precisely, we need that for any vertex $v \in V(\mathscr{H})$ and large subset of vertices $W \subset V(\mathscr{H})$, there is an edge of $\mathscr{H}$ containing v and vertices in W. As we did with the small orchards, our first step is to apply this theory of large matchings via perfect fractional matchings, to obtain a condition on an orchard which guarantees shrinkability. This is the content of Section 3.6.1 and in Lemma 3.6.1 we derive such a condition (condition (3.6.1)) which allows us to greedily apply the results above to find the relevant edges and hence perfect fractional matchings in sub-hypergraphs of the K_r-hypergraph $\mathscr{H}(\mathcal{O})$.

For the rest of the section, we thus focus on finding orchards which satisfy the condition (3.6.1) given in Lemma 3.6.1. This condition is essentially a density condition similar to the condition of Lemma 2.6.3 above (indeed this lemma is used in the proof of Lemma 3.6.1) saying that we need every vertex of our K_r-hypergraph $\mathscr{H}$ to be contained in many edges. However given that we need to apply Lemma 2.6.3 in many different sub-hypergraphs of $\mathscr{H}$, the condition (3.6.1) is much stronger and requires us to find edges containing any vertex $v \in V(\mathscr{H})$ and vertices from arbitrary sets some of which are sublinear in size. A detailed discussion of the condition (3.6.1) is given in Section 3.6.1.

We will then show in Section 3.6.2 that we can appeal to Proposition 3.3.1 to generate diamond trees whose removable vertices are contained in many copies of K_r and hence the condition (3.6.1) will be satisfied. This will then allow us to prove the existence of shrinkable orchards of large order in Section 3.6.3 . As in Section 3.5 however, this first argument will fall short of the range of orders needed in Proposition 3.4.2. The rest of the section is thus concerned with extending our methods to capture more orders. This leads us to a process which generates an orchard in two rounds. The outcome of the first round is discussed in Section 3.6.4 and building on this, in Section 3.6.5 we detail properties of the orchard after a second round of generation. Finally in Section 3.6.6, we show that by generating orchards via this two-phase process, we end up with orchards which are shrinkable. This allows us to complete the proof of Proposition 3.4.2.

3.6.1 A density condition which guarantees shrinkability

We begin by applying Lemma 2.6.3 and Theorem 2.7.3 to give a density condition which we can use to show that an orchard is shrinkable. This transforms our problem into finding orchards which satisfy this condition.

Lemma 3.6.1. *For all $3 \leq r \in \mathbb{N}$ and $0 < \gamma < \frac{1}{2r^3}$, there exists a $k_0 \in \mathbb{N}$ such that the following holds. Suppose that $\mathcal{O}$ is a $(k, m)_r$-orchard in a graph G with $k_0 \leq k \in \mathbb{N}$ and $m \in \mathbb{N}$. For a diamond tree $\mathcal{D} \in \mathcal{O}$, let $R_{\mathcal{D}}$ denote its removable vertices and for a suborchard $\mathcal{O}' \subset \mathcal{O}$, let $R(\mathcal{O}') := \cup_{\mathcal{D}\in\mathcal{O}'} R_{\mathcal{D}}$ denote the union of the sets of removable vertices of diamond trees in $\mathcal{O}'$. Suppose that the following condition holds:*

For any $\mathcal{D} \in \mathcal{O}$ and $\mathcal{P} \subset \mathcal{O}\setminus\{\mathcal{D}\}$ such that $|\mathcal{P}| \geq \frac{k}{4r}$, there exists a suborchard $\mathcal{P}^* = \mathcal{P}^*(\mathcal{D}, \mathcal{P}) \subset \mathcal{P}$ such that $|\mathcal{P}^*| \leq k^{1-r^3\gamma}$ and for any *disjoint suborchards $\mathcal{O}_1, \ldots, \mathcal{O}_{r-2} \subset \mathcal{P} \setminus \mathcal{P}^*$, with $|\mathcal{O}_i| \geq k^{1-r^3\gamma}$ for $i \in [r-3]$ and $|\mathcal{O}_{r-2}| \geq \gamma k$, there is a copy of K_r in G traversing $R_{\mathcal{D}}$, $R(\mathcal{P}^*)$ and $R(\mathcal{O}_i)$ for $i \in [r-2]$.* (3.6.1)

Then $\mathcal{O}$ is γ-shrinkable.

Let us take a moment to digest the density condition (3.6.1). For simplicity, one can think of $\mathcal{P}^*$ being a single diamond tree $\mathcal{D}^* = \mathcal{D}^*(\mathcal{D}, \mathcal{P})$. Indeed this is the setting that we will work in first when applying Lemma 3.6.1. Simplifying further and just focusing on the case that $r = 3$, the condition (3.6.1) then translates as having that for any K_3-diamond tree $\mathcal{D}$ in the orchard and large suborchard $\mathcal{P} \subset \mathcal{O}$, there is some diamond tree $\mathcal{D}^* \in \mathcal{P}$ so that the pair $\{\mathcal{D}, \mathcal{D}^*\}$ has high degree in the K_3-hypergraph generated by $\mathcal{P}$. Indeed for any small linear sized $\mathcal{O}_1 \subset \mathcal{P}$, there is a hyperedge in $\mathscr{H}(\mathcal{O})$ containing $\mathcal{D}, \mathcal{D}^*$ and a diamond tree in $\mathcal{O}_1$. In general, when $r \geq 4$, we need to guarantee traversing K_rs when some of the sets we look to traverse are smaller than linear (size $k^{1-r^3\gamma}$). Also later on we will need the full power of Lemma 3.6.1 which allows us to choose the $\mathcal{P}^*$ as a small suborchard as opposed to a single diamond tree. We now prove the lemma.

Proof of Lemma 3.6.1. Let $\mathcal{Q} \subset \mathcal{O}$ be an arbitrary suborchard of $\mathcal{O}$ of size γk. We will show that $\mathcal{O}$ is shrinkable with respect to $\mathcal{Q}$. So fix some arbitrary suborchard $\mathcal{Q}' \subset \mathcal{Q}$ and let $\mathscr{H} := \mathscr{H}(\mathcal{O} \setminus \mathcal{Q}')$ be the K_r-hypergraph generated by $\mathcal{O} \setminus \mathcal{Q}'$. We have to show that $\mathscr{H}$ has a matching covering all but at most $k^{1-\gamma}$ vertices of $\mathscr{H}$.

In order to show the existence of a large matching in $\mathscr{H}$, as we did in Lemma 3.5.3, we appeal to Theorem 2.7.3. So let us fix $N = |V(\mathscr{H})|$ and note that as $N \geq (1-\gamma)k$, by choosing k_0 to be large, we can assume that N is sufficiently large in what follows. Now fix some 2-uniform graph J on $V(\mathscr{H})$ of maximum degree at most $N^{r^2\gamma}$. If we can show that $\mathscr{H} \setminus \mathscr{H}_J$ contains a perfect fractional matching, then we are done by Theorem 2.7.3 as, because J was arbitrary, the theorem guarantees a matching covering all but at most $N^{1-\gamma} \leq k^{1-\gamma}$ vertices of $\mathscr{H}$.

In order to prove the existence of a perfect fractional matching in $\mathscr{H} \setminus \mathscr{H}_J$, we appeal to Lemma 2.6.3, fixing $M := \frac{N}{2r}$. Thus, we need to show that given any K_r-diamond tree $\mathcal{D} \in V(\mathscr{H}) = \mathcal{O} \setminus \mathcal{Q}'$ and suborchard $\mathcal{P}_0 \subset V(\mathscr{H}) \setminus \{\mathcal{D}\}$ with $|\mathcal{P}_0| \geq M$, there is an edge in $\mathscr{H} \setminus \mathscr{H}_J$ containing $\mathcal{D}$ and $r-1$ K_r-diamond trees in $\mathcal{P}_0$. So fix such a $\mathcal{D}$ and $\mathcal{P}_0$. Let $\mathcal{P} := \mathcal{P}_0 \setminus N_J(\mathcal{D})$. Therefore, we have that

$$|\mathcal{P}| \geq |\mathcal{P}_0| - |N_J(\mathcal{D})| \geq \frac{N}{2r} - N^{r^2\gamma} \geq \frac{(1-\gamma)k}{2r} - k^{r^2\gamma} \geq \frac{k}{4r},$$

for k sufficiently large. Hence by condition (3.6.1), we have the existence of some $\mathcal{P}^* = \mathcal{P}^*(\mathcal{D}, \mathcal{P}) \subset \mathcal{P}$ as in the hypothesis. Now we will iteratively define $\mathcal{O}_i$ for $1 \leq i \leq r-2$ as follows. We begin by fixing $\mathcal{P}' = \mathcal{P}$ and defining $\mathcal{Q}_0 := \cup_{C \in \mathcal{P}^*}(N_J(C) \cup \{C\})$. Now for $1 \leq i \leq r-2$, we update $\mathcal{P}'$ by removing any diamond trees in $\mathcal{Q}_{i-1}$ from $\mathcal{P}'$ and then define $\mathcal{O}_i$ to be an arbitrary suborchard of $\mathcal{P}'$ of size $k^{1-r^3\gamma}$ if $i \in [r-3]$, and of size γk if $i = r-2$. If $i = r-2$ we then end this process. If $i < r-2$, we define $\mathcal{Q}_i := \cup_{C \in \mathcal{O}_i}(N_J(C) \cup \{C\})$ and move to the next index.

Let us check that we are successful in each round. Indeed this follows because at the beginning of step i in the process, $\mathcal{P}'$ has size

$$|\mathcal{P}'| \geq \frac{k}{4r} - ik^{1-r^3\gamma}(1 + N^{r^2\gamma}) \geq \frac{k}{4r} - rk^{1-r^3\gamma+r^2\gamma} \geq \gamma k \geq k^{1-r^3\gamma},$$

for large k. Therefore there is always space in $\mathcal{P}'$ to choose our suborchard $\mathcal{O}_i$ at each step i. Now the condition (3.6.1) gives a copy of K_r in G traversing $R_{\mathcal{D}}$, $R(\mathcal{P}^*)$ and $R(\mathcal{O}_i)$ for $i \in [r-2]$. This thus gives a hyperedge e in the K_r-hypergraph $\mathscr{H} = \mathscr{H}(\mathcal{O} \setminus \mathcal{Q}')$ which has one vertex as $\mathcal{D}$, one vertex in $\mathcal{P}^* \subset \mathcal{P}_0$ and one vertex in each of the $\mathcal{O}_i \subset \mathcal{P}_0$. Moreover this edge e lies in $\mathscr{H} \setminus \mathscr{H}_J$. Indeed, by our construction of $\mathcal{P}^*$ and the $\mathcal{O}_i$, there is no edge in J between any pair of distinct sets in the family $\{\{\mathcal{D}\}, \mathcal{P}^*, \mathcal{O}_1, \ldots, \mathcal{O}_{r-2}\}$. We have therefore established the existence of a perfect fractional matching in $\mathscr{H} \setminus \mathscr{H}_J$ due to Lemma 2.6.3 which implies that $\mathcal{O}$ is γ-shrinkable as detailed above. □

Lemma 3.6.1 gives a route to proving the existence of shrinkable orchards. Indeed, if the sets of vertices which arise as pools of removable vertices of suborchards are sufficiently large, then appealing to Corollary 3.2.5 can give the required transversal copy of K_r in G, so that (3.6.1) is satisfied. However, we cannot immediately derive such results because the size of the sets required in (3.6.1) are too small. In particular, (3.6.1) forces only one set (namely $R(\mathcal{O}_{r-2})$) to be linear in size whilst all other sets that feature can have sublinear size. This is troublesome because the examples we have from Corollary 3.2.5 to generate transversal copies of K_r, require at least two of the sets involved to be linear. Indeed, it can be seen from the more general Lemma 3.2.4 that we cannot do any better than this. That is, in order to use Definition 1.4.1 and our condition on β to derive the existence of a copy of K_r that traverses a family of sets, at least two of the sets in the family must be linear in size. Therefore in order to apply Lemma 3.6.1 and derive the existence of shrinkable orchards, we have to obtain orchards with some additional structure. We start by exploring properties of singular diamond tress that we can guarantee.

3.6.2 Popular diamond trees

As was the case when we were interested in proving the existence of shrinkable orchards with small order, Proposition 3.3.1 gives a powerful tool for proving the existence of diamond trees with additional desired properties. Here we show that we can choose a diamond tree so that there are many copies of K_r formed with its removable vertices.

Lemma 3.6.2. *For any $3 \leq r \in \mathbb{N}$ and $0 < \alpha < \frac{1}{2^{12r}}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$ and any vertex subset $U \subseteq V(G)$ with $|U| \geq \frac{n}{4}$. Suppose that $m \in \mathbb{N}$ with*

$$\max\{p^{1-r}, p^{r-1}n\} \leq m \leq n^{7/8}$$

and we have set families $\mathcal{W}_0, \mathcal{W}_1, \ldots, \mathcal{W}_{r-2} \subset 2^{V(G)}$ such that:

1. *$|W_0| \geq \alpha p^{r-1} n$ for all $W_0 \in \mathcal{W}_0$;*
2. *$|W_i| \geq \alpha p n$ for all $W_i \in \mathcal{W}_i$, $1 \leq i \leq r-3$;*
3. *$|W_{r-2}| \geq \alpha n$ for all $W_{r-2} \in \mathcal{W}_{r-2}$;*
4. *$\prod_{i=0}^{r-2} |\mathcal{W}_i| \leq 2^{m/4}$.*

Then there exists a K_r-diamond tree $\mathcal{D} = (T, R, \Sigma)$ in $G[U]$ of order at least m and at most $2m$ such that for any *choice of sets $\mathbf{W} = (W_0, \ldots, W_{r-2}) \in \mathcal{W}_0 \times \cdots \times \mathcal{W}_{r-2}$, there is a copy of K_r in G traversing R and the sets $W_0, \ldots, W_{r-2}$.*

Proof. Let us fix $\varepsilon > 0$ small enough to apply Proposition 3.3.1 with $\alpha_{3.3.1} = \alpha' := \frac{1}{2^{3r}}$ and Corollary 3.2.5 with $\tau = \alpha$ as well as being small enough to force n to be sufficiently large. Note that our lower bound of $\Omega(p^{r-1}n)$ on m and Fact 3.2.2 imply that m tends to infinity as n tends to infinity and so we can also assume m is sufficiently large in what follows. We begin by splitting U into disjoint subsets U' and W' arbitrarily so that $|U'|, |W'| \geq \frac{n}{8} = 4\alpha' r n$, noting that this is possible due to our definition of α'. We further fix $\delta := \alpha'^2 p^{r-1} n$.

Now we apply Proposition 3.3.1 with $z := \frac{\alpha'^2 n}{4} = \frac{n}{2^{6r+2}}$ and fix the sets $X \subset U'$ and $Y \subset U'$ which are output. Note that

$$|X| \leq \max \begin{Bmatrix} \frac{2z}{\delta} & = \frac{1}{2p^{r-1}} \\ & 1 \end{Bmatrix} \leq \frac{m}{2} \quad \text{and} \quad |Y| = z - |X| \geq \frac{z}{2} = \frac{n}{2^{6r+3}},$$

for n large.

Now for each choice of $\mathbf{W} = (W_0, \ldots, W_{r-2}) \in \mathcal{W}_0 \times \cdots \times \mathcal{W}_{r-2}$, we find some subset $Y(\mathbf{W}) \subset Y$ of size $\frac{|Y|}{2}$ such that for every $v \in Y(\mathbf{W})$, there is a copy of K_{r-1} in the neighbourhood of v which traverses $W_0, \ldots, W_{r-2}$. In other words, for every $v \in Y(\mathbf{W})$, there is a copy of K_r traversing $W_0, \ldots, W_{r-2}$ and $\{v\}$. We can find $Y(\mathbf{W})$ by repeated applications of Corollary 3.2.5 (2). In more detail, we initiate with $Y_0 = Y$ and $Y(\mathbf{W})$ empty and in each step we find a copy of K_r traversing $W_0, \ldots, W_{r-2}$ and Y_0. Taking v to be the[9] vertex of this K_r that lies in Y_0, we add v to $Y(\mathbf{W})$, delete it from Y_0 and move to the next step. We continue for $\frac{|Y|}{2}$ steps using that the conditions of Corollary 3.2.5 (2) are satisfied at each step. Indeed this is due to the lower bounds on the sizes of W_i in conditions (1), (2) and (3) of this lemma and the fact that $|Y_0| \geq |Y| - |Y(\mathbf{W})| \geq \frac{|Y|}{2} \geq \alpha n$ throughout, using our upper bound on α and our lower bound on $|Y|$ here.

Similarly to the proof of Lemma 3.5.6, we now take Q to be a random subset of Y by taking each vertex of Y into Q independently with probability $p' := \frac{5m}{4|Y|}$. Thus $\mathbb{E}\big[|Q|\big] = \frac{5m}{4}$ and by Theorem 2.1.1, we have that $m \leq |Q| \leq \frac{3m}{2}$ with probability at least $1 - 2e^{-m/60}$. Furthermore, for any fixed $\mathbf{W} \in \mathcal{W}_0 \times \cdots \times \mathcal{W}_{r-2}$, we have that

$$\mathbb{E}\big[|Q \cap Y(\mathbf{W})|\big] = p'|Y(\mathbf{W})| = \frac{5m}{8}.$$

Applying Theorem 2.1.1 again implies that the probability that $|Q \cap Y(\mathbf{W})| = 0$ is less than $e^{-5m/16}$. Therefore using that $\prod_{i=0}^{r-2} |\mathcal{W}_i| \leq 2^{m/4}$ and appealing to a union bound, we can conclude that whp as n (and hence m) tend to infinity, we have that $m \leq |Q| \leq \frac{3m}{2}$ and $Q \cap Y(\mathbf{W}) \neq \emptyset$ for *all* choices of $\mathbf{W} \in \mathcal{W}_0 \times \cdots \times \mathcal{W}_{r-2}$. So for sufficiently large n we can fix such an instance $Q \subset Y$ and taking $R := X \cup Q$ we have that a K_r-diamond tree $\mathcal{D} = (T, R, \Sigma)$ with removable set of vertices R is guaranteed by Proposition 3.3.1. We claim that $\mathcal{D}$ satisfies all the necessary conditions. Indeed, the fact that the order of $\mathcal{D}$ lies between m and $2m$ follows from the fact that $m \leq |Q| \leq \frac{3m}{2}$ and $|X| \leq \frac{m}{2}$ whilst the fact that $Q \cap Y(\mathbf{W}) \neq \emptyset$ for each choice of $\mathbf{W} = (W_0, \ldots, W_{r-2})$ guarantees that we have a copy of K_r traversing $Q \subset R$ and the sets $W_0, \ldots, W_{r-2}$. □

[9]Here we refer to *the* vertex that lies in Y_0 although there may be several (if the W_i intersect the Y_0). What we mean here is the vertex v in the copy of K_r which is assigned to Y_0 by virtue of the copy being traversing.

3.6.3 The existence of shrinkable orchards of large order

Using Lemma 3.6.2 to generate the diamond trees that form our orchard, we can prove that the orchard generated satisfies the condition of Lemma 3.6.1 and hence is shrinkable. This gives the following.

Proposition 3.6.3. *For any $3 \leq r \in \mathbb{N}$ and $0 < \alpha, \gamma < \frac{1}{2^{12r}}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p,β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$ and any vertex subset $U \subseteq V(G)$ with $|U| \geq \frac{n}{2}$. For any $m \in \mathbb{N}$ with*

$$\max\{p^{1-r}, p^{r-1}n\} \leq m \leq n^{7/8},$$

there exists a γ-shrinkable $(k,m)_r$-orchard $\mathcal{O}$ in $G[U]$ with $k \in \mathbb{N}$ such that $\alpha n \leq km \leq 2\alpha n$.

Proof. Fix $\varepsilon > 0$ small enough to apply Lemma 3.6.1 with $\gamma_{3.6.1} = \gamma$ and Lemma 3.6.2 with $\alpha_{3.6.2} = \alpha' = \alpha\gamma$. Fix some $k \in \mathbb{N}$ such that $\alpha n \leq km \leq 2\alpha n$. We also ensure that ε is small enough to force n (and hence k, due to our upper bound on m) to be sufficiently large in what follows. Now we begin by noticing that $k \leq \frac{m}{8r}$. Indeed we have that if $p \geq n^{-1/(2r-2)}$, then

$$p^{1-r} \leq \sqrt{n} \leq p^{r-1} n \leq m,$$

while if $p \leq n^{-1/(2r-2)}$, then

$$p^{r-1} n \leq \sqrt{n} \leq p^{1-r} \leq m.$$

Therefore, for any p we have that $m \geq \sqrt{n}$ and $k \leq \frac{2\alpha n}{m} \leq 2^{-11r}\sqrt{n} \leq \frac{m}{8r}$.

Now we turn to finding our $(k,m)_r$-orchard in $G[U]$. We do this by finding one diamond tree at a time as follows. For $1 \leq i \leq k$, fix $U_i := U \setminus (\bigcup_{i'<i} V(\mathcal{D}_{i'}))$ and note that $|U_i| \geq |U| - 2\alpha r n \geq \frac{n}{4}$ throughout due to our condition on α. We then apply Lemma 3.6.2 to find a diamond tree $\mathcal{D}_i = (T_i, R_i, \Sigma_i)$ such that $V(\mathcal{D}_i) \subset U_i$ and for any choice of $i' \in [i-1]$ and disjoint subsets $I_1, \dots I_{r-2} \subset [i-1] \setminus \{i'\}$ with $|I_j| \geq pk$ for $1 \leq j \leq r-3$, and $|I_{r-2}| \geq \gamma k$ we have that there is a copy of K_r traversing R_i, $R_{i'}$ and the sets $\bigcup_{\ell \in I_j} R_\ell$ for $j \in [r-2]$. The existence of such a $\mathcal{D}_i$ follows from Lemma 3.6.2. Indeed, we define $\mathcal{W}_0 = \{R_{i'} : i' \in [i-1]\}$, $\mathcal{W}_j = \{\bigcup_{\ell \in I'} R_\ell : I' \subset [i-1], |I'| \geq pk\}$ for $1 \leq j \leq r-3$ and finally we define $\mathcal{W}_{r-2} = \{\bigcup_{\ell \in I'} R_\ell : I' \subset [i-1], |I'| \geq \gamma k\}$. We need to check that conditions (1)-(4) of Lemma 3.6.2 are satisfied. Indeed condition (1) follows from our lower bound on m whilst (2) and (3) follow from the fact that $km \geq \alpha n$ and our definition of α'. Finally note that each choice of a set in any of the $\mathcal{W}_j$ comes from a subset of $[i-1]$. Hence we can upper bound $\prod_{j=0}^{r-2} |\mathcal{W}_j|$ by $(2^i)^{r-1} \leq 2^{rk}$. As discussed in the opening paragraph, we have that $k \leq \frac{m}{8r}$ and so condition (4) of Lemma 3.6.2 is also satisfied. Thus Lemma 3.6.2 succeeds in finding the necessary K_r-diamond tree at every step of this process.

Let $\mathcal{O} = \{\mathcal{D}_1, \ldots, \mathcal{D}_k\}$ be the orchard obtained by this process. We claim that $\mathcal{O}$ is γ-shrinkable and to show this we appeal to Lemma 3.6.1 and so need to show that the density condition (3.6.1) is satisfied by $\mathcal{O}$. So fix some arbitrary $\mathcal{D}_i \in \mathcal{O}$ and $\mathcal{P} \subset \mathcal{O} \setminus \{\mathcal{D}_i\}$ with $|\mathcal{P}| \geq \frac{k}{4r}$. We then define $\mathcal{D}^* = \mathcal{D}^*(\mathcal{D}_i, \mathcal{P})$ (this plays the rôle of $\mathcal{P}^*$ in (3.6.1)) to be the diamond tree in $\mathcal{P}$ with the highest index. That is we define $i^* := \max\{i' : \mathcal{D}_{i'} \in \mathcal{P}^*\}$ and set $\mathcal{D}^* = \mathcal{D}_{i^*}$. Note that we may have that $i^* < i$ but this will not be a problem. We claim that condition (3.6.1) is satisfied with this choice of $\mathcal{D}^*$. Indeed, let $\mathcal{O}_1, \ldots, \mathcal{O}_{r-2} \subset \mathcal{P}^* \setminus \{\mathcal{D}^*\}$ be disjoint suborchards satisfying the lower bounds on the sizes given by (3.6.1). For each $j \in [r-2]$, define $I_j := \{i' : \mathcal{D}_{i'} \in \mathcal{O}_j\}$. Therefore we have that $|I_{r-2}| \geq \gamma k$. For $1 \leq j \leq r-3$ we have that $|I_j| \geq k^{1-r^3\gamma} \geq pk$. This follows from the fact that

$$k^{-r^3\gamma} \geq k^{-1/2^r} \geq n^{-1/2^{(r+1)}} \geq n^{-1/(8(r-1))} \geq p,$$

where we used the upper bound on γ in the first inequality, the fact that $k \leq \sqrt{n}$ in the second inequality (see the opening paragraph of the proof), and the fact that $p^{r-1}n \leq m \leq n^{7/8}$ in the last inequality. Now relabelling $\{i, i^*\}$ as $\{\ell_0, \ell_1\}$ so that $\ell_0 < \ell_1$, we have that at the point of choosing $\mathcal{D}_{\ell_1}$, we guaranteed that there was a K_r traversing R_{ℓ_1}, R_{ℓ_0} and the sets $R(\mathcal{O}_j) = \cup_{i' \in I_j} R_{i'}$ for $j \in [r-2]$. By Lemma 3.6.1 this completes the proof that $\mathcal{O}$ is γ-shrinkable. □

Proposition 3.6.3 establishes Proposition 3.4.2 when G is very dense. However when G is sparse (when $p \leq n^{-1/(2r-2)}$ to be specific), the lower bound of $m \geq p^{1-r}$ takes over and we are left with a gap between the range covered by Proposition 3.6.3 and the desired range of Proposition 3.4.2. Tracing the condition that $m = \Omega(p^{1-r})$ back through the proof, we can see that this was necessary in order to prove Lemma 3.6.2. There, we used our key Proposition 3.3.1 to generate a diamond tree where we had a large pool Y of vertices which were candidates for being removable vertices. In order to establish the existence of the cliques we need in Lemma 3.6.2, we needed Y to be linear in size. The sticking point then comes from the fact that Proposition 3.3.1 can only guarantee a maximum factor of $O(p^{r-1}n)$ between the size of the pool of vertices Y and the order of the diamond tree that we generate. Indeed, in Proposition 3.3.1 we are forced to include the set X in the removable vertices of the diamond tree we generate and when Y is linear in size, X could have size as large as $\Omega(p^{1-r})$. It is unclear how one would improve on this and find diamond trees with smaller order that are still contained in sufficiently many copies of K_r.

Thankfully, there is a way to circumvent this issue and apply our methods to close the gap in the range of orders nonetheless. The key idea is to replace the diamond tree generated by Lemma 3.6.2 with a *set of diamond trees*, that is, a small suborchard. Indeed, by grouping together diamond trees, we can decrease their order but guarantee that the collective pool of potential removable vertices for the group is still linear in size. Through following a similar proof to that of Lemma 3.6.2, this has the outcome of being able to guarantee many copies of K_r which contain a vertex in the removable vertices of *one of* the diamond trees in the group. Moreover, in the proof of Proposition 3.6.3, we crucially used that we could generate diamond trees from Lemma 3.6.2 to establish the density condition (3.6.1) of Lemma 3.6.1. We chose an appropriate $\mathcal{D}^*$ and used that it had been generated by Lemma 3.6.2 to prove the required existence of transversal copies of K_r. However, Lemma 3.6.1 allows for us to use a much larger suborchard $\mathcal{P}^*$ for this condition as opposed to a single diamond tree. Therefore there is hope to incorporate the idea of using a suborchard instead of a single diamond tree in Lemma 3.6.2 whilst maintaining the overall scheme of the proof. There are some further difficulties to overcome but on a high level, this is the approach we follow in the next sections to establish Proposition 3.4.2.

3.6.4 Preprocessing the orchard

As discussed above, in order to prove Proposition 3.4.2 and remove the condition that $m = \Omega(p^{1-r})$ from Proposition 3.6.3, we need to replace the rôle played by $\mathcal{D}^*$ in the proof by a small suborchard $\mathcal{P}^*$. This allows us to prove an analogue of Lemma 3.6.2, where one now finds an orchard whose collective set of removable vertices lie in many copies of K_r. Our shrinkable orchard then, will be formed as the union of many of these smaller orchards. Indeed, in what follows we will split k as $k = \ell t$ and will aim to have t smaller $(\ell, m)_r$-orchards contributing to our shrinkable orchard $\mathcal{O}$. Each of the $(\ell, m)_r$-orchards will have strong connectivity to the rest of the orchard $\mathcal{O}$.

In order to work with the fact that we are splitting k into t sets of size ℓ, we introduce a two-coordinate index system, with $(i, j) \in [t] \times [\ell]$ indicating that we are referring to the j^{th} object in the i^{th} subset and we will work through these indices lexicographically. In more detail, we let $<_L$ denote the lexicographic order on the pairs $(i, j) \in [t] \times [\ell]$. That is $(i', j') <_L (i, j)$ if and only if either $1 \leq i' \leq i - 1$ and $1 \leq j' \leq \ell$ or $i' = i$ and $1 \leq j' \leq j - 1$. Furthermore for each $1 \leq i \leq t$ and $1 \leq j \leq \ell$, we define

$$I_{<ij} := \{(i', j') \in [t] \times [\ell] : (i', j') <_L (i, j)\},$$

to be the indices (i', j') which come before (i, j) in the lexicographic order.

A hurdle that arises with our new approach is that we lose the symmetry provided by the fact that both $\mathcal{D}$ and $\mathcal{D}^*$ in our applications of Lemma 3.6.1 were given by singular diamond trees. Indeed, in our proof of Proposition 3.6.3, when verifying the condition (3.6.1) of Lemma 3.6.1, we use that both the arbitrary diamond tree $\mathcal{D} = \mathcal{D}_i$ and the diamond tree $\mathcal{D}^* = \mathcal{D}^*(\mathcal{D}_i, \mathcal{P})$ that we can choose, were generated using Lemma 3.6.2. We now hope to generate our suborchards $\mathcal{P}^*$ using an equivalent to Lemma 3.6.2 and this will mean that we can no longer switch the rôles of $\mathcal{D}$ and $\mathcal{P}^*$ when appealing to the conclusion of (the proof method of) Lemma 3.6.2. In particular, this places a higher demand on the properties we need to conclude of our $(\ell, m)_r$-suborchards.

In more detail, we need to generate suborchards which are highly connected to *all* the other vertices of the K_r-hypergraph $\mathscr{H}(\mathcal{O})$. Therefore it no longer suffices to build our orchard in a linear fashion, choosing diamond trees (or indeed suborchards) to be well connected (in terms of the K_r-hypergraph) with previously chosen diamond trees. We will instead generate our orchard in two rounds. In the first round we fix a part of each diamond tree and using Proposition 3.3.1, provide large pools of vertices which can extend the parts of the diamond trees chosen so far, which we will then do in the second round. Lemma 3.6.4 details the outcome we draw from this preprocessing first round.

Lemma 3.6.4. *For any $3 \le r \in \mathbb{N}$ and $0 < \alpha < \frac{1}{2^{12r}}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \le \varepsilon p^{r-1} n$, any vertex subset $U \subseteq V(G)$ with $|U| \ge \frac{n}{2}$ and any $k, m, t, \ell \in \mathbb{N}$ such that $k = t\ell$,*

$$\alpha n \le km \le 2\alpha n \quad \text{and} \quad \ell m \ge p^{1-r}.$$

There exists vertex sets $Z_{ij}, Y_{ij} \subset U$ and K_{r-1}-matchings $\Pi_{ij}, \Upsilon_{ij} \subset K_{r-1}(G[U])$ for each $i \in [t]$ and $j \in [\ell]$ such that the copies of K_{r-1} in each $\Upsilon_{ij} =: \{S_v : v \in Y_{ij}\}$ are indexed by the vertices in Y_{ij} and such that the conditions (1_{ij}) through (5_{ij}) below are satisfied for all $1 \le i \le t$ and $1 \le j \le \ell$.

(1_{ij}) *We have that $|Z_{ij}| = m$ and $|\Pi_{ij}| = |Z_{ij}| - 1$.*

(2_{ij}) *We have that $|Y_{ij}| = |\Upsilon_{ij}| = \frac{\sqrt{\alpha} n}{\ell}$.*

(3_{ij}) *We have that the vertex sets Z_{ij}, Y_{ij}, $V(\Pi_{ij})$ and $V(\Upsilon_{ij})$ are all disjoint from each other.*

(4_{ij}) *We have that $A \cap A' = \emptyset$ for any choice of $A \in \{Z_{ij}, V(\Pi_{ij}), Y_{ij}, V(\Upsilon_{ij})\}$ and*[10]

$$A' \in \{Z_{i'j'}, V(\Pi_{i'j'}) : (i', j') \in I_{<ij}\} \cup \{Y_{ij'}, V(\Upsilon_{ij'}) : 1 \le j' \le j - 1\}.$$

(5_{ij}) *For any choice of $\tilde{Y}$ such that $\tilde{Y} \subseteq Y_{ij}$, there exists a K_r-diamond tree $\mathcal{D} = (T, R, \Sigma)$ such that $R = Z_{ij} \cup \tilde{Y}$ and $\Sigma = \Pi_{ij} \cup \tilde{\Upsilon}_{ij}$, where $\tilde{\Upsilon}_{ij} \subset \Upsilon_{ij}$ is defined to be*

$$\tilde{\Upsilon}_{ij} := \{S_{\tilde{v}} : \tilde{v} \in \tilde{Y} \subset Y_{ij}\}.$$

As mentioned above, in this first round we put aside part of every single diamond tree in the $(k, m)_r$-orchard we are going to generate, thus partially defining the orchard. We also put aside large pools of vertices which will be used to extend these diamond trees in the second round of generating our orchard. The fixed parts of the diamond trees chosen in Lemma 3.6.4

[10] Crucially, we do not require that A is disjoint from *all* $Y_{i'j'}$ and $V(\Upsilon_{i'j'})$, only those that are in the same subfamily indexed by i.

are the sets Z_{ij} and the interior cliques Π_{ij} whilst the pools of potential removable vertices and interior cliques that can be used to extend the diamond trees chosen are given by the sets Y_{ij} and Υ_{ij}, respectively. We make sure through the conditions (1_{ij}) that these fixed sub-diamond trees contribute a substantial portion of the final diamond trees that we are shooting for (which will have order between m and $2m$). We also guarantee through the conditions (4_{ij}), that the parts of the diamond trees that we put aside in this preprocessing round do not interfere with each other, in that they are vertex-disjoint. Notice also that if we fix $i \in [t]$, then the conditions (4_{ij}) for all $j \in [\ell]$ guarantee that the sets $Y_{ij}, V(\Upsilon_{ij})$, $j \in [\ell]$ do not intersect each other. This is important because in the second round of generating our orchard, we will want to extend all the diamond trees in the i^{th} $(\ell, m)_r$-suborchard simultaneously and so we do not want any interference between the choices of the extensions within such a suborchard. Also note that the conditions (2_{ij}) for fixed $i \in [t]$ and all $j \in [\ell]$, guarantee that the collective pool of potential removable vertices for the i^{th} $(\ell, m)_r$-suborchard (the set $\cup_{j\in[\ell]} Y_{ij}$) is linear in size, as required. Finally, the conditions (5_{ij}) contain the heart of Proposition 3.3.1, allowing us to arbitrarily extend any of the diamond trees we have so far using any subsets of the pools (the Y_{ij}) of potential removable vertices and interior cliques (the Υ_{ij}) we have put aside.

Our final remark on the statement of Lemma 3.6.4 is that we do not require e.g. Y_{ij} and $Y_{i'j'}$ for $i \neq i'$, to be disjoint. Indeed as we have t suborchards and each has a linear collective pool of potential removable vertices, there would not be enough space in the graph to keep these pools disjoint. However, by requiring that the collective pool is much larger than all the vertices in our orchard (that is, much larger than km), we guarantee that we will be able to proceed greedily in our second round (Lemma 3.6.5) of defining the orchard, always having a large enough set of potential removable vertices at each step.

Proof of Lemma 3.6.4. Let us fix $\varepsilon > 0$ small enough to apply Proposition 3.3.1 with $\alpha_{3.3.1} = \alpha' := \frac{1}{2^{2r+1}}$. We will find these vertex sets and K_{r-1}-matchings algorithmically working through the pairs $(i, j) \in [t]\times[\ell]$ in lexicographic order. So let us fix some $(i^*, j^*) \in [t]\times[\ell]$ and suppose that we have already found Z_{ij}, Y_{ij}, Π_{ij} and Υ_{ij} such that the conditions (1_{ij}) through (5_{ij}) are satisfied for all $(i, j) \in I_{<i^*j^*}$. We fix $W^* \subset U$ to be

$$W^* := \left(\bigcup \{Z_{ij} \cup V(\Pi_{ij}) : (i, j) \in I_{<i^*j^*}\}\right) \bigcup \left(\bigcup \{Y_{i^*j} \cup V(\Upsilon_{i^*j}) : 1 \leq j \leq j^* - 1\}\right),$$

and let $U^* := U \setminus W^*$. We use conditions (1_{ij}) and (2_{ij}) to upper bound the size of W^* as follows. We have that

$$|W^*| \leq rm((i^* - 1)\ell + j^* - 1) + \frac{\sqrt{\alpha} rn}{\ell}(j^* - 1) \leq rmt\ell + \sqrt{\alpha} rn \leq (2\alpha + \sqrt{\alpha})rn,$$

using that $mt\ell = mk \leq 2\alpha n$. Hence $|U^*| \geq \frac{n}{4}$ from our upper bound on α. We will find $Z_{i^*j^*}, Y_{i^*j^*} \subset U^*$ and $\Pi_{i^*j^*}, \Upsilon_{i^*j^*} \subset K_{r-1}(G[U^*])$ and so condition $(4_{i^*j^*})$ will be satisfied. The required vertex sets $Z_{i^*j^*}$ and $Y_{i^*j^*}$ are found by an application of Proposition 3.3.1. So let us split U^* into disjoint subsets U' and W' arbitrarily so that $|U'|, |W'| \geq \frac{n}{8} \geq 4\alpha' rn$, noting that this is possible due to our definition of α'. We further fix $\delta := \alpha'^2 p^{r-1} n$ and $z := m + \frac{\sqrt{\alpha} n}{\ell}$ and note that $z \leq \alpha' n$ due to the fact that $m \leq \frac{2\alpha n}{k} \leq 2\alpha n$ and our upper bound on α.

So Proposition 3.3.1 gives us that there exists disjoint vertex subsets $X, Y \subset U' \subset U^*$ such that $|X| + |Y| = z$ and $|X| = 1 \leq m$ or

$$|X| \leq \frac{2z}{\delta} \leq \frac{2m}{\delta} + \frac{2\sqrt{\alpha} n}{\delta \ell} \leq \frac{m}{2} + \frac{2\sqrt{\alpha}}{\alpha'^2 p^{r-1} \ell} \leq m,$$

using our upper bound on α and lower bound on ℓm in the last inequality. As $|X| \leq m$, we can fix some $Z_{i^*j^*} \subset X \cup Y$ such that $X \subseteq Z_{i^*j^*}$ and $|Z_{i^*j^*}| = m$. Therefore letting $Y_{i^*j^*} := Y \setminus Z_{i^*j^*}$, we have that $|Y_{i^*j^*}| = z - m = \frac{\sqrt{\alpha} n}{\ell}$ and so the size requirements on $Z_{i^*j^*}$ in $(1_{i^*j^*})$ and on $Y_{i^*j^*}$ in $(2_{i^*j^*})$ are both satisfied. Moreover, we also have that part of condition $(5_{i^*j^*})$ is satisfied. Indeed, for some $\tilde{Y} \subset Y_{i^*j^*}$, taking $Y' = \tilde{Y} \cup (Z_{i^*j^*} \setminus X)$, Proposition 3.3.1 gives that there is a diamond tree $\mathcal{D} = (T, R, \Sigma)$ with removable vertices $R = X \cup Y' = Z_{i^*j^*} \cup \tilde{Y}$ and Σ a K_{r-1}-matching in $G[U^*]$.

Now in order to complete the proof of the lemma, we need to define the K_{r-1}-matchings $\Pi_{i^*j^*}$ and $\Upsilon_{i^*j^*}$ and reason that the remaining conditions of the lemma are satisfied. This comes from recalling how we proved Proposition 3.3.1 in Section 3.3.1 (see also Figure 3.5). There, we applied Lemma 3.3.5 to find a large δ-scattered K_r-diamond tree $\mathcal{D}_{sc} = (T_{sc}, R_{sc}, \Sigma_{sc})$. We had that $R_{sc} = X \cup Y$ was the set of removable vertices of $\mathcal{D}_{sc}$ and $Y \subset R_{sc}$ was the set of leaves in $\mathcal{D}_{sc}$. The conclusion of Proposition 3.3.1 then followed readily as we could choose which leaves in Y to include in a diamond subtree $\mathcal{D}$ of $\mathcal{D}_{sc}$. From this proof we see that we can partition Σ_{sc} into $\Sigma_{sc} =: \Pi_{i^*j^*} \cup \Upsilon_{i^*j^*}$ where the $(r-1)$-cliques $\Pi_{i^*j^*}$ are interior cliques of the K_r-diamond subtree of $\mathcal{D}_{sc}$ spanned by the removable vertices $Z_{i^*j^*}$. Furthermore, we can label $\Upsilon_{i^*j^*}$ with the vertices in $Y_{i^*j^*}$ so that $(5_{i^*j^*})$ is satisfied. Indeed each vertex v in $Y_{i^*j^*}$ corresponds to a leaf of the diamond tree $\mathcal{D}_{sc}$ and so there is an interior clique $S_v \in \Sigma_{sc}$ such that any sub diamond-tree which contains the non-leaves X of $\mathcal{D}_{sc}$ can be extended by adding v to the set of removable vertices and S_v to the set of interior cliques. As $\mathcal{D}_{sc}$ is a well-defined K_r-diamond tree, we also have that condition $(3_{i^*j^*})$ is satisfied and the size constraints of $\Pi_{i^*j^*}$ and $\Upsilon_{i^*j^*}$ in $(1_{i^*j^*})$ and $(2_{i^*j^*})$ are also immediate, noting that $|\Pi_{i^*j^*}| = |Z_{i^*j^*}| - 1$ as the set of interior cliques of a diamond tree with removable vertices $Z_{i^*j^*}$. □

3.6.5 Completing the orchard

We will now use Lemma 3.6.4 to generate our orchard. This can be thought of as extending the parts of the diamond trees (the Z_{ij} and Π_{ij}) which were fixed in Lemma 3.6.4. The strategy is very similar to that of Lemma 3.6.2 and Proposition 3.6.3. Indeed we take random subsets of the pools of potential vertices in order to guarantee that the K_r-hypergraph generated by our final orchard is sufficiently dense. The key difference here is that, as opposed to fixing our orchard one diamond tree at a time, we appeal to Lemma 3.6.4 to fix part of all the diamond trees in our orchard and then carry out the extensions on $(\ell, m)_r$-suborchards. That is, we apply the approach of Lemma 3.6.2 on the whole suborchard as opposed to a singular K_r-diamond tree. After doing this process for all suborchards we end up with an orchard which generates a dense K_r-hypergraph. This is detailed in the following lemma.

Lemma 3.6.5. *For any $3 \leq r \in \mathbb{N}$, $0 < \alpha < \frac{1}{2^{12r}}$ and $0 < \gamma < 1$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$, any vertex subset $U \subseteq V(G)$ with $|U| \geq \frac{n}{2}$ and any $k, m, t, \ell \in \mathbb{N}$ such that*

$$k = t\ell, \qquad m \geq p^{r-1}n, \qquad \ell m \geq p^{1-r}, \qquad \text{and} \qquad \alpha n \leq km \leq 2\alpha n. \tag{3.6.2}$$

Then there exists a $(k, m)_r$-orchard $\mathcal{O}$ in G such that $V(\mathcal{O}) \subset U$ and $\mathcal{O}$ can be partitioned into suborchards $\mathcal{Q}_1, \ldots, \mathcal{Q}_t$ such that each $\mathcal{Q}_i$ with $1 \leq i \leq t$ is an $(\ell, m)_r$-orchard and we have the following property. For any $i \in [t]$, any $\mathcal{D}' \in \mathcal{O}$, any suborchard $\mathcal{Q}' \subseteq \mathcal{Q}_i$ with $|\mathcal{Q}'| \geq \frac{\ell}{4r}$ and any set of disjoint suborchards $\mathcal{O}'_1, \ldots, \mathcal{O}'_{r-2} \subset \mathcal{O}$ with $|\mathcal{O}'_{i'}| \geq pk$ for $i' \in [r-3]$ and $|\mathcal{O}'_{r-2}| \geq \gamma k$, there exists a copy of K_r traversing[11] *$R_{\mathcal{D}'}$, $R(\mathcal{Q}')$ and $R(\mathcal{O}_{i'})$ for $i' \in [r-2]$.*

Proof. Fix $\varepsilon > 0$ small enough to apply Corollary 3.2.5 with $\tau := \gamma\alpha$, small enough to apply Lemma 3.6.4 with α and small enough to force n to be sufficiently large in what follows. We begin by applying Lemma 3.6.4 to get vertex sets Z_{ij}, Y_{ij} and K_{r-1}-matchings Π_{ij} and Υ_{ij} for $(i, j) \in [t] \times [\ell]$ satisfying (1_{ij}) through (5_{ij}) as listed in that lemma. Now we turn to finding the diamond trees $\mathcal{D}_{ij}$ for $(i, j) \in [t] \times [\ell]$ which will form our orchard $\mathcal{O}$, so that the suborchard $\mathcal{Q}_i$ is defined to be $\mathcal{Q}_i := \{\mathcal{D}_{ij} : j \in [\ell]\}$ for each $i \in [t]$. We will appeal in particular to condition (5_{ij}) of Lemma 3.6.4 to find each $\mathcal{D}_{ij} = (T_{ij}, R_{ij}, \Sigma_{ij})$. In more detail, for each $(i, j) \in [t] \times [\ell]$, we will find $\tilde{Y}_{ij} \subset Y_{ij}$ and apply (5_{ij}) to find a diamond tree with removable set of vertices $R_{ij} := Z_{ij} \cup \tilde{Y}_{ij}$.

Now for a set of indices $I' \subseteq [t] \times [\ell]$, we let $Z(I') = \cup_{(i,j) \in I'} Z_{ij}$. In order to guarantee the key property of $\mathcal{O}$ it suffices that for each $i \in [t]$ we have the following. For any choice of $J \subseteq [\ell]$ with $|J| \geq \frac{\ell}{4r}$ and any choice of $(i_0, j_0) \in [t] \times [\ell]$ and subsets $I_1, \ldots, I_{r-2} \subset [t] \times [\ell]$

[11] Here as before, for a diamond tree $\mathcal{D}$, $R_{\mathcal{D}}$ denotes the set of removable vertices of $\mathcal{D}$ and for a suborchard $\mathcal{O}' \subseteq \mathcal{O}$, $R(\mathcal{O}')$ denotes the union of the set of removable vertices of diamond trees in $\mathcal{O}'$. That is, $R(\mathcal{O}') := \cup_{\mathcal{D} \in \mathcal{O}'} R_{\mathcal{D}}$.

with $|I_{i'}| \geq pk$ for $i' \in [r-2]$ and $|I_{r-2}| \geq \gamma k$, the following holds. There exists a copy of K_r traversing $\cup_{j\in J}\tilde{Y}_{ij}$, $Z_{i_0j_0}$ and the sets $Z(I_{i'})$ for $i' \in [r-2]$. This is what we prove in what follows as we select our sets $\tilde{Y}_{ij}$.

We work through the $i \in [t]$ in order. Let $W_0 := \bigcup_{(i,j)\in[t]\times[\ell]}(Z_{ij} \cup V(\Pi_{ij}))$ and initiate with $U_0 = U \setminus W_0$. Suppose that we are at some step $i^* \in [t]$ and we have fixed $\mathcal{D}_{ij} = (T_{ij}, R_{ij}, \Sigma_{ij})$ for all $i < i^*$. In this step, we will fix $\mathcal{D}_{i^*j}$ for all $j \in [\ell]$. We define $W_{i^*} := \left(\bigcup_{(i,j):i<i^*} V(\mathcal{D}_{ij})\right) \cup W_0$. We further define for each $J \subseteq [\ell]$,

$$Y_J^{i^*} := \left\{v \in U \setminus W_{i^*} : v \in Y_{i^*j} \text{ for some } j \in [J] \text{ and } S_v \in \Upsilon_{i^*j} \cap K_{r-1}\left(G[U \setminus W_{i^*}]\right)\right\}.$$

In words, $Y_J^{i^*}$ keeps track of the vertices v which lie in one of the Y_{i^*j} with $j \in J$ which we can still use, in that the vertex v has not been used in a previous diamond tree and neither have the vertices of its associated copy of K_{r-1}, S_v. Note that $|W_{i^*}| \leq 4\alpha rn$ as a subset of vertices of a (k,m)-orchard with $km \leq 2\alpha n$. Hence if $|J| \geq \frac{\ell}{4r}$, we have that

$$|Y_J^{i^*}| \geq \frac{\ell}{4r}\left(\frac{\sqrt{\alpha}n}{\ell}\right) - 4\alpha rn \geq \left(\frac{\sqrt{\alpha}}{4r} - 4\alpha r\right)n \geq 2\alpha n,$$

using conditions (2_{ij}) and (4_{ij}) of Lemma 3.6.4 for $i = i^*$ and our upper bound on α.

We now define a random subset $\tilde{Y}^{i^*}$ by taking each vertex $v \in Y_{[\ell]}^{i^*}$ into $\tilde{Y}^{i^*}$ independently with probability $q := \frac{\ell m}{2\sqrt{\alpha}n}$, noting that $0 < q \leq 1$ due to the fact that $\ell m \leq km \leq 2\alpha n \leq 2\sqrt{\alpha}n$. For $j \in [\ell]$, we define $\tilde{Y}_{i^*j} := \tilde{Y}^{i^*} \cap Y_{i^*j}$. It follows from (2_{ij}) that $\mathbb{E}[|\tilde{Y}_{i^*j}|] = q|Y_{i^*j}| \leq \frac{m}{2}$ for all $j \in [\ell]$ and an application of Theorem 2.1.1 as well as a union bound gives that with high probability, $|\tilde{Y}_{i^*j}| \leq m$ for all $j \in [\ell]$. Note that in order to show that the upper bound on the failure probability given by Theorem 2.1.1 is strong enough to beat a union bound of the ℓ events, we use our lower bound on m and Fact 3.2.2. Furthermore, we have that with high probability, for any choice of $J \subseteq [\ell]$ with $|J| \geq \frac{\ell}{4r}$ and any choice of $(i_0, j_0) \in [t] \times [\ell]$ and subsets $I_1, \ldots, I_{r-2} \subset [t] \times [\ell]$ with $|I_{i'}| \geq pk$ for $i' \in [r-2]$ and $|I_{r-2}| \geq \gamma k$, there exists a copy of K_r traversing $\cup_{j\in J}\tilde{Y}_{i^*j}$, $Z_{i_0j_0}$ and the sets $Z(I_{i'})$ for $i' \in [r-2]$. Indeed, this follows from an application of Theorem 2.1.1 very similar to the proof of Lemma 3.6.2. For a fixed J, (i_0, j_0) and $I_{i'}$ for $i' \in [r-2]$ as above, we have that there is some subset X of αn vertices of $Y_J^{i^*}$ such that each vertex in X has a copy of K_{r-1} in its neighbourhood which traverses the sets $Z_{i_0j_0}$ and $Z(I_{i'})$ for $i' \in [r-2]$. Indeed, X can be found by repeated applications of Corollary 3.2.5 (2), deleting vertices from $Y_J^{i^*}$ and adding them to X on each application. Therefore $\mathbb{E}[|X \cap \tilde{Y}^{i^*}|] = q|X| = \frac{\sqrt{\alpha}\ell m}{2}$ and by Theorem 2.1.1, the probability that $X \cap \tilde{Y}^{i^*} = \emptyset$ is less than $e^{-\sqrt{\alpha}\ell m/4}$. Due to the fact that $\ell m \geq m \geq p^{r-1}n \geq n^{(r-2)/(2r-3)}$, because of our lower bound on m and Fact 3.2.2, we have that this probability tends to 0 as n tends to infinity. Moreover as there are less than $k \cdot (2^k)^{r-2} \cdot 2^\ell \leq 2^{rk}$ choices for such a (i_0, j_0), $I_{i'}$ for $i' \in [r-2]$ and J we have that a union bound gives the traversing copies of K_r for all choices with high probability. Indeed, we have that

$$rk \leq \frac{2\alpha rn}{m} \leq \frac{2\alpha r}{p^{r-1}} \leq 2\alpha r\ell m \leq \frac{\sqrt{\alpha}\ell m}{8},$$

using our conditions on km, m, ℓm and our upper bound on α from the hypotheses.

Therefore, we can fix an instance of Y^{i^*} which satisfies the desired conclusions that we have shown happen with high probability. For each $j \in [\ell]$, taking $\tilde{Y}_{i^*j} = \tilde{Y}^{i^*} \cap Y_{i^*j}$ and defining $\tilde{\Upsilon}_{i^*j} := \{S_v \in \Upsilon_{i^*j} : v \in \tilde{Y}_{i^*j}\}$, we apply condition (5_{ij}) for $i = i^*$ to get a family $\mathcal{Q}_{i^*}$ of $\mathcal{D}_{i^*j} = (T_{i^*j}, R_{i^*j}, \Sigma_{i^*j})$ for $j \in [\ell]$ such that for each j we have that $R_{i^*j} = Z_{i^*j} \cup \tilde{Y}_{i^*j}$ and $\Sigma_{i^*j} = \Pi_{i^*j} \cup \tilde{\Upsilon}_{i^*j}$. This completes the step for i^* and we move to $i^* + 1$ and repeat. Doing this for each $1 \leq i^* \leq t$ completes the proof. □

3.6.6 The existence of shrinkable orchards of smaller order

With Lemma 3.6.5 in hand, we can now complete the proof of Proposition 3.4.2 as follows.

Proof of Proposition 3.4.2. Fix $\varepsilon > 0$ small enough to apply Lemmas 3.6.1 and 3.6.5 and Proposition 3.6.3 all with the same α and γ and small enough that to guarantee that $p \geq Cn^{-1/(2r-3)}$ with $C := \left(\frac{2}{\alpha}\right)^{1/r}$ (see Fact 3.2.2). We also guarantee that $\varepsilon > 0$ is small enough to ensure that n is sufficiently large in what follows.

Now note that Proposition 3.6.3 directly implies the existence of the desired shrinkable orchard if $p^{1-r} \leq p^{r-1}n$ or if $p^{r-1}n \leq p^{1-r} \leq m \leq n^{7/8}$ and so we can assume from now on that

$$p^{r-1}n \leq m \leq \min\{p^{1-r}, n^{7/8}\}. \tag{3.6.3}$$

We are therefore in a position (due to our lower bound on m) to apply Lemma 3.6.5 but we first need to fix $k, t, \ell \in \mathbb{N}$ so that the conditions (3.6.2) are satisfied. We first fix $\ell \in \mathbb{N}$ so that $p^{1-r} \leq \ell m \leq 2p^{1-r}$. This is possible as $m \leq p^{1-r}$ and so there is a multiple of m in the desired range. Next we fix $t \in \mathbb{N}$ to be any integer such that $\alpha n \leq t\ell m \leq 2\alpha n$. Again, this is possible because $\ell m \leq 2p^{1-r} \leq \alpha n^{(r-1)/(2r-3)} \leq \alpha n$ using Fact 3.2.2. So there is indeed a multiple of ℓm in the desired range. Finally, we fix $k = t\ell$ and so we have that all the conditions in (3.6.2) are satisfied with our choice of parameters. Before analysing the conclusion of Lemma 3.6.5, let us point out a few further implications of our choices of parameters. Firstly, we have that

$$k^{-r^3\gamma} \geq \left(\frac{1}{k}\right)^{1/2^r} \geq \left(\frac{m}{n}\right)^{1/2^r} \geq \left(p^{r-1}\right)^{1/2^r} \geq p. \tag{3.6.4}$$

Moreover

$$\frac{\ell}{k} \leq \frac{2}{p^{r-1}mk} \leq \frac{2}{\alpha p^{r-1}n} \leq p, \tag{3.6.5}$$

where we use the upper bound on ℓm in the first inequality, the lower bound on km in the second inequality and the fact that $p \geq Cn^{-1/(2r-3)} \geq Cn^{-1/r}$ from Fact 3.2.2 in the last inequality. Putting (3.6.4) and (3.6.5) together then gives that

$$k^{1-r^3\gamma} \geq pk \geq \ell. \tag{3.6.6}$$

Now we apply Lemma 3.6.5 and let $\mathcal{O}$ be the resulting $(k,m)_r$-orchard partitioned into $(\ell,m)_r$-suborchards $\mathcal{Q}_1,\ldots,\mathcal{Q}_t$. We will show that $\mathcal{O}$ is γ-shrinkable by appealing to Lemma 3.6.1. Firstly note that due to the upper bound of $m \leq n^{7/8}$ and the fact that $k = \Theta(\frac{n}{m})$, by forcing n to be sufficiently large, we can assume that k is also sufficiently large to apply Lemma 3.6.1. We therefore just need to check the density condition (3.6.1) of the lemma. So fix some arbitrary $\mathcal{D} \in \mathcal{O}$ and suborchard $\mathcal{P} \subset \mathcal{O} \setminus \{\mathcal{D}\}$ such that $\mathcal{P} \geq \frac{k}{4r}$. By the pigeonhole principle, there exists an $i \in [t]$ such that $|\mathcal{P} \cap \mathcal{Q}_i| \geq \frac{\ell}{4r}$. So fix such an i and define $\mathcal{P}^* := \mathcal{P} \cap \mathcal{Q}_i$, noting that we have that $|\mathcal{P}^*| \leq k^{1-r^3\gamma}$ due to (3.6.6). Now we simply need to check that for any choice of suborchards $\mathcal{O}_1,\ldots,\mathcal{O}_{r-2} \subset \mathcal{P} \setminus \mathcal{P}^*$, with $|\mathcal{O}_{i'}| \geq k^{1-r^3\gamma}$ for $i' \in [r-3]$ and $|\mathcal{O}_{r-2}| \geq \gamma k$, there is a copy of K_r in G traversing $R_{\mathcal{D}}$, $R(\mathcal{P}^*)$ and the sets $R(\mathcal{O}_{i'})$ for $i' \in [r-2]$. This is verified by the conclusion of Lemma 3.6.5, setting $\mathcal{D}' = \mathcal{D}$, $\mathcal{Q}' = \mathcal{P}^*$ and $\mathcal{O}'_{i'} = \mathcal{O}_{i'}$ for $i' \in [r-2]$, noting that the lower bounds on the sizes of the $\mathcal{O}'_{i'}$ are guaranteed by (3.6.6). Hence $\mathcal{O}$ is indeed γ-shrinkable by Lemma 3.6.1 and this concludes the proof. □

3.7 The final absorption

The aim of this section is to prove Proposition 3.1.9. In order to prove this, in Section 3.7.1 we first define an absorbing structure whose vertex set will play the rôle of A in Proposition 3.1.9. We then prove that it has the required absorbing property. Next, in Section 3.7.2, we prove that we can find the absorbing structure in a suitably pseudorandom graph and show that this implies Proposition 3.1.9.

3.7.1 Defining an absorbing structure

Recall from Section 2.8 the definition of a template and the fact that templates of flexibility t with maximum degree 40 exist for all large enough t (Theorem 2.8.2). We will use a template as an auxiliary graph to define an *absorbing structure* which can contribute to a K_r-factor in many ways.

Definition 3.7.1. Let $\mathcal{T}$ be a template with flexibility t on vertex sets I and $J := J_1 \cup J_2$ with $|I| = 3t$ and $|J_1| = |J_2| = 2t$. A *K_r-absorbing structure* $\mathbb{A}$ of *order* M with respect to $\mathcal{T}$ in G consists of a labelled K_{r-1}-matching in G, $\Xi(\mathbb{A}) := \{S_i : i \in I\} \subset K_{r-1}(G)$ and a labelled $(4t, M)_r$-orchard $\mathcal{J}(\mathbb{A}) := \{\mathcal{D}_j : j \in J\}$ such that the following holds for each $i \in I$ and $j \in J$:

- $S_i \cap V(\mathcal{D}_j) = \emptyset$;
- if $ij \in E(\mathcal{T})$ then there is a vertex in the removable set of vertices R_j of $\mathcal{D}_j$ which forms a K_r with S_i in G.

We say that $\mathbb{A}$ has *flexibility* t, which is inherited from the template by which $\mathbb{A}$ is defined. We refer to the vertices of the absorbing structure, denoted $V(\mathbb{A})$, which is all vertices which feature in cliques in $\Xi(\mathbb{A})$ or diamond trees in $\mathcal{J}(\mathbb{A})$.

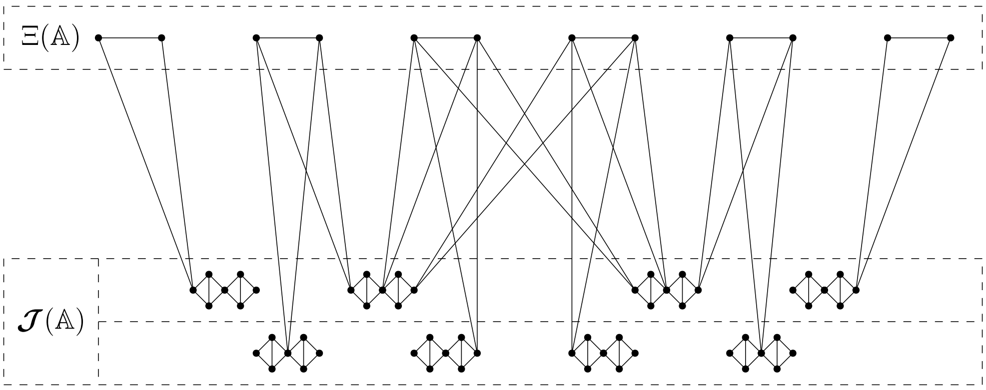


FIGURE 3.6: A K_3-absorbing structure of order 3 and flexibility 2, whose defining template is the template $\mathcal{T}$ displayed in Figure 2.1.

See Figure 3.6 for an example of an absorbing structure. Note that a K_r-absorbing structure $\mathbb{A}$ of flexibility t and order M has less than

$$3t(r-1) + 4t((2M-1)r+1) \leq 8rtM - rt + t \leq 8rtM \tag{3.7.1}$$

vertices. The absorbing structure is defined in such a way that it inherits the robust property that the template has with respect to perfect matchings but has such a property with respect to K_r-factors. The following lemma demonstrates this and reduces Proposition 3.1.9 to finding an appropriate absorbing structure in G.

Lemma 3.7.2. *For any $3 \leq r \in \mathbb{N}$ and $0 < \zeta, \eta < \frac{1}{2^{2r}}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$. Suppose that $t, M \in \mathbb{N}$ such that $tM \geq \zeta n$ and there exists an K_r-absorbing structure $\mathbb{A}$ of flexibility t (with respect to some template $\mathcal{T}$) and order M in G. Let $A := V(\mathbb{A})$.*

Then there exists some vertex subset $B \subset V(G)$, such that $|B| \leq \eta p^{2r-4} n$ and for any $(k, m)_r$-orchard $\mathcal{R}$ whose vertices lie in $V(G) \setminus (A \cup B)$ with $|A| + |V(\mathcal{R})| \in r\mathbb{N}$, $k \leq \frac{t}{4r}$ and $m \geq M$, we have that $G[A \cup V(\mathcal{R})]$ has a K_r-factor.

Proof. Fix $\varepsilon > 0$ small enough to apply Lemma 3.1.4 with ζ and η as defined here and small enough to apply Corollary 3.2.5 with $\tau := \frac{\zeta}{r+1}$. Let $\mathcal{O} = \{\mathcal{D}_j : j \in J_2\}$ be the suborchard of $\mathcal{J}(\mathbb{A})$ defined by those indices which lie in the flexible set J_2 of the template $\mathcal{T}$ which defines $\mathbb{A}$. Thus $\mathcal{O}$ is a $(2t, M)_r$-orchard. Therefore, applying Lemma 3.1.4, we have that there exists a set $B \subset V(G)$ with $|B| \leq \eta p^{2r-4} n$ and for any $(k, m)_r$-orchard $\mathcal{R}$ as in the statement of the lemma, $\mathcal{O}$ absorbs $\mathcal{R}$. Indeed, note that in the notation of Lemma 3.1.4, we have that k, m and M are as defined here while $K = 2t$. Hence the condition that $k \leq \frac{K}{8r}$ is precisely the same as our presumption that $k \leq \frac{t}{4r}$ whilst the condition that $kM \leq mK$ is guaranteed by the fact that $m \geq M$ and $k \leq \frac{K}{8r}$. Unpacking the conclusion of Lemma 3.1.4, we thus have that for any such $\mathcal{R}$ there exists some subfamily $\mathcal{P}_1 \subseteq \mathcal{O}$ such that $|\mathcal{P}_1| = (r-1)k \leq \frac{t}{2}$ and $G[V(\mathcal{P}_1) \cup V(\mathcal{R})]$ has a K_r-factor. We will show that $G[A \setminus V(\mathcal{P}_1)]$ also has a K_r-factor which will complete the proof.

Now note that for any $\mathcal{P} \subset \mathcal{O}$ such that $|\mathcal{P}| = t$, we have that $G[A \setminus V(\mathcal{P})]$ has a K_r-factor. Indeed let $\bar{J} := \{j \in J_2 : \mathcal{D}_j \in \mathcal{P}\}$ be the indices of diamond trees that feature in the suborchard $\mathcal{P}$. By the definition of the template $\mathcal{T}$, Definition 2.8.1, we know that there is a perfect matching $F \subset E(\mathcal{T})$ in $\mathcal{T}[V(\mathcal{T}) \setminus \bar{J}]$. Now for $ij \in F$, we can take a K_r-factor on $S_i \cup V(\mathcal{D}_j)$ in G guaranteed by the fact that S_i forms a copy of K_r with a removable vertex of $\mathcal{D}_j$ (Definition 3.7.1) and the key property of the removable vertices of a K_r-diamond tree (Observation 3.1.2). As F is a perfect matching in $\mathcal{T}[V(\mathcal{T}) \setminus \bar{J}]$, we see that by taking these K_r-factors for each $ij \in F$, we obtain a K_r-factor in $G[A \setminus V(\mathcal{P})]$ as required.

If we had that $|\mathcal{P}_1| = t$, this would complete the proof. However we have that $\mathcal{P}_1$ is actually much smaller than this. Indeed $|\mathcal{P}_1| \leq \frac{t}{2}$. We will proceed by finding some $\mathcal{P}_2 \subset \mathcal{O} \setminus \mathcal{P}_1$ such that $G[V(\mathcal{P}_2)]$ has a K_r-factor and $|\mathcal{P}_1| + |\mathcal{P}_2| = t$. We build $\mathcal{P}_2$ by the following greedy process. We initiate with $\mathcal{O}' = \mathcal{O} \setminus \mathcal{P}_1$ and $\mathcal{P}_2 = \emptyset$. Then at each time step, as long as $|\mathcal{P}_1| + |\mathcal{P}_2| + r \leq t$ we partition $\mathcal{O}'$ into r parts $\mathcal{O}' = \mathcal{O}_1 \cup \cdots \cup \mathcal{O}_r$ such that the sizes of the parts are as equal as possible. We let R_x be the union of the removable vertices of diamond trees in the orchard $\mathcal{O}_x$ for $x \in [r]$. We have that each R_x has size at least $\frac{tM}{r+1} \geq \frac{\zeta n}{r+1} = \tau n$. Therefore, by Corollary 3.2.5 (2), there is a copy of K_r traversing the R_x, $x \in [r]$. This gives some r-tuple of diamond trees $\mathcal{D}_1, \ldots, \mathcal{D}_r$ such that $\mathcal{D}_x \in \mathcal{O}_x$ for all $x \in [r]$ and there is a K_r-factor in $G[V(\mathcal{D}_1) \cup \cdots \cup V(\mathcal{D}_r)]$, given by taking the copy of K_r that traverses their sets of removable vertices and applying Observation 3.1.2. We add $\mathcal{D}_1, \ldots, \mathcal{D}_r$ to $\mathcal{P}_2$ and delete these diamond trees from the orchard $\mathcal{O}'$ which completes this time step. Clearly at all points in this process there is a K_r-factor in $G[V(\mathcal{P}_2)]$ and we claim that this process terminates when $|\mathcal{P}_1| + |\mathcal{P}_2|$ is exactly equal to t. Indeed if this is not the case, as we increase the size of $|\mathcal{P}_2|$ by exactly r in each step, we have that $|\mathcal{P}_1| + |\mathcal{P}_2| = t - s$ for some $s \in [r-1]$ at the end of the process. Let $\mathcal{P}_3 \subset \mathcal{O} \setminus (\mathcal{P}_1 \cup \mathcal{P}_2)$ be a set of s K_r-diamond trees. Now as $V_1 := V(\mathcal{R}) \cup V(\mathcal{P}_1) \cup V(\mathcal{P}_2)$ hosts a K_r-factor, we have that $r||V_1|$. Likewise, we know from above that $V_2 = A \setminus (\cup_{i=1}^3 V(\mathcal{P}_i))$ hosts a K_r-factor and so $r||V_2|$. Due to the fact that r divides the size of $A \cup V(\mathcal{R})$ and $A \cup V(\mathcal{R}) = V_1 \cup V_2 \cup V(\mathcal{P}_3)$, we can defer that $r||V(\mathcal{P}_3)|$. This is a contradiction as $\mathcal{P}_3$ is a set of s K_r-diamond trees for some $1 \leq s \leq r-1$ and the number of vertices in any diamond tree is $1 \mod r$. Therefore we can find $\mathcal{P}_2$ as claimed.

Finally, taking $\mathcal{P} := \mathcal{P}_1 \cup \mathcal{P}_2$, we are then done by taking our K_r-factor in $G[A \cup V(\mathcal{R})]$ to be the union of the K_r-factor in $G[V(\mathcal{R}) \cup V(\mathcal{P}_1)]$, the K_r-factor in $G[V(\mathcal{P}_2)]$ and the K_r-factor in $G[A \setminus V(\mathcal{P})]$. □

3.7.2 Finding an absorbing structure

Lemma 3.7.2 reduces Proposition 3.1.9 to finding an appropriate absorbing structure in G. In this section we prove that this is possible by proving the following proposition.

Proposition 3.7.3. *For any $3 \leq r \in \mathbb{N}$ and $0 < \alpha < \frac{1}{2^{3r}}$ there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p,β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$ and any vertex subset $W \subseteq V(G)$ with $|W| \geq \frac{n}{2}$. There exists an K_r-absorbing structure $\mathbb{A}$ in G of flexibility $t = \alpha n^{1/8}$ and order $M = n^{7/8}$ such that $V(\mathbb{A}) \subseteq W$.*

With Lemma 3.7.2 and Proposition 3.7.3, the proof of Proposition 3.1.9 follows readily as we now show.

Proof of Proposition 3.1.9. Fix $\zeta := \frac{\alpha}{8r}$ and $\varepsilon > 0$ small enough to apply Lemma 3.7.2 with ζ and η as defined here and small enough to apply Proposition 3.7.3 with $\alpha_{3.7.3} = \zeta$. We can therefore apply Proposition 3.7.3 to get an absorbing structure $\mathbb{A}$ in G with flexibility $t = \zeta n^{1/8}$ and order $M = n^{7/8}$. We have that $A = V(\mathbb{A}) \subset W$ has size $|A| \leq 8rtM = \alpha n$ (see (3.7.1)). The conclusion then follows from Lemma 3.7.2 noting that $k \leq \alpha^2 n^{1/8}$ implies that $k \leq \frac{t}{4r} = \frac{\zeta n^{1/8}}{4r} = \frac{\alpha n^{1/8}}{48r^2}$ due to our upper bound on α. □

Now in order to prove the existence of an absorbing structure as in Proposition 3.7.3, we will first fix some template $\mathcal{T}$ which will define $\mathbb{A}$. Next, we will set aside a large matching $\Pi \subset K_{r-1}(G)$ of $(r-1)$-cliques. These will be candidates for the K_{r-1}-matching $\Xi(\mathbb{A})$ in our absorbing structure but we start with a much bigger set Π of size $\Omega(n^{2/3})$. Moreover, to each $(r-1)$-clique $S \in \Pi$ we will associate a set of vertices $X_S \subset G$ such that $X_S \subset N_G(S)$, $|X_S| = \Omega(n^{1/3})$ and, crucially, the sets $\{X_S : S \in \Pi\}$ are *disjoint*. We will find this K_{r-1}-matching Π with a simple greedy procedure, appealing to Corollary 3.2.5 (3) to find each $S \in \Pi$ (and its corresponding neighbourhood set X_S), one by one.

After finding Π, we then turn to constructing the $(4t, M)$-orchard $\mathcal{J}(\mathbb{A})$ for the absorbing structure $\mathbb{A}$. Again, this will be done greedily, fixing the diamond trees $\mathcal{D} \in \mathcal{J}(\mathbb{A})$ one at a time. Let us consider fixing some diamond tree $\mathcal{D}_j \in \mathcal{J}(\mathbb{A})$. Note that as we fix $\mathcal{D}_j$, we immediately get restrictions on which $S \in \Pi$ remain as candidates to play the rôle of certain $S_i \in \Xi(\mathbb{A})$. Indeed, if the removable vertices of $\mathcal{D}_j$ are disjoint from $N_G(S)$ and ij is an edge in the template $\mathcal{T}$ defining $\mathbb{A}$, then there is no way S can play the rôle of S_i in $\Xi(\mathbb{A})$. Therefore as we fix our diamond trees, we will aim to have that their sets of removable vertices intersect as many of the X_S (and hence neighbourhoods $N_G(S)$) for $S \in \Pi$, as possible.

In order to do this, we will use the following lemma, which shows that we can find diamond trees whose removable vertices intersect many prescribed sets (in our case this will be the sets X_S). The proof of this lemma is a simple application of Proposition 3.3.1.

Lemma 3.7.4. *For any $3 \leq r \in \mathbb{N}$ and $0 < \alpha < \frac{1}{2^{2r}}$, there exists an $\varepsilon > 0$ such that the following holds for any n-vertex (p, β)-bijumbled graph G with $\beta \leq \varepsilon p^{r-1} n$.*

Suppose $\frac{\alpha^2}{2} n^{2/3} \leq \ell \leq \alpha n^{2/3}$ and we have disjoint vertex subsets $W, U_1, \ldots, U_\ell$ such that $|W| \geq \frac{n}{4}$ and $|U_i| \geq n^{1/3}$ for all $i \in [\ell]$. Then there exists a diamond tree $\mathcal{D} = (T, R, \Sigma)$ in G such that the following conditions hold:

(i) $\Sigma \subset K_{r-1}(G[W])$ is a K_{r-1}-matching in W;

(ii) $R \subset \cup_{i=1}^{\ell} U_i$ and R intersects ℓ' of the sets U_i for some $\ell' \geq \frac{\ell}{4r}$;

(iii) The order of $\mathcal{D}$ is at most $n^{2/3}$;

(iv) For all but at most $n^{1/2}$ of the indices $i \in [\ell]$, we have that $|V(\mathcal{D}) \cap U_i| \leq n^{1/6}$.

Proof. We begin by fixing $\gamma := \frac{\ell}{n^{2/3}}$ so that $\frac{\alpha^2}{2} \leq \gamma \leq \alpha$ and we fix ε small enough to apply Proposition 3.3.1 with $\alpha_{3.3.1} = \alpha' := \frac{\gamma}{4r}$ and small enough to guarantee that $p \geq C n^{-1/(2r-3)}$ with $C = \frac{4}{\alpha'}$ (see Fact 3.2.2). Now shrink each set U_i so that it has exactly $n^{1/3}$ vertices and define $U := \cup_{i=1}^{\ell} U_i$. Furthermore fix $\delta := \alpha'^2 p^{r-1} n$ and apply Proposition 3.3.1 with U, W and $z = \alpha' n$. So we get disjoint subsets $X, Y \subset U$ as in the outcome of Proposition 3.3.1.

Now firstly note that as $|X| + |Y| = z = \alpha' n$, $|U| = \ell n^{1/3} = \gamma n = 4rz$ and each of the U_i have equal size, we must have that $X \cup Y$ intersects at least $\frac{\ell}{4r}$ of the sets U_i. We will choose our $\mathcal{D} = (T, R, \Sigma)$ so that R intersects all the sets U_i that $X \cup Y$ intersects, thus guaranteeing condition (ii). Indeed, if we let $Y' \subset Y$ be the *minimal* subset of Y such that there exists no $i \in [\ell]$ with $Y \cap U_i \neq \emptyset$ and $Y' \cap U_i = \emptyset$, Proposition 3.3.1 gives the existence of a diamond tree $\mathcal{D} = (T, R, \Sigma)$ so that $R = X \cup Y'$ and $\Sigma \subset K_{r-1}(G[W])$ and so conditions (i) and (ii) are satisfied.

In order to establish condition (iii), note that $|Y'| \leq \ell \leq \frac{n^{2/3}}{2}$ and if $|X| > 1$ then

$$|X| \leq \frac{2z}{\delta} \leq \frac{2}{\alpha' p^{r-1}} \leq \frac{2n^{(r-1)/(2r-3)}}{\alpha' C^{r-1}} \leq \frac{n^{2/3}}{2},$$

due to our definition of C and the fact that $\frac{r-1}{2r-3} \leq \frac{2}{3}$ for all $r \geq 3$. Finally, condition (iv) is a simple consequence of (iii). Indeed, if (iv) were not true, then as the U_i are pairwise disjoint, we would have that $\mathcal{D}$ has order greater than $n^{1/2} \cdot n^{1/6} \geq n^{2/3}$, a contradiction. □

Let us return to sketching the proof of Proposition 3.7.3, considering now that we can use Lemma 3.7.4 to find diamond trees $\mathcal{D} \in \mathcal{J}(\mathbb{A})$. As discussed above, the key property of diamond trees generated by Lemma 3.7.4 is (ii), allowing us to find diamond trees that intersect many of sets $\{X_S : S \in \Pi\}$ which we begin the proof with. The property (iv) will also be useful as it shows that in the process of building $\mathcal{J}(\mathbb{A})$ one by one, we do not destroy many of the sets X_S and most of them remain large and can be used by other $\mathcal{D} \in \mathcal{J}(\mathbb{A})$.

One potentially troublesome consequence of Lemma 3.7.4 is that the diamond trees it finds are far too small (iii). Indeed the diamond trees in our orchard $\mathcal{J}(\mathbb{A})$ are supposed to be of order $M = n^{7/8}$. It turns out that this is not such a big hurdle as we can find a large diamond tree disjoint from all the X_S and connect it to the diamond tree C output by Lemma 3.7.4. In more detail, we can apply Proposition 3.3.1 to create a large (linear) pool Y of vertices that can be removable vertices of some diamond tree which will be disjoint from all the vertices in the sets X_S. We also consider the large (linear) pool of vertices Z that lie in some $X_S \setminus V(C)$ with $S \in \Pi$ such that the removable vertices of C intersect X_S. It is not hard to show (see for example Corollary 3.2.5 (3)) that there is a copy of K^-_{r+1} with one degree $r-1$ vertex in Y and the other in $X_{S^*} \subset Z$ for some $S^* \in \Pi$. By also taking S^* into $\mathcal{D}$ and choosing an appropriate $Y' \subset Y$ to apply the key property of Proposition 3.3.1, we can obtain a diamond tree $\mathcal{D}$ of the correct size that contains the diamond tree C output by Lemma 3.7.4.

More troublesome is the fact that the condition (ii) which gives that the removable vertices of C intersect many of the desired sets X_S is, in fact, not strong enough. Indeed, consider some fixed $i \in I$ for which we want to find a copy S_i of K_{r-1} to lie in $\Xi(\mathbb{A})$. If $j, j' \in N_{\mathcal{T}}(i)$ and the sets $\{X_S : S \in \Pi, R_{\mathcal{D}_j} \cap X_S \neq \emptyset\}$ and $\{X_S : S \in \Pi, R_{\mathcal{D}_{j'}} \cap X_S \neq \emptyset\}$ (here, as usual, we use $R_{\mathcal{D}}$ to denote the removable vertices of $\mathcal{D}$) are disjoint, then already there are no candidates for S_i in Π. To fix this, we actually need that when we choose a diamond tree $\mathcal{D} \in \mathcal{J}(\mathbb{A})$, we want $R_{\mathcal{D}}$ to intersect *almost all* of the sets $\{X_S : S \in \Pi\}$. We achieve this by iterating Lemma 3.7.4, creating constantly many disjoint diamond trees C that together hit almost all of the X_S with their removable vertices. We then connect all of these diamond trees C with a large diamond tree disjoint from the sets X_S to obtain the desired diamond tree $\mathcal{D} \in \mathcal{J}(\mathbb{A})$. This connecting process is similar (although slightly more involved) to the connecting process outlined in the previous paragraph. We now give the full details for the proof of Proposition 3.7.3, concluding this section and chapter.

Proof of Proposition 3.7.3. We begin by fixing $\varepsilon > 0$ small enough to apply Corollary 3.2.5 with $\tau := \frac{\alpha^2}{16r}$ and to apply Proposition 3.3.1 and Lemma 3.7.4 each with α as in the statement.

We also make sure that $\varepsilon > 0$ is small enough to force n to be sufficiently large in what follows and small enough to guarantee that $p \geq C'n^{-1/(2r-3)}$ for $C' := \frac{2}{\tau^2}$ using Fact 3.2.2. We further fix some template $\mathcal{T}$ with vertex sets I and $J = J_1 \cup J_2$ of flexibility t and maximum degree 40 which we know exists for n (and hence t) sufficiently large due to Montgomery [136] (see Theorem 2.8.2).

We will find an absorbing structure with respect to $\mathcal{T}$ and so must prove the existence of a K_{r-1}-matching $\Xi(\mathbb{A}) = \{S_i : i \in I\} \subset K_{r-1}(G[W])$ of $3t$ copies of K_{r-1}, and a $(4t, M)$-orchard $\mathcal{J} = \mathcal{J}(\mathbb{A}) = \{\mathcal{D}_j : j \in J\}$ such that the conditions of Definition 3.7.1 are satisfied. We will do this in three stages. In Claim 3.7.5, we fix some large matching $\Pi \subset K_{r-1}(G[W])$ of $(r-1)$-cliques which will be candidates for the $(r-1)$-cliques which will feature in $\Xi(\mathbb{A})$. We will guarantee that the cliques in Π are contained in many copies of K_r which will help as we proceed to build our absorbing structure. In Claim 3.7.6, we will fix the K_r-diamond trees which will form our orchard $\mathcal{J}$ for our K_r-absorbing structure. We will carefully control how these diamond trees intersect the cliques in our candidate set Π and their neighbourhoods. Finally, we will show that we can find a suitable $\Xi(\mathbb{A}) \subset \Pi$ so that we obtain the desired absorbing structure.

Claim 3.7.5. *There exists a matching $\Pi = \{S_1, \dots, S_\ell\} \subset K_{r-1}(G[W])$ of $\ell := \alpha n^{2/3}$ copies of K_{r-1} and sets $X_h \subset W \setminus V(\Pi)$ for each $h \in [\ell]$, such that the X_h are pairwise disjoint, each has size $|X_h| = 2n^{1/3}$ and for all $h \in [\ell]$ we have that $X_h \subset N_G(S_h; W)$.*

<u>Proof of Claim:</u> We can do this by way of a simple greedy process choosing such an $(r-1)$-clique S_h and set X_h in order for $h = 1, \dots, \ell$. When choosing S_h and X_h, we look at the set of vertices $V_h \subset W$ which have not been used in previous choices of $S_{h'}$ or $X_{h'}$. We have that

$$|V_h| \geq |W| - |\cup_{h'<h} (X_{h'} \cup S_{h'})| \geq \frac{n}{2} - (\ell-1)(r-1+2n^{1/3}) \geq \left(\frac{1}{2} - 2\alpha\right) n \geq \frac{n}{4}$$

and an application of Corollary 3.2.5 (3) with $W_0 = W_1 = W_2 = V_h$ gives the desired S_h and X_h in V_h using that

$$\tau^2 p^{r-1} n \geq \tau^2 C' n^{1-(r-1)/(2r-3)} \geq 2n^{1/3},$$

due to Fact 3.2.2. ■

Next we turn to fixing our $(4t, M)$-orchard $\mathcal{J}$.

Claim 3.7.6. *Let S_h and X_h for $h = 1, \dots, \ell$ be as in Claim 3.7.5. Then there exists a $(4t, M)$-orchard $\mathcal{J} = \{\mathcal{D}_1, \dots, \mathcal{D}_{4t}\}$ such that $V(\mathcal{J}) \subset W$ and the following properties hold for each $\mathcal{D}_j = (T_j, R_j, \Sigma_j)$ with $j \in [4t]$:*

1. *The set of removable vertices R_j intersects at least $(1-\alpha)\ell$ of the sets X_h with $h \in [\ell]$;*
2. *$V(\mathcal{D}_j)$ intersects at most $C := \frac{\log(\frac{2}{\alpha})}{\log(\frac{4r}{4r-1})}$ of the S_h with $h \in [\ell]$.*

Before proving the claim, let us see how it implies the proposition. Indeed, taking the $(4t, M)_r$ orchard $\mathcal{J}$ from Claim 3.7.6 as $\mathcal{J}(\mathbb{A})$, we just need to choose a K_{r-1}-matching $\Xi(\mathbb{A}) = \{S_i : i \in I\}$ so that $S_i \cap V(\mathcal{J}) = \emptyset$ for all $i \in I$ and whenever $ij \in E(\mathcal{T})$, we have that there is a vertex in R_j which forms a copy of K_r with S_i. We do this greedily, showing that for each $i = 1, 2, \ldots, 3t$ in order, there is a suitable choice for S_i in Π. We initiate by fixing $L \subseteq [\ell]$ to be the indices $h \in [\ell]$ such that $S_h \cap V(\mathcal{J}) = \emptyset$. By condition (2) in Claim 3.7.6, for large n we have that

$$|L| \geq \ell - 4Ct \geq (1 - \alpha)\ell,$$

at the beginning of this process, recalling that $\ell = \alpha n^{2/3}$ and $t = \alpha n^{1/8}$. Now for $i = 1, \ldots, 3t$, we find an index $h = h(i) \in L$ such that S_h forms a copy of K_r with a vertex in R_j for all j such that $ij \in E(\mathcal{T})$. We fix $S_i = S_h$ and delete h from L. If this process succeeds in finding a suitable $h = h(i)$ for each $i \in I$ then the resulting $\Xi(\mathbb{A}) = \{S_i : i \in I\}$ along with $\mathcal{J}$ form the desired K_r-absorbing structure.

It remains to check that we are successful at each step. So consider step $i^* \in [3t]$. We have that $|L| \geq (1-\alpha)\ell - (i^* - 1) \geq (1-2\alpha)\ell$ at the beginning of the step, using here that n (and hence ℓ) is sufficiently large in the second inequality here. Now for each $j \in J$ which is a neighbour of i^* in the template $\mathcal{T}$, we have by Claim 3.7.6 (1) that there are at most $\alpha\ell$ indices $h \in [\ell]$ such that no vertex of R_j forms a K_r with S_h in G. Indeed for almost all choices of h, we have that $R_j \cap X_h \neq \emptyset$ and $X_h \subset N_G(S_h)$. Given that $\mathcal{T}$ has maximum degree 40, we have that this gives at most $40\alpha\ell$ indices $h \in L$ that would not be a good choice for $h(i^*)$. Therefore there are at least $(1 - 42\alpha)\ell$ indices $h \in L$ which can be chosen as $h(i^*)$ and we simply choose one arbitrarily.

This shows that the algorithm is successful in generating the desired absorbing structure and so it only remains to prove Claim 3.7.6, which we do now.

<u>Proof of Claim 3.7.6:</u> We will find the diamond trees $\mathcal{D}_j$, $j = 1, \ldots, 4t$ one by one so that they are vertex-disjoint and satisfy the two conditions in the statement of the claim as well as the further following condition:

3. $V(\mathcal{D}_j)$ intersects less than $Cn^{1/2}$ of the X_h with $h \in [\ell]$ in more than $2Cn^{1/6}$ vertices.

We will initiate the process with $\mathbb{L} := [\ell]$ and $U_h = X_h$ for all $h \in [\ell]$. These sets U_h will keep track of vertices in X_h that we are still allowed to use, that is, those vertices which have not been used in previously chosen diamond trees. Furthermore, the set $\mathbb{L} \subset [\ell]$ will keep track of all indices which are *alive*. When we choose a $\mathcal{D}_j$ for some $j \in [4t]$, we *kill* (and remove from $\mathbb{L}$) all the indices $h \in [\ell]$ such that $V(\mathcal{D}_j)$ intersects X_h in more than $2Cn^{1/6}$ vertices. We also kill any index h such that $V(\mathcal{D}_j)$ intersects S_h. Due to our conditions (2) and (3), we have that throughout the process,

$$|\mathbb{L}| \geq \ell - 4t(C + Cn^{1/2}) \geq \left(1 - \frac{\alpha}{2}\right)\ell,$$

for n large recalling that $\ell = \alpha n^{2/3}$ and $t = \alpha n^{1/8}$. Moreover, due to condition (3), at any point in the process, for all alive indices h in $\mathbb{L}$, the size of $U_h \subset X_h$ is at least

$$|U_h| \geq |X_h| - \sum_j |V(\mathcal{D}_j) \cap X_h| \geq 2n^{1/3} - 8tCn^{1/6} \geq n^{1/3},$$

for n sufficiently large. We remark that it is crucial in the previous two calculations that $t = n^{1/8}$ and so when choosing our diamond trees, we do not kill too many indices or make too many of the sets X_h too small to be used by subsequent diamond trees. In fact any t polynomially smaller than $n^{1/6}$ would suffice for this.

So let us suppose that we are at step $j^* \in [4t]$ where we look to find $\mathcal{D}_{j^*}$ and we have some fixed set $\mathbb{L}$ of alive indices and subsets $U_h \subset X_h$ for $h \in \mathbb{L}$. We run a sub-algorithm that finds $\mathcal{D}_{j^*}$ in two phases. We begin by setting $\mathbb{K} = \mathbb{L}$ and $\mathcal{C} = \emptyset$. The first phase of the sub-algorithm works by finding at most C small order diamond trees whose removable vertices intersect many of the U_h for $h \in \mathbb{L}$. The family $\mathcal{C}$ will collect these small order diamond trees and the set $\mathbb{K}$ will keep track of the indices h in $\mathbb{L}$ for which we have not yet intersected U_h. In the second phase of the algorithm, we will form $\mathcal{D}_{j^*}$ by joining together the diamond trees in $\mathcal{C}$ so that they form one diamond tree. By guaranteeing that our diamond trees in $\mathcal{C}$ have removable vertices that intersect most of the sets U_h, we will guarantee condition (1) of the claim. Before starting, we also initiate by setting $W' \subset W$ to be

$$W' = W \setminus \left(\cup_{h\in[\ell]}(S_h \cup X_h) \cup_{j<j^*} V(\mathcal{D}_j)\right).$$

In words, W' is the subset of vertices of W that has not been used in any of the structures that we have found so far. Finally we initiate a counter by setting $s = 1$.

At step s, we apply Lemma 3.7.4 on the sets W' and $\{U_h : h \in \mathbb{K}\}$. We thus find a K_r-diamond tree $C_s = (T, R, \Sigma)$ which we add to $\mathcal{C}$, which has the following properties guaranteed by Lemma 3.7.4:

(i) $\Sigma \subset K_{r-1}(G[W'])$ and we delete $V(\Sigma)$ from W';

(ii) $R \subset \cup_{h\in\mathbb{K}} U_h$ and defining $\mathbb{K}_s \subset \mathbb{K}$ to be $\mathbb{K}_s := \{h' : R \cap U_{h'} \neq \emptyset\}$, we have that $|\mathbb{K}_s| \geq \frac{|\mathbb{K}|}{4r}$. We delete $\mathbb{K}_s$ from $\mathbb{K}$;

(iii) The order of C_s is at most $n^{2/3}$;

(iv) There is a set $\mathbb{J}_s \subset \mathbb{K}_s \subset \mathbb{K} \subset [\ell]$ of at most $n^{1/2}$ indices, such that for all $h \in [\ell] \setminus \mathbb{J}_s$ we have that $|V(C_s) \cap U_h| \leq n^{1/6}$.

Finding such a C_s concludes this step s. If $|\mathbb{K}| < \frac{\alpha\ell}{2}$, we terminate this phase and move on to the next phase. If $|\mathbb{K}| \geq \frac{\alpha\ell}{2}$, we move to step $s+1$.

We must check that the conditions for Lemma 3.7.4 are satisfied throughout this phase in order to find the required diamond trees C_s at each step. Indeed this follows because

$$\frac{\alpha^2}{2}n^{2/3} = \frac{\alpha}{2}\ell \leq |\mathbb{K}| \leq \ell = \alpha n^{2/3}$$

throughout and we have that $|U_h| \geq n^{1/3}$ for all $h \in \mathbb{K}$ as $\mathbb{K} \subset \mathbb{L}$ is a subset of alive indices. Finally we have $|W'| \geq \frac{n}{4}$ throughout this process. Indeed, note that due to condition (ii) and the fact that we only continue until $|\mathbb{K}| < \frac{\alpha\ell}{2}$, we have that the process runs for a maximum of C steps, recalling the definition of C from condition (2) of the claim. That is, we have that $|\mathcal{C}| \leq C$ throughout and so

$$\begin{aligned}
|W'| &\geq |W| - \sum_{h\in[\ell]}(|S_h| + |X_h|) - \sum_{j<j^*}|V(\mathcal{D}_j)| - \sum_{C\in\mathcal{C}} V(C) \\
&\geq \frac{n}{2} - \ell \cdot 3n^{1/3} - 8trM - Cn^{2/3} \\
&\geq \left(\frac{1}{2} - (4+8r)\alpha\right)n \geq \frac{n}{4},
\end{aligned} \tag{3.7.2}$$

due to our upper bound on α, for n sufficiently large. This verifies that we find C_s at every step s of this process and so we finish this phase with $|\mathbb{K}| < \frac{\alpha\ell}{2}$ and some family $\mathcal{C} = \{C_1, \ldots, C_c\}$ of $c \leq C$ vertex-disjoint K_r-diamond trees.

Now we describe how we generate $\mathcal{D}_{j^*}$ which will have all the diamond trees $C_s \in \mathcal{C}$ as sub-diamond trees. We refer the reader to Figure 3.7 to keep on track of the many components that contribute to our $\mathcal{D}_{j^*}$. One thing to note is that the sum of the orders of the diamond trees in $\mathcal{C}$ is far too small for us to just build $\mathcal{D}_{j^*}$ from the diamond trees in $\mathcal{C}$. Indeed the sum of the orders is $O(n^{2/3})$ and we want $\mathcal{D}_{j^*}$ to have order $M = n^{7/8}$. Therefore we will have to find the majority of the K_r-diamond tree $\mathcal{D}_{j^*}$ elsewhere. In order to prepare for this, we first split W' arbitrarily into U_0, W_0 and Z_0 of roughly equal size and note that due to our lower bound (3.7.2) on $|W'|$, we have that each of these sets has size at least $\frac{n}{16}$. Next we fix $\delta := \alpha^2 p^{r-1} n$ and $z = \alpha^2 n$ and apply Proposition 3.3.1 with respect to the sets U_0 and W_0 to get disjoint sets $X, Y \subset U_0$ as detailed there. Note that $|X| \leq 2n^{2/3}$. Indeed if $|X| > 1$, then we have that $|X| \leq \frac{2z}{\delta} \leq 2p^{1-r} \leq 2n^{2/3}$ due to Fact 3.2.2.

Now for $1 \leq s \leq c$, define $Z_s := \cup_{h\in\mathbb{K}_s\setminus\mathbb{J}_s}(U_h \setminus V(C_s))$. In words, Z_s is the union of the sets U_h which C_s intersects, after removing the sets $U_{h'}$ which C_s intersects in too many vertices and then removing the vertices of C_s. Now we have that for each $s \in [c]$,

$$|Z_s| \geq (|\mathbb{K}_s| - |\mathbb{J}_s|)(n^{1/3} - n^{1/6}) \geq \frac{\alpha\ell n^{1/3}}{8r} - 2n^{5/6} \geq \frac{\alpha^2 n}{16r} \geq \tau n,$$

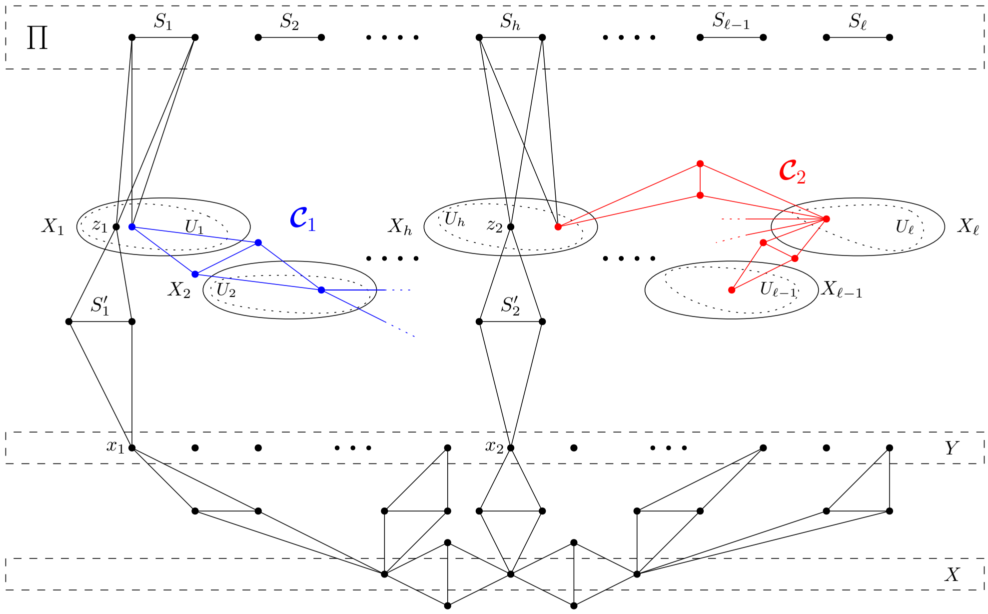


FIGURE 3.7: An example of $\mathcal{D}_{j^*}$ and its components. In this case, we have $c = 2$, $h_1 = 1$ and $h_2 = h$.

for n large, as the U_h are pairwise disjoint. Note also that as the $\mathbb{K}_s$ are pairwise disjoint, the Z_s are also pairwise disjoint for $s \in [c]$. Now for $1 \le s \le c$, apply Corollary 3.2.5 (3) to find an $(r-1)$-clique $S'_s \subset K_{r-1}(G[Z_0])$ such that there is a vertex $z_s \in Z_s \cap N_G(S'_s)$ and a vertex $x_s \in (X \cup Y) \cap N_G(S'_s)$. We delete the S'_s from Z_0 and move to the next index $s+1$ or finish if $s = c$.

Now choose some $Y' \subset Y$ such that $x_s \in X \cup Y'$ for all $s \in [c]$ and

$$|Y'| + |X| + \sum_{s \in [c]} (|R_{C_s}| + 1) = M.$$

This is easily done as $|X| + |Y| = \alpha^2 n$ is linear and $|X|, |R_{C_s}| \le 2n^{2/3}$ for all $s \in [c]$ which is much smaller than $M = n^{7/8}$. By Proposition 3.3.1, there is a K_r-diamond tree $\tilde{\mathcal{D}} = (\tilde{T}, \tilde{R}, \tilde{\Sigma})$ with $\tilde{R} = X \cup Y'$ and $\tilde{\Sigma} \subset K_{r-1}(G[W_0])$ a K_{r-1}-matching in $W_0 \subset W'$. Our diamond tree $\mathcal{D}_{j^*}$ is then obtained by connecting $\tilde{\mathcal{D}}$ and all the $C_s \in \mathcal{C}$. In more detail, for each $s \in [c]$, there exists some $h_s \in \mathbb{K}_s$ such that $z_s \in U_{h_s} \subset X_{h_s}$. We define

$$R_{j^*} := \tilde{R} \cup_{s \in [c]} (R_{C_s} \cup \{z_s\}) \qquad \text{and} \qquad \Sigma_{j^*} := \tilde{\Sigma} \cup_{s \in [c]} (\Sigma_{C_s} \cup \{S'_s\} \cup \{S_{h_s}\}),$$

where Σ_{C_s} is the set of interior $(r-1)$-cliques of C_s, we have that $S'_s \in K_{r-1}(G[Z_0])$ is the $(r-1)$-clique which forms a clique with both z_s and x_s defined above and S_{h_s} is the $(r-1)$-clique corresponding to the set X_{h_s} (which contains z_s) in Claim 3.7.5. We claim that there exists a diamond tree $\mathcal{D}_{j^*}$ of order M which has R_{j^*} as a set of removable vertices and Σ_{j^*} as a set of interior $(r-1)$-cliques. Indeed we can form the defining auxiliary tree T_{j^*} by starting with the forest of the disjoint union of $\tilde{T}$ and the T_{C_s} for $s \in [c]$, where T_{C_s} denotes the defining tree for the K_r-diamond tree C_s. For each $s \in [c]$, we then add a path of length two (with two edges) between some vertex in $V(T_{C_s})$ and $V(\tilde{T})$. The edges of this path correspond exactly to the internal $(r-1)$-cliques S_{h_s} and S'_s and thus the vertices of this path correspond to x_s, z_s and some vertex in $R_{C_s} \cap U_{h_s}$ for each $s \in [c]$.

This thus defines $\mathcal{D}_{j^*}$ and so we update all the U_h to be $U_h \setminus V(\mathcal{D}_{j^*})$ for $h \in [\ell]$ and kill any indices $h \in \mathbb{L}$ such that either $V(\mathcal{D}_{j^*})$ intersects S_h or $|X_h \cap V(\mathcal{D}_{j^*})| \geq 2Cn^{1/6}$. We now need to check that the conditions (1), (2) and (3) hold for $\mathcal{D}_{j^*}$. To see (1), note that R_{j^*} contains all the R_{C_s} for $s \in [c]$ and so intersects X_h for all $h \in \cup_{s \in [c]} \mathbb{K}_s$. Moreover taking $\mathbb{K}$ as defined at the end of finding the C_s, we have that $|\mathbb{K} \cup (\cup_{s\in[c]} \mathbb{K}_s)| \geq \left(1 - \frac{\alpha}{2}\right)\ell$ and $|\mathbb{K}| \leq \frac{\alpha \ell}{2}$ and so this confirms (1). To see (2), note that the only times we used vertices of the S_h with $h \in [\ell]$ to construct $\mathcal{D}_{j^*}$ was when we added the S_{h_s} for $s \in [c]$ to the set of interior cliques. Thus we intersected exactly $c \leq C$ of these with $V(\mathcal{D}_{j^*})$. Finally we have that (3) for $\mathcal{D}_{j^*}$ is implied by the conditions (iv) when we found the C_s. Indeed, we have that $R_{j^*} \cap (\cup_{h \in [\ell]} X_h) = \cup_{s \in [c]} (R_{C_s} \cup \{z_s\})$ and so for any index h that does not lie in $\cup_{s\in[c]} \mathbb{J}_s$ (which has size at most $Cn^{1/2}$), we have that

$$|V(\mathcal{D}_{j^*}) \cap X_h| \leq \sum_{s \in [c]} |(V(C_s) \cup \{z_s\}) \cap X_h| + \leq C(n^{1/6} + 1) \leq 2Cn^{1/6}.$$

This concludes the process for finding $\mathcal{D}_{j^*}$ and doing this for all $j^* \in [4t]$ gives the desired claim and hence the proposition. ■ □

Chapter 4

Robustness for triangle factors

In this chapter, we prove Theorem II, which we restate below for convenience, showing that n-vertex graphs G with $n \in 3\mathbb{N}$ and minimum degree at least $\frac{2n}{3}$ are robust with respect to containing triangle factors, in that a random sparsification of such a graph G contains a triangle factor whp.

Theorem II. *(Restated) There is a constant $C > 0$ such that for all $n \in 3\mathbb{N}$ and $p \geq C(\log n)^{1/3}n^{-2/3}$ the following holds. If G is an n-vertex graph with $\delta(G) \geq \frac{2n}{3}$ then whp G_p has a triangle factor.*

The chapter is organised as follows. In Section 4.1 we explain that the main instrument for proving Theorem II is a result on triangle factors in random sparsifications of super-regular tripartite graphs, Theorem II*. We then give an overview of the proof of this main technical theorem, state the main propositions and lemmas needed for this and show how these imply Theorem II*. More precisely, we shall formulate one proposition, Proposition 4.1.1, allowing us to count certain partial triangle factors, one proposition, Proposition 4.1.2, allowing us to extend a partial triangle factor by one triangle, and a key lemma, which we call the Local Distribution Lemma (Lemma 4.1.3).

After this, we provide some results on triangle counts in Section 4.2, which will be useful in the proofs of our propositions. In Section 4.3, we prove Proposition 4.1.1 and Proposition 4.1.2, using Lemma 4.1.3 as a black box. In Section 4.4, we then show Lemma 4.1.3. An important ingredient of this proof is a lemma which we call the Entropy Lemma (Lemma 4.4.4).

This will complete the proof of the main technical theorem, Theorem II*, and it will remain to deduce Theorem II from Theorem II*. Before embarking on this, we need to build some more theory. We begin in Section 4.5 by providing a stability statement of a fractional version of the Hajnal–Szemerédi theorem. Next, in Section 4.6, we derive a sequence of probabilistic lemmas which imply the existence of K_3-matchings in various random sparsification settings. In Section 4.7, finally, we show how Theorem II* implies Theorem II. The basic approach we use is a combination of the regularity method with an analysis of the extremal cases, as is common in the area.

4.1 The main technical result and its proof overview

The main technical result we reduce Theorem II to is the following partite version with the minimum degree condition replaced by regularity.

Theorem II* (main technical theorem)**.** *For every $0 < d \leq 1$ there exists constants $\varepsilon > 0$ and $C > 0$ such that the following holds for every $n \in \mathbb{N}$ and $p \in (0,1)$ such that $p \geq C(\log n)^{1/3} n^{-2/3}$. If Γ is an (ε, d^+)-super-regular tripartite graph with parts of size n then Γ_p whp contains a triangle factor.*

The reduction of Theorem II to this partite version uses the regularity method together with a stability result for the fractional Hajnal–Szemerédi Theorem developed in Section 4.5 and an analysis of the extremal cases. We give the full details in Section 4.7.

The main challenge of this chapter is proving Theorem II*, and in this section we will reduce Theorem II* further to two intermediate propositions. We will then discuss the proof of these propositions, outlining the remainder of the chapter and some of the key ideas involved. We encourage the reader to recall the relevant terminology from the Notation Section on embedding partial factors in tripartite graphs, in particular the definition of Ψ^t.

The first proposition counts partial triangle factors.

Proposition 4.1.1 (counting partial-factors)**.** *For all $0 < \eta, d \leq 1$ there exists $\varepsilon > 0$ and $C > 0$ such that the following holds for all sufficiently large $n \in \mathbb{N}$ and for any $p \geq C(\log n)^{1/3} n^{-2/3}$. If Γ is an (ε, d)-regular tripartite graph with parts of size n, then whp we have that*

$$|\Psi^t(\Gamma_p)| \geq (1-\eta)^t (pd)^{3t} (n!_t)^3 \,, \tag{4.1.1}$$

for all $t \in \mathbb{N}$ with $t \leq (1-\eta)n$.

Here the condition (4.1.1) should be read as Γ_p having roughly the 'correct' number of embeddings of D_t, the graph with t labelled disjoint triangles. Indeed, the term $(pd)^{3t} (n!_t)^3$ is the expected number of embeddings of D_t in a random sparsification of the complete tripartite graph with probability pd, which provides a sensible benchmark for our model Γ_p. The $(1-\eta)^t$ factor is then an error term which we can control. In order to go beyond Proposition 4.1.1 to counting subgraphs D_t with larger t, we need different techniques. Our second proposition allows us to extend partial triangle factors by embedding further triangles one by one.

Proposition 4.1.2 (extending by one triangle)**.** *For all $0 < d \leq 1$ there exists $\alpha, \eta, \varepsilon > 0$ and $C > 0$ such that for all sufficiently large $n \in \mathbb{N}$ and for any $p \geq C(\log n)^{1/3} n^{-2/3}$, if Γ is an (ε, d)-super-regular tripartite graph with parts of size n, then whp the following holds in Γ_p for every $t \in \mathbb{N}$ with $(1-\eta)\, n \leq t < n$. If*

$$|\Psi^t(\Gamma_p)| \geq (1-\eta)^n \,(pd)^{3t}\, (n!_t)^3, \tag{4.1.2}$$

then

$$|\Psi^{t+1}(\Gamma_p)| \geq \alpha (pd)^3 (n-t)^3 |\Psi^t(\Gamma_p)|. \tag{4.1.3}$$

Again the condition (4.1.2) in Proposition 4.1.2 should be read as Γ_p having roughly the 'correct' number of embeddings of D_t and condition (4.1.3) then implies that Γ_p has roughly the 'correct' number of embeddings of D_{t+1}. In contrast to Proposition 4.1.1 we now lose control of the error term (given by α) but as we will only apply Proposition 4.1.2 for large t, we can make sure the error term does not accumulate too much. Indeed, recall that our goal is merely to obtain one triangle factor in the end.

We now show how Theorem II* follows from these two intermediate propositions before outlining the proofs of these propositions.

Proof of Theorem II.* Given d choose $0 < \varepsilon, \frac{1}{C} \ll \eta \ll \alpha \ll d$ and note that by choosing $C > 0$ sufficiently large, we can assume that n is sufficiently large in what follows, as otherwise the statement is trivially true. Let us fix Γ to be an (ε, d^+)-super-regular tripartite graph with parts of size n. We can assume that $dn^2 \in \mathbb{N}$. Indeed, if this is not the case, then replace d with the minimum $d' > d$ such that $d'n^2 \in \mathbb{N}$ and note that, after redefining d (if necessary), we maintain that Γ is (ε, d^+)-super-regular. Now let Γ' be the $(4\sqrt{\varepsilon}, d)$-super-regular tripartite graph obtained by applying Lemma 2.2.8 between each of the parts of Γ. As Γ' is a spanning subgraph of Γ it suffices to find our triangle factor in Γ'.

Note that by our choice of constants, we have that whp both the conclusion of Proposition 4.1.1 (with η replaced by η^2) and the conclusion of Proposition 4.1.2 hold in Γ' simultaneously. We will now assume they hold and show that this implies

$$|\Psi^t(\Gamma'_p)| \geq \left(1-\eta^2\right)^n \alpha^{t-(1-\eta^2)n} (pd)^{3t}\, (n!_t)^3\,, \tag{4.1.4}$$

for all $\left(1-\eta^2\right) n \leq t \leq n$. Indeed, for $t = \left(1-\eta^2\right) n$, (4.1.4) readily follows from (the assumed conclusion of) Proposition 4.1.1. Assume now (4.1.4) holds for some $\left(1-\eta^2\right) n \leq t < n$.

Since $\eta \ll \alpha$, we have that

$$\left(1-\eta^2\right)^n \alpha^{t-(1-\eta^2)n} \geq \left(1-\eta^2\right)^n \alpha^{\eta^2 n} = \left(1-\eta^2\right)^n e^{-\log(1/\alpha)\eta^2 n}$$
$$\geq \left(1-\eta^2\right)^n \left(1-\log\left(\tfrac{1}{\alpha}\right)\eta^2\right)^n \geq (1-\eta)^n .$$

It follows from (the assumed conclusion of) Proposition 4.1.2 that (4.1.4) holds for $t+1$. In particular, we have

$$|\Psi^n(\Gamma_p)| \geq |\Psi^n(\Gamma'_p)| \geq \left(1-\eta^2\right)^n \alpha^{\eta^2 n}(pd)^{3n}(n!)^3 \geq 1,$$

completing the proof. □

Thus it remains to prove Propositions 4.1.1 and 4.1.2. Proving Proposition 4.1.1 is relatively straightforward: It follows from embedding the triangles of D_t one by one greedily and counting in how many ways we can embed each such triangle by using that all large enough vertex sets whp induce roughly the 'correct' number of triangles in Γ_p, which we establish in Lemma 4.2.1 using regularity and Janson's inequality (Lemma 2.1.3). The details for deriving Proposition 4.1.1 are provided in Section 4.3.1.

The proof of Proposition 4.1.2 is much more involved and the main challenge of this chapter. In order to count embeddings of partial triangle factors in $\Psi^{t+1}(\Gamma_p)$, one naïve idea would be to proceed as follows: We fix any triple $\underline{u} = (u_1, u_2, u_3) \in \mathcal{V}$ of vertices and count in how many partial triangle factors from $\Psi^t(\Gamma_p)$ these are isolated. If this number would be roughly the same for each triple of vertices then we would be able to bound the size of $\Psi^{t+1}(\Gamma_p)$ using bounds on how many triples actually form triangles in Γ_p to extend a partial triangle factor from $\Psi^t(\Gamma_p)$ by one triangle. However, we do not know how to prove that all triples of vertices behave similarly in this sense. Hence, we need to resort to a more refined strategy, still considering embeddings which leave certain vertices isolated, but doing so in stages, growing our set of isolated vertices one vertex at a time. This step by step process is made precise in the following Local Distribution Lemma, which is a key step of our argument. We will show that this lemma implies Proposition 4.1.2 in Section 4.3.2.

Lemma 4.1.3 (Local Distribution Lemma)**.** *For all* $0 < \alpha, d \leq 1$ *and* $K > 0$ *there exists* $\eta, \varepsilon > 0$ *and* $C > 0$ *such that for all sufficiently large* $n \in \mathbb{N}$ *and for any* $p \geq C(\log n)^{1/3} n^{-2/3}$*, if* Γ *is an* (ε, d)*-super-regular tripartite graph with parts of size* n*,* $t \in \mathbb{N}$ *such that* $(1-\eta)n \leq t < n$*,* $\ell \in [3]$ *and* $\underline{u} = (u_1, \ldots, u_{\ell-1}) \in \mathcal{V}$ *then the following holds in* Γ_p *with probability at least* $1 - n^{-K}$*. If*

$$|\Psi^t_{\underline{\hat{u}}}(\Gamma_p)| \geq (1-\eta)^n (pd)^{3t} ((n-1)!_t)^{\ell-1} (n!_t)^{4-\ell}, \tag{4.1.5}$$

then for all but at most αn vertices $u_\ell \in V^\ell$ we have, with $\underline{v} = (\underline{u}, u_\ell) \in \mathcal{V}$, that

$$|\Psi^t_{\hat{\underline{v}}}(\Gamma_p)| \geq \left(\frac{d}{10}\right)^2 \left(\frac{n-t}{n}\right) |\Psi^t_{\hat{\underline{u}}}(\Gamma_p)|\,. \tag{4.1.6}$$

Again, (4.1.5) should be read as Γ_p having roughly the 'correct' number (up to the error term $(1-\eta)^n$) of embeddings of D_t that avoid using vertices in $\underline{u}$, where correct means what we expect in a random sparsification of the complete tripartite graph with probability pd. The conclusion of Lemma 4.1.3 then tells us that that for most choices of extending $\underline{u}$ to $\underline{v}$, we have roughly the correct number of embeddings of D_t that avoid using the vertices in $\underline{v}$.

For proving Proposition 4.1.2 in Section 4.3.2, we shall use Lemma 4.1.3 with $\ell = 2$ and $\ell = 3$ to prove a lemma, Lemma 4.3.1, which states that if for a vertex $w \in V^1$ we have

$$|\Psi^t_{\hat{w}}(\Gamma_p)| \geq (1-\eta)^n\,(pd)^{3t}(n-1)!_t(n!_t)^2, \tag{4.1.7}$$

then

$$|\Psi^{t+1}_w(\Gamma_p)| \geq \alpha(pd)^3(n-t)^2|\Psi^t_{\hat{w}}(\Gamma_p)|\,, \tag{4.1.8}$$

where we recall that $\Psi^t_w(G)$ is the set of embeddings $\psi \in \Psi^t(G)$ for which $\psi((1,1)) = w$, that is, the first triangle is embedded so that its first vertex is w. Indeed, using Lemma 4.1.3, we can see that if there are many embeddings of D_t avoiding w (4.1.7), then for almost all choices of further vertices $w_2 \in V^2$ and $w_3 \in V^3$, there will be many embeddings of D_t avoiding all 3 vertices w, w_2, w_3. Intuitively, (4.1.8) then follows due to the fact that we can expect many of these triangles w, w_2, w_3 to feature in Γ_p and each triangle that does, gives an embedding of D_{t+1} which maps w to a triangle. We have to be very careful with the dependence of these different random variables here and the essence of the proof of Lemma 4.3.1 (which is done in Section 4.3.2) is to work with random variables that are independent from each other. Now together with Lemma 4.1.3 for $\ell = 1$ and our assumption (4.1.2), using the conclusion of Lemma 4.3.1 (namely (4.1.8)), Proposition 4.1.2 follows readily as almost all choices of $w \in V^1$ satisfy (4.1.7).

We will now sketch some of the ideas involved in proving Lemma 4.1.3. To ease the discussion, let us fix $\ell = 1$ and hence $\underline{u} = \emptyset$; the other cases are similar. In this case our assumption (4.1.5) simply states that Γ_p has roughly the correct number of embeddings of D_t and a simple averaging argument will find some $u = u_\ell$ for which (4.1.6) holds with $\underline{v} = u$. Fix some such vertex u. The challenge now is to show that (4.1.6) holds for *almost all* choices of u_ℓ.

In order to do this, we fix some typical vertex $v \in V^1 \setminus \{u\}$. We will aim to lower bound the size of $\Psi^t_{\hat{v}}(\Gamma_p)$ by comparing it to the size of $\Psi^t_{\hat{u}}(\Gamma_p)$. Let us suppose, momentarily, that $\mathrm{Tr}_u(\Gamma_p) = \mathrm{Tr}_v(\Gamma_p)$. In such a case, we can easily compare the sizes of $\Psi^t_{\hat{v}}(\Gamma_p)$ and $\Psi^t_{\hat{u}}(\Gamma_p)$. Indeed, for every embedding $\psi \in \Psi^t_{\hat{u}}(\Gamma_p)$ there are two cases. Firstly, if v is not in a triangle in $\psi(D_t)$ then $\psi \in \Psi^t_{\hat{v}}(\Gamma_p)$ already. Secondly, if v is in a triangle $\{v, w_2, w_3\}$ of $\psi(D_t)$, then $\mathrm{Tr}_u(\Gamma_p) = \mathrm{Tr}_v(\Gamma_p)$ implies that $\{u, w_2, w_3\}$ is also a triangle, hence we can switch the triangle $\{v, w_2, w_3\}$ with $\{u, w_2, w_3\}$ in ψ to get an embedding $\psi' \in \Psi^t_{\hat{v}}(\Gamma_p)$. This gives an injection from $\Psi^t_{\hat{u}}(\Gamma_p)$ to $\Psi^t_{\hat{v}}(\Gamma_p)$, proving that $\Psi^t_{\hat{v}}(\Gamma_p)$ is also of roughly the 'correct' size.

Of course, the situation that $\mathrm{Tr}_u(\Gamma_p) = \mathrm{Tr}_v(\Gamma_p)$ is wildly unrealistic. Let us loosen this and suppose instead that

$$|\mathrm{Tr}_u(\Gamma_p) \cap \mathrm{Tr}_v(\Gamma_p)| = \Omega(p^3 n^2)\,. \tag{4.1.9}$$

As we expect every vertex to be in $\Theta(p^3n^2)$ triangles, (4.1.9) can be interpreted as saying that a constant fraction of the set of edges that form a triangle with v, also form a triangle with u. We can only expect this to happen when p is constant and this is also a gross oversimplification of our setting but serves to demonstrate a key idea of the proof. So for now, we take (4.1.9) to be the case and note that as above, we can perform a switching, replacing triangles containing v with triangles containing u to map embeddings in $\Psi^t_{\hat{u}}(\Gamma_p)$ to embeddings in $\Psi^t_{\hat{v}}(\Gamma_p)$, whenever the embedding $\psi \in \Psi^t_{\hat{u}}(\Gamma_p)$ has v in a triangle $\{v, w_2, w_3\}$ such that $\{w_2, w_3\} \in \mathrm{Tr}_u(\Gamma_p)$. We have, by (4.1.9), that a constant proportion of the triangles containing v can be switched in this way but we do *not* know that this translates to having a constant proportion of the *embeddings* in $\Psi^t_{\hat{u}}(\Gamma_p)$ being switchable. It could well be that almost all (or even all) of the embeddings in $\Psi^t_{\hat{u}}(\Gamma_p)$ map v to a triangle $\{v, w_2, w_3\}$ such that $\{u, w_2, w_3\} \notin K_3(\Gamma_p)$. What we need then, is to be able to discount such a situation and show that each triangle containing v contributes to roughly the same number of embeddings $\psi \in \Psi^t_{\hat{u}}(\Gamma_p)$. Put differently, when we consider a *uniformly random embedding* $\psi^* \in \Psi^t_{\hat{u}}(\Gamma_p)$, we want that the random variable T_v, which encodes the triangle containing v in $\psi^*(D_t)$, induces a roughly uniform distribution on the set $\mathrm{Tr}_v(\Gamma_p)$. Note that it is possible that ψ^* leaves v isolated but this is unlikely (as t is large) and so we ignore this possibility for this discussion.

We can now see how entropy enters the picture as it provides a tool for studying distributions, and how far they are from being uniform. Let us now consider v as not fixed anymore. Our argument will take a uniformly random $\psi^* \in \Psi^t_{\hat{u}}(\Gamma_p)$ and consider the random variables T_v which describe the triangle containing each vertex $v \in V^1$. Due to the fact that $\Psi^t_{\hat{u}}(\Gamma_p)$ is roughly the 'correct' size, we have that ψ^* has large entropy. Moreover, ψ^* is completely described (up to labelling) by the set $\{T_v : v \in V^1\}$ and so the entropy of ψ^* can be decomposed as a sum of individual entropy values $h(T_v)$ of the T_v, using the chain rule (Lemma 2.3.5) for example. We will be able to use random properties of Γ_p (for example that no vertex is in too many triangles) to conclude that no single T_v has too large entropy. This will thus imply that for almost all vertices $v \in V^1$, the entropy of T_v is large. Therefore, by applying Lemma 2.3.9, we will be able to conclude that for a typical vertex $v \in V^1$, the random variable T_v induces a roughly uniform distribution on $\mathrm{Tr}_v(\Gamma_p)$, as desired. This idea is formalised in what we call the Entropy Lemma (Lemma 4.4.4).

Our discussion above is premised on (4.1.9). In reality, a typical vertex v will have $\mathrm{Tr}_v(\Gamma_p)$ completely disjoint from $\mathrm{Tr}_u(\Gamma_p)$ and so the switching argument outlined above cannot possibly work. However, we can still compare the sizes of $\Psi^t_{\hat{u}}(\Gamma_p)$ and $\Psi^t_{\hat{v}}(\Gamma_p)$ by noting that a constant proportion of $\mathrm{Tr}_u(\Gamma_p)$ and $\mathrm{Tr}_v(\Gamma_p)$ are *drawn from the same distribution*. By this we mean the following. For a typical v, by using regularity properties, there will be $\Omega(n^2)$ edges $F \subset E(\Gamma)$ in the joint neighbourhood (with respect to Γ) of u and v. Consider revealing all edges in Γ_p apart from those incident to u or v. After this, $F_p := F \cap E(\Gamma_p)$ is revealed and whp has size $|F_p| = \Omega(pn^2)$; each edge $e \in F_p$ has the potential to land in both $\mathrm{Tr}_v(\Gamma_p)$ and $\mathrm{Tr}_u(\Gamma_p)$, depending on which random edges incident to u and v appear.

Moreover, without having revealed the random edges incident to u or v yet, we can associate a *weight* to the edges e in F_p, which encodes the number of embeddings of D_{t-1} in Γ_p, which avoid u, v and the vertices of e. Now, revealing the edges incident to v, we have that for every $e \in \mathrm{Tr}_v(\Gamma_p) \cap F_p$, the probability that a uniformly random embedding $\psi^* \in \Psi^t_{\hat{u}}(\Gamma_p)$ uses the triangle $\{v\} \cup e$, is directly proportional to the weight of e in F_p. The Entropy Lemma (Lemma 4.4.4) discussed above tells us that the random variable $T_v \in \mathrm{Tr}_v(\Gamma_p)$, encoding the triangle containing v in a uniformly random $\psi^* \in \Psi^t_{\hat{u}}(\Gamma_p)$, has a roughly uniform distribution in $\mathrm{Tr}_v(\Gamma_p)$. From this, we can deduce that the weights of edges in F_p are 'well-behaved' in that many of the edges in F_p have a sufficiently large weight. This in turn gives that $\Psi^t_{\hat{v}}(\Gamma_p)$ will be large, as when we reveal the edges incident to u, we can expect that $\mathrm{Tr}_u(\Gamma_p)$ contains many (i.e. $\Omega(p^3n^2)$) edges of large weight from F_p. Each such edge e contributes many embeddings in $\Psi^t_{\hat{v}}(\Gamma_p)$ which map u to a triangle with e.

In order for all of this to work, we need our Entropy Lemma (Lemma 4.4.4) to be very strong, due to the fact that the edges in the F_p defined above contribute only a small fraction of edges in $\mathrm{Tr}_v(\Gamma_p)$. Pushing the strength of the Entropy Lemma is one of the main novelties of the current work, in comparison to previous arguments for triangle factors in random graphs [6, 96], and requires a delicate analysis.

Remark: Many of the results in this chapter state that Γ_p whp satisfies some statement of the form "if $\mathcal{A}$, then $\mathcal{B}$", where $\mathcal{A}$ and $\mathcal{B}$ are certain graph properties. Indeed, we see examples of this Propositions 4.1.1 and 4.1.2 and Lemma 4.1.3. In symbols, these results posit that $\mathbb{P}[\mathcal{A} \implies \mathcal{B}] = 1 - o(1)$. We remark that this is *not* equivalent to showing that $\mathbb{P}[\mathcal{B}|\mathcal{A}] = 1 - o(1)$. Indeed,

$$\mathbb{P}[\mathcal{A} \cap \neg\mathcal{B}] = \mathbb{P}[\neg B|\mathcal{A}]\,\mathbb{P}[\mathcal{A}]\,,$$

and so if $\mathbb{P}[\mathcal{B}|\mathcal{A}] = 1 - o(1)$, then we have that $\mathbb{P}[\mathcal{A} \implies \mathcal{B}] = 1 - o(1)$. That is, the statement $\mathbb{P}[\mathcal{A} \implies \mathcal{B}] = 1 - o(1)$ is *weaker* than the statement $\mathbb{P}[\mathcal{B}|\mathcal{A}] = 1 - o(1)$ and thus (potentially) easier to prove. As is usual in random graph theory, we will often separate the randomness in our proofs, showing first that in *any* subgraph $G \subseteq \Gamma$ that satisfies some set of properties $\boldsymbol{\mathcal{P}} = \{\mathcal{P}\}$, we have that if $\mathcal{A}$ occurs then $\mathcal{B}$ must occur also. We then show that $G = \Gamma_p$ whp satisfies all the properties in $\boldsymbol{\mathcal{P}}$. At other times, we directly upper bound the probability $\mathbb{P}[\mathcal{A} \cap \neg\mathcal{B}]$.

4.2 Counting triangles in Γ_p

The purpose of this section is to prove that certain properties of Γ_p hold with high probability when Γ is a (super-)regular tripartite graph and p is sufficiently large. These properties regard triangle counts in Γ_p and their proofs use the properties of regular tuples given in Section 2.2 and the probabilistic tools outlined in Section 2.1. Our first lemma gives an estimate on the number of triangles induced on vertex subsets.

Lemma 4.2.1. *For all $0 < \varepsilon' < d \leq 1$ and $L > 0$ there exists $\varepsilon > 0$ and $C > 0$ such that the following holds for all sufficiently large $n \in \mathbb{N}$ and for any $p \geq C(\log n)^{1/3}n^{-2/3}$. If Γ is an (ε, d)-regular tripartite graph with parts V^1, V^2, V^3 of size n, then with probability at least $1 - n^{-L}$ we have that*

$$|K_3(\Gamma_p[X_1 \cup X_2 \cup X_3])| = (pd)^3|X_1||X_2||X_3| \pm \varepsilon' p^3 n^3, \tag{4.2.1}$$

for all $X_1 \subseteq V^1$, $X_2 \subseteq V^2$ and $X_3 \subseteq V^3$.

Proof. Choose $0 < \varepsilon, \frac{1}{C} \ll \varepsilon', d, \frac{1}{L}$ and fix Γ and $p \geq C(\log n)^{1/3}n^{-2/3}$. We first show (a stronger version of) the lower bound holds using Janson's inequality.

Claim 4.2.2. *With probability at least $1 - e^{-n}$, we have*

$$|K_3(\Gamma_p[X_1 \cup X_2 \cup X_3])| \geq (pd)^3|X_1||X_2||X_3| - \frac{\varepsilon' p^3 n^3}{8}, \tag{4.2.2}$$

for all $X_1 \subseteq V^1$, $X_2 \subseteq V^2$ and $X_3 \subseteq V^3$.

<u>Proof of Claim:</u> Fix $X_1 \subseteq V^1$, $X_2 \subseteq V^2$ and $X_3 \subseteq V^3$ and let $Y := K_3(\Gamma[X_1 \cup X_2 \cup X_3])$. We may assume that

$$|X_1||X_2||X_3| \geq \frac{\varepsilon' n^3}{8d^3} \geq \varepsilon n^3, \tag{4.2.3}$$

with the first inequality holding as otherwise (4.2.2) is trivially true and the second inequality holding by our choice of ε. In particular, we have $|X_i| \geq \varepsilon n$ for all $i \in [3]$ and thus Lemma 2.2.7 implies $|Y| \geq d^3|X_1||X_2||X_3| - 10\varepsilon n^3$. Consider now the random variable

$$X := |K_3(\Gamma_p[X_1 \cup X_2 \cup X_3])| = \sum_{T \in Y} I_T,$$

where for each triangle $T \in Y$, I_T is the indicator random variable for the event that T is present in Γ_p. Let

$$\lambda := \mathbb{E}[X] = p^3|Y| \geq (pd)^3|X_1||X_2||X_3| - 10\varepsilon p^3 n^3, \tag{4.2.4}$$

which in combination with (4.2.3) implies $\lambda \geq \varepsilon p^3 n^3$. Furthermore, we have

$$\bar{\Delta} := \sum_{T,T' \in Y:\, T \cap T' \neq \emptyset} \mathbb{E}[I_T I_{T'}] \leq p^5 \cdot |Y| \cdot 3n + p^3 \cdot |Y| = \lambda(3np^2 + 1), \tag{4.2.5}$$

where the inequality follows from the fact that there are at most $|Y| \cdot 3n$ pairs of triangles intersecting in exactly one edge, no pairs intersecting in exactly two edges and $|Y|$ pairs intersecting in three edges. Hence Janson's inequality (Lemma 2.1.3) implies

$$\begin{aligned}
\mathbb{P}[X \leq (1-\varepsilon)\lambda] \leq \exp\left(-\frac{\varepsilon^2\lambda^2}{2\bar{\Delta}}\right) &\leq \exp\left(-\frac{\varepsilon^3 p^3 n^3 \lambda}{2\bar{\Delta}}\right) \\
&\leq \exp\left(-\frac{\varepsilon^3 p^3 n^3}{12np^2}\right) + \exp\left(-\frac{\varepsilon^3 p^3 n^3}{4}\right) \\
&\leq \exp(-4n)
\end{aligned}$$

for all large enough n. Here, we used that $\lambda \geq \varepsilon p^3 n^3$ (see (4.2.4)) in the second inequality, and (4.2.5) in the third (more precisely, we used that (4.2.5) implies that $\bar{\Delta} \leq 6\lambda np^2$ or $\bar{\Delta} \leq 2\lambda$).

By (4.2.4), we have $(1-\varepsilon)\lambda \geq (pd)^3|X_1||X_2||X_3| - 11\varepsilon p^3 n^3 \geq (pd)^3|X_1||X_2||X_3| - (\frac{\varepsilon'}{8})p^3 n^3$. Hence, taking a union bound over all choices of $X_1 \subseteq V^1, X_2 \subseteq V^2, X_3 \subseteq V^3$, we deduce that, (4.2.2) holds with probability at least $1 - 2^{3n} \cdot e^{-4n} \geq 1 - e^{-n}$ for all $X_1 \subseteq V^1, X_2 \subseteq V^2, X_3 \subseteq V^3$. ■

We now show that the upper bound holds in the case when $X_i = V^i$ for all $i \in [3]$.

Claim 4.2.3. *With probability at least* $1 - n^{-2L}$ *we have*

$$|K_3(\Gamma_p)| \leq (pd)^3 n^3 + \frac{\varepsilon' p^3 n^3}{8}.$$

Proof of Claim: Let $Y = K_3(\Gamma)$ and let $X = |K_3(\Gamma_p)| = \sum_{T \in Y} I_T$ with I_T being the indicator random variable for the event that a triangle T appears in Γ_p, as above. By Lemma 2.2.7, we have $|Y| = d^3 n^3 \pm 10\varepsilon n^3$. It follows that

$$\lambda := \mathbb{E}[X] = (pd)^3 n^3 \pm 10\varepsilon p^3 n^3. \tag{4.2.6}$$

Using notations from the Kim–Vu inequality (Lemma 2.1.4), we have $E_1 \leq np^2$, $E_2 = p$ and $E_3 = 1$. Hence $E' = \max\{1, np^2\} \leq \lambda^{1/2}$ and $E = \lambda$. Let $\mu = \lambda^{1/16}$ and let $c = c(3)$ be the constant from Lemma 2.1.4. Then, for large enough n,

$$c(EE')^{1/2}\mu^3 \leq c\lambda^{3/4} \cdot \lambda^{3/16} \leq \varepsilon\lambda.$$

Hence, we have

$$\mathbb{P}[X \geq (1+\varepsilon)\lambda] \leq 10cn^4 e^{-\mu} \leq e^{-n^{1/16}} \leq n^{-2L}$$

for all large enough n. Here, the middle inequality follows from (4.2.6) which implies $\lambda \geq n \log n$, due to our choice of ε and C. This finishes the proof of the claim as $(1+\varepsilon)\lambda \leq (pd)^3 n^3 + \left(\frac{\varepsilon'}{8}\right) p^3 n^3$ by (4.2.6) and our choice of ε. ∎

We now conclude the proof of the lemma. With probability at least $1 - n^{-L}$ both claims above hold simultaneously. Suppose now both claims hold and fix $X_1 \subseteq V^1, X_2 \subseteq V^2, X_3 \subseteq V^3$. Let $\mathcal{U} = \left(\{X_1, V^1 \setminus X_1\} \times \{X_2, V^2 \setminus X_2\} \times \{X_3, V^3 \setminus X_3\}\right) \setminus \{(X_1, X_2, X_3)\}$ and observe that

$$\begin{aligned} |K_3(\Gamma_p[X_1 \cup X_2 \cup X_3])| &= |K_3(\Gamma_p)| - \sum_{(U_1,U_2,U_3)\in\mathcal{U}} |K_3(\Gamma_p[U_1 \cup U_2 \cup U_3])| \\ &\leq (pd)^3 |X_1||X_2||X_3| + \varepsilon' p^3 n^3. \end{aligned}$$

Here we used Claim 4.2.3 to bound $|K_3(\Gamma_p)|$ and (4.2.2) to bound each $|K_3(\Gamma_p[U_1 \cup U_2 \cup U_3])|$. This completes the proof. □

As a corollary, we can conclude that we have the expected count of triangles at almost all vertices.

Corollary 4.2.4. *For all $0 < \varepsilon' < d \leq 1$ and $L > 0$ there exists $\varepsilon > 0$ and $C > 0$ such that the following holds for all sufficiently large $n \in \mathbb{N}$ and for any $p \geq C(\log n)^{1/3} n^{-2/3}$. If Γ is an (ε, d)-regular tripartite graph with parts of size n, then with probability at least $1 - n^{-L}$ we have that*

$$|\mathrm{Tr}_v(\Gamma_p)| = (1 \pm \varepsilon')(pd)^3 n^2,$$

for all but at most $\varepsilon' n$ vertices $v \in V(\Gamma)$.

Proof. Choose $0 < \varepsilon, \frac{1}{C} \ll \tilde{\varepsilon} \ll \varepsilon', d, \frac{1}{L}$ and let $G \subseteq \Gamma$ be any graph with

$$|K_3(G[X_1 \cup X_2 \cup X_3])| = (pd)^3 |X_1||X_2||X_3| \pm \tilde{\varepsilon} p^3 n^3, \tag{4.2.7}$$

for all $X_1 \subseteq V^1$, $X_2 \subseteq V^2$ and $X_3 \subseteq V^3$. Since (by Lemma 4.2.1 and our choice of constants) this is satisfied by Γ_p with probability $1 - n^{-L}$, it suffices to show that G satisfies the conclusion of Corollary 4.2.4. For $i \in [3]$, let X_i be the set of vertices $v \in V^i$ with $|\mathrm{Tr}_v(G)| \leq (1-\varepsilon')(pd)^3 n^2$, and let Y_i be the set of vertices $v \in V^i$ with $|\mathrm{Tr}_v(G)| \geq (1+\varepsilon')(pd)^3 n^2$. We claim that $|X_1| \leq \frac{\varepsilon' n}{10}$.

Indeed, assuming the contrary, we have

$$|K_3(G[X_1 \cup V^2 \cup V^3])| \leq (pd)^3|X_1||V^2||V^3| - \frac{\varepsilon'^2(pd)^3n^3}{10} < (pd)^3|X_1||V^2||V^3| - \tilde{\varepsilon}p^3n^3,$$

by our choice of $\tilde{\varepsilon}$. This contradicts (4.2.7). Similarly, we can bound the sizes of X_2 and X_3, and Y_1, Y_2 and Y_3, completing the proof. □

Sometimes, we will need an upper bound on $|\mathrm{Tr}_v(\Gamma_p)|$ which works for all $v \in V(\Gamma)$. For this we simply upper bound this quantity by the number of triangles in $G(3n, p)$ containing a specific vertex using a result of Spencer [162] (see also [159]).

Lemma 4.2.5. *For all* $L > 0$ *there exists* $C > 0$ *such that the following holds for all sufficiently large* $n \in \mathbb{N}$ *and for any* $p \geq C(\log n)^{1/3}n^{-2/3}$*. If* Γ *is a tripartite graph with parts of size* n*, then with probability at least* $1 - n^{-L}$ *we have that*

$$|\mathrm{Tr}_v(\Gamma_p)| \leq 10p^3n^2,$$

for all vertices $v \in V(\Gamma)$.

In the remainder of this section we prove some more technical properties of Γ_p which will be useful in the proofs of Proposition 4.1.2 and Lemma 4.1.3. The ultimate goal will be to lower bound the number of triangles at a fixed vertex but we will need this lower bound to hold in a robust way, allowing us to apply the count with respect to various prescribed sets of edges and vertices which we either want to avoid or want to be included in the triangles.

Our next lemma follows simply from well-known concentration bounds but we wish to highlight the slightly subtle (in-)dependencies of the random variables involved. Given a vertex u of our graph Γ, by saying that a random subset of vertices/edges is *determined by* $(\Gamma_{\hat{u}})_p$, we mean that the set, as a random variable, is completely determined by the random status of edges in $\Gamma_{\hat{u}}$. In other words, the random set is independent of the status of edges adjacent to u in Γ_p.

Lemma 4.2.6. *For any* $0 < \alpha \leq 1$ *and* $L > 0$*, there exists a* $C > 0$ *such that the following holds for all sufficiently large* $n \in \mathbb{N}$ *and for any* $p \geq C(\log n)^{1/3}n^{-2/3}$*. Suppose* Γ *is a tripartite graph with parts of size* n *and* $u \in V(\Gamma)$*. Then we have the following.*

1. *Suppose* $X \subseteq N_\Gamma(u)$ *is a random subset of vertices determined by* $(\Gamma_{\hat{u}})_p$*. Then with probability at least* $1 - n^{-L}$ *we have that the following statement holds in* Γ_p*.*

   If $|X| \geq \alpha n$ *then* $|X \cap N_{\Gamma_p}(u)| \geq \frac{\alpha pn}{2}$.

2. *Suppose* $F \subseteq \mathrm{Tr}_u(\Gamma)$ *is a random subset of edges determined by* $(\Gamma_{\hat{u}})_p$*. Then with probability at least* $1 - n^{-L}$ *we have that the following statement holds in* Γ_p*.*

If $|F| \geq \alpha pn^2$ *then* $|F \cap \mathrm{Tr}_u(\Gamma_p)| \geq \frac{\alpha p^3 n^2}{2}$.

Proof. Choose $\frac{1}{C} \ll \frac{1}{L}, \alpha$. Let $G_1 \subset \Gamma_p$ be the graph on $V(\Gamma)$ consisting of all edges adjacent to u and $G_2 = (\Gamma_{\hat{u}})_p = \Gamma_p \setminus G_1$. For all $w \in N_\Gamma(u)$, let I_w be the indicator random variable for the event that the edge uw appears. By assumption, our random sets X and F depend only on G_2 and clearly the random variables I_w depend only on G_1.

Part 1 now follows from Chernoff's inequality (Theorem 2.1.1). Indeed we have that

$$\mathbb{P}\left[|X \cap N_{\Gamma_p}(u)| < \frac{\alpha pn}{2} \text{ and } |X| \geq \alpha n\right] \leq \mathbb{P}\left[|X \cap N_{\Gamma_p}(u)| < \frac{\alpha pn}{2} \,\middle|\, |X| \geq \alpha n\right]$$

and it suffices to show that $\mathbb{P}\left[|X \cap N_{\Gamma_p}(u)| < \frac{\alpha pn}{2}\right] \leq n^{-L}$ holds for any instance of G_2 and X with $|X| \geq \alpha n$. Fixing such an instance and letting $Y = |X \cap N_{\Gamma_p}(u)| = \sum_{w \in X} I_w$, we have that Y is a sum of independent random variables with expectation $\lambda = \mathbb{E}[Y] = p|X|$ and so

$$\mathbb{P}\left[Y < \frac{\alpha pn}{2}\right] \leq \mathbb{P}\left[Y < \frac{\lambda}{2}\right] \leq e^{-\lambda/8} \leq e^{-\alpha pn/8} \leq n^{-L},$$

for sufficiently large n, as required.

For part 2, we start by noting that $\Delta(G_2) \leq 4pn$ with probability at least $1 - n^{-2L}$ by another simple application of Chernoff's bound (Theorem 2.1.1) and a union bound over all vertices. We have that

$$\begin{aligned}
&\mathbb{P}\left[|F \cap \mathrm{Tr}_u(\Gamma_p)| < \frac{\alpha p^3 n^2}{2} \text{ and } |F| \geq \alpha pn^2\right] \leq \\
&\mathbb{P}\left[|F \cap \mathrm{Tr}_u(\Gamma_p)| < \frac{\alpha p^3 n^2}{2}, |F| \geq \alpha pn^2 \text{ and } \Delta(G_2) \leq 4pn\right] + \mathbb{P}\left[\Delta(G_2) > 4pn\right] \leq \\
&\mathbb{P}\left[|F \cap \mathrm{Tr}_u(\Gamma_p)| < \frac{\alpha p^3 n^2}{2} \,\middle|\, |F| \geq \alpha pn^2 \text{ and } \Delta(G_2) \leq 4pn\right] + n^{-2L}.
\end{aligned}$$

Thus it suffices to prove that $\mathbb{P}\left[|F \cap \mathrm{Tr}_u(\Gamma_p)| < \frac{\alpha p^3 n^2}{2}\right] \leq n^{-2L}$ for any instance of G_2 such that $\Delta(G_2) \leq 4pn$ and $|F| \geq \alpha pn^2$. So let us fix such an instance of G_2 and $F \subseteq \mathrm{Tr}_u(\Gamma)$. Let $\mathcal{F} = \{\{uw_1, uw_2\} : w_1 w_2 \in F\}$ and for $A = \{uw_1, uw_2\} \in \mathcal{F}$, let $I_A = I_{w_1} I_{w_2}$ be the indicator random variable for the event that both edges of A appear in G_1. We will now use Janson's inequality to show that many pairs of edges in $\mathcal{F}$ are present in G_1. Let

$$Z = |F \cap \mathrm{Tr}_u(\Gamma_p)| = \sum_{A \in \mathcal{F}} I_A$$

be the random variable counting the number of triangles containing u and an edge in F. Then

$$\lambda := \mathbb{E}[Z] = p^2 |\mathcal{F}| \geq \alpha p^3 n^2 \geq C^2 \log n. \tag{4.2.8}$$

Furthermore, we have that

$$\bar{\Delta} := \sum_{(A,A')\in\mathcal{F}^2:\ A\cap A'\neq\emptyset} \mathbb{E}\left[I_A I_{A'}\right] \leq 8p^4|\mathcal{F}|n + p^2|\mathcal{F}| = \lambda(1+8p^2n). \tag{4.2.9}$$

Here, the inequality follows from the fact that there are at most $|\mathcal{F}|\cdot 2\cdot\Delta(G_2) = |\mathcal{F}|\cdot 8pn$ pairs $(A,A')\in\mathcal{F}^2$ intersecting in exactly one edge, and $|\mathcal{F}|$ pairs intersecting in two edges. Hence Janson's inequality (Lemma 2.1.3) implies

$$\begin{aligned}\mathbb{P}\left[Z\leq\frac{\lambda}{2}\right] \leq \exp\left(-\frac{\lambda^2}{8\bar{\Delta}}\right) &\leq \exp\left(-\frac{\lambda}{8(1+8p^2n)}\right)\\ &\leq \exp\left(-\frac{\lambda}{16}\right)+\exp\left(-\frac{\lambda}{128p^2n}\right)\\ &\leq n^{-C}+e^{-n^{1/3}} \leq n^{-2L},\end{aligned}$$

for all large enough n. Here, we used (4.2.9) in the second inequality, the fact that $1+8pn^2\leq 2$ or $1+8pn^2\leq 16pn^2$ in the third, (4.2.8) in the fourth and our choice of C in the final inequality. This completes the proof. □

Finally, we show that for most pairs of vertices u and v in the same part, there are many edges appearing in Γ_p that lie in their common neighbourhood (with respect to Γ). We need this to hold even when we forbid certain vertices from being used. This leads to the following statement, for which we direct the reader to the Notation Section for the relevant definitions of e.g. $\mathcal{V}$ and $\mathrm{Tr}_u(G)$.

Lemma 4.2.7. *For all $0<d\leq 1$ there exists $\varepsilon>0$ and $C>0$ such that the following holds for all sufficiently large $n\in\mathbb{N}$ and for any $p\geq C(\log n)^{1/3}n^{-2/3}$. If Γ is an (ε,d)-super-regular tripartite graph with parts V^1,V^2,V^3 of size n, $\ell\in[3]$, $\underline{u}=(u_1,\dots,u_{\ell-1})\in\mathcal{V}$ and $u\in V^\ell$ then with probability at least $1-e^{-n}$ we have that*

$$|\mathrm{Tr}_u(\Gamma_{\hat{\underline{u}}})\cap\mathrm{Tr}_v(\Gamma_{\hat{\underline{u}}})\cap E(\Gamma_p)| \geq \frac{d^5pn^2}{4},$$

for all but at most $2\varepsilon n$ vertices $v\in V^\ell$.

Proof. Choose $0<\varepsilon,\frac{1}{C}\ll d$ and fix Γ, $\ell\in[3]$, $\underline{u}=(u_1,\dots,u_{\ell-1})\in\mathcal{V}$ and $u\in V^\ell$ as in the statement of the lemma.oof We first use regularity to show that there are many edges in the deterministic graph.

Claim. *We have*

$$|\mathrm{Tr}_u(\Gamma_{\hat{\underline{u}}})\cap\mathrm{Tr}_v(\Gamma_{\hat{\underline{u}}})| \geq \frac{d^5n^2}{2},$$

for all but at most $2\varepsilon n$ vertices $v\in V^\ell$.

<u>Proof of Claim:</u> We will prove the claim in the case that $\ell = 3$, the other cases are identical. For $i \in [2]$, let $X_i = N_\Gamma\big(u; V^i_{\underline{\hat{u}}}\big)$ and for $v \in V^3 \setminus \{u\}$, let $Y_i(v) = N_\Gamma\big(u, v; V^i_{\underline{\hat{u}}}\big) \subseteq X_i$. Since Γ is (ε, d)-super-regular, we have $|X_i| \geq (d - 2\varepsilon)n$ for both $i \in [2]$ (we need the factor of 2 in front of the ε here to take account of the fact that we are potentially missing a vertex in $\underline{u}$). For $i \in [2]$, let $R_i \subset V^3$ be the set of vertices $v \in V^3$ for which $|Y_i(v)| < (d - 2\varepsilon)^2 n$ and let $R = R_1 \cup R_2$. It follows from the ε-regularity of (V^i, V^3) and Lemma 2.2.3, that $|R_i| \leq \varepsilon n$ for both $i \in [2]$ and hence $|R| \leq 2\varepsilon n$. Furthermore, for every $v \in V^3 \setminus R$, it follows from the ε-regularity of the pair (V^2, V^3) that $|E(\Gamma) \cap (Y_1(v) \cup Y_2(v))| \geq (d - 2\varepsilon)^5 n^2$. This completes the proof by our choice of ε. ■

Observe now that each edge in $E(\Gamma) \cap N_{\Gamma_{\underline{\hat{u}}}}(u, v) = \mathrm{Tr}_u(\Gamma_{\underline{\hat{u}}}) \cap \mathrm{Tr}_v(\Gamma_{\underline{\hat{u}}})$ is present independently in Γ_p and hence it follows from Chernoff's inequality (Theorem 2.1.1) that for all vertices v satisfying the conclusion of the claim, we have that

$$\mathbb{P}\left[|\mathrm{Tr}_u(\Gamma_{\underline{\hat{u}}}) \cap \mathrm{Tr}_v(\Gamma_{\underline{\hat{u}}}) \cap E(\Gamma_p)| < \frac{d^5 p n^2}{4}\right] \leq \exp\left(-\frac{d^5 p n^2}{16}\right) \leq e^{-2n},$$

for sufficiently large n. This completes the proof after a union bound over choices of $v \in V^\ell$. □

4.3 Embedding (partial) triangle factors

In this section we will prove Proposition 4.1.1 and reduce Proposition 4.1.2 to Lemma 4.1.3. As we have already shown in Section 4.1 that Theorem II* follows from Propositions 4.1.1 and 4.1.2, it will only remain to prove Lemma 4.1.3 after this section, in order to establish Theorem II*.

4.3.1 Counting almost triangle factors

Here we prove Proposition 4.1.1.

Proof of Proposition 4.1.1. Choose $\varepsilon, \frac{1}{C} \ll \varepsilon' \ll \eta, d$ and fix some Γ and p as in the statement of the proposition. By Lemma 4.2.1, we have whp that

$$|K_3(\Gamma_p[X_1 \cup X_2 \cup X_3])| = (pd)^3 |X_1||X_2||X_3| \pm \varepsilon' p^3 n^3, \tag{4.3.1}$$

for all $X_1 \subseteq V^1$, $X_2 \subseteq V^2$ and $X_3 \subseteq V^3$. We will show by induction on t that if Γ_p satisfies (4.3.1), then it satisfies

$$|\Psi^t(\Gamma_p)| \geq (1 - \eta)^t (pd)^{3t} \left(n!_t\right)^3, \tag{4.3.2}$$

for all integers $t \leq (1 - \eta)n$, as claimed. Firstly, note that (4.3.2) is trivial for $t = 0$, recalling that by definition $n!_0 = 1$. Suppose now (4.3.2) holds for some integer $0 \leq t \leq (1 - \eta)n$. Fix some $\psi \in \Psi^t(\Gamma_p)$ and let $X_i \subseteq V^i$, $i \in [3]$, be the sets of vertices which are not in $\psi(D_t)$. Note that $|X_i| = n - t$ for all $i \in [3]$. Now the number of triangles which extend ψ to an embedding in $\Psi^{t+1}(\Gamma_p)$ is precisely $|K_3(\Gamma_p[X_1 \cup X_2 \cup X_3])|$ and by (4.3.1), we have

$$\begin{aligned}|K_3(\Gamma_p[X_1 \cup X_2 \cup X_3])| &\geq (pd)^3|X_1||X_2||X_3| - \varepsilon' p^3 n^3 \\ &\geq (pd)^3(n-t)^3 - \frac{\varepsilon'}{\eta^3 d^3}(pd)^3(n-t)^3 \\ &\geq (1-\eta)(pd)^3(n-t)^3,\end{aligned}$$

by our choice of constants. It follows from the induction hypothesis that

$$\begin{aligned}|\Psi^{t+1}(\Gamma_p)| &\geq |\Psi^t(\Gamma_p)|(1-\eta)(pd)^3(n-t)^3 \\ &\geq (1-\eta)^{t+1}(pd)^{3(t+1)}\left(n!_{t+1}\right)^3,\end{aligned}$$

finishing the proof. □

4.3.2 Extending almost triangle factors

In this subsection, we will prove Proposition 4.1.2 using the Local Distribution Lemma (see Lemma 4.1.3) as a black box for now. We first reduce Proposition 4.1.2 to the following lemma, which concentrates on adding a triangle at a fixed vertex. Recall that given $G \subseteq \Gamma$, a vertex $v \in V^1$ and some $t \in \mathbb{N}$, we denote by $\Psi^t_v(G) \subseteq \Psi^t(G)$ the set of embeddings $\psi \in \Psi^t(G)$ for which $\psi((1,1)) = v$.

Lemma 4.3.1 (adding a triangle at a fixed vertex)**.** *For all $0 < d \leq 1$ there exists $\alpha, \eta, \varepsilon > 0$ and $C > 0$ such that for all sufficiently large $n \in \mathbb{N}$ and for any $p \geq C(\log n)^{1/3} n^{-2/3}$, if Γ is an (ε, d)-super-regular tripartite graph with parts of size n, then whp the following holds in Γ_p for all $t \in \mathbb{N}$ with $(1-\eta)n \leq t < n$ and for all $v \in V^1$. If*

$$|\Psi^t_{\hat{v}}(\Gamma_p)| \geq (1-\eta)^n\,(pd)^{3t}(n-1)!_t(n!_t)^2,$$

then

$$|\Psi^{t+1}_v(\Gamma_p)| \geq \alpha(pd)^3(n-t)^2|\Psi^t_{\hat{v}}(\Gamma_p)|.$$

We first show how Proposition 4.1.2 follows from this.

Proof of Proposition 4.1.2. Choose $0 < \varepsilon, \frac{1}{C} \ll \eta \ll \eta' \ll \alpha \ll \alpha' \ll d$. Now by our choice of constants (also choosing $K \geq 5$) and taking a union bound over all choices of t with $(1-\eta)n \leq t < n$, $\ell \in [3]$ and $\underline{u} = (u_1, \ldots, u_{\ell-1}) \in \mathcal{V}$ we have whp that the conclusion of Lemma 4.1.3 holds in Γ_p for all such choices and also the conclusion of Lemma 4.3.1 holds with η' and α' replacing η and α. We will now show that given these conclusions hold in Γ_p, we have the desired statement of Proposition 4.1.2. So fix some $t \in \mathbb{N}$ with $(1-\eta)\,n \leq t < n$ and suppose that

$$|\Psi^t(\Gamma_p)| \geq (1-\eta)^n\,(pd)^{3t}\,(n!_t)^3.$$

Let $U_1 \subseteq V^1$ be the set of vertices $u_1 \in V^1$ for which

$$|\Psi^t_{\hat{u}_1}(\Gamma_p)| \geq \left(\frac{d}{10}\right)^2\left(\frac{n-t}{n}\right)|\Psi^t(\Gamma_p)|. \tag{4.3.3}$$

It follows from (the assumed conclusion of) Lemma 4.1.3 (with $\ell = 1$) that we have $|U_1| \geq \frac{n}{2}$. Now as

$$\left(\frac{d}{10}\right)^2\left(\frac{n-t}{n}\right)|\Psi^t(\Gamma_p)| \geq \left(\frac{d}{10}\right)^2(1-\eta)^n\,(pd)^{3t}\,(n-1)!_t\,(n!_t)^2,$$

and $\left(\frac{d}{10}\right)^2(1-\eta)^n \geq (1-\eta')^n$, we have that

$$|\Psi^{t+1}_{u_1}(\Gamma_p)| \geq \alpha'\left(\frac{d}{10}\right)^2(pd)^3\frac{(n-t)^3}{n}|\Psi^t(\Gamma_p)|,$$

for every $u_1 \in U_1$, from (the assumed conclusion of) Lemma 4.3.1 and (4.3.3). Therefore, we have that

$$\begin{aligned}
|\Psi^{t+1}(G)| &\geq \sum_{u_1\in U_1}|\Psi^{t+1}_{u_1}(\Gamma_p)| \\
&\geq \frac{\alpha'}{2}\left(\frac{d}{10}\right)^2(pd)^3(n-t)^3|\Psi^t(\Gamma_p)| \\
&\geq \alpha(pd)^3(n-t)^3|\Psi^t(\Gamma_p)|,
\end{aligned}$$

by our choice of constants. This finishes the proof. □

It remains to prove Lemma 4.3.1. Before embarking on this, we sketch some of the key ideas involved. For this discussion, we fix some $t \in [n]$ and $v \in V^1$ that we think of as satisfying the conditions of Lemma 4.3.1 (including the "if" statement). We will say that a pair $(w_2, w_3) \in V^2 \times V^3$ is *good* if

$$|\Psi^t_{\hat{\underline{w}}}(\Gamma_p)| = \Omega\left(\left(\frac{n-t}{n}\right)^2|\Psi^t_{\hat{v}}(\Gamma_p)|\right),$$

where $\underline{w} = (v, w_2, w_3)$. Note that we can appeal to the Local Distribution Lemma (Lemma 4.1.3) twice (once with $\ell = 2$ and once with $\ell = 3$) to conclude that almost all pairs $(w_2, w_3) \in V^2 \times V^3$ are good. That is, for almost all choices of $(w_2, w_3) \in V^2 \times V^3$, we have that there are roughly the 'correct' number of embeddings of D_t that avoid $\underline{w} = (v, w_2, w_3)$. Moreover, due to Γ being super-regular, there will be some proportion of these (w_2, w_3) (say, at least $\frac{1}{2}d^3n^2$) that form triangles with v in Γ. So we have some set $W \subset V^2 \times V^3$ of size at least $\frac{1}{2}d^3n^2$ such that all $(w_2, w_3) \in W$ are good and have that $\{v, w_2, w_3\} \in K_3(\Gamma)$. The conclusion of Lemma 4.3.1 will then follow if we can prove that at least, say, $\frac{p^3}{2}|W|$ triangles $\{v, w_2, w_3\}$ with $(w_2, w_3) \in W$, appear in Γ_p. Of course, every triangle in Γ appears in Γ_p with probability p^3 and so this is something we can expect to be true but we cannot appeal to standard tools to prove this.

The issue here is that W itself is a random set as the property of *being good* depends on the random edges that appear in Γ_p. Indeed, in order to determine whether an edge $(w_2, w_3) \in V^2 \times V^3$ is good or not, we need to count the number of embeddings of D_t in Γ_p that avoid $\underline{w} = (v, w_2, w_3)$ and so certainly need to know that random status of edges in Γ_p to carry out this count. However, crucially, W does not depend on *all* the random edges. Indeed, for *any* $(w_2, w_3) \in V^2 \times V^3$, we can determine whether (w_2, w_3) is in our set W without knowing the random status of edges adjacent to v. Indeed, as the property of being good only depends on counting embeddings that *avoid* v, the random status of edges adjacent to v has no bearing on whether an edge $(w_2, w_3) \in V^2 \times V^3$ is good or not. Therefore, by appealing to a two-stage revealing process (see Lemma 4.2.6 (2)), we will be able to prove Lemma 4.3.1 if we know that at least, say, $\frac{p}{2}|W|$ of the pairs $(w_2, w_3) \in W$ host edges in Γ_p, as then we will be able to conclude that roughly a p^2 proportion of these edges in $W \cap E(\Gamma_p)$ extend to triangles with v in Γ_p.

Again, requiring that $\frac{p}{2}|W|$ edges in W appear in Γ_p is certainly a natural thing to expect as each edge appears with probability p, but again the set W containing good edges, depends heavily on the random status of edges in $\Gamma[V^2, V^3]$. Our aim is to use a two-stage revealing process (Lemma 4.2.6), manipulating independence, as above. Again here, it is crucial that we are counting embeddings that avoid vertices. That is, if $e = \{w_2, w_3\} \in E(\Gamma[V^2, V^3])$, then in order to determine the number of embeddings that avoid (v, w_2, w_3), we do not need to know the random status of e and in fact more is true. The number of embeddings of D_t avoiding (v, w_2, w_3) is independent of the random status of *all* (w_2, u_3) with $u_3 \in N_\Gamma(w_2; V^3)$. Therefore our approach is to lower bound the number of edges in $|W \cap \Gamma_p|$ by grouping together edges in W according to their V^2-endpoint. This gives hope to use a two-stage random revealing argument (appealing to Lemma 4.2.6 (1)) to conclude that roughly the expected number of good edges appear in Γ_p.

However, there is an oversight in the discussion above. The point is that our definition of whether an edge $(w_2, w_3) \in V^2 \times V^3$ is good or not does not only rely on counting embeddings avoiding $\underline{w} = (v, w_2, w_3)$, we *also* need to know the size of $|\Psi^t_{\hat{v}}(\Gamma_p)|$. Therefore, if $e = \{w_2, w_3\} \in E(\Gamma[V^2, V^3])$, then in order to determine if (w_2, w_3) is *good*, we actually need to reveal the random status of e itself as well as all the random edges between V^2 and V^3 (to determine $|\Psi^t_{\hat{v}}(\Gamma_p)|$). To remedy this, we adjust our definition of good to be independent of $|\Psi^t_{\hat{v}}(\Gamma_p)|$. We will therefore give a *grading* of the possible range of $|\Psi^t_{\hat{v}}(\Gamma_p)|$ and show that the desired conclusion holds with respect to each *grade* (see Claim 4.3.3 in the proof). In order to be able to perform a union bound over all of the possible grades, we need an upper bound on how large $|\Psi^t_{\hat{v}}(\Gamma_p)|$ can be (whp) and this is provided by Claim 4.3.2. This idea allows us to remove $|\Psi^t_{\hat{v}}(\Gamma_p)|$ from the definition of being good, leading to the definition of being *sound* in the proof. Hence we have that for $e = \{w_2, w_3\} \in \Gamma[V^2, V^3]$, whether (w_2, w_3) is sound or not relies only on counting embeddings avoiding $\underline{w} = (v, w_2, w_3)$ and so is independent of whether e appears in Γ_p and in fact, as sketched above, the 'soundness' of (w_2, w_3) is independent of the random status of *all* (w_2, u_3) with $u_3 \in N_\Gamma(w_2; V^3)$. We therefore consider potential triangles one vertex at a time and we refine our definition of sound to handle this, leading to the definition of *sound tuples* in the proof. We now give the full details of the proof of Lemma 4.3.1.

Proof of Lemma 4.3.1. Choose $0 \ll \varepsilon, \frac{1}{C} \ll \eta \ll \eta' \ll \alpha \ll d$ and fix p and Γ as in the statement of the lemma. We begin by showing the following simple claim which gives a weak upper bound on the number of embeddings that avoid a fixed vertex v_1.

Claim 4.3.2. *We have that the following statement holds whp in Γ_p. For any $t \in \mathbb{N}$ such that $(1-\eta)\, n \le t < n$ and $v_1 \in V^1$, we have that*

$$|\Psi^t_{\hat{v}_1}(\Gamma_p)| \le n^3 p^{3t} (n-1)!_t (n!_t)^2. \tag{4.3.4}$$

Proof of Claim: Fix some $t \in \mathbb{N}$ and $v_1 \in V^1$ as in the statement of the claim. Then

$$|\Psi^t_{\hat{v}_1}(\Gamma)| \le |\Psi^t_{\hat{v}_1}(K_{n,n,n})| \le (n-1)!_t (n!_t)^2,$$

and so, as each embedding of D_t in Γ appears in Γ_p with probability p^{3t}, we have that $\lambda := \mathbb{E}\left[|\Psi^t_{\hat{v}_1}(\Gamma_p)|\right] \le p^{3t}(n-1)!_t (n!_t)^2$. Therefore, appealing to Markov's inequality gives that

$$\mathbb{P}\left[|\Psi^t_{\hat{v}_1}(\Gamma_p)| > n^3 p^{3t} (n-1)!_t (n!_t)^2\right] \le \mathbb{P}\left[|\Psi^t_{\hat{v}_1}(\Gamma_p)| > n^3 \lambda\right] \le \frac{1}{n^3}.$$

Taking a union bound over the choices of $v \in V^1$ and $t \in \mathbb{N}$ with $(1-\eta)n \le t \le n$ completes the proof of the claim. ■

Claim 4.3.2 gives us an upper bound on the size of $\Psi^t_{\hat{v}_1}(\Gamma_p)$ that holds whp, whilst the statement of the lemma gives a lower bound. Our next claim replaces the lower bound in the statement of the lemma, with lower bounds independent of $|\Psi^t_{\hat{v}_1}(\Gamma_p)|$. These lower bounds will depend on a parameter $s \in \mathbb{Z}$ and we make the following definitions which will define the range of s we are interested in. Firstly let s_0 be the largest $s \in \mathbb{Z}$ such that $2^s \le (1-\eta)^n$. Further, let s_1 be the minimum integer $s \in \mathbb{N}$ such that $2^s d^{3t} \ge n^3$. So we have that

$$s_0 \ge \frac{\log(1-\eta)\, n}{\log 2} - 1 \ge -n \qquad \text{and} \qquad s_1 \le \frac{3\log n - 3t\log(d)}{\log 2} + 1 \le \frac{n}{\alpha}.$$

Finally, let $\mathbb{S} := \{s \in \mathbb{Z} : s_0 \le s \le s_1\}$. We now state our second claim.

Claim 4.3.3. *For any $t \in \mathbb{N}$ with $(1-\eta)n \le t < n$, $s \in \mathbb{S}$ and $v_1 \in V^1$, with probability at least $1 - n^{-4}$, the following statement holds in Γ_p. If*

$$|\Psi^t_{\hat{v}_1}(\Gamma_p)| \ge 2^s (pd)^{3t} (n-1)!_t (n!_t)^2, \tag{4.3.5}$$

then

$$|\Psi^t_{v_1}(\Gamma_p)| \ge 2^{s+1} \alpha (pd)^{3(t+1)} (n-1)!_t (n!_{t+1})^2.$$

Before proving Claim 4.3.3, we show how the lemma follows from the two claims. Taking a union bound, we can conclude that whp the conclusion of Claim 4.3.3 holds for all choices of t, s and v_1 (noting that $|\mathbb{S}| \le n \log n$), as well as the conclusion of Claim 4.3.2. Now suppose that this is the case and let $t \in \mathbb{N}$ with $(1-\eta)n \le t < n$ and $v \in V^1$. If

$$|\Psi^t_{\hat{v}}(\Gamma_p)| \ge (1-\eta)^n (pd)^{3t} (n-1)!_t (n!_t)^2, \tag{4.3.6}$$

then, letting $s^* \in \mathbb{Z}$ be the maximum integer $s \in \mathbb{Z}$ such that

$$|\Psi^t_{\hat{v}}(\Gamma_p)| \ge 2^s (pd)^{3t} (n-1)!_t (n!_t)^2,$$

we have from (the assumed conclusion of) Claim 4.3.2 and (4.3.6), that $s^* \in \mathbb{S}$. Therefore, from (the assumed conclusion of) Claim 4.3.3, we have that that

$$|\Psi^t_{v}(\Gamma_p)| \ge 2^{s^*+1} \alpha (pd)^{3(t+1)} (n-1)!_t (n!_{t+1})^2 \ge \alpha (pd)^3 (n-t)^2 |\Psi^t_{\hat{v}}(\Gamma_p)|,$$

as required, where we used that $|\Psi^t_{\hat{v}}(\Gamma_p)| \le 2^{s^*+1} (pd)^{3t} (n-1)!_t (n!_t)^2$, by the definition of s^*. Thus it remains to prove Claim 4.3.3.

<u>Proof of Claim 4.3.3:</u> Let us fix some choice of $t \in \mathbb{N}$ with $(1-\eta)n \le t < n$, $s \in \mathbb{S}$ and $v_1 \in V^1$. Given some $\ell \in [3]$, we call a sequence of vertices $\underline{u} = (u_1, \dots, u_\ell) \in \mathcal{V}$ *sound* if

$$|\Psi^t_{\hat{\underline{u}}}(\Gamma_p)| \ge \left(8\sqrt{\alpha}\right)^{\ell-1} 2^s (pd)^{3t} ((n-1)!_t)^\ell (n!_t)^{3-\ell}.$$

Note that (4.3.5) holds if and only if (v_1) is sound. Also note that for any $\underline{u} = (u_1, \dots, u_\ell)$ and $i \in [\ell]$, we can determine whether $\underline{u}$ is sound or not without knowing the random status of edges adjacent to u_i in Γ, as determining whether $\underline{u}$ is sound relies on counting embeddings that *avoid* u_i.

We now formulate a sequence of steps, that we will prove later, claiming that certain properties hold. Let $X_2(v_1) \subseteq N_\Gamma(v_1; V^2)$ be the set of vertices $u_2 \in N_\Gamma(v_1; V^2)$ such that (v_1, u_2) is sound and $\deg_\Gamma(v_1, u_2; V^3) \ge \frac{d^2 n}{2}$.

Step 1. *With probability at least $1 - n^{-6}$, the following statement holds in Γ_p.*

If (v_1) is sound, then $|X_2(v_1)| \geq \frac{dn}{2}$.

Given $v_2 \in V^2$, let $X_3(v_1, v_2) \subseteq N_\Gamma(v_1, v_2; V^3)$ be the set of vertices $u_3 \in N_\Gamma(v_1, v_2; V^3)$ such that (v_1, v_2, u_3) is sound. Furthermore, let $Y_3(v_1, v_2) \subseteq X_3(v_1, v_2)$ be the set of those u_3 such that $v_2 u_3 \in E(\Gamma_p)$.

Step 2. *With probability at least $1 - n^{-6}$ the following statement holds in Γ_p for every $v_2 \in V^2$.*

If (v_1) is sound and $v_2 \in X_2(v_1)$, then we have $|Y_3(v_1, v_2)| \geq \frac{pd^2n}{8}$.

Let now $Z'(v_1) = \left\{(u_2, u_3) \in V^2 \times V^3 : u_2 \in X_2(v_1), u_3 \in Y_3(v_1, u_2)\right\}$ and $Z(v_1) = \{(u_2, u_3) \in Z'(v_1) : \{v_1, u_2, u_3\}$ is a triangle in $\Gamma_p\} = \mathrm{Tr}_{v_1}(\Gamma_p) \cap Z'(v_1)$. We will use Steps 1 and 2 to deduce the following.

Step 3. *With probability at least $1 - n^{-5}$, the following statement holds in Γ_p.*

If (v_1) is sound, then $|Z'(v_1)| \geq \frac{pd^3n^2}{16}$.

Step 4. *With probability at least $1 - n^{-5}$, the following statement holds in Γ_p.*

If $|Z'(v_1)| \geq \frac{pd^3n^2}{16}$, then we have $|Z(v_1)| \geq \frac{(pd)^3n^2}{32}$.

Before we prove the claims in Steps 1 to 4, let us deduce Claim 4.3.3. Note that assuming the statements in Steps 3 and 4 hold in Γ_p we have that if (v_1) is sound then $|Z(v_1)| \geq \frac{(pd)^3n^2}{32}$ by combining both the claims. Furthermore, for all $(u_2, u_3) \in Z(v_1)$, the vector (v_1, u_2, u_3) is sound and hence $|\Psi^t_{\hat{v}_1, \hat{u}_2, \hat{u}_3}(\Gamma_p)| \geq 64\alpha 2^s (pd)^{3t}((n-1)!_t)^3$. Therefore

$$
\begin{aligned}
|\Psi^{t+1}_{v_1}(\Gamma_p)| &\geq \sum_{(u_2,u_3)\in Z(v_1)} |\Psi^t_{\hat{v}_1,\hat{u}_2,\hat{u}_3}(\Gamma_p)| \\
&\geq \frac{(pd)^3 n^2}{32} \cdot 64\alpha 2^s (pd)^{3t}((n-1)!_t)^3 \\
&\geq 2^{s+1}\alpha (pd)^{(3t+1)}((n-1)!_t)(n!_{t+1})^2,
\end{aligned}
$$

as required. Now as the statements of Steps 3 and 4 hold simultaneously in Γ_p with probability at least $1 - n^{-4}$, this concludes the proof of the claim. It remains to prove Steps 1 to 4.

Proof of Step 1. For $i = 2, 3$, let $A_i := N_\Gamma(v_1; V^i)$. Furthermore, let $A_2' \subseteq V^2$ be the set of vertices $u_2 \in V^2$ for which (v_1, u_2) is sound and let $A_2'' \subseteq V^2$ be the set of vertices $u_2 \in V^2$ for which $\deg(v_1, u_2; V^3) \geq \frac{d^2n}{2}$. Note that $X_2(v_1) = A_2 \cap A_2' \cap A_2''$. Since (V^1, V^i) is (ε, d)-super-regular, we have $|A_i| \geq (d-\varepsilon)n$ for $i = 2, 3$. Since (V^2, V^3) is ε-regular, we have $|A_2''| \geq (1-\varepsilon)n$

by Lemma 2.2.3. Finally, it follows from Lemma 4.1.3 with $\ell = 2$ and η replaced by η', that with probability at least $1 - n^{-6}$, if (v_1) is sound then we have that $|A_2'| \geq (1-\alpha)n$ and hence

$$|X_2(v_1)| = |A_2 \cap A_2' \cap A_2''| \geq \frac{dn}{2},$$

as claimed. Here we used that (v_1) being sound implies that

$$|\Psi^t_{\hat{v}_1}(\Gamma_p)| \geq 2^{s_0}(pd)^{3t}(n-1)!_t(n!_t)^2 \geq (1-\eta')^n(pd)^{3t}(n-1)!_t(n!_t)^2,$$

in our application of Lemma 4.1.3. ■

Proof of Step 2. Fix some $v_2 \in V^2$. Let $X_3 = X_3(v_1, v_2)$ and $Y_3 = Y_3(v_1, v_2) \subseteq X_3$. It follows from an application of Lemma 4.1.3 with $\ell = 3$ and η' replacing η, that the following statement holds in Γ_p with probability at least $1 - n^{-8}$.

If (v_1) is sound and $v_2 \in X_2(v_1)$, then $|X_3| \geq \frac{d^2n}{4}$.

Here we used here that $v_2 \in X_2(v_1)$ implies that $\deg_\Gamma(v_1, v_2; V^3) \geq \frac{d^2n}{2}$ as well as the fact that (v_1, v_2) being sound implies that

$$|\Psi^t_{\hat{v}_1,\hat{v}_2}(\Gamma_p)| \geq 8\sqrt{\alpha}2^{s_0}(pd)^{3t}((n-1)!_t)^2(n!_t) \geq (1-\eta')^n(pd)^{3t}((n-1)!_t)^2(n!_t),$$

in order to appeal to Lemma 4.1.3.

Now note that, in order to determine X_3, we do not need to reveal edges adjacent to v_2. That is, the random set of vertices X_3 is determined by $(\Gamma_{\hat{v}_2})_p$. Therefore, by Lemma 4.2.6 (1) we have that with probability at least $1 - n^{-8}$ the following statement holds in Γ_p.

If $|X_3| \geq \frac{d^2n}{4}$, then $|Y_3| \geq \frac{pd^2n}{8}$.

Therefore with probability at least $1 - n^{-7}$, both the above statements hold in Γ_p and so by combining them we have the desired statement of this step for $v_2 \in V^2$. Taking a union bound over all $v_2 \in V^2$ then completes the proof. ■

Proof of Step 3. This is a simple case of combining Steps 1 and 2. Indeed with probability at least $1 - n^{-5}$ both the statements of Steps 1 and 2 hold in Γ_p. Taking this to be the case, if (v_1) is sound, we then have that

$$|Z'(v_1)| = \sum_{u_2 \in X_2(v_1)} |Y_3(v_1, u_2)| \geq \frac{dn}{2} \cdot \frac{pd^2n}{8} = \frac{pd^3n^2}{16},$$

as required. ■

Proof of Step 4. This is a direct application of Lemma 4.2.6 (2). Indeed, note that $Z'(v_1) \subseteq \mathrm{Tr}_{v_1}(\Gamma)$ is a random subset of edges determined by $(\Gamma_{\hat{v}_1})_p$. The conclusion of Step 4 then follows immediately from the lemma. ■

This concludes the proof of Claim 4.3.3 and hence the lemma. ■ □

4.4 Proof of the Local Distribution Lemma

The purpose of this section is to prove the Local Distribution Lemma, Lemma 4.1.3. We will begin by reducing Lemma 4.1.3 to another lemma, Lemma 4.4.1 below, using a simple averaging argument. Before proving Lemma 4.4.1, we will then take a detour, establishing an Entropy Lemma (Lemma 4.4.4) which will be crucial for the remaining proof.

4.4.1 A simplification

Given some t, ℓ and $\underline{u} = (u_1, \ldots, u_{\ell-1})$ as in the statement of Lemma 4.1.3, we aim to prove a lower bound on the size of $\Psi^t_{\underline{\hat{u}},\hat{u}_\ell}$ for almost all of the $u_\ell \in V^\ell$. Given that $\Psi^t_{\underline{\hat{u}}}$ is large, a simple averaging argument shows that (4.1.6) is true 'on average' (i.e. if we take the average of $|\Psi^t_{\underline{\hat{u}},\hat{u}_\ell}(\Gamma_p)|$ over all $u_\ell \in V^\ell$). The challenge comes in proving that (4.1.6) holds for almost all choices of u_ℓ. In order to do this, we compare the difference in the sizes of $\Psi^t_{\underline{\hat{u}},\hat{u}_\ell}$ for different choices of $u_\ell \in V^\ell$. The key step is given in the following lemma.

Lemma 4.4.1. *For all $0 < \alpha, d \leq 1$ and $K > 0$ there exists $\eta, \varepsilon > 0$ and $C > 0$ such that for all sufficiently large $n \in \mathbb{N}$ and for any $p \geq C(\log n)^{1/3} n^{-2/3}$, if Γ is an (ε, d)-super-regular tripartite graph with parts of size n, $t \in \mathbb{N}$ such that $(1-\eta)n \leq t < n$, $\ell \in [3]$, $\underline{u} = (u_1, \ldots, u_{\ell-1}) \in \mathcal{V}$ and $u \in V^\ell$ then the following holds in Γ_p with probability at least $1 - n^{-K}$. If*

$$|\Psi^t_{\underline{\hat{u}},\hat{u}}(\Gamma_p)| \geq (1-\eta)^n (pd)^{3t} ((n-1)!_t)^\ell (n!_t)^{3-\ell},$$

then

$$|\Psi^t_{\underline{\hat{u}},\hat{v}}(\Gamma_p)| \geq \left(\frac{d}{10}\right)^2 \cdot |\Psi^t_{\underline{\hat{u}},\hat{u}}(\Gamma_p)|$$

for at least $(1-\alpha)n$ vertices $v \in V^\ell$.

Indeed, with Lemma 4.4.1 in hand, Lemma 4.1.3 follows easily.

Proof of Lemma 4.1.3. Fix $\varepsilon, \frac{1}{C} \ll \eta \ll d, \alpha$. Fix Γ, $t \in \mathbb{N}$ with $(1-\eta)n \leq t < n$, $\ell \in [3]$ and $\underline{u} = (u_1 \ldots, u_{\ell-1}) \in \mathcal{V}$. By applying Lemma 4.4.1 with $K+1$ replacing K and taking a union bound, we have that with probability at least $1 - n^{-K}$, the conclusion of Lemma 4.4.1 holds in $G = \Gamma_p$ for all $u \in V^\ell$. So suppose that this is the case and further suppose that

$$|\Psi^t_{\underline{\hat{u}}}(G)| \geq (1-\eta)^n (pd)^{3t} ((n-1)!_t)^{\ell-1} (n!_t)^{4-\ell}.$$

Now, for each $\psi \in \Psi^t_{\underline{\hat{u}}}(G)$, we have $\psi \in \Psi^t_{\underline{\hat{u}},\hat{u}_\ell}(G)$ for exactly $n-t$ choices of $u_\ell \in V^\ell$. Therefore, we have that

$$\sum_{u \in V^\ell} |\Psi^t_{\underline{\hat{u}},\hat{u}}(G)| = (n-t)|\Psi^t_{\underline{\hat{u}}}(G)|.$$

By averaging, there must be some $u^* \in V^\ell$ such that

$$\begin{aligned} |\Psi^t_{\underline{\hat{u}},\hat{u}^*}(G)| &\geq \left(\frac{n-t}{n}\right) |\Psi^t_{\underline{\hat{u}}}(G)| \\ &\geq \left(\frac{n-t}{n}\right) (1-\eta)^n (pd)^{3t} ((n-1)!_t)^{\ell-1} (n!_t)^{4-\ell} \\ &= (1-\eta)^n (pd)^{3t} ((n-1)!_t)^{\ell} (n!_t)^{3-\ell}. \end{aligned}$$

The result now follows from applying the assumed conclusion of Lemma 4.4.1 with u^* playing the rôle of u. □

4.4.2 The Entropy Lemma

In this section, we will prove a key lemma, Lemma 4.4.4, which we call the Entropy Lemma. We start with some definitions. Given some tripartite Γ with parts of size n, some $\ell \in [3]$, $t \in [n]$ and some $\psi \in \Psi^t(\Gamma)$, we define $I^\ell(\psi) \subset V^\ell$ to be the vertices in V^ℓ which are isolated in the embedded subgraph $\psi(D_t)$. If ℓ is clear from context, we will drop the superscript. If we are further given some $v \in V^\ell$, we define

$$\psi_v = \begin{cases} \emptyset & \text{if } v \in I(\psi), \\ \left(N_{\psi(D_t)}\left(v; V^j\right) : j \in J\right) & \text{if } v \notin I(\psi), \end{cases}$$

where $J = [3] \setminus \{\ell\}$. So ψ_v either returns an empty set, indicating that the vertex v is isolated in $\psi(D_t)$, or it returns the pair of vertices which are contained in the triangle containing v in $\psi(D_t)$. We also define the function

$$Y_v(\psi) = \mathbb{1}[\{\psi_v \neq \emptyset\}] = \begin{cases} 1 & \text{if } \psi_v \neq \emptyset, \\ 0 & \text{if } \psi_v = \emptyset, \end{cases}$$

which returns 1 if $v \notin I(\psi)$ and 0 otherwise. Note that for any $\ell \in [3]$ the set $\{\psi_v : v \in V^\ell\}$ completely determines the (unordered) subgraph $\psi(D_t)$.

For a fixed $u \in V^\ell$ and $v \in V^\ell \setminus \{u\}$, we will be interested in the distribution of ψ^*_v if ψ^* is chosen randomly among a set of embeddings we wish to extend. In order to analyse this, we use entropy. See Section 2.3 for the definition and basic properties. We remark that there will be two independent stages of randomness in the argument. First, there is the random subgraph $\Gamma_p \subseteq \Gamma$, and second, there will be a randomly chosen $\psi^* \in \Psi^t(\Gamma_p)$. In particular, the values of the entropy function $h(\psi^*), h(\psi^*_v)$ are random variables themselves. However, once we fix a particular instance $G = \Gamma_p$, these values are deterministic. We proceed with the following definition which will be convenient to ease notation in what follows.

Definition 4.4.2. For $n \in \mathbb{N}$, $p = p(n) \in (0, 1)$ and $0 < d \leq 1$, we define

$$H = H(n, p, d) := \log\left((pd)^3 \cdot n^2\right).$$

To see the relevance of this function, note that in a random sparsification of the complete tripartite graph $K_{n,n,n}$ with probability pd, we would expect a given vertex to lie in $(pd)^3 n^2$ triangles. Therefore if we fix a vertex v and take a uniformly random triangle containing v, we expect the entropy of the random variable which chooses this triangle, to be roughly $H(n, p, d)$. The function H can thus be seen as benchmark for the maximum entropy (recalling Lemma 2.3.1) of a randomly chosen triangle containing a fixed vertex. Our aim will be to show that, for most choices of fixed vertex v, H is a good approximation for the entropy of the random variable ψ^*_v discussed above.

We begin with observing that the function H provides an appropriate upper bound on the entropy we will be interested in.

Observation 4.4.3. For all $0 < \varepsilon' < d \leq 1$ and $L > 0$ there exists $\varepsilon > 0$ and $C > 0$ such that for all sufficiently large $n \in \mathbb{N}$ and for any $p \geq C(\log n)^{1/3} n^{-2/3}$, if Γ is an (ε, d)-super-regular tripartite graph with parts of size n, $t \in [n]$, $\ell \in [3]$, $\underline{u} = (u_1, \ldots, u_{\ell-1}) \in \mathcal{V}$ and $u \in V^\ell$ then the following holds in Γ_p with probability at least $1 - n^{-L}$.

For ψ^* chosen uniformly from $\Psi^t_{\underline{\hat{u}},\hat{u}}(\Gamma_p)$, we have that $h(\psi^*_v | Y_v(\psi^*) = 1) \leq H(n, p, d) + \varepsilon'$ for all but at most $\varepsilon' n$ vertices $v \in V^\ell$.

Proof. Choose $0 < \varepsilon, \frac{1}{C} \ll \varepsilon', d, \frac{1}{L}$. By Corollary 4.2.4, we have that with probability at least $1 - n^{-L}$,

$$|\mathrm{Tr}_v(\Gamma_p)| = (1 \pm \varepsilon')(pd)^3 n^2,$$

for all but at most $\varepsilon' n$ vertices $v \in V^\ell$. In particular, for each such v, we have $\log|\mathrm{Tr}_v((\Gamma_p)_{\underline{\hat{u}},\hat{u}})| \leq H(n, p, d) + \varepsilon'$. Therefore, by Lemma 2.3.1, we have $h(\psi^*_v | Y_v(\psi^*) = 1) \leq H(n, p, d) + \varepsilon'$ for all v as above and for $\psi^* \in \Psi^t_{\underline{\hat{u}},\hat{u}}(\Gamma_p)$ chosen uniformly at random. ◻

The main purpose of this section is to provide a partial converse to the above observation, showing that for almost all vertices $v \in V^\ell$, H is a good approximation for the entropy $h(\psi_v^*|Y_v(\psi^*) = 1)$. The full statement is as follows.

Lemma 4.4.4 (Entropy Lemma)**.** *For all $0 < \beta, d \leq 1$ and $L > 0$ there exists $\eta, \varepsilon > 0$ and $C > 0$ such that for all sufficiently large $n \in \mathbb{N}$ and for any $p \geq C(\log n)^{1/3}n^{-2/3}$, if Γ is an (ε, d)-super-regular tripartite graph with parts of size n, $t \in \mathbb{N}$ such that $(1-\eta)n \leq t < n$, $\ell \in [3]$, $\underline{u} = (u_1, \dots, u_{\ell-1}) \in \mathcal{V}$ and $u \in V^\ell$ then the following holds in Γ_p with probability at least $1 - n^{-L}$. If*

$$|\Psi^t_{\hat{\underline{u}},\hat{u}}(\Gamma_p)| \geq (1-\eta)^n (pd)^{3t}((n-1)!_t)^\ell (n!_t)^{3-\ell},$$

and ψ^* is chosen uniformly from $\Psi^t_{\hat{\underline{u}},\hat{u}}(\Gamma_p)$, then we have that $h(\psi_v^*|Y_v(\psi^*) = 1) \geq H(n,p,d) - \beta$ for all but at most βn vertices $v \in V^\ell$.

In the remainder of this section, we will prove Lemma 4.4.4. Recall that we have $V(\Gamma) = V(\Gamma_p) = V^1 \cup V^2 \cup V^3$ with each V^i of size n. As above, for $t \in [n]$, an embedding $\psi \in \Psi^t(\Gamma)$ and some $\ell \in [3]$, we denote by $I(\psi) = I^\ell(\psi)$ the vertices in V^ℓ which are not contained in the subgraph $\psi(D_t)$. In the proof, we will describe ψ by revealing the status of ψ_v one by one for each $v \in V^\ell$ according to some linear order σ of V^ℓ. In order to do so, we need to make some further definitions. Firstly we denote by $w <_\sigma v$ that w occurs before v in the ordering σ. Now given some fixed t, ψ and ℓ as above and an ordering σ of V^ℓ, we will be interested in revealing $\psi \in \Psi^t(\Gamma)$ according to the ordering σ as follows. We imagine processing the vertices $v \in V^\ell$ in order and as we process each vertex v we reveal its status in ψ by revealing ψ_v. Either v is not in a triangle in $\psi(D_t)$ or v is in a triangle, in which case, we are given the other vertices of the triangle containing v in $\psi(D_t)$. Now consider the moment before processing some vertex $v \in V^\ell$. At this point, we know all the triangles in $\psi(D_t)$ that contain vertices $w \in V^\ell$ such that $w <_\sigma v$. We are interested in which vertices are candidates to feature in ψ_v at this point and the following definition captures this.

For some fixed t, ψ and ℓ as above, an ordering σ of V^ℓ, some $\underline{u} \in \mathcal{V}$, some $j \in [3] \setminus \{\ell\}$ and some $v \in V^\ell$ we define

$$A_v^j(\psi, \sigma, \underline{u}) := \left\{ a \in V^j_{\hat{\underline{u}}} : a \notin \bigcup_{w \in V^\ell :\, w <_\sigma v} \psi_w \right\}$$

and $A_v(\psi, \sigma, \underline{u}) := \bigcup_{j \in J} A_v^j(\psi, \sigma, \underline{u})$, where $J := [3] \setminus \{\ell\}$. We think of these vertices as being *'alive'* at the point just before processing v (when we are about to reveal ψ_v). By 'alive', we mean that it is still possible that ψ_v reveals that $a \in A_v^j(\psi, \sigma, \underline{u})$ is in a triangle with v. All other vertices $a \in V^j \setminus A_v^j(\psi, \sigma, \underline{u})$ are already embedded in triangles with vertices $w \in V^\ell$ which come before v in the ordering σ (or lie in $\underline{u}$ in which case we are forbidden from including them in a triangle in ψ).

Triangles with alive vertices

In this subsection, we will prove that most vertices $v \in V^\ell$ are in the expected number of triangles with the other two vertices still being 'alive'. This will be useful in the proof of the Entropy Lemma, Lemma 4.4.4.

Lemma 4.4.5. *For all $0 < \tau < d \leq 1$ and $L > 0$ there exists $\varepsilon > 0$ and $C > 0$ such that for all sufficiently large $n \in \mathbb{N}$ and for any $p \geq C(\log n)^{1/3} n^{-2/3}$, if Γ is an (ε, d)-regular tripartite graph with parts of size n then the following holds in Γ_p with probability at least $1 - n^{-L}$. If $t \in [n-1]$, $\ell \in [3]$, $\underline{u} = (u_1, \ldots, u_{\ell-1}) \in \mathcal{V}$, $u \in V^\ell$, $\psi \in \Psi^t_{\hat{\underline{u}}, \hat{u}}(\Gamma_p)$ and σ is an ordering of V^ℓ, then there are at most τn vertices $v \in V^\ell$ for which*

$$|\mathrm{Tr}_v(\Gamma_p) \cap E(\Gamma[A_v(\psi, \sigma, \underline{u})])| > (pd)^3 \prod_{j \in J} |A^j_v(\psi, \sigma, \underline{u})| + \tau (pd)^3 n^2, \tag{4.4.1}$$

where, as above, $J = [3] \setminus \{\ell\}$.

Proof. Choose $0 < \varepsilon, \frac{1}{C} \ll \varepsilon' \ll \tau, d, \frac{1}{L}$. Let $G \subseteq \Gamma$ be any subgraph satisfying

$$|K_3(G[X_1 \cup X_2 \cup X_3])| \leq (pd)^3 |X_1||X_2||X_3| + \varepsilon' p^3 n^3, \tag{4.4.2}$$

for all $X_1 \subseteq V^1$, $X_2 \subseteq V^2$, $X_3 \subseteq V^3$ and note that Γ_p is such a subgraph with probability at least $1 - n^{-L}$ by Lemma 4.2.1. We will show that G already satisfies the conclusion of Lemma 4.4.5. Let $\ell \in [3]$, $t \in [n-1]$, $\underline{u} = (u_1 \ldots, u_{\ell-1}) \in \mathcal{V}$, $u_\ell \in V^\ell$, $\psi \in \Psi^t_{\hat{\underline{u}}, \hat{u}}(G)$ and let σ be an ordering of V^ℓ. Enumerate $V^\ell = \{v^\ell_1, \ldots, v^\ell_n\}$ according to the ordering σ, that is, in such a way that $v^\ell_1 <_\sigma \cdots <_\sigma v^\ell_n$. Define $U \subseteq V^\ell$ to be the set of vertices satisfying (4.4.1). We will show that $|U| < \tau n$. We split V^ℓ into intervals as follows. Let $\tau' := \frac{\tau}{4}$, $K := \lceil \frac{1}{\tau'} \rceil$ and for $k = 1, \ldots, K$, let

$$W_k = \{v^\ell_i : 1 + (k-1) \cdot \tau' n \leq i < 1 + k \cdot \tau' n\}$$

and $U_k := U \cap W_k$. Fix some $k \in [K]$ and let $i_k := 1 + \lceil (k-1) \cdot \tau' n \rceil$ and $w_k := v^\ell_{i_k}$ (that is, w_k is the first vertex in W_k). Let $X_\ell = U_k$ and $X_j = A^j_{w_k}(\psi, \sigma, \underline{u})$ for $j \in J = [3] \setminus \{\ell\}$. It follows that, for any $z \in U_k$,

$$\begin{aligned}
|\mathrm{Tr}_z(G[\cup_{i \in [3]} X_i])| &\geq |\mathrm{Tr}_z(G[X_\ell \cup A_z(\psi, \sigma, \underline{u})])| \\
&\geq (pd)^3 \prod_{j \in J} |A^j_z(\psi, \sigma, \underline{u})| + \tau (pd)^3 n^2 \\
&\geq (pd)^3 \prod_{j \in J} \big(|X_j| - \tau' n\big) + \tau (pd)^3 n^2 \\
&\geq (pd)^3 \prod_{j \in J} |X_j| + \frac{\tau}{2} (pd)^3 n^2.
\end{aligned}$$

Here, the first inequality follows from the fact that $z >_\sigma w_k$ and thus $A_z(\psi, \sigma, \underline{u}) \subseteq A_{w_k}(\psi, \sigma, \underline{u})$ for every $z \in U_k$. The second inequality follows from the fact that $z \in U$ and the third from the fact that $|A_z^j(\psi, \sigma, \underline{u})| \geq |A_{w_k}^j(\psi, \sigma, \underline{u})| - \tau' n$ for all $z \in U_k$ since z and w_k are close in the ordering σ. By summing over all $z \in U_k$, it follows that

$$|K_3(G[X_1 \cup X_2 \cup X_3])| \geq (pd)^3 |X_1||X_2||X_3| + \frac{\tau}{2}(pd)^3 |X_\ell| n^2.$$

Combining this with (4.4.2) gives $|U_k| = |X_\ell| \leq \frac{2\varepsilon'}{\tau d^3} n < \frac{\tau^2}{8} n$, by our choice of constants. It follows that $|U| = \sum_{k=1}^K |U_k| < \tau n$, as claimed. □

Proof of the Entropy Lemma

Here, we will prove Lemma 4.4.4. The proof is quite long and so we will break it up into smaller claims along the way. Our proof works by contradiction. As $|\Psi^t_{\underline{\hat{u}},\hat{u}}(\Gamma_p)|$ is large, we know that $h(\psi^*)$ is large as ψ^* is chosen uniformly at random from $\Psi^t_{\underline{\hat{u}},\hat{u}}(\Gamma_p)$. Moreover, using the chain rule (Lemma 2.3.5), we can decompose $h(\psi^*)$ as the sum of local entropy values depending on the ψ^*_v. Now we assume that there are a significant number of bad vertices v for which the local entropy value $h(\psi^*_v | Y_v(\psi^*) = 1)$ is too small. We will then apply the chain rule (Lemma 2.3.5) using an ordering on the vertices which places these bad vertices at the beginning of the ordering. This has the effect that the shortcoming of their contribution to the overall entropy $h(\psi^*)$ is felt the most. We then upper bound the contribution of the entropy values at other (good) vertices, and hence conclude that the overall entropy $h(\psi^*)$ is too small, giving a contradiction. In order to achieve this upper bound, we rely on random properties of Γ_p and we have to split the entropy values further, delving into the average that outputs the entropy values and looking at individual embeddings.

Proof of Lemma 4.4.4. Choose $0 < \varepsilon, \frac{1}{C} \ll \tau \ll \eta \ll \delta \ll \gamma \ll \beta, d, \frac{1}{L}$. Fix Γ, $t \in \mathbb{N}$, $\ell \in [3]$, $\underline{u} = (u_1 \ldots, u_{\ell-1}) \in \mathcal{V}$, and $u \in V^\ell$ as in the statement of Lemma 4.4.4. Assume $G \subseteq \Gamma$ is a subgraph of Γ with $V(G) = V(\Gamma)$ which satisfies the following properties for all $\psi \in \Psi^t_{\underline{\hat{u}},\hat{u}}(G)$ and every ordering σ of V^ℓ.

(P.1) For all vertices $v \in V(G)$, we have

$$|\mathrm{Tr}_v(G)| \leq 10 p^3 n^2.$$

(P.2) There are at most τn vertices $v \in V^\ell$ for which

$$|\mathrm{Tr}_v(G) \cap E(G[A_v(\psi, \sigma, \underline{u})])| > (pd)^3 \prod_{j \in [3] \setminus \{\ell\}} |A_v^j(\psi, \sigma, \underline{u})| + \tau (pd)^3 n^2.$$

By Lemmas 4.4.5, 4.2.5 and a union bound, Γ_p satisfies these properties with probability at least $1 - n^{-L}$ and therefore it suffices to show that any G satisfying the above properties, satisfies the conclusion of Lemma 4.4.4.

To ease notation, let $\Psi := \Psi^t_{\underline{\hat{u}},\hat{u}}(G)$. Furthermore, let ψ^* be chosen uniformly from Ψ. We may assume that

$$|\Psi| \geq (1-\eta)^n (pd)^{3t} ((n-1)!_t)^{\ell} (n!_t)^{3-\ell},$$

as otherwise there is nothing to prove. In particular, by Lemma 2.3.1, we have

$$\begin{aligned} h(\psi^*) &\geq n\log(1-\eta) + 3t\log(pd) + 3\log(n!_t) - 3\log(n) \\ &\geq 3t\log(pd) + 3\log(n!_t) - \delta n, \end{aligned} \tag{4.4.3}$$

where we used $\eta \ll \delta$ and that n is large enough in the last step.

Assume for a contradiction that there are at least βn vertices $v \in V^\ell$ such that $h(\psi^*_v | Y_v(\psi^*) = 1) < H(n,p,d) - \beta$ and let $U \subset V^\ell$ be a set of these exceptional vertices of size $|U| = \gamma n$. We will derive an upper bound on $h(\psi^*)$ which contradicts (4.4.3). Recall that $I(\psi) = I^\ell(\psi) \subset V^\ell$ is the set of vertices which are isolated in $\psi(D_t)$. We begin as follows

$$\begin{aligned} h(\psi^*) &= h\left(\psi^*, \{\psi^*_v\}_{v\in V^\ell}, I(\psi^*)\right) && (4.4.4)\\ &= h\left(\{\psi^*_v\}_{v\in V^\ell}, I(\psi^*)\right) + h\left(\psi^* | \{\psi^*_v\}_{v\in V^\ell}, I(\psi^*)\right) && (4.4.5)\\ &\leq h\left(\{\psi^*_v\}_{v\in V^\ell}, I(\psi^*)\right) + \log(t!) && (4.4.6)\\ &= h\left(\{\psi^*_v\}_{v\in V^\ell} | I(\psi^*)\right) + h\left(I(\psi^*)\right) + \log(t!) && (4.4.7)\\ &\leq h\left(\{\psi^*_v\}_{v\in V^\ell} | I(\psi^*)\right) + \log(t!) + \log\left(\binom{n}{t}\right) && (4.4.8)\\ &= h\left(\{\psi^*_v\}_{v\in V^\ell} | I(\psi^*)\right) + \log(n!_t). && (4.4.9) \end{aligned}$$

Here, we used Lemma 2.3.3 in (4.4.4) and the chain rule (Lemma 2.3.5) in (4.4.5) and (4.4.7). In (4.4.6), we used Lemma 2.3.6 coupled with the fact that the set $\{\psi_v\}_{v\in V^\ell}$ completely determines the (unordered) subgraph $\psi(D_t)$. Indeed, note that there are $t!$ embeddings $\psi \in \Psi$ which map to the same subgraph $\psi(D_t)$, namely one for each choice of ordering of the triangles. Finally, in (4.4.8) we used Lemma 2.3.1.

Now, in order to estimate this sum further, we fix some ordering σ of V^ℓ in which the vertices in U come first, that is $w <_\sigma w'$ for all $w \in U$ and $w' \in V^\ell \setminus U$. We then reveal vertices in that order and apply the conditional chain rule (Lemma 2.3.8). That is,

$$h\left(\{\psi_v^*\}_{v\in V^\ell}|I(\psi^*)\right) = \sum_{v\in V^\ell} h\left(\psi_v^*|\{\psi_w^* : w <_\sigma v\}, I(\psi^*)\right)$$
$$\leq \sum_{v\in U} h\left(\psi_v^*|I(\psi^*)\right) + \sum_{v\in V^\ell\setminus U} h\left(\psi_v^*|\{\psi_w^* : w <_\sigma v\}, I(\psi^*)\right), \quad (4.4.10)$$

where we applied Lemma 2.3.4 in the second step. We treat the vertices in U separately to those in $V^\ell \setminus U$. To ease notation, we make the following definition. For $\psi \in \Psi$ and $v \in V^\ell$, we let $t_v(\psi)$ denote the number of vertices $w \in V^\ell$ such that $w <_\sigma v$ and $w \notin I(\psi)$. Let us first address the vertices in U.

Claim 4.4.6. *For all $v \in U$, we have that*

$$h(\psi_v^*|I(\psi^*)) \leq \frac{1}{|\Psi|}\sum_{\psi\in\Psi} Y_v(\psi)\left(\log\left((pd)^3(n-t_v(\psi))^2\right) - \frac{\beta}{2}\right).$$

<u>Proof of Claim:</u> Now, for each $v \in U$, we have

$$\begin{aligned} h(\psi_v^*|I(\psi^*)) &\leq h(\psi_v^*|Y_v(\psi^*)) \\ &= \mathbb{P}\left[Y_v(\psi^*) = 1\right] h(\psi_v^*|Y_v(\psi^*) = 1) + \mathbb{P}\left[Y_v(\psi^*) = 0\right] h(\psi_v^*|Y_v(\psi^*) = 0) \\ &\leq \mathbb{P}\left[Y_v(\psi^*) = 1\right] (H(n,p,d) - \beta) \\ &= \frac{1}{|\Psi|}\sum_{\psi\in\Psi} Y_v(\psi)\,(H(n,p,d) - \beta)\,. \end{aligned}$$

Here we used Lemma 2.3.4 and the fact that $I(\psi^*)$ determines $Y_v(\psi^*)$, the definition of conditional entropy (2.3.1), and the definition of U. Furthermore, we have $t_v(\psi) \leq \gamma n$ for all $v \in U$ and $\psi \in \Psi$ since U comes at the beginning of the ordering σ. Therefore,

$$\begin{aligned} \log\left((pd)^3(n-t_v(\psi))^2\right) &\geq \log\left((pd)^3(1-\gamma)^2 n^2\right) \\ &= H(n,p,d) + 2\log(1-\gamma) \\ &\geq H(n,p,d) - 4\gamma \\ &\geq H(n,p,d) - \frac{\beta}{2}. \end{aligned}$$

Combining this with our upper bound on $h(\psi_v^*|I(\psi^*))$ above completes the proof of the claim. ■

We will now deal with the vertices outside U. Given $v \in V^\ell$ and $\psi \in \Psi$, we write

$$h'(v,\psi) := h\left(\psi_v^*|I(\psi^*) = I(\psi), \{\psi_w^* = \psi_w\}_{w<_\sigma v}\right).$$

Claim 4.4.7. *The following is true for all $\psi \in \Psi$.*

(i) For all $v \in V^\ell$, we have

$$h'(v,\psi) \le \log\left((pd)^3(n-t_v(\psi))^2\right) + \log\left(\frac{10}{d^3}\right) + \log\left(\frac{n^2}{(n-t_v(\psi))^2}\right).$$

(ii) There exists a set $B(\psi) \subset V^\ell$ with $|B(\psi)| \le \delta n$, such that for all $v \in V^\ell \setminus B(\psi)$, we have

$$h'(v,\psi) \le \log\left((pd)^3(n-t_v(\psi))^2\right) + \delta.$$

Proof of Claim: The first inequality follows from **(P.1)** and Lemma 2.3.6. Indeed, for all $v \in V^\ell$, we have

$$\begin{aligned} h'(v,\psi) &\le \log\left(|\mathrm{Tr}_v(G)|\right) \\ &\le \log(10p^3n^2) \\ &= \log\left((pd)^3(n-t_v(\psi))^2\right) + \log\left(\frac{10}{d^3}\right) + \log\left(\frac{n^2}{(n-t_v(\psi))^2}\right). \end{aligned}$$

For the second inequality, we will use **(P.2)** in combination with Lemma 2.3.6. We have that for all but at most τn vertices,

$$\begin{aligned} h'(v,\psi) &\le \log\left(|\mathrm{Tr}_v(G) \cap E(G[A_v(\psi,\sigma,\underline{u})])|\right) \\ &\le \log\left((pd)^3\prod_{j\in J}|A_v^j(\psi,\sigma,\underline{u})| + \tau(pd)^3n^2\right) \\ &\le \log\left((pd)^3(n-t_v(\psi))^2 + \tau(pd)^3n^2\right). \end{aligned} \tag{4.4.11}$$

Observe that $t_v(\psi) \le \left(1-\frac{\delta}{2}\right)n$ for all but at most $\frac{\delta n}{2}$ vertices $v \in V^\ell$. In particular, we have

$$(n-t_v(\psi))^2 \ge \frac{\delta^2n^2}{4} \ge \frac{\delta^2}{4\tau}\cdot\tau n^2 \ge \frac{1}{\delta}\cdot\tau n^2,$$

for all but at most $\frac{\delta n}{2}$ vertices $v \in V^\ell$ (we used that $\tau \ll \delta$ here). Plugging this back into (4.4.11), we get

$$h'(v,\psi) \le \log\left((1+\delta)\cdot(pd)^3(n-t_v(\psi))^2\right) \le \delta + \log\left((pd)^3(n-t_v(\psi))^2\right)$$

for all but at most $\left(\tau+\frac{\delta}{2}\right)n \le \delta n$ vertices $v \in V^\ell$. ■

We will now use Claims 4.4.6 and 4.4.7 to finish the proof. Indeed, it follows from Claim 4.4.6 that

$$\sum_{v\in U} h(\psi_v^*|I(\psi^*)) \le \frac{1}{|\Psi|}\sum_{\psi\in\Psi}\sum_{v\in U} Y_v(\psi)\left(\log\left((pd)^3(n-t_v(\psi))^2\right) - \frac{\beta}{2}\right). \tag{4.4.12}$$

Furthermore, using Claim 4.4.7, the definition of conditional entropy (2.3.2) (and Lemma 2.3.6 to conclude that $h'(v,\psi)=0$ if $Y_v(\psi)=0$), we have

$$\begin{aligned}\sum_{v\in V^\ell\setminus U} h\left(\psi_v^*|\{\psi_w^*: w<_\sigma v\}, I(\psi^*)\right) &= \sum_{v\in V^\ell\setminus U}\frac{1}{|\Psi|}\sum_{\psi\in\Psi} Y_v(\psi)h'(v,\psi)\\ &\le \frac{1}{|\Psi|}\sum_{\psi\in\Psi}\left(\delta n+N_1(\psi)+\sum_{v\in V^\ell\setminus U} Y_v(\psi)\log\left((pd)^3(n-t_v(\psi))^2\right)\right), \end{aligned}\tag{4.4.13}$$

where

$$N_1(\psi)=\sum_{v\in B(\psi)} Y_v(\psi)\left(\log\left(\frac{10}{d^3}\right)+2\log\left(\frac{n}{n-t_v(\psi)}\right)\right).$$

Let now

$$M(\psi):=\sum_{v\in V^\ell} Y_v(\psi)\log\left((pd)^3(n-t_v(\psi))^2\right),\qquad \text{and}\qquad N_2(\psi):=\sum_{v\in U} Y_v(\psi)\cdot\frac{\beta}{2}.$$

Then, combining (4.4.10), (4.4.12) and (4.4.13), we get

$$h\left(\{\psi_v^*\}_{v\in V^\ell}|I(\psi^*)\right)\le\frac{1}{|\Psi|}\sum_{\psi\in\Psi}\left(M(\psi)+N_1(\psi)+\delta n-N_2(\psi)\right).\tag{4.4.14}$$

We will bound each of these terms one by one.

Claim 4.4.8. *For all $\psi\in\Psi$, we have that*

$$M(\psi)=3t\log(pd)+2\log(n!_t),\qquad N_1(\psi)\le\sqrt{\delta}n\qquad \textit{and}\qquad N_2(\psi)\ge\gamma^2 n.$$

Before we prove Claim 4.4.8, let us finish the main proof. Combining Claim 4.4.8 with (4.4.14), we get (using $\delta\ll\gamma$) that

$$\begin{aligned}h\left(\{\psi_v^*\}_{v\in V^\ell}|I(\psi^*)\right)&\le 3t\log(pd)+2\log(n!_t)+(\delta+\sqrt{\delta}-\gamma^2)n\\ &\le 3t\log(pd)+2\log(n!_t)-2\delta n.\end{aligned}$$

Plugging this back into (4.4.9), we get that $h(\psi^*)\le 3t\log(pd)+3\log(n!_t)-2\delta n$, contradicting (4.4.3). Hence it remains to prove Claim 4.4.8.

<u>Proof of Claim:</u> Let $\psi\in\Psi$ and observe that $\{t_v(\psi): v\in V^\ell\setminus I(\psi)\}=[t-1]_0$. Thus

$$\begin{aligned}M(\psi)&=\sum_{v\in V^\ell\setminus I(\psi)}\log\left((pd)^3(n-t_v(\psi))^2\right)\\ &=\sum_{k=0}^{t-1}\log\left((pd)^3(n-k)^2\right)=3t\log(pd)+2\log(n!_t).\end{aligned}$$

We now turn to bounding $N_1(\psi)$. We define $B' =: B(\psi) \setminus I(\psi)$ and observe that $|B'| \le |B(\psi)| \le \delta n$. Further, let $\mathbb{K} = \{t_v(\psi) : v \in B'\}$. Enumerate $\mathbb{K} = \{k_1, \ldots, k_{|B'|}\}$ so that $k_1 \ge \ldots \ge k_{|B'|}$ and observe that $k_i \le n - i$ for all $i \in [|B'|]$, by virtue of the the fact that $t_v(\psi) \le t \le n-1$ for all $v \in B'$ and, as $B' \cap I(\psi) = \emptyset$, we cannot have that $t_v(\psi) = t_{v'}(\psi)$ for $v \ne v' \in B'$. Hence

$$
\begin{aligned}
N_1(\psi) &= \sum_{v \in B'} Y_v(\psi) \left(\log \left(\frac{10}{d^3} \right) + 2 \log \left(\frac{n}{(n - t_v(\psi))} \right) \right) \\
&\le \delta n \log \left(\frac{10}{d^3} \right) + \sum_{\ell=1}^{\delta n} 2 \log \left(\frac{n}{\ell} \right) \\
&\le \delta n \log \left(\frac{10}{d^3} \right) + 2\delta n \log(n) - 2 \log((\delta n)!) \\
&\le \delta n \log \left(\frac{10}{d^3} \right) + 2\delta n \left(\log(n) - \log \left(\frac{\delta n}{e} \right) \right) \\
&\le \sqrt{\delta} n,
\end{aligned}
$$

where we used $(\delta n)! \ge \left(\frac{\delta n}{e}\right)^{\delta n}$ in the second to last line. Finally, let $U' = U \setminus I(\psi)$ and observe that, since $\eta \ll \gamma$, we have $|U'| \ge \frac{\gamma n}{2}$. Therefore,

$$
N_2(\psi) = \sum_{v \in U'} \frac{\beta}{2} \ge \gamma^2 n,
$$

as claimed. ■ □

4.4.3 Counting via comparison

In this subsection, we will prove Lemma 4.4.1 which, as we have shown in Section 4.4.1, completes the proof of the Local Distribution Lemma (Lemma 4.1.3) and hence our main technical theorem, Theorem II*. Elements of the proof of Lemma 4.4.1 were already sketched in Section 4.1 but before embarking on the details, we outline and reiterate some of the key ideas, ignoring the technicalities in order to elucidate the general proof scheme. For this discussion, we fix some (ε, d)-super-regular tripartite graph Γ, fix $\ell = 1$ and some $t \in [n]$ close to n. We also fix a vertex $u \in V^\ell$ which we think of as satisfying the "if" statement in Lemma 4.4.1 and some *typical* $v \in V^\ell$ which we aim to show satisfies the conclusion of Lemma 4.4.1. By typical, we mean that $v \in V^\ell$ satisfies certain conditions that we have shown whp almost all vertices in V^ℓ satisfy. For example, we can assume that ψ^*_v has large entropy, when ψ^* is a uniformly random embedding in $\Psi^t_{\hat{u}}(\Gamma_p)$, from the Entropy Lemma (Lemma 4.4.4).

Now our aim is to lower bound the number of embeddings ψ of D_t that leave v isolated and we concentrate on the subset of embeddings that place u in some triangle (as t is large we can expect that almost all embeddings do place u in a triangle). Refining further, we will only count embeddings that place u in a triangle with an edge that lies in some special set $F \subset E(\Gamma[V^2, V^3])$.

To define F, we begin by concentrating on edges in $\mathrm{Tr}_u(\Gamma) \cap \mathrm{Tr}_v(\Gamma)$. That is, any edge in F will form a triangle with *both* u and v. We then take F to be the edges in $\mathrm{Tr}_u(\Gamma) \cap \mathrm{Tr}_v(\Gamma)$ which appear in Γ_p. Note that we do *not* require that for an edge $w_2w_3 \in F$, any of the edges vw_i or uw_i with $i = 2, 3$, lie in Γ_p, just that they lie in Γ.

To motivate this definition, we consider a multi-stage revealing process as in Lemma 4.2.6. First, we reveal all edges of Γ_p that are *not* adjacent to u or v. The definition of F comes from the fact that at this point in the process, any edge in F has the *potential* to lie in $\mathrm{Tr}_u(\Gamma_p)$ and also $\mathrm{Tr}_v(\Gamma_p)$, depending on which random edges are adjacent to the vertices u and v. Now note that, in particular, if an edge $e = w_2w_3 \in F$ *does* end up in $\mathrm{Tr}_u(\Gamma_p)$, then it will contribute to embeddings that avoid v and place u in a triangle. We introduce a weight function ζ on F (we will in fact define it more generally on $E(\Gamma[V^2, V^3])$) which precisely counts the contribution to our desired lower bound, from embeddings which use the triangle $u \cup e = \{u, w_2, w_3\}$. That is, for all $w_2w_3 \in F$, we have that $\zeta(w_2w_3)$ encodes the number of embeddings of D_{t-1} (with $t-1$ triangles) in Γ, that avoid v *and* the vertices u, w_2, w_3. Therefore, as we can assume F is large (as v is typical, using Lemma 4.2.7), our desired conclusion will follow if we can lower bound the ζ values in (some subset of) F.

The central idea of the proof is that we *can* lower bound ζ values in F by reasoning about embeddings that place v in a triangle (and avoid u). Indeed, if we consider a uniformly random embedding $\psi^* \in \Psi^t_{\hat{u}}(\Gamma_p)$, as v is typical, we know from Lemma 4.4.4, that the random variable ψ^*_v, which encodes the triangle containing v in $\psi(D_t)$, has high entropy. Appealing to Lemma 2.3.9 then implies that the distribution of ψ^*_v in $\mathrm{Tr}_v(\Gamma_p)$ is close to uniform and hence for almost all edges $f \in \mathrm{Tr}_v(\Gamma_p)$, we have that $\mathbb{P}\left[\psi^*_v = f\right]$ is large (in that it is close to the average). Moreover, we have that $\mathbb{P}\left[\psi^*_v = f\right]$ is directly proportional to $\zeta(f)$ by the definition of ζ. Therefore, using Lemma 4.2.6 (2) (and observing that the ζ values do not depend on random edges adjacent to u or v), we can see that we must have a significant proportion of the edges in F having large ζ values. Indeed, if this were not the case, then it would be very unlikely that almost all edges in $\mathrm{Tr}_v(\Gamma_p)$ have large ζ values.

We can therefore conclude that there is some subset $F_L \subset F$ of half the edges in F such that $\zeta(f)$ is large for all $f \in F_L$. Finally, through another application of Lemma 4.2.6 (2), we can show that many edges in F_L end up in $\mathrm{Tr}_u(\Gamma_p)$ and therefore contribute to the lower bound on the number of embeddings that leave v isolated. We now give the full details of the proof.

Proof of Lemma 4.4.1. Choose $0 < \varepsilon, \frac{1}{C} \ll \varepsilon' \ll \eta \ll \beta' \ll \beta \ll \frac{1}{L} \ll \alpha, d, \frac{1}{K}$. Fix Γ, $p = p(n)$, $\ell \in [3]$, $(1-\eta)n \leq t < n$, $\underline{u} = (u_1 \ldots, u_{\ell-1}) \in \mathcal{V}$ and $u \in V^\ell$ as in the statement of Lemma 4.4.1. We define $J := [3] \setminus \{\ell\}$ and label the indices of J by $j_1, j_2 \in [3]$ so that $J = \{j_1, j_2\}$.

Now for a subgraph G of Γ, we will make some definitions relative to G and posit certain properties of G. Our proof will then proceed by first proving that any G satisfying all the properties, satisfies the desired conclusion of the lemma. After this we will show that whp we can take that Γ_p satisfies all the defined properties, which will complete the proof. Herein, we fix some subgraph G of Γ for the discussion. Our first property comes from the statement of the lemma.

(Q.1) We have

$$|\Psi^t_{\underline{\hat{u}},\hat{u}}(G)| \geq (1-\eta)^n (pd)^{3t}((n-1)!_t)^\ell (n!_t)^{3-\ell}.$$

For $v \in V^\ell$, we now define the set of edges which lie in G and in the common neighbourhood (with respect to Γ) of both u and v. In symbols,

$$F(v) := \mathrm{Tr}_u(\Gamma_{\underline{\hat{u}}}) \cap \mathrm{Tr}_v(\Gamma_{\underline{\hat{u}}}) \cap E(G) \subseteq V^{j_1}_{\underline{\hat{u}}} \times V^{j_2}_{\underline{\hat{u}}}. \tag{4.4.15}$$

Note that here (and throughout this proof), for convenience, we will think of edges in $e = \{y_1, y_2\} \in E(\Gamma[V^{j_1} \cup V^{j_2}])$ as *ordered* pairs $(y_1, y_2) \in V^{j_1} \times V^{j_2}$.

Now let ψ^* be chosen uniformly from $\Psi^t_{\underline{\hat{u}},\hat{u}}(G)$. We define the following subsets of V^ℓ, recalling the definition of $H(n,p,d)$ from Definition 4.4.2.

$$\begin{aligned}
Z_1 &:= \{v \in V^\ell : h(\psi^*_v | Y_v(\psi^*) = 1) \geq H(n,p,d) - \beta'\}, \\
Z_2 &:= \left\{v \in V^\ell : |\mathrm{Tr}_v(G)| = (1 \pm \varepsilon')(pd)^3 n^2\right\}, \\
Z_3 &:= \left\{v \in V^\ell : |F(v)| \geq \frac{d^5 p n^2}{4}\right\}, \\
Z &:= Z_1 \cap Z_2 \cap Z_3.
\end{aligned}$$

Our second property of G posits that Z is large.

(Q.2) If **(Q.1)** holds in G then

$$|Z| \geq (1-\alpha)n.$$

We now define the weight functions we will be interested in. For $v \in V^\ell \setminus \{u\}$ and $(w_1, w_2) \in V^{j_1}_{\underline{\hat{u}}} \times V^{j_2}_{\underline{\hat{u}}}$, define $\zeta_v(w_1, w_2)$ to be t times the number of labelled embeddings of D_{t-1} into $G_{\underline{\hat{u}},\hat{u},\hat{v}}$ in which both w_1 and w_2 are isolated vertices. That is,

$$\zeta_v(w_1, w_2) := t \cdot |\Psi^{(t-1)}_{\hat{w}_1,\hat{w}_2}(G_{\underline{\hat{u}},\hat{u},\hat{v}})|. \tag{4.4.16}$$

For our last property of G, we need a further definition. For $v \in V^\ell$, consider $F(v)$ as in (4.4.15). We split $F(v)$ in half according to the values of the weight function ζ_v. That is we partition $F(v)$

into $F_S(v)$ and $F_L(v)$ so that $\zeta(y_1, y_2) \le \zeta(z_1, z_2)$ for all $(y_1, y_2) \in F_S(v)$ and $(z_1, z_2) \in F_L(v)$, and $|F_S(v)| = |F_L(v)| \pm 1$. Our final property gives that G has many triangles containing u (resp. v) and the edges of $F_L(v)$ (resp. $F_S(v)$).

(Q.3) If $v \in Z$, then

$$|F'| \ge \frac{d^5 p^3 n^2}{20},$$

for both $F' = F_L(v) \cap \mathrm{Tr}_u(G)$ and $F' = F_S(v) \cap \mathrm{Tr}_v(G)$.

We now proceed by taking that G satisfies **(Q.2)** and **(Q.3)** and showing that it then satisfies the desired conclusion of the lemma. We will do this by proving that if G satisfies **(Q.1)** then every $v \in Z$ satisfies

$$|\Psi^t_{\underline{\hat{u}},\hat{v}}(G)| \ge \left(\frac{d}{10}\right)^2 \cdot |\Psi^t_{\underline{\hat{u}},\hat{u}}(G)|,$$

which in combination with the fact that G satisfies **(Q.2)**, gives what is needed. So let us fix some $v \in Z$. We define the following sets of embeddings.

$$\begin{aligned}
\Psi_{\hat{u}\hat{v}} &:= \Psi^t_{\underline{\hat{u}},\hat{u}}(G) \cap \Psi^t_{\underline{\hat{u}},\hat{v}}(G),\\
\Psi_{v\hat{u}} &:= \Psi^t_{\underline{\hat{u}},\hat{u}}(G) \setminus \Psi_{\hat{u}\hat{v}} \text{ and}\\
\Psi_{u\hat{v}} &:= \Psi^t_{\underline{\hat{u}},\hat{v}}(G) \setminus \Psi_{\hat{u}\hat{v}}.
\end{aligned}$$

In words, $\Psi_{\hat{u}\hat{v}}$ consists of those embeddings which leave both u and v isolated whilst embeddings in $\Psi_{v\hat{u}}$ leave u isolated but have v contained in a triangle, and vice versa for $\Psi_{u\hat{v}}$. Clearly, we have

$$\begin{aligned}
|\Psi^t_{\underline{\hat{u}},\hat{u}}(G)| &= |\Psi_{\hat{u}\hat{v}}| + |\Psi_{v\hat{u}}|, \text{ and}\\
|\Psi^t_{\underline{\hat{u}},\hat{v}}(G)| &= |\Psi_{\hat{u}\hat{v}}| + |\Psi_{u\hat{v}}|.
\end{aligned}$$

If $|\Psi_{\hat{u}\hat{v}}| \ge \left(\frac{d}{10}\right)^2 |\Psi^t_{\underline{\hat{u}},\hat{u}}(G)|$, we are done and so we may assume that

$$|\Psi_{v\hat{u}}| \ge \left(1 - \left(\frac{d}{10}\right)^2\right) |\Psi^t_{\underline{\hat{u}},\hat{u}}(G)| \ge \frac{1}{2} |\Psi^t_{\underline{\hat{u}},\hat{u}}(G)|. \tag{4.4.17}$$

In what remains, we will compare the sizes of $\Psi_{v\hat{u}}$ and $\Psi_{u\hat{v}}$. Let $\zeta = \zeta_v$ be the weight function as defined in (4.4.16). Observe that

$$\begin{aligned}
|\Psi_{v\hat{u}}| &= \sum_{(y_1,y_2)\in \mathrm{Tr}_v(G_{\underline{\hat{u}}})} \zeta(y_1, y_2), \text{ and}\\
|\Psi_{u\hat{v}}| &= \sum_{(y_1,y_2)\in \mathrm{Tr}_u(G_{\underline{\hat{u}}})} \zeta(y_1, y_2).
\end{aligned}$$

Recall that we took ψ^* to be a uniformly random embedding in $\Psi^t_{\underline{\hat{u}},\hat{u}}(G)$. Note that $\psi^*_v|Y_v(\psi^*) = 1$ is a random variable taking values in $S := \mathrm{Tr}_v\left(G_{\underline{\hat{u}}}\right)$ and the distribution of $\psi^*_v|Y_v(\psi^*) = 1$ is determined by ζ. That is, for all $(z_1, z_2) \in S$,

$$\mathbb{P}\left[\psi^*_v = (z_1, z_2)|Y_v(\psi^*)\right] = \frac{\zeta(z_1, z_2)}{\sum_{(y_1,y_2)\in S}\zeta(y_1, y_2)} = \frac{\zeta(z_1, z_2)}{|\Psi_{v\hat{u}}|}. \tag{4.4.18}$$

Moreover, as $v \in Z \subseteq Z_2$, we have that $\log(|S|) \leq \log(1 + \varepsilon') + H(n, p, d)$ and therefore, using also that $v \in Z \subseteq Z_1$, we can apply Lemma 2.3.9 (with $2\beta'$ replacing β') to obtain some set $W^* \subseteq S = \mathrm{Tr}_v\left(G_{\underline{\hat{u}}}\right)$ with the following properties (using (4.4.18) to unpack the conclusions here):

(i) $\displaystyle\sum_{(w_1,w_2)\in W^*}\zeta(w_1, w_2) \geq (1 - \beta)|\Psi_{v\hat{u}}|$;

(ii) There exists some value $\bar{\zeta}$ such that for each $(w_1, w_2) \in W^*$, we have that

$$\zeta(w_1, w_2) = (1 \pm \beta)\bar{\zeta};$$

(iii) We have $(1 - \beta)|S| \leq |W^*| \leq |S|$.

Therefore we can estimate the size of $\Psi_{v\hat{u}}$ using (i) to (iii) in that order, as follows:

$$\begin{aligned}
|\Psi_{v\hat{u}}| &\leq \left(\frac{1}{1-\beta}\right)\sum_{(w_1,w_2)\in W^*}\zeta(w_1, w_2)\\
&\leq \left(\frac{1+\beta}{1-\beta}\right)|W^*|\bar{\zeta}\\
&\leq \left(\frac{1+\beta}{1-\beta}\right)|S|\bar{\zeta} \leq 2\bar{\zeta}(pd)^3n^2.
\end{aligned} \tag{4.4.19}$$

In the last inequality, we used that $|S| = |\mathrm{Tr}_v\left(G_{\underline{\hat{u}}}\right)| \leq (1 + \varepsilon')(pd)^3n^2$ since $v \in Z \subseteq Z_2$.

We are now going to lower bound $|\Psi_{u\hat{v}}|$ in a similar way. However, the entropy argument above only shows that ζ is 'well-behaved' on $S = \mathrm{Tr}_v\left(G_{\underline{\hat{u}}}\right)$ but nothing about $\mathrm{Tr}_u\left(G_{\underline{\hat{u}}}\right)$. Using **(Q.3)** though, we can infer though that ζ is 'well-behaved' on a large part of $F(v)$, as defined in (4.4.15). Recall also our definitions of $F_L(v)$ and $F_S(v)$.

Claim 4.4.9. *We have $\zeta(y_1, y_2) \geq (1 - \beta)\bar{\zeta}$ for all $(y_1, y_2) \in F_L(v)$.*

<u>Proof of Claim:</u>

By **(Q.3)**, we have that

$$|\mathrm{Tr}_v\left(G_{\underline{\hat{u}}}\right) \cap F_S(v)| \geq \frac{d^5p^3n^2}{20},$$

noting that $\mathrm{Tr}_v\left(G_{\underline{\hat{u}}}\right) \cap F_S(v) = \mathrm{Tr}_v\left(G\right) \cap F_S(v)$ due to the fact that $F_S(v) \subset E(\Gamma_{\underline{\hat{u}}})$. Furthermore, it follows from (iii) and the fact that $v \in Z \subseteq Z_2$, that

$$|\mathrm{Tr}_v\left(G_{\underline{\hat{u}}}\right) \setminus W^*| \leq \beta |\mathrm{Tr}_v\left(G_{\underline{\hat{u}}}\right)| \leq 2\beta (pd)^3 n^2.$$

Hence, as $\beta \ll d$, we can conclude that $W^* \cap F_S(v) \neq \emptyset$ and so

$$(1-\beta)\bar{\zeta} \leq \min_{(y_1,y_2)\in W^*} \zeta(y_1,y_2) \leq \max_{(y_1,y_2)\in F_S(v)} \zeta(y_1,y_2) \leq \min_{(y_1,y_2)\in F_L(v)} \zeta(y_1,y_2),$$

using (ii) in the first inequality. ∎

We now appeal to **(Q.3)** to lower bound the size of $|\Psi_{u\hat{v}}|$ as follows:

$$\begin{aligned}
|\Psi_{u\hat{v}}| &= \sum_{(y_1,y_2)\in \mathrm{Tr}_u\left(G_{\underline{\hat{u}}}\right)} \zeta(y_1,y_2) \\
&\geq \sum_{(y_1,y_2)\in \mathrm{Tr}_u\left(G_{\underline{\hat{u}}}\right)\cap F_L(v)} \zeta(y_1,y_2) \\
&\geq (1-\beta)\bar{\zeta}|\mathrm{Tr}_u\left(G_{\underline{\hat{u}}}\right) \cap F_L(v)| \\
&\geq \frac{\bar{\zeta} d^5 p^3 n^2}{25},
\end{aligned} \tag{4.4.20}$$

where we used Claim 4.4.9. Putting (4.4.17), (4.4.19) and (4.4.20) together, we get that

$$|\Psi^t_{\underline{\hat{u}},\hat{v}}(G)| \geq |\Psi_{u\hat{v}}| \geq \frac{\bar{\zeta} d^5 p^3 n^2}{25} \geq \frac{d^2}{50}|\Psi_{v\hat{u}}| \geq \frac{d^2}{100}|\Psi^t_{\underline{\hat{u}},\hat{u}}(G)|,$$

as required.

It remains to verify that for $G = \Gamma_p$ the statements in **(Q.2)** and **(Q.3)** hold with probability at least $1 - n^{-K}$. We start with **(Q.2)**, which follows simply from Corollary 4.2.4 and Lemmas 4.2.7 and 4.4.4. Indeed, from those results (using that $\frac{1}{L} \ll \frac{1}{K}$) and a union bound, with probability at least $1 - n^{-2K}$, we have that $|Z_2| \geq (1-\varepsilon')n$, $|Z_3| \geq (1-2\varepsilon)n$ and if **(Q.1)** holds in $G = \Gamma_p$ then $|Z_1| \geq (1-\beta')n$. It then follows easily by our choice of constants that the statement of **(Q.2)** holds in $G = \Gamma_p$ with probability at least $1 - n^{-2K}$.

For **(Q.3)**, we will appeal to Lemma 4.2.6 (2). Note that for a fixed $v \in V^\ell \setminus \{u\}$ the value of $\zeta_v(w_1,w_2)$ for $(w_1,w_2) \in V^{j_1}_{\underline{\hat{u}}} \times V^{j_2}_{\underline{\hat{u}}}$ does not depend on the random status of any of the edges containing u or v. Indeed, our definition of ζ_v counts only embeddings that leave both u and v isolated. We also have that the random set of edges $F(v)$, as defined in (4.4.15), is independent of the random status of any edges adjacent to u or v. Consequently, in the language of Lemma 4.2.6, we have that the random sets of edges $F_L(v)$ and $F_S(v)$ are *determined* by $(\Gamma_{\hat{u}})_p$ (resp. $(\Gamma_{\hat{v}})_p$). Therefore, for a fixed $v \in V^\ell$, two applications of Lemma 4.2.6 (2) (once for u and $F_L(v)$ and once for v and $F_S(v)$) give that with probability at least $1 - n^{-(2K+1)}$, we have that **(Q.3)** holds for v. Here we used that $v \in Z \subseteq Z_3$ implies that $|F_L(v)|, |F_S(v)| \geq \frac{d^5 pn^2}{10}$. Taking a union bound over all $v \in V^\ell$, we have that **(Q.3)** holds in $G = \Gamma_p$ for all $v \in V^\ell$, with probability at least $1 - n^{-2K}$. A final union bound gives that with probability at least $1 - n^{-K}$, both **(Q.2)** and **(Q.3)** hold in $G = \Gamma_p$ which completes the proof. □

4.5 Stability for a fractional version of the Hajnal–Szemerédi theorem

In this section we discuss some fractional variants of the Hajnal–Szemerédi theorem (Theorem 1.1.2). We will use the results here in our proof reducing Theorem II to Theorem II* in Section 4.7. The starting point is to relax the notion of a K_k-factor to that of a *fractional K_k-factor* in an analogous fashion to the fractional perfect matchings introduced in Section 2.6. That is, for a graph G, a fractional K_k-factor in G is a weighting $\omega : K_k(G) \to \mathbb{R}_{\geq 0}$ such that $\sum_{K \in K_k(G,u)} \omega(K) = 1$ for all $u \in V(G)$. If all cliques $K \in K_k(G)$ are assigned weights in $\{0, 1\}$, we recover the notion of a K_k-factor and so the definition of a fractional K_k-factor is more general. However, from an extremal point of view, the same minimum degree condition is needed to force both objects. Indeed, focusing on the case when $n \in k\mathbb{N}$, the Hajnal–Szemerédi theorem (Theorem 1.1.2) gives that graphs G with n vertices and minimum degree at least $\left(\frac{k-1}{k}\right)n$ have K_k-factors and hence fractional K_k-factors whilst the same construction proving the tightness of Theorem 1.1.2 can be used to show tightness for fractional factors, as we now show. Take a graph G to be a complete graph with $n \in k\mathbb{N}$ vertices with a clique of size $\frac{n}{k} + 1$ removed to leave an independent set of vertices I. Therefore G has minimum degree $\delta(G) = \left(\frac{k-1}{k}\right)n - 1$ and suppose for a contradiction that G has a fractional K_k-factor given by a weight function $\omega : K_k(G) \to \mathbb{R}_{\geq 0}$. Then we have that $\sum_{K \in K_k(G,u)} \omega(K) = 1$ for all $u \in V(G)$ and note that for $w \neq w' \in I$, we have that $K_k(G, w) \cap K_k(G, w') = \emptyset$ as I is an independent set. Therefore

$$\sum_{K \in K_k(G)} \omega(K) \geq \sum_{w \in I} \sum_{K \in K_k(G,w)} \omega(K) \geq |I| = \frac{n}{k} + 1$$

but we also have that

$$\sum_{K \in K_k(G)} \omega(K) = \frac{1}{k} \sum_{u \in V(G)} \sum_{K \in K_k(G,u)} \omega(K) = \frac{n}{k},$$

a contradiction. The results of this section, which may be of independent interest, will give stability for this phenomenon, showing that if we avoid the construction detailed above (and other similar constructions), by imposing that $\alpha(G) \leq \left(\frac{1}{k} - \eta\right)n$ for some $\eta > 0$, then a weaker minimum degree condition of $\delta(G) \geq \left(\frac{k-1}{k} - \gamma\right)n$ for some $\gamma = \gamma(\eta) > 0$, suffices to force a fractional K_k-factor. Here we will only focus on the cases $k = 2$ and $k = 3$, as these are all we will need. However all results shown in this section hold more generally for all $k \geq 2$, as shown in [5].

As with our results on fractional perfect matchings in hypergraphs in Section 2.6, here we will also use that the existence of a fractional K_k-factor can be encoded by a linear program whose dual is a covering linear program which assigns weights to vertices such that every clique is sufficiently 'covered'. The duality theorem from linear programming will then be used to transfer between the two settings.

Theorem 4.5.1 (stability for fractional Hajnal–Szemerédi)**.** *For every $\eta > 0$ and $k \in \{2, 3\}$, there is some $\gamma > 0$ such that the following is true for all $n \in \mathbb{N}$. Let G be an n-vertex graph with $\delta(G) \geq \left(\frac{k-1}{k} - \gamma\right)n$ and $\alpha(G) < \left(\frac{1}{k} - \eta\right)n$. Then G contains a fractional K_k-factor.*

Proof. We begin by making some observations. First, note that the existence of a fractional K_k-factor in a graph G is the same as saying that there exists a *perfect fractional matching* in the k-uniform hypergraph $\mathcal{H} = \mathcal{H}(G)$ with vertex set $V(\mathcal{H}) = V(G)$ and the edge set of $\mathcal{H}$ corresponding to the set of k-vertex cliques in G. Note that for $k = 2$, $\mathcal{H}(G) = G$. Now, as discussed in Section 2.6, we can encode the existence of a perfect fractional matching (and hence a fractional K_k-factor in G) as a linear program which calculates the fractional matching number $\vartheta^*(\mathcal{H})$. Indeed from Proposition 2.6.1 (1), we have that G has a K_k-matching if and only if $\vartheta^*(\mathcal{H}) = \frac{n}{k}$. Moreover, taking the dual of the linear program that calculates $\vartheta^*(\mathcal{H})$ gives a *covering* linear program in which we place non-negative weights on the vertices of G, with minimum sum, subject to the constraint that the total weight on the vertices of each element of $K_k(G) = E(\mathcal{H})$ is at least 1. The optimal outcome of the covering linear program, we call the fractional cover number of $\mathcal{H}$, denoted $\tau^*(\mathcal{H})$, and the strong duality theorem from linear programming implies that to prove the existence of a K_k-factor, it suffices to prove that $\tau^*(\mathcal{H}) \geq \frac{n}{k}$. Unpacking this and returning to the graph perspective, we have that Theorem 4.5.1 follows if we prove the following claim.

Claim 4.5.2. *For every $\eta > 0$ and $k \in \{2, 3\}$, there is some $\gamma > 0$ such that the following is true for all $n \in \mathbb{N}$. Let G be an n-vertex graph with $\delta(G) \geq \left(\frac{k-1}{k} - \gamma\right)n$ and $\alpha(G) < \left(\frac{1}{k} - \eta\right)n$. Suppose $c : V(G) \to \mathbb{R}_{\geq 0}$ is any weight function such that for each $Q \in K_k(G)$ we have $\sum_{v \in Q} c(v) \geq 1$. Then $\sum_{v \in V(G)} c(v) \geq \frac{n}{k}$.*

<u>Proof of Claim:</u> We begin by addressing the $k = 2$ case and prove that for this we can simply take $\gamma = \frac{\eta}{4}$. So fix a graph G as in the claim. Let $g : V(G) \to \mathbb{R}_{\geq 0}$ be an optimal fractional cover of G (see Section 2.6 for the definition). It suffices to show that $\sum_{v \in V(G)} g(v) \geq \frac{n}{2}$ as we have that $\sum_{v \in V(G)} c(v) \geq \tau^*(G) = |g|$ for all weight functions $c : V(G) \to \mathbb{R}_{\geq 0}$ such that for each $e \in E(G)$ we have $\sum_{v \in e} c(v) \geq 1$. Now by Proposition 2.6.2, if $g(v) > 0$ for

all $v \in V(G)$ then $|g| \geq \frac{n}{2}$ and we are done. So assume this is not the case and let the vertices of G be $v_1, \dots, v_n$ in order of decreasing weight, i.e. $g(v_i) \geq g(v_j)$ if $i \leq j$.

Therefore $g(v_n) = 0$ and since v_n has at least $\left(\frac{1}{2} - \frac{\eta}{4}\right)n$ neighbours, we see that for each i such that $v_i v_n \in E(G)$, we have $g(v_i) \geq 1$. In particular, $g(v_i) \geq 1$ for each $i \leq \left(\frac{1}{2} - \frac{\eta}{4}\right)n$. Furthermore, the vertices $\left\{v_i : i \geq \left(\frac{1}{2} + \frac{\eta}{2}\right)n\right\}$ do not form an independent set, so there is an edge within this set; at least one endpoint of this edge has weight at least $\frac{1}{2}$, and in particular each vertex v_i with $\left(\frac{1}{2} - \frac{\eta}{4}\right)n < i < \left(\frac{1}{2} + \frac{\eta}{2}\right)n$ has weight at least $\frac{1}{2}$. Summing the weights $g(v_i)$, we obtain that $|g|$ is at least $\frac{n}{2}$ as desired.

Next, we prove the claim for $k = 3$. So let $\eta > 0$ and choose $0 < \gamma \ll \eta$. Fix some graph G as in the claim and let $g : V(G) \to \mathbb{R}_{\geq 0}$ be an optimal fractional cover of the 3-uniform hypergraph $\mathscr{H} = \mathscr{H}(G)$ defined at the beginning of the proof. As in the $k = 2$ case, it suffices to prove the claim for the weight function $c = g$. Again, we order the vertices of G as $v_1, \dots, v_n$ so that $g(v_i) \geq g(v_j)$ if $i \leq j$, and by Proposition 2.6.2, we can assume that $g(v_n) = 0$. For convenience of notation, we relabel v_n as $u_1 = v_n$. Now from the minimum degree condition, at least $\left(\frac{2}{3} - \gamma\right)n$ vertices are neighbours of u_1. Consider the last $\left(\frac{1}{3} - 5\gamma\right)n$ of these neighbours according to the order on the vertices. Since $\gamma \ll \eta$ they do not form an independent set and so contain an edge u_2u_3. Since $\sum_{i=1}^{k} g(u_i) \geq 1$, and $g(u_1) = 0$, one of these vertices has weight at least $\frac{1}{2}$. In particular, $g(v_i) \geq \frac{1}{2}$ whenever $i \leq \left(\frac{1}{3} + 4\gamma\right)n$.

Now let $g^* := g(v_{(1/3-2\gamma)n})$. If $g^* \geq 1$ then we have

$$\sum_{i \in [n]} g(v_i) \geq \left(\tfrac{1}{3} - 2\gamma\right)n + \tfrac{1}{2} \cdot 6\gamma n > \tfrac{n}{3}$$

and we are done; so we can assume $g^* < 1$. Next, we let G' denote the subgraph of G induced by vertices v_i with $i \geq \left(\frac{1}{3} + 4\gamma\right)n$ and let $\tilde{n} := v(G') = \left(\frac{2}{3} - 4\gamma\right)n$. If e is any edge in G', then e has a common neighbourhood in G of size at least $\left(\frac{1}{3} - 2\gamma\right)n$, and so in particular e extends to a copy of K_3 in G by adding a vertex whose weight is at most g^*. Thus the function $c'(u) := \frac{1}{1-g^*} g(u)$ on $V(G')$ is a weight function on $V(G')$ taking values in $\mathbb{R}_{\geq 0}$ and such that $\sum_{u \in e} c'(u) \geq 1$ for each $e \in E(G')$. Furthermore every vertex in G' has at most $\left(\frac{1}{3} + \gamma\right)n$ non-neighbours in G, at most all of which are in G', so the minimum degree of G' is at least $\tilde{n} - \left(\frac{1}{3} + \gamma\right)n \geq \left(\frac{1}{2} - \frac{\eta}{8}\right)\tilde{n}$ as $\gamma \ll \eta$. We also have that $\alpha(G') \leq \left(\frac{1}{3} - \eta\right)n \leq \left(\frac{1}{2} - \frac{\eta}{2}\right)\tilde{n}$.

We are therefore in a position to apply the the case $k = 2$ of Claim 4.5.2 to G', with $\frac{\eta}{2}$ replacing η (recalling that taking γ as $\frac{\eta}{8}$ suffices in this case). We conclude that

$$\sum_{u \in V(G')} c'(u) \geq \frac{\tilde{n}}{2} = \left(\tfrac{1}{3} - 2\gamma\right)n$$

and so

$$\sum_{i\in[n]} g(v_i) \geq g^*\big(\tfrac{1}{3} - 2\gamma\big)n + \tfrac{1}{2}\cdot 6\gamma n + (1-g^*)\cdot\big(\tfrac{1}{3} - 2\gamma\big)n > \frac{n}{3},$$

as desired. ■ □

We will in fact need some more general results. First we want to be able to set (potentially different but close to uniform) weights $\lambda(u)$ for each $u \in V$ and obtain a weighting $\omega : K_k(G) \to \mathbb{R}_{\geq 0}$ such that $\sum_{K\in K_k(G,u)} \omega(K) = \lambda(u)$ for all $u \in V(G)$. The case of fractional K_k-factors corresponds to setting $\lambda(u) = 1$ for all $u \in V(G)$.

Corollary 4.5.3. *For every $\eta > 0$ and $k \in \{2,3\}$, there is some $\gamma > 0$ such that the following is true for all $n \in \mathbb{N}$. Let G be an n-vertex graph with $\delta(G) \geq \big(\frac{k-1}{k} - \gamma\big)n$ and $\alpha(G) < \big(\frac{1}{k} - \eta\big)n$. Let $\lambda : V(G) \to \mathbb{N}$ be a weight function with $\lambda(u) = (1\pm\gamma)\frac{1}{n}\sum_{v\in V(G)}\lambda(v)$ for all $u \in V(G)$. Then there is a weight function $\omega : K_k(G) \to \mathbb{R}_{\geq 0}$ such that $\sum_{K\in K_k(G,u)}\omega(K) = \lambda(u)$ for all $u \in V(G)$.*

Proof. Fix some $k \in \{2,3\}$ and $\eta > 0$. Choose $0 \ll \gamma \ll \gamma' \ll \eta$. Now let G and λ be as in the statement of the corollary. We define an auxiliary graph H by blowing-up every $v \in V(G)$ to an independent set of size $\lambda(v)$ (that is, every edge is replaced by a complete bipartite graph). Then, with $N := v(H) = \sum_{v\in V(G)}\lambda(v)$, we have $\delta(H) \geq \big(\frac{k}{k-1} - \gamma'\big)N$ and $\alpha(H) \leq \big(\frac{1}{k} - \frac{\eta}{2}\big)N$. Hence, we can apply Theorem 4.5.1 to H and obtain a weight function $\omega_H : K_k(H) \to \mathbb{R}_{\geq 0}$ such that $\sum_{K'\in K_k(G,x)}\omega_H(K') = 1$ for all $x \in V(H)$. We define $\omega : K_k(G) \to \mathbb{R}_{\geq 0}$ by $\omega(K) = \sum_{K'\in K_k(H[K])}\omega_H(K')$, where $H[K]$ is the subgraph of H induced by the blown-up vertices of K. This weight function ω satisfies the desired conditions. □

Next, we extend yet further to guarantee an integer-valued weight-function $\omega : K_k(G) \to \mathbb{N}$. In order for this to work, we need that our function λ assigns each vertex a sufficiently large weight. In applications this will be guaranteed as our weights λ will be proportional to the number of vertices n of a host graph but Theorem 4.5.4 will actually be applied to the reduced graph R after applying the regularity lemma to the host graph and hence the number of vertices of R (the parameter n in Theorem 4.5.4) will be bounded by some constant.

Theorem 4.5.4 (stability for fractional Hajnal–Szemerédi with integer weights)**.** *For every $\eta > 0$ and $k \in \{2,3\}$, there is some $\gamma > 0$ such that the following is true for all sufficiently large $n \in \mathbb{N}$. Let G be a connected n-vertex graph with $\delta(G) \geq \big(\frac{k-1}{k} - \gamma\big)n$ and $\alpha(G) < \big(\frac{1}{k} - \eta\big)n$. Let $\lambda : V(G) \to \mathbb{N}$ be a weight function such that $\lambda(u) = \big(1 \pm \frac{\gamma}{2}\big)\frac{1}{n}\sum_{v\in V(G)}\lambda(v)$, $\lambda(u) \geq n^{2k}$ for all $u \in V(G)$ and k divides $\sum_{v\in V(G)}\lambda(v)$. Then there is a weight function $\omega : K_k(G) \to \mathbb{N}$ such that $\sum_{K\in K_k(G,u)}\omega(K) = \lambda(u)$ for all $u \in V(G)$.*

Note that for $k = 3$, the requirement that G is connected is readily implied by the minimum degree condition in this theorem.

Proof of Theorem 4.5.4. Fix some $k \in \{2, 3\}$ and $\eta > 0$. Choose $\gamma \ll \eta$ and suppose G and λ are given as in the statement. We will construct ω in three steps. Define $\lambda' : V(G) \to \mathbb{N}$ by $\lambda'(u) = \lambda(u) - k|K_k(G,u)|n^k \geq 0$. By Corollary 4.5.3, there is some weight function $\omega' : K_k(G) \to \mathbb{R}_{\geq 0}$ such that $\sum_{K \in K_k(G,u)} \omega'(K) = \lambda'(u)$ for all $u \in V(G)$. We define $\omega'' : V(G) \to \mathbb{N}$ such that, for each $K \in K_k(G)$,

(i) $\omega''(K) \in \left\{\lfloor \omega'(K) + kn^k \rfloor, \lceil \omega'(K) + kn^k \rceil\right\}$, and

(ii) $k \sum_{K \in K_k(G)} \omega''(K) = \sum_{v \in V(G)} \lambda(v)$.

Note that this is possible since by construction the unrounded sum satisfies (ii) and since k divides $\sum_{v \in V(G)} \lambda(v)$. Furthermore, for each $u \in V(G)$, we have $\sum_{K \in K_k(G,u)} \omega''(K) = \lambda(u) \pm n^{k-1}$ (since the unrounded sum would be exactly correct and $|K_k(G,u)| \leq n^{k-1}$).

Finally, we obtain ω from ω'' via the following iterative process. As long as possible, we identify pairs $u, v \in V(G)$ such that $\sum_{K \in K_k(G,u)} \omega''(K) > \lambda(u)$ and $\sum_{K \in K_k(G,v)} \omega''(K) < \lambda(v)$. If $k = 3$, we claim that there is an edge in the common neighbourhood of u and v. Indeed, since $\delta(G) \geq \left(\frac{2}{3} - \gamma\right)n$, we have that the common neighbourhood of u, v has size at least $\left(\frac{1}{3} - 2\gamma\right)n > \left(\frac{1}{3} - \eta\right)n$ and so contains an edge w_1w_2. Let $K_u = \{u, w_1, w_2\}$ and $K_v = \{v, w_1, w_2\}$, and decrease the weight of K_u by 1 and increase the weight of K_v by 1. If $k = 2$, we do the following: Since $\alpha(G) < \frac{n}{2}$, G is not bipartite and hence contains an odd cycle. Since G is connected, this implies that there is a walk from u to v of even length (even number of edges). We take a shortest such walk (in terms of edges) and note that every edge is traversed at most twice by this walk. We decrease the weight of the edge at u and then alternate increasing and decreasing the weight of the edges along the walk. Note that in both cases the total weight at u decreases by 1 and the total weight at v increases by 1, while the total weight at any other vertex remains unchanged.

Note that $\sum_{v \in V(G)} |\lambda(v) - \sum_{K \in K_k(v,G)} \omega(K)|$ decreases by 2 in every step. So this process finishes after at most n^k steps. Clearly, at this time, we have $\sum_{K \in K_k(v,G)} \omega(K) = \lambda(v)$ for all $v \in V(G)$ and $\omega(K) \geq \omega''(K) - 2n^k \geq 0$ for all $K \in K_k(G)$, completing the proof. □

4.6 Triangle matchings

In this section, we detail some probabilistic lemmas which allow us to find a *triangle matching*, that is, a collection of vertex-disjoint triangles, in various settings. These will be useful in proving Theorem II in Section 4.7. Recall that the *size* of a triangle matching is the number of triangles it contains and we write $V(\mathcal{T})$ for the set of vertices covered by a triangle matching $\mathcal{T}$. The first lemma allows us to find a triangle matching in G_p if G contains many triangles. We refer the reader to the Notation Section for any notational conventions (for example, the definition of $G[X_1, X_2, X_3]$).

Lemma 4.6.1. *For all $\mu > 0$ there exists $C > 0$ such that the following holds. Let $k, n \in \mathbb{N}$, $p \geq Cn^{-2/3}$ and let G be an n-vertex graph.*

(i) Assume that for every set $X \subseteq V(G)$ with $|X| \geq 3k$, $G[X]$ contains at least μn^3 triangles. Then, whp, G_p contains a triangle matching of size at least $\frac{n}{3} - k$.

(ii) Assume that $n_0 \geq k$ and $V(G) = V_1 \cup V_2 \cup V_3$ is a partition into sets of size at least n_0 so that for every $X_i \subseteq V_i$ with $|X_i| \geq k$ for all $i \in [3]$, $G[X_1, X_2, X_3]$ contains at least μn^3 triangles. Then, whp, G_p contains triangle matching of size at least $n_0 - k$.

Proof. Let $\mu > 0$ and set $C = 50\mu^{-2}$. Let p, k, n, G be given as in the statement. We will deduce the lemma from the following claim.

Claim 4.6.2. *The following holds whp for all $X \subseteq V(G)$. If $G[X]$ contains at least $t \geq \mu n^3$ copies of K_3, then the number of triangles in $G_p[X]$ is at least $\frac{1}{2}p^3 t$.*

Proof of Claim: This is a straightforward application of Janson's inequality (Lemma 2.1.3) and the union bound. Note that the total number of choices of X is at most 2^n. Fix one such choice. The expected number of triangles in $G_p[X]$ is $p^3 t \geq \mu p^3 n^3$, and we have $\bar{\Delta} \leq 2\max(p^5 n^4, p^3 n^3)$. Hence Janson's inequality tells us that the probability of having less than $\frac{1}{2}p^3 t$ triangles is at most

$$\exp\left(-\frac{\mu^2 p^6 n^6}{16\max(p^5 n^4, p^3 n^3)}\right) \leq \exp\left(-\frac{\mu^2}{16}\min(pn^2, p^3 n^3)\right) \leq \exp\left(-\frac{C\mu^2}{16}n\right)$$

and by our choice of C and the union bound, the claim follows. ■

We only prove (i) as (ii) is similar. Suppose that $\mathcal{T}$ is a maximal collection of vertex-disjoint triangles with $|\mathcal{T}| < \frac{n}{3} - k$. Then $X := V(G) \setminus V(\mathcal{T})$ has size at least $3k$ but $G_p[X]$ does not contain a triangle. Thus, the claimed result follows from the above claim. □

The next lemma allows us to find triangles which cover a given small set of vertices, using edges in specified places.

Lemma 4.6.3. *For any $0 < \mu < \frac{1}{100}$, there exists $C > 0$ such that the following holds for every $n \in \mathbb{N}$ and $p \geq Cn^{-2/3}(\log n)^{1/3}$. Let G be an n-vertex graph, and let $v_1, \ldots, v_\ell \in V(G)$ be distinct vertices with $\ell \leq \mu^2 n$. For each $i \in [\ell]$, let $E_i \subseteq \mathrm{Tr}_{v_i}(G)$ be a set of edges that form a triangle with v_i such that $|E_i| \geq \mu n^2$. Moreover, suppose $A_1, \ldots, A_t \subset V(G) \setminus \{v_1, \ldots, v_\ell\}$ are disjoint sets for some $t \in \mathbb{N}$. Then, whp, there is a triangle matching $\mathcal{T} = \{T_1, \ldots, T_\ell\}$ in G_p*

such that for each $i \in [\ell]$ the triangle T_i consists of v_i joined to an edge of E_i and $|A_k \cap V(\mathcal{T})| \le 12\mu|A_k| + 1$ for all $k \in [t]$.

Proof. Given $0 < \mu < \frac{1}{100}$, we set $C = 1000\mu^{-1}$. We can assume $p = Cn^{-2/3}(\log n)^{1/3}$, since the probability of any given collection of triangles of G appearing in G_p is monotone increasing in p.

We use a careful step-by-step revealing argument and choose $T_1, \dots, T_\ell$ one at a time. We will call an edge $e \in E(G)$ *alive* if its random status is yet to be revealed. Given $k \in [t]$ and $i \in [\ell]$, say that A_k is *full* at time i if $|A_k \cap V(\{T_1, \dots, T_{i-1}\})| \ge 12\mu|A_k|$. Let X_i be the union of the sets A_k that are full at time i. For each step $i \in [\ell]$ in succession, we will reveal certain edges of G_p and then choose a triangle T_i among the edges revealed. Specifically, we first reveal the random status of all edges in G adjacent to v_i, which do not go to $v_1, \dots, v_\ell$, X_i or a vertex of $T_1, \dots, T_{i-1}$. Let the edges amongst these that appear in G_p be denoted by S_i. We then reveal all *alive* edges of E_i which form a triangle with v_i using two edges of S_i. From these edges we pick any that appears, fixing the resulting triangle T_i, and move on to the next i.

Observe that by definition we do not reveal any edge of G_p twice; and if we successfully choose a triangle at each step we indeed obtain the desired triangle matching. To begin with, we argue that when we come to v_i, most edges of E_i are potential candidates to be in T_i. Note that any edge of E_i which is adjacent to any v_j or T_j will not be a candidate; there are at most $3\mu^2 n$ such vertices, which are adjacent to at most $3\mu^2 n^2$ edges of E_i. Any edge adjacent to X_i is also not a candidate; we have $|X_i| \le \frac{3\ell}{12\mu} \le \frac{\mu}{4}n$ and hence there are at most $\frac{\mu}{4}n^2$ edges adjacent to X_i. We also have that any candidate edge of E_i must be alive. When we reveal edges at some v_j, with probability at least $1 - n^{-2}$ by Chernoff's inequality (Theorem 2.1.1), we reveal at most $2pn = 2Cn^{1/3}(\log n)^{1/3}$ edges, and hence we reveal at most $4C^2n^{2/3}\log^{2/3} n$ edges of E_i in this step. Since there are at most $\mu^2 n$ steps, in total we will have revealed less than $n^{7/4}$ edges of E_i whp. Note that any edge in E_i which has not been ruled out for reasons outlined above, is a candidate at the beginning of step i, for forming T_i with v_i. Putting this together then, we have that whp, for each i there remains at least $\frac{1}{2}\mu n^2$ candidate edges of E_i at the beginning of step i. We denote this set of candidate edges by F_i.

When we reveal edges at v_i, for each edge of F_i we keep the edges from v_i to the endpoints of F_i with probability p^2, and so the expected number of edges of F_i whose ends are both adjacent to v_i in G_p is $p^2|F_i| \ge \frac{1}{2}p^2\mu n^2$. Applying Janson's inequality (Lemma 2.1.3), we have $\bar{\Delta} \le p^3n^3$, which is tiny compared to the square of the expectation, so with probability at least $1 - n^{-2}$, at least $\frac{1}{4}p^2\mu n^2$ edges of F_i are revealed to lie in $N_{G_p}(v_i)$. We now reveal which of these edges survive in G_p; by Chernoff's inequality (Theorem 2.1.1) and by our choice of C, with probability at least $1 - n^{-2}$, at least $\frac{1}{8}p^3\mu n^2$ of these edges survive in G_p, and in particular T_i exists.

Taking a union bound, the probability of failure at any step is $o(1)$. □

The next lemma allows us to find a reasonably large triangle matching using a possibly sparse set of edges, each of which extends to many triangles; we will use this to deal with nearly independent sets which have size larger than $\frac{1}{3}n$. Recall that we denote by $\deg_G(e; X)$ the size of the common neighbourhood of an edge e inside a set X. Recall also that given a set of edges E, we will sometimes think of E as the graph $H_E := (V(E), E)$ where $V(E)$ denotes the set of vertices contained in edges in E. We use notation like $\delta(E) := \delta(H_E)$ and $\deg_E(v) := \deg_{H_E}(v)$. Furthermore, given a set of vertices $A \subseteq V(G)$, $E[A]$ is used to denote the set of edges in E that are contained in A, that is, $E[A] := \{e \in E : e \subset A\}$.

Lemma 4.6.4. *For any $0 < \mu < \frac{1}{1000}$ there exists $C > 0$ such that the following holds for all $n, \delta, \delta_1, \delta_2 \in \mathbb{N}$, every n-vertex graph G and every $p \geq Cn^{-2/3}(\log n)^{1/3}$.*

(i) Let $X_1, X_2, X_3 \subset V(G)$ be disjoint sets of size at least $\frac{n}{10}$, and let $E \subseteq E(G[X_1])$ be a set of edges such that $\deg_E(v) \geq \delta$ for all $v \in X_1$ and $\deg_G(e; X_i) \geq \mu n$ for all $e \in E$ and $i = 2, 3$. Let $n_2, n_3 \in \mathbb{N}$ with $n_2 + n_3 \leq \min(\delta, \mu^5 n)$. Then, whp, there is a triangle matching $\mathcal{T} = \{T_1, \ldots, T_{n_2+n_3}\}$ in G_p with n_i triangles consisting of an edge $e \in E$ together with a vertex of X_i for each $i = 2, 3$.

(ii) Let $X_1, X_2 \subset V(G)$ be disjoint sets of size at least $\frac{n}{10}$. Let $E_i \subseteq E(G[X_i])$ be sets of edges such that $\deg_{E_i}(v) \geq \delta_i$ for all $v \in X_i$ and $\deg(e; X_{3-i}) \geq \mu n$ for all $e \in E_i$ and $i \in [2]$. Let $n_i \in \mathbb{N}$ with $n_i \leq \min(\delta_i, \mu^5 n)$ for each $i \in [2]$. Then, whp, there is a triangle matching $\mathcal{T} = \{T_1, \ldots, T_{n_1+n_2}\}$ in G_p with n_i triangles consisting of an edge $e \in E_i$ together with a vertex of X_{3-i} for each $i \in [2]$.

Observe that, unlike other lemmas in this section, both cases of this lemma are very tight and we can't even guarantee more vertex-disjoint triangles in the underlying graph G. If the edges E have small maximum degree however, the situation is somewhat easier and we will make use of this in the proof of Lemma 4.6.4. We obtain the following lemma.

Lemma 4.6.5. *For all $\mu > 0$ there exists $C > 0$ such that the following holds for all $n \in \mathbb{N}$, every n-vertex graph G and every $p \geq Cn^{-2/3}(\log n)^{1/3}$. Suppose that E is a subset of $E(G)$ with $\Delta(E) \leq \mu n$ and $\mu n \leq |E| \leq \mu^2 n^2$. Suppose in addition that for each edge $e \in E$ there is a given set X_e of size $|X_e| \geq \mu n$ consisting of vertices $v \in V(G) \setminus V(E)$ such that $e \in \mathrm{Tr}_v(G)$. Then, whp, there is a triangle matching $T_1, \ldots, T_\ell$ in G_p, where each T_i consists of an edge $e \in E$ together with a vertex of X_e, such that $\ell \geq \frac{|E|}{10\mu n}$.*

Proof. Let $0 < \frac{1}{C} \ll \mu$. We may assume that $p = Cn^{-2/3}(\log n)^{1/3}$ and that n is large enough for the following arguments. We will deduce the lemma from the following claim.

Claim 4.6.6. *Whp the following is true for all $X \subset V(G)$ with $|X| \leq \frac{|E|}{\mu n}$. If $|E[V(G) \setminus X]| \geq \frac{|E|}{2}$ and $|X_e \setminus X| \geq \frac{\mu n}{2}$ for all $e \in E$, then there is a triangle in $G_p[V(G) \setminus X]$ consisting of an edge $e \in E$ together with a vertex of X_e.*

Proof of Claim: This is a straightforward application of Janson's inequality and the union bound. Note that the total number of choices of X is at most $n^{|E|/(\mu n)}$. Fix one such choice. Let Y denote the number of suitable triangles in $G_p[V(G) \setminus X]$ and note that $\lambda := \mathbb{E}[Y] \geq \frac{p^3\mu|E|n}{4} \geq C^2 \log(n)\frac{|E|}{n}$. Furthermore, we have $\bar{\Delta} \leq 2\max(p^5|E|n^2, \lambda) \leq 2\max(\frac{4}{\mu}p^2 n\lambda, \lambda) \leq 2\lambda$. Hence, by Janson's inequality (see Lemma 2.1.3), the probability of having less than $\frac{\lambda}{2}$ triangles is at most

$$\exp\left(-\frac{\lambda^2}{8\bar{\Delta}}\right) \leq n^{-C|E|/n}.$$

The claim now follows by taking a union bound and noting $C \gg \frac{1}{\mu}$. ■

Assume now the high probability event in the claim occurs and let $T_1, \ldots, T_\ell$ be a maximal triangle matching as in the statement of the lemma. Suppose for contradiction that $\ell < \frac{|E|}{10\mu n}$ and let X be the set of vertices covered by $T_1, \ldots, T_\ell$. We have $|E[V(G) \setminus X]| \geq |E| - |X|\mu n \geq \frac{|E|}{2}$ and $|X_e \setminus X| \geq \mu n - \frac{3|E|}{10\mu n} \geq \frac{\mu n}{2}$ for all $e \in E$, and hence there is a suitable triangle in $G_p[V(G) \setminus X]$ which extends the triangle matching, a contradiction. □

We are now ready to prove Lemma 4.6.4.

Proof of Lemma 4.6.4. Let $0 < \frac{1}{C} \ll \mu$. We begin by proving (i). We may assume that $\delta \leq \mu^5 n$ and that n is large enough for the following arguments.

Let G_1, G_2, G_3 be independent copies of $G_{p/3}$. Observe that $G_1 \cup G_2 \cup G_3$ is distributed like $G_{p'}$ for some $p' \leq p$ and therefore it suffices to show that $G_1 \cup G_2 \cup G_3$ contains our desired triangle matching $\mathcal{T}$ whp. In what follows we will find $\mathcal{T}$ as the disjoint union of three triangle matchings $\mathcal{T}_1, \mathcal{T}_2$ and $\mathcal{T}_3$. For $i \in [3]$, the edges of G_i will be used to find the triangles in $\mathcal{T}_i$ and we will reveal G_1, G_2 and G_3 at different stages of our process, making use of their independence.

Let $B := \{v \in X_1 : \deg_E(v; X_1) \geq \mu n\}$, and let $S := X_1 \setminus B$. If $|B| \geq n_2 + n_3$, let $\mathcal{T}_1 = \mathcal{T}_2 = \emptyset$, $n_2' = n_3' = 0$, and move to the last stage of the process, in which we find $\mathcal{T}_3$. Otherwise, fix $n_2' := \min(n_2 + n_3 - |B|, n_2)$ and $n_3' := n_2 + n_3 - |B| - n_2' = \max(0, n_3 - |B|)$. In a first round of probability we find a triangle matching $\mathcal{T}_1$ of size n_2' in G_1, each triangle containing an edge in $E[S]$ and a vertex in X_2. This triangle matching exists whp due to Lemma 4.6.5. Indeed we have that $\Delta(E[S]) \leq \mu n$ (by the definition of S) and $\deg(e; X_2) \geq \mu n$ for all $e \in E[S]$. It remains to estimate $|E[S]|$. For this, note that

$$\begin{aligned} |E[S]| &\geq \tfrac{1}{2}|S|(\delta - |B|) \\ &\geq \tfrac{1}{2}\left(\tfrac{n}{10} - \delta\right)(n_2 + n_3 - |B|) \\ &\geq \tfrac{n}{40}(n_2 + n_3 - |B|) \geq \mu n. \end{aligned} \tag{4.6.1}$$

Furthermore, if $|E[S]| > \mu^2 n^2$ then we can shrink $E[S]$ to some subset having size exactly $\mu^2 n^2$. Applying Lemma 4.6.5 then gives a triangle matching of size at least $t \geq \frac{|E[S]|}{10\mu n}$. If $E[S]$ was shrunk to have size $\mu^2 n^2$, then $t \geq \frac{\mu}{10} n \geq n_2'$ and if not, then

$$t \geq \frac{n(n_2 + n_3 - |B|)}{400\mu n} \geq n_2 + n_3 - |B| \geq n_2',$$

using (4.6.1). In either case we can pick a sub-triangle matching $\mathcal{T}_1$ of the desired size n_2'.

We now fix $S' = S \setminus V(\mathcal{T}_1)$. Similarly to the previous stage, we will use G_2 to find a triangle matching $\mathcal{T}_2$ of size n_3' such that each triangle contains an edge in $E[S']$ and a vertex in X_3. We still clearly have that $\Delta(E[S']) \leq \mu n$ and $\deg(e; X_3) \geq \mu n$ for all $e \in E[S']$. Moreover, we have that

$$|E[S']| \geq |E[S]| - \mu n \cdot 2n_2' \geq \left(\tfrac{n}{40} - 2\mu n\right)(n_2 + n_3 - |B|) \geq \mu n,$$

where we used (4.6.1) and the fact that $2n_2'$ vertices of S were used in $\mathcal{T}_1$, each of which has degree at most μn in $E[S]$. Therefore, as in the previous phase, Lemma 4.6.5 gives the existence of at least n_3' vertex-disjoint triangles in G_3, each of which contain an edge of $E[S']$ and a vertex in X_3. From this, we choose our triangle matching $\mathcal{T}_2$ of size n_3'.

In our final phase we find a triangle matching $\mathcal{T}_3$ in G_3 to complete $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}_3$ as desired. Let $X_i'' = X_i \setminus (V(\mathcal{T}_1 \cup \mathcal{T}_2))$ for $i \in [3]$ and note that $B \subset X_1''$. Further, for $i \in [2]$, let $n_i'' = n_i - n_i'$ and note that each $n_i'' \geq 0$ and $n_2'' + n_3'' = \min(|B|, n_2 + n_3)$. Pick disjoint subsets $B_i \subset B$ of size n_i'' for each $i = 2, 3$. Since $n_2'' + n_3'' \leq n_2 + n_3 \leq \delta \leq \mu^5 n$, it follows from Lemma 4.6.3, that whp there is a triangle matching $\mathcal{T}_3$ of size $n_2'' + n_3''$ in $G_3[X_1'' \cup X_2'' \cup X_3'']$ consisting of n_i'' triangles which contain an edge in $E[X_1'']$ and one vertex in X_i'', for each $i = 2, 3$. Indeed, in applying Lemma 4.6.3, we can fix $t = 0$ (we do not need to use the full extent of the lemma here) and for $i = 1, 2$ and $v \in B_i$, we choose a collection of at least $\frac{\mu^2 n^2}{4}$ edges f in $\mathrm{Tr}_v(G)$ such that $|f \cap X_1''| = |f \cap X_i''| = 1$. These edges exist as

$$\deg_E(v; X_1'') \geq \deg_E(v; X_1) - |V(\mathcal{T}_1 \cup \mathcal{T}_2)| \geq \mu n - 4\delta \geq \tfrac{\mu n}{2}$$

and for each edge $e \in E[X_1'']$, $\deg_G(e, X_i'') \geq \mu n - |V(\mathcal{T}_1 \cup \mathcal{T}_2)| \geq \frac{\mu n}{2}$ for $i = 1, 2$. To conclude, we have that whp all three stages of the process above succeed and we have a triangle matching $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}_3$ in G_p as in (i).

Part (ii) is similar to part (i). We begin again by noting that we can assume $\delta_i \leq \mu^5 n$ for $i = 1, 2$. We will again find three triangle matchings $\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3$ whose union will give us our desired triangle matching $\mathcal{T}$ and we again use three independent copies G_1, G_2, G_3 of $G_{p/3}$, finding the triangles in $\mathcal{T}_i$ using the edges of G_i for $i \in [3]$. For convenience, let us also fix $\mu' = \frac{\mu}{2}$. Now for $i = 1, 2$, let $B_i := \{v \in X_i : \deg_{E_i}(v; X_i) \geq \mu' n\}$ and if $|B_i| \geq n_i$, then shrink B_i to have size n_i (that is, take B_i to be a subset of $\{v \in X_i : \deg_{E_i}(v; X_i) \geq \mu' n\}$ of size n_i). Further, for $i \in [2]$, let $S_i := X_i \setminus B_i$ and define $n_i' = n_i - |B_i|$. Let us assume for now that $n_1' \leq n_2'$.

In G_1, we now find $\mathcal{T}_1$, a triangle matching of size n_1' with each triangle containing an edge in $E[S_1]$ and a vertex of S_2. If $n_1' = 0$ there is nothing to prove here and in the case that $n_1' \geq 1$ (and so $|B_1| < n_1$), such a triangle matching exists whp due to Lemma 4.6.5 (applied with μ' replacing μ). Indeed, the verification of the conditions of Lemma 4.6.5 is almost identical to our proof of the existence of $\mathcal{T}_1$ in part (i). One slight difference is that, for an edge $e \in E_1[S_1]$ we cannot use all of $N(e; X_2)$ to give the set X_e needed in Lemma 4.6.5. Indeed, we need to discount vertices in B_2 but as $|B_2| \leq n_2 \leq \mu^5 n$ and $|N(e; X_2)| \geq \mu n$, we can certainly have at least $\mu' n$ vertices in $N(e; S_2)$.

Given that we succeed in finding $\mathcal{T}_1$, we now turn to finding $\mathcal{T}_2$ in G_2. For this we define $S_i' = S_i \setminus V(\mathcal{T}_1)$ for $i = 1, 2$ and we aim to find n_2' vertex-disjoint triangles, each containing an edge in $E_2[S_2']$ and a vertex of S_1'. If $n_2' = 0$, then the existence of $\mathcal{T}_2$ is immediate. For the case when $n_2' \geq 1$, we again appeal to Lemma 4.6.5 (with μ' replacing μ). Note that due to the fact that $n_2' \geq 1$, we have that B_2 contains all high degree vertices and so, in particular, $\Delta(E_2[S_2']) \leq \mu' n$. Also using this, we have that

$$\begin{aligned}
|E[S_2']| &\geq |E[S_2]| - |V(\mathcal{T}_1) \cap S_2|\mu' n \\
&\geq \tfrac{1}{2}(|X_2| - |B_2|)(\delta_2 - |B_2|) - n_1' \mu' n \\
&\geq \tfrac{n}{40} n_2' - n_1' \mu' n \\
&\geq n\left(\tfrac{1}{40} - \mu'\right) n_2' \geq \mu' n,
\end{aligned}$$

where in the last two inequalities, we used that $n_1' \leq n_2'$ and that we are in the case that $n_2' \geq 1$. Finally, it is not hard to see that $|N(e; S_1')| \geq \mu' n$ for all $e \in E_2[S_2']$ and so the conditions of Lemma 4.6.5 are indeed satisfied and whp we get our desired triangle matching $\mathcal{T}_2$. For the above, we needed that $n_1' \leq n_2'$. In the case that $n_2' > n_1'$, we can run exactly the same proof except that we first find $\mathcal{T}_2$ and then find $\mathcal{T}_1$ after.

Finally, we find $\mathcal{T}_3$ in G_3 by applying Lemma 4.6.3. Indeed, similarly to our proof for part (i), we fix $S_i'' = S_i' \setminus V(\mathcal{T}_2)$ for $i \in [2]$ and we know that for each $i \in [2]$ and $v \in B_i$, we have at least $\frac{\mu'^2 n^2}{8}$ edges $f \in \mathrm{Tr}_v(G)$ such that $|f \cap S_i''| = |f \cap S_{3-i}''| = 1$. Therefore, as $|B_1| + |B_2| = n_1 + n_2 - n_1' - n_2' \leq 2\mu^5 n$, Lemma 4.6.3 gives that whp, there exists a triangle matching $\mathcal{T}_3$ in G_3, of size $|B_1| + |B_2|$, such that for each $i \in [2]$ and $v \in B_i$, there is a triangle in $\mathcal{T}_3$ containing v, some vertex in S_i'' and a vertex in S_{3-i}''. Altogether, we have that whp, we can find all the triangle matchings $\mathcal{T}_i$ and $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}_3$ provides the desired triangle matching, completing the lemma. □

4.7 Reduction

We are now in a position to prove Theorem II, assuming Theorem II*. Our proof relies on the use of the Regularity Lemma (Lemma 2.2.1), we refer the reader to Section 2.2 for the relevant definitions. Before giving the details, let us briefly sketch the approach. Given G with $n \in 3\mathbb{N}$ vertices and minimum degree at least $\frac{2}{3}n$, we separate three cases.

Our first case is that there is no set S of about $\frac{n}{3}$ vertices such that $G[S]$ has small maximum degree. In this case, we apply the Regularity Lemma (Lemma 2.2.1) and observe that the (ε, d)-reduced graph R has no large independent set. By the Hajnal–Szemerédi Theorem for K_3-matchings (Theorem 2.5.1), we find a large triangle matching $\mathcal{T}^*$ in R, and make the corresponding pairs of clusters super-regular by removing a few vertices to obtain a subgraph T of G. If T were spanning in G, and the clusters were balanced, we would be done by Theorem II*. In order to fix this, we need to remove a few more triangles covering the vertices outside T (which we do using Lemma 4.6.3) and then further triangles to balance the clusters of T (using Lemma 4.6.1). For the latter we use the $k = 3$ case of Theorem 4.5.4 to find a fractional triangle factor which tells us where to remove triangles: this is the point where we use the fact that G has no large sparse set. We obtain the following lemma (whose proof we defer to later).

Lemma 4.7.1 (No large sparse set)**.** *For every sufficiently small $\mu > 0$ there exists constants $C > 0$ and $0 < d \leq \mu$ such that the following holds. Let $n \in 3\mathbb{N}$, $p \geq Cn^{-2/3}(\log n)^{1/3}$ and suppose G is an n-vertex graph with $\delta(G) \geq \left(\frac{2}{3} - \frac{d}{2}\right)n$ such that there is no $S \subseteq V(G)$ of size at least $\left(\frac{1}{3} - 2\mu\right)n$ with $\Delta\big(G[S]\big) \leq 2dn$. Then whp G_p contains a triangle factor.*

Note that in the case that there is no large sparse set, we can reduce the minimum degree necessary slightly.

Our second case is that there is a set S of about $\frac{n}{3}$ vertices such that $G[S]$ has maximum degree at most $2dn$, but there is no second such set in $G - S$. The idea here is that we will remove a few triangles from G in order to obtain a subgraph of G which can be partitioned into sets X_1, X_2 of sizes $|X_2| = 2|X_1| \approx \frac{2n}{3}$, such that all vertices of X_1 are adjacent to almost all vertices of X_2 and vice versa (here Lemma 4.6.4 will be very useful). Note that, with this degree condition, X_2 can be very close to the union of two cliques of size about $\frac{n}{3}$; this leads to a 'parity case' in which we have to be very careful, which is something of a complication. If we can arrange for the correct parities however, it will be easy to split X_1 into two sets, each of which induces a super-regular triple with one of the 'near-cliques' and apply our Theorem II*. If we are not in the parity case, we will apply the Regularity Lemma to X_2 and find an almost-spanning matching $\mathcal{M}^*$ in the reduced graph R. We proceed similarly as in the previous case, making these pairs super-regular, removing 'atypical' vertices and then balancing the pairs. Here, we make sure that every triangle we remove has two vertices in X_2 and one in X_1 to keep the right balance between the two parts.

Finally we can partition X_1 into smaller sets and form balanced super-regular triples with the edges of $\mathcal{M}^*$ in order to apply our Theorem II*. We obtain the following lemma (whose proof we defer to later).

Lemma 4.7.2 (One large sparse set)**.** *For every sufficiently small $\mu > 0$, there exists constants $C > 0$ and $0 < \tau, d \leq \mu$ such that the following holds for all $n \in 3\mathbb{N}$ and $p \geq Cn^{-2/3}(\log n)^{1/3}$. Suppose G is an n-vertex graph with $\delta(G) \geq \frac{2}{3}n$, and suppose S is a subset of $V(G)$ with $|S| \geq (\frac{1}{3} - \tau)n$ and $\Delta(G[S]) \leq \tau n$. Suppose further that there is no $S' \subseteq V(G) \setminus S$ of size at least $\big(\frac{1}{3} - 2\mu\big)n$ with $\Delta\big(G[S']\big) \leq 2dn$. Then whp G_p contains a triangle factor.*

Our third and final case is that there are two vertex-disjoint sets S_1, S_2 each of which has size about $\frac{n}{3}$ in G and small maximum degree. In this case G must be very close to a balanced complete tripartite graph. We start by partitioning $V(G)$ into sets X_1, X_2 and X_3 of size around $\frac{n}{3}$, so that (X_1, X_2, X_3) is an $(\varepsilon, d^+, \delta)$-super-regular triple, where d is close to 1, but δ can be quite small (we need $\delta \gg \varepsilon$ in order to apply Theorem II*). We remove some carefully chosen vertex-disjoint triangles in order to balance the X_i and to remove some 'atypical' vertices. This leaves us with a balanced (ε, d^+)-super-regular triple for some d close to 1, and Theorem II* finds the required triangle factor, giving the following.

Lemma 4.7.3 (Two large sparse sets)**.** *There exists constants $C, \tau > 0$ such that the following holds for all $n \in 3\mathbb{N}$ and $p \geq Cn^{-2/3}(\log n)^{1/3}$. Suppose G is an n-vertex graph with $\delta(G) \geq \frac{2}{3}n$, and suppose S_1 and S_2 are disjoint subsets of $V(G)$ with $|S_i| \geq (\frac{1}{3} - \tau)n$ and $\Delta(G[S_i]) \leq \tau n$ for each $i = 1, 2$. Then whp G_p contains a triangle factor.*

Before we give proofs of these three lemmas, we show how they imply Theorem II.

Proof of Theorem II. Choose $0 < \mu_2 \ll \tau_3 \ll 1$ where τ_3 is chosen small enough to apply Lemma 4.7.3. Let $\tau_2, d_2 \leq \mu_2$ be the constants returned by Lemma 4.7.2 with input μ_2 and choose $0 < \mu_1 \ll \tau_2, d_2$. Finally, let $d_1 \leq \mu_1$ be the constant returned by Lemma 4.7.1 with input μ_1 and choose $0 < \frac{1}{C} \ll d_1$. Let $n \in 3\mathbb{N}$ and let $p \geq Cn^{-2/3}(\log n)^{1/3}$ and suppose that G is an n-vertex graph with $\delta(G) \geq \frac{2n}{3}$.

If G contains no subset of size at least $\big(\frac{1}{3} - 2\mu_1\big)n$ vertices with maximum (induced) degree at most $2d_1 n$, then by Lemma 4.7.1, G_p contains a triangle factor whp. We may therefore suppose G contains a subset S_1 of vertices of size at least $\big(\frac{1}{3} - 2\mu_1\big)n \geq \big(\frac{1}{3} - \tau_2\big)n$ with maximum degree $\Delta(G[S_1])$ at most $2d_1 n \leq \tau_2 n$. If there is no $S_2 \subseteq V(G) \setminus S_1$ of size at least $\big(\frac{1}{3} - 2\mu_2\big)n$ with maximum degree $\Delta(G[S_2])$ at most $2d_2 n$, then by Lemma 4.7.2, G_p contains a triangle factor whp. We can therefore suppose that G contains a subset S_2 disjoint from S_1 of size at least $\big(\frac{1}{3} - 2\mu_2\big)n \geq \big(\frac{1}{3} - \tau_3\big)n$ with maximum (induced) degree at most $2d_2 n \leq \tau_3 n$. So by Lemma 4.7.3, G_p contains a triangle factor whp. □

The remainder of the section (and indeed the chapter) is devoted to proving the three lemmas.

4.7.1 Case: No large sparse set

Proof of Lemma 4.7.1. Fix some $0 < \mu \ll 1$ and choose $0 < \frac{1}{m_0} \ll \varepsilon \ll d \ll \mu$. Let $M_0 \geq m_0$ be returned by Lemma 2.2.1 with input m_0, ε and fix $\gamma = \frac{2}{3} - \frac{d}{2}$ and $0 < \frac{1}{C} \ll \frac{1}{M_0}$. Assume also that $n \gg M_0$. Let p and G be as in the statement and let G_1, G_2, G_3 be independent copies of $G_{p/3}$; we will show that $G_1 \cup G_2 \cup G_3$ satisfies the desired properties whp.

We apply Lemma 2.2.1 to G, and obtain an (ε, d)-reduced graph R on m vertices with $m_0 \leq m \leq M_0$ and minimum degree at least $\left(\frac{2}{3} - \frac{d}{2} - d - 2\varepsilon\right)m \geq \left(\frac{2}{3} - 2d\right)m$. Recall that we identify the vertex set of R as $[m]$ with each $i \in [m]$ corresponding to a cluster V_i in the ε-regular partition of $V(G)$.

Claim 4.7.4. *We have $\alpha(R) < \left(\frac{1}{3} - \mu\right)m$.*

Proof of Claim: Suppose for a contradiction that R contains an independent set I of size $\left(\frac{1}{3} - \mu\right)m$. Now call an index $i \in I$ *bad* if there are more than $\sqrt{\varepsilon}m$ indices $j \in [m] \setminus \{i\}$ such that (V_i, V_j) is *not* ε-regular. Due to the fact that the V_i form an ε-regular partition, we have that there are at most $2\sqrt{\varepsilon}m$ bad indices. Let I' be the set obtained from I after removing bad indices and so $|I'| \geq \left(\frac{1}{3} - \frac{3\mu}{2}\right)m$. Now in $\bigcup_{i \in I'} V_i$ there must exist at least $\frac{\mu}{4}n$ vertices, each of whose degree into $\bigcup_{i \in I'} V_i$ exceeds $2dn$, otherwise removing all such vertices would leave a set S whose existence is forbidden in the lemma statement. By averaging, there is some $i^* \in I'$ such that $\frac{\mu}{4}|V_{i^*}|$ of these vertices are in V_{i^*}. Let $U_{i^*} \subseteq V_{i^*}$ be this subset of high degree vertices. Now vertices of V_{i^*} can have at most $|V_{i^*}|$ neighbours in V_{i^*}, and at most $\sqrt{\varepsilon}m \cdot \frac{n}{m} \leq \sqrt{\varepsilon}n$ neighbours in sets V_j such that $j \in I$ and (V_{i^*}, V_j) is not ε-regular (as $i^* \in I'$). So the vertices of U_{i^*} all have at least $\frac{3d}{2}n$ neighbours in total in sets V_j such that $j \in I$, $j \neq i^*$ and (V_{i^*}, V_j) is ε-regular. By averaging, there is one of these sets V_j such that the density between U_{i^*} and V_j exceeds $\frac{3}{2}d$. But (V_{i^*}, V_j) is ε-regular and has density less than d; this is a contradiction. ■

We apply the Hajnal–Szemerédi Theorem for K_3-matchings (Theorem 2.5.1) to R, which gives us a triangle matching $\mathcal{T}^*$ in R covering at least $(1 - 13d)m$ vertices. We denote by $T^* := V(\mathcal{T}^*)$ the set of indices in triangles of $\mathcal{T}^*$. By Lemma 2.2.4, there are $V'_i \subset V_i$ for each $i \in T^*$ such that $|V'_i| = \lceil (1 - 3\varepsilon)|V_i| \rceil$ and, for every triangle $ijk \in \mathcal{T}^*$, the triple (V'_i, V'_j, V'_k) is $(2\varepsilon, (d - \varepsilon)^+, d - 3\varepsilon)$-super-regular. Let $T = \bigcup_{i \in T^*} V'_i$ be the set of vertices in G which are in a cluster V'_i corresponding to a triangle of $\mathcal{T}^*$. Let $X = V(G) \setminus T$. Observe that $|X| \leq \varepsilon n + 13dn + 3\varepsilon n \leq 14dn$. Let $W \subset T$ be a set such that

(i) $|W \cap V'_i| = \left(\frac{1}{2} \pm \frac{1}{20}\right)\frac{n}{m}$ for each $i \in T^*$,

(ii) $\deg_G(v; W) \geq \frac{3}{5}|W|$ for each $v \in V(G)$, and

(iii) we have that $\deg_G(v; V_i' \cap W) = \left(\frac{1}{2} \pm \frac{1}{4}\right)\deg_G(v; V_i')$ for each $i \in T^*$ and $v \in V(G)$ with $\deg_G(v; V_i') \geq \varepsilon|V_i'|$.

Such a set W can be found by choosing each vertex of T independently with probability $\frac{1}{2}$ and applying Chernoff's inequality (Theorem 2.1.1) and a union bound.

We will start by covering X. We will not use vertices that belong to $T \setminus W$ in order to maintain super-regularity properties.

Claim 4.7.5. *Whp in G_1, there is a triangle matching $\mathcal{T}_1 \subset K_3(G_1[W \cup X])$ so that $X \subset V(\mathcal{T}_1)$ and $|V(\mathcal{T}_1) \cap V_i'| \leq 50\sqrt{d}|V_i'|$ for all $i \in T^*$.*

Proof of Claim: Let $\tilde{\mu} := 4\sqrt{d}$ and enumerate $X = \{v_1, \dots, v_\ell\}$, noting that $\ell \leq \tilde{\mu}^2 n$. For each $i \in [\ell]$, let $E_i := E(G[W]) \cap \mathrm{Tr}_{v_i}(G)$. Note that, since $\deg(v; W) \geq \frac{3}{5}|W|$ for all $v \in V(G)$, we have $|E_i| \geq \tilde{\mu} n^2$ for all $i \in [\ell]$. Finally, let $A_i = V_i'$ for each $i \in T^*$. The claim now follows readily from Lemma 4.6.3. ∎

Let now $V_i'' = V_i' \setminus V(\mathcal{T}_1)$ for each $i \in T^*$. We would like to apply Theorem II* to the super-regular triples (V_i'', V_j'', V_k'') for each $ijk \in \mathcal{T}^*$. However, these triples are not necessarily balanced. The next claim corrects this.

Claim 4.7.6. *Whp in G_2, there is a triangle matching $\mathcal{T}_2 \subset K_3(G_2[W \setminus V(\mathcal{T}_1)])$ so that $|V_i'' \setminus V(\mathcal{T}_2)| = \lfloor \frac{9}{10}\frac{n}{m} \rfloor$ for all $i \in T^*$.*

Proof of Claim: The key idea in this proof is to use fractional factors to dictate how we remove triangles in order to balance the parts. More specifically, we will apply our stability theorem for the fractional Hajnal-Szemerédi theorem with integer weights (Theorem 4.5.4), using that the reduced graph has large minimum degree and no large independent sets. In detail, let $R' = R[T^*]$ and let $\lambda : T^* \to \mathbb{N}$ be given by $\lambda(i) = |V_i''| - \lfloor \frac{9}{10}\frac{n}{m} \rfloor$. Note that $\left(\frac{1}{10} - 60\sqrt{d}\right)\frac{n}{m} \leq \lambda(i) \leq \lceil \frac{1}{10}\frac{n}{m} \rceil$, and that $\sum_{i \in T^*} \lambda(i) = n - 3|\mathcal{T}_1| - 3|\mathcal{T}^*|\lfloor \frac{9}{10}\frac{n}{m} \rfloor$ is divisible by 3. Also, we have that $\delta(R') \geq \delta(R) - 13dm \geq \left(\frac{2}{3} - 15d\right)|R'|$ and $\alpha(R') \leq \left(\frac{1}{3} - \mu\right)m \leq \left(\frac{1}{3} - \frac{\mu}{2}\right)|R'|$. Hence, by Theorem 4.5.4 (and the fact that $d \ll \mu$), there is a weight function $\omega : K_3(R') \to \mathbb{N}$ such that for each $i \in T^*$ we have $\sum_{K \in K_3(R',i)} \omega(K) = \lambda(i)$. We claim that we can remove $\omega(ijk)$ triangles from $G_2[V_i'' \cap W, V_j'' \cap W, V_k'' \cap W]$ for each triangle ijk of R', making sure that all our choices are vertex-disjoint. Indeed, observe that for any choice of $X_h \subset V_h'' \cap W$ such that $|X_h| \geq d\frac{n}{m}$ for $h \in \{i, j, k\}$, we have $|K_3(G[X_i, X_j, X_k])| \geq \frac{d^6}{10m^3}n^3$ due to Lemma 2.2.7 and the (ε, d^+) regularity of $G[V_i, V_j, V_k]$. Furthermore, observe that $|V_i'' \cap W| \geq \frac{2}{5} \cdot \frac{n}{m}$ for each $i \in T^*$. Hence, Lemma 4.6.1 (ii) implies that whp there are $\frac{7}{20} \cdot \frac{n}{m} > 3 \cdot \lceil \frac{1}{10}\frac{n}{m} \rceil$ vertex-disjoint triangles in $G_2[V_i'' \cap W, V_j'' \cap W, V_k'' \cap W]$ for each $ijk \in K_3(R')$, so we can select the desired number of triangles for each $K \in K_3(R')$ one at a time. ∎

Let now $V_i''' = V_i'' \setminus V(\mathcal{T}_2)$ for all $i \in T^*$ and observe that we have covered all vertices except for those in $\bigcup_{i\in T^*} V_i'''$. We claim that (V_i''', V_j''', V_k''') is $(5\varepsilon, (d/2)^+, d/8)$-super-regular for all $ijk \in \mathcal{T}^*$. Indeed, this follows from the Slicing Lemma (Lemma 2.2.2), and from $\deg(v; V_j''') \geq \deg(v; V_j' \setminus W) \geq \frac{1}{4}\deg_G(v; V_j') \geq \frac{d}{8}|V_j'|$ for all $v \in V_i$ and the analogous inequalities for other pairs. Finally, we apply Theorem II* to each of these triples individually in G_3 to obtain (whp) a triangle matching $\mathcal{T}_3$ covering exactly $\bigcup_{i\in T^*} V_i'''$. □

4.7.2 Case: Two large sparse sets

Next, we deal with the case when G has two large sparse sets; i.e. it looks similar to the extremal complete tripartite graph. This is the easiest case; we will not need the regularity lemma.

Proof of Lemma 4.7.3. Choose $0 < \frac{1}{C} \ll \tau \ll \rho \ll \frac{1}{1000}$. Let $n \in 3\mathbb{N}$ be large enough for the following arguments and let $p \geq Cn^{-2/3}(\log n)^{1/3}$. Let G and sparse sets S_1 and S_2 be given as in the statement. Let G_1, G_2, G_3 be independent copies of $G_{p/3}$. We will find a triangle factor in $G_1 \cup G_2 \cup G_3$.

Claim 4.7.7. *There is a partition $V(G) = X_1 \cup X_2 \cup X_3$ such that*

(i) $|X_i| = \left(\frac{1}{3} \pm \rho^6\right)n$ *for all* $i \in [3]$,

(ii) $\deg(v; X_j) \geq \rho n$ *for all* $i \neq j \in [3]$ *and* $v \in X_i$,

(iii) $d(X_i, X_j) \geq 1 - \rho^6$ *for all* $1 \leq i < j \leq 3$,

(iv) For each $i \in [3]$, *if* $|X_i| \geq \frac{n}{3}$, *then* $\deg(v; X_j) \geq |X_j| - 4\rho n$ *for all* $v \in X_i$ *and* $j \in [3] \setminus \{i\}$.

<u>Proof of Claim:</u> For $i \in [2]$, let $Z_i = \{v \in V(G) \setminus (S_1 \cup S_2) : \deg(v; S_i) \leq \rho n\}$. Let $U_i = S_i \cup Z_i$ for $i \in [2]$ and $U_3 = \left\{v \in V(G) : \deg(v; S_i) \geq \left(\frac{1}{3} - 2\rho\right)n \text{ for each } i \in [2]\right\}$. Note that, since $\delta(G) \geq \frac{2}{3}n$, Z_1 and Z_2 are disjoint and hence U_1 and U_2 are disjoint as well. Furthermore, by definition, U_3 is disjoint from U_1 and U_2. Let $Z' := V(G) \setminus (U_1 \cup U_2 \cup U_3)$ be the set of remaining vertices. Partition $Z' = Z_1' \cup Z_2' \cup Z_3'$ so that $Z_i' = \emptyset$ if $|U_i| \geq \frac{n}{3}$ and $|U_i| + |Z_i'| \leq \frac{n}{3}$ otherwise. Finally, let $X_i = U_i \cup Z_i'$ for all $i \in [3]$. Note that $V(G) = X_1 \cup X_2 \cup X_3$ is indeed a partition.

We will first show that the sets Z_1, Z_2 and Z' are small. Let $i \in [2]$. Since $|S_i| \geq \left(\frac{1}{3} - \tau\right)n$, each vertex of S_i has at least $\left(\frac{1}{3} - 2\tau\right)n$ non-neighbours in S_i, and so at most $2\tau n$ non-neighbours outside S_i. Therefore, the total number of non-edges between S_i and $V(G) \setminus S_i$ is at most τn^2 (using here that we certainly have $|S_i| \leq \frac{n}{2}$ for $i = 1, 2$). Since every $v \in Z_i$ has at least $\frac{n}{4}$ non-neighbours in S_i, this implies $|Z_i| \leq 4\tau n$. Moreover, the number of non-edges between $U_1 \cup U_2$ and Z' is at most $2\tau n^2 + (|Z_1| + |Z_2|)n \leq 10\tau n^2$. Observe that every $v \in Z'$ has at least ρn

non-neighbours in $U_1 \cup U_2$ (otherwise it would be in U_3), and therefore $|Z'| \le \rho^8 n$, by our choice of τ. We now show that this implies condition (i). Indeed, we have that $|S_1|, |S_2| = \left(\frac{1}{3} \pm \tau\right)n$ where the lower bounds are directly from our assumption and the upper bounds are due to the fact that every vertex in S_i has $\left(\frac{2}{3} - \tau\right)n$ neighbours outside of S_i for $i = 1, 2$. For each i, we add at most $(4\tau + \rho^8)n$ vertices to S_i to obtain X_i and so we have that $|X_i| = \left(\frac{1}{3} \pm \rho^7\right)n$ for $i = 1, 2$. Finally, the bounds on $|X_3|$ can be deduced from the fact that the X_i partition $V(G)$.

Furthermore, for each $v \in Z'$, we have $\deg(v; S_i) \ge \rho n$ since $v \notin Z_i$ for $i \in [2]$, and $\deg(v; U_3) \ge \rho n$ for otherwise v would be in U_3. Clearly, we also have that $\deg(v; X_j) \ge \rho n$ for all $i \in [2]$, $j \in [3] \setminus \{i\}$ and $v \in X_i$ and so (ii) holds. Moreover, we have $\deg(v; X_i) \ge |X_i| - 2\tau n$ for all $v \in S_1$ and $i = 2, 3$ as v already has at least $\left(\frac{1}{3} - 2\tau\right)n$ non-neighbours in S_1. Since $|Z_1 \cup Z_1'| \le \rho^7 n$, this implies $d(X_1, X_i) \ge 1 - \rho^6$ for $i = 2, 3$. Similarly $d(X_2, X_3) \ge 1 - \rho^6$.

Finally, let $i, j \in [3]$ be distinct. If $|X_i| \ge \frac{n}{3}$, then $X_i \cap Z' = \emptyset$ by construction. Now if $i = 1$ or $i = 2$, then it is clear that $\deg(v; X_j) \ge |X_j| - 4\rho n$ for all $v \in X_i$ as v as $\deg(v; X_i) \le 2\rho n$ and so v already has many non-neighbours in X_i (considering the size of X_i given in (i)). If $i = 3$, then for any $v \in X_i$, we have that $\deg(v; X_j) \ge \deg(v; S_j) \ge \left(\frac{1}{3} - 2\rho\right)n \ge |X_j| - 4\rho n$. This establishes (iv). ∎

We now perform a stage of removing some vertex-disjoint triangles in order to obtain a balanced tripartite graph.

Claim 4.7.8. *Whp in G_1, there is triangle matching $\mathcal{T}_1 \subset K_3(G_1)$ so that $|X_1 \setminus V(\mathcal{T}_1)| = |X_2 \setminus V(\mathcal{T}_1)| = |X_3 \setminus V(\mathcal{T}_1)| \ge (\frac{1}{3} - \rho^6)n$.*

Proof of Claim: If all three sets X_1, X_2, X_3 have size exactly $\frac{n}{3}$, we are done. Otherwise, one or two of these sets has size exceeding $\frac{n}{3}$.

Case 1. Assume first that only one set exceeds $\frac{n}{3}$ in size and, without loss of generality, this set is X_1. Let $n_2 := \frac{n}{3} - |X_3|$ and $n_3 := \frac{n}{3} - |X_2|$, and let $E = E(G[X_1])$. Observe that $\delta(E) \ge |X_1| - \frac{n}{3} = n_2 + n_3$. Furthermore, we have $\deg(e; X_i) \ge |X_i| - 10\rho n \ge \frac{n}{4}$ for both $i = 2, 3$. Therefore, by Lemma 4.6.4 (i), there is a triangle matching $\mathcal{T}_1$ of size $n_2 + n_3$ in G_1 such that the triangles in $\mathcal{T}_1$ all have two vertices in X_1, n_2 of them have their third vertex in X_2, and n_3 of them have their third vertex in X_3. We then have $|X_1 \setminus V(\mathcal{T}_1)| = |X_2 \setminus V(\mathcal{T}_1)| = |X_3 \setminus V(\mathcal{T}_1)| = \frac{2n}{3} - |X_1| \ge (\frac{1}{3} - \rho^6)n$, as claimed, by our definitions of n_2 and n_3.

Case 2. Assume now that there are two sets (say X_1 and X_2) exceeding $\frac{n}{3}$ in size. For $i \in [2]$, let $n_i := |X_i| - \frac{n}{3}$ and $E_i = E(G[X_i])$. Observe that, for $i \in [2]$, $\delta(E_i) \ge n_i$ and $\deg(e; X_{3-i}) \ge |X_{3-i}| - 10\rho n \ge \frac{n}{4}$ for all $e \in E_i$. Therefore, by Lemma 4.6.4 (ii), there is a triangle matching $\mathcal{T}_1$ of size $n_1 + n_2$ in G_1, with n_1 triangles having two vertices in X_1 and one in X_2, and n_2 triangles having two vertices in X_2 and one in X_1. Therefore, we have $|X_1 \setminus V(\mathcal{T}_1)| = |X_2 \setminus V(\mathcal{T}_1)| = |X_3 \setminus V(\mathcal{T}_1)| = |X_3| \ge (\frac{1}{3} - \rho^6)n$, as claimed. ∎

Let now $X_i' = X_i \setminus V(\mathcal{T}_1)$ and observe that $|X_1'| = |X_2'| = |X_3'|$. Define

$$Y_i' := \left\{ v \in X_i' : \deg(v; X_j') \le \left(1 - \tfrac{\rho}{2}\right)|X_j'| \text{ for some } j \in [3] \setminus \{i\} \right\}.$$

Since $d(X_i', X_j') \ge 1 - 4\rho^6$ for all $1 \le i < j \le 3$, we have $|Y_i'| \le 4\rho^5 n$ for each $i \in [3]$. Furthermore, for each $i \in [3]$ and vertex $v \in Y_i'$ there are at least $\frac{1}{8}\rho^2 n^2$ triangles of G containing v and one vertex in each $X_j' \setminus Y_j'$ for $j \in [3] \setminus \{i\}$. Indeed, we have that

$$\deg(v; X_j' \setminus Y_j') \ge \deg(v; X_j) - 2|V(\mathcal{T}_1)| - |Y_j'| \ge \tfrac{3\rho}{4} n,$$

for each $j \in [3] \setminus \{i\} =: \{j_1, j_2\}$. Due to the defining condition of the Y_j', we then have that for each $x \in N(v; X_{j_1}' \setminus Y_{j_1}')$, we have that $\deg(v, x; X_{j_2}' \setminus Y_{j_2}') \ge \frac{\rho}{4} n$. This implies the claimed lower bound on the number of triangles containing $v \in Y_i'$.

By applying Lemma 4.6.3 (with $t = 0$), whp in G_2, we can find a triangle matching $\mathcal{T}_2 \subset K_3(G_2)$ with each triangle using one vertex from each part and such that $Y_1' \cup Y_2' \cup Y_3' \subset V(\mathcal{T}_2) \subset X_1' \cup X_2' \cup X_3'$ and $|V(\mathcal{T}_2)| \le 3(|Y_1'| + |Y_2'| + |Y_3'|) \le \rho^4 n$.

Let now $X_i'' := X_i' \setminus V(\mathcal{T}_2)$ for each $i \in [3]$ and observe that $|X_1''| = |X_2''| = |X_3''| \ge (\frac{1}{3} - 2\rho^4)n$. Furthermore, (X_1'', X_2'', X_3'') is $(\sqrt{\rho}, (1-\rho)^+)$-super-regular by Lemma 2.2.5. Hence, by Theorem II*, whp there is a triangle matching $\mathcal{T}_3$ in G_3 covering the X_i''. Together with $\mathcal{T}_1$ and $\mathcal{T}_2$ this gives a full triangle factor in G_p. □

4.7.3 Case: One large sparse set

Finally, we deal with the second case sketched in the discussion at the beginning of Section 4.7, when there is one large sparse set but not a further disjoint one. We will use several of the ideas from the previous two lemmas, and so will abbreviate the details in places.

Proof of Lemma 4.7.2. Fix some $0 < \mu \ll 1$ and choose $0 < \frac{1}{m_0} \ll \varepsilon \ll d \ll \mu$. Let $M_0 \ge m_0$ be returned by Lemma 2.2.1 with input m_0, ε and fix $0 < \frac{1}{C} \ll \tau \ll \rho \ll \frac{1}{M_0}$. Assume that $n \in 3\mathbb{N}$ is large enough for the following arguments. Let p, G and S be as in the statement of the lemma and let $G_1, \dots, G_5$ be independent copies of $G_{p/5}$. We will show that $G_1 \cup \ldots \cup G_5$ contains a triangle factor whp.

We begin with a claim that gives us a lot of structure. For $\eta > 0$ we will call a set $X \subseteq V(G)$ *η-strongly connected*[1] if $\overline{e}(X', X \setminus X') \le \frac{|X|^2}{4} - \eta n^2$ for all $X' \subseteq X$, where we denote by $\overline{e}(X, Y) = |X||Y| - e(X, Y)$ the number of non-edges between X and Y. Furthermore, we say that X is *η-close to complete* if $e(G[X]) \ge \left(\frac{1}{2} - \eta\right)|X|^2$ and $\deg(v; X) \ge \frac{1}{10}|X|$ for all $v \in X$.

[1] This definition might appear somewhat strange now but will assure that the reduced graph in this proof is connected.

Claim 4.7.9. *Whp there is a triangle matching $\mathcal{T}_1$ in $G_1 \cup G_2$ and disjoint sets $X_1, X_2 \subset V(G)$ so that*

(i) $X_1 \cup X_2 = V(G) \setminus V(\mathcal{T}_1)$ *and* $|X_1| = \frac{|X_2|}{2} = \left(\frac{1}{3} \pm \rho\right)n$,

(ii) $\deg(v; X_{3-i}) \geq (1-4\rho)|X_{3-i}|$ *for all* $i \in [2]$ *and* $v \in X_i$,

(iii) X_2 *is* $8d$*-strongly connected or there is a partition* $X_2 = X_{2,1} \cup X_{2,2}$ *so that, for each* $j \in [2]$, *we have that* $|X_{2,j}| \geq \frac{n}{4}$ *is even and* $X_{2,j}$ *is* $200d$*-close to complete.*

Proof of Claim: Let $Y_1 = \{v \in V(G) \setminus S : \deg(v; S) \leq \rho n\}$. Let $U_1 = S \cup Y_1$ and $U_2 = V(G) \setminus U_1$. With a similar (and simpler) proof to that of Claim 4.7.7, one can show that

(P1) $\deg(v; U_2) \geq |U_2| - 2\rho n$ for all $v \in U_1$ and $\deg(v; U_1) \geq \rho n$ for all $v \in U_2$,

(P2) $|U_1| = (\frac{1}{3} \pm \rho^6)n$ and $|U_2| = (\frac{2}{3} \pm \rho^6)n$, and

(P3) $d(U_1, U_2) \geq 1 - \rho^6$.

Let $\sigma = 10d$ and let $U_2 = U_{2,1} \cup U_{2,2}$ be the partition of U_2 which maximises $\overline{e}(U_{2,1}, U_{2,2})$. Throughout this proof, we will have to distinguish between two cases: either U_2 is σ-strongly-connected (this we will call the *connected case* from now on) or $\overline{e}(U_{2,1}, U_{2,2}) \geq \frac{|U_2|^2}{4} - \sigma n^2$ (which we call the *disconnected case*). Although the process is very similar for both, we will handle them separately, starting with the disconnected case.

The disconnected case. We claim that

(Q1) $|U_{2,j}| = \left(\frac{1}{3} \pm 2\sigma\right)n$ and $e(U_{2,j}) \geq \frac{1}{2}|U_{2,j}|^2 - 2\sigma n^2$ for both $j \in [2]$, and

(Q2) $\deg(v; U_{2,j}) \geq \frac{n}{10}$ for any $j \in [2]$ and $v \in U_{2,j}$.

Indeed, (Q1) follows from the case assumption and the fact that $\delta(G) \geq \frac{2n}{3}$, and (Q2) since $U_{2,1}, U_{2,2}$ are chosen to maximise non-edges in between (otherwise, moving a vertex violating (Q2) to the other set increases the count).

In a first round of probability (G_1), our goal is to balance the sizes. Assume first that $|U_1| > \frac{n}{3}$. Let $n_2 = 0$ if $|U_{2,1}|$ is even and $n_2 = 1$ otherwise, and let $n_3 = |U_1| - \frac{n}{3} - n_2 \geq 0$. Let $E = E(G[U_1])$, and observe that $\delta(E) \geq n_2 + n_3$. Furthermore, we have $\deg(e; U_{2,j}) \geq |U_{2,j}| - 10\rho n \geq \frac{n}{4}$ for both $j \in [2]$ by (P1) and (Q1). Therefore, by Lemma 4.6.4 (i), whp there is a triangle matching $\mathcal{T}_1'$ of size $n_2 + n_3 = |U_1| - \frac{n}{3}$ in G_1 with each triangle having two vertices in U_1 and one vertex in U_2 (n_2 have their third vertex in $U_{2,1}$ and n_3 have their third vertex in $U_{2,2}$). Let $U_i' = U_i \setminus V(\mathcal{T}_1')$ and $U_{2,j}' = U_{2,j} \setminus V(\mathcal{T}_1')$ for $i, j \in [2]$. By construction, we have $|U_2'| = 2|U_1'| = \frac{4n}{3} - 2|U_1| \geq 2\big(\frac{1}{3} - \rho^5\big)n$ and $|U_{2,j}'|$ is even for both $j \in [2]$.

Assume now that $|U_2| > \frac{2n}{3}$. Observe that for each $j \in [2]$ and $X \subseteq U_{2,j}$ of size $|X| \geq \frac{n}{9}$, we have $|K_3(G[X])| \geq \frac{n^3}{1000}$ by (Q1). Thus, by Lemma 4.6.1 (i), there are triangle matchings of size $\frac{n}{15}$ in each of $G_1[U_{2,j}]$ whp for both $j = 1, 2$. Thus, we can pick a triangle matching $\mathcal{T}_1'$ of exactly $\frac{n}{3} - |U_1|$ from these, again taking either one or no triangle in $U_{2,1}$ depending on its parity. By construction, we then have $|U_2'| = 2|U_1'| = 2|U_1| \geq 2\big(\frac{1}{3} - \rho^5, \big)n$ and $|U_{2,j}'|$ is even for both $j \in [2]$ (where U_i' and $U_{2,j}'$ are defined as above by removing the vertices of $\mathcal{T}_1'$ from the sets U_i and $U_{2,j}$).

Finally it remains to deal with the case that $|U_2| = 2|U_1| = \frac{2n}{3}$. Note that as $|U_2|$ is even in this case, we have that $|U_{2,1}|$ and $|U_{2,2}|$ have the same parity. If they are both even, there is no need to take any triangles in $\mathcal{T}_1'$ and we can move to the next stage. However, if they are odd in size, we have to do a little more work. We say a triangle T is *transversal* if $|V(T) \cap U_1| = |V(T) \cap U_{2,1}| = |V(T) \cap U_{2,2}| = 1$. We aim to prove the existence of a single transversal triangle in G_1. In order to do this, we first show that there are at least τn^2 transversal triangles in G. Indeed, without loss of generality suppose that $|U_{2,1}| < |U_{2,2}|$ and let $Y_0 \subseteq U_{2,1}$ be the set of vertices y in $U_{2,1}$ such that $\deg(y; U_1) \geq (1 - \rho^2)|U_1|$. Due to (P3), we have that $|Y_0| \geq \frac{n}{10}$. Now for each vertex $y \in Y_0$, as $|U_{2,1}| < |U_{2,2}|$ we have that y has some neighbour z in $U_{2,2}$ and due to (P1) and the fact that $y \in Y_0$, we have that $\deg(y, z; U_1) \geq \frac{\rho}{2}n$ and hence y is contained in at least $\frac{\rho}{2}n$ transversal triangles. Considering all $y \in Y_0$ thus gives the existence of τn^2 transversal triangles in G. A simple application of Janson's inequality (Lemma 2.1.3) gives that whp at least one of these transversal triangles survives in G_1 and so taking $\mathcal{T}_1'$ to be this single triangle, $U_i' = U_i \setminus V(\mathcal{T}_1')$ for $i \in [2]$ and $U_{2,j}' = U_{2,j} \setminus V(\mathcal{T}_1')$ for $j \in [2]$, we again have in this case that $|U_2'| = 2|U_1'| = 2(|U_1| - 1) \geq 2\big(\frac{1}{3} - \rho^5\big)n$ and $|U_{2,j}'|$ is even for both $j \in [2]$.

In a second round of probability (G_2), we will remove 'atypical' vertices in U_2'. From this point onwards, we will only remove triangles with one vertex in U_1' and two vertices in $U_{2,j}'$ for some $j \in [2]$, thus maintaining the right balance between U_1' and U_2' and the parity of $U_{2,1}'$ and $U_{2,2}'$. For $j \in [2]$, let $Y_{2,j} := \{v \in U_{2,j}' : \deg(v; U_1') \leq |U_1'| - \frac{\rho}{2}n\}$ and for each $v \in Y_{2,j}$ let $E_v := \{u_1u_2 : u_1 \in U_1', u_2 \in U_{2,j}' \setminus Y_{2,j}, vu_1u_2 \in K_3(G)\}$. It follows from (P3) (and counting non-edges between U_1 and U_2) that $|Y_{2,j}| \leq 2\rho^5 n$ for both $j \in [2]$. Furthermore, (P1) and (Q2) imply that $|E_v| \geq (\rho - \frac{\rho}{2})n \cdot \big(\frac{1}{10} - \rho^4\big)n \geq \rho^2 n$ for all $v \in Y_{2,1} \cup Y_{2,2}$. Thus, by Lemma 4.6.3, whp there is a triangle matching $\mathcal{T}_1''$ of size at most $4\rho^5 n$ in $G_2[U_1' \cup U_2']$ of the desired form (each triangle having one vertex in U_1' and two vertices in $U_{2,j}'$ for some $j \in [2]$) such that $Y_{2,1} \cup Y_{2,2} \subset V(\mathcal{T}_1'')$. Let $\mathcal{T}_1 = \mathcal{T}_1' \cup \mathcal{T}_1''$, $X_i = U_i' \setminus V(\mathcal{T}_1'')$ and $X_{2,j} = U_{2,j}' \setminus V(\mathcal{T}_1'')$ for each $i, j \in [2]$. These resulting sets have all the desired properties (i)-(iii).

The connected case. This case is very similar but less technical since we do not have to worry about the sets $U_{2,1}$ and $U_{2,2}$. We will therefore skip some details.

In a first round of probability (G_1), our goal is to balance the sizes. The case $|U_1| > \frac{n}{3}$ is completely analogous to the disconnected case and we find a triangle matching $\mathcal{T}_1'$ of size $|U_1| - \frac{n}{3}$ in G_1 with each triangle having two vertices in U_1 and one vertex in U_2. Let $U_i' = U_i \setminus V(\mathcal{T}_1')$ for $i \in [2]$. By construction, we have $|U_2'| = 2|U_1'| = \frac{4n}{3} - 2|U_1| \geq 2\left(\frac{1}{3} - \rho^5\right)n$.

Assume now that $|U_2| \geq \frac{2n}{3}$. Observe that for every set $Z \subset U_2$ with $|Z| \leq dn$ and every $v \in U_2 \setminus Z$, we have $\deg(v; U_2 \setminus Z) \geq \left(\frac{1}{3} - d\right)n$ and thus there are at least dn^2 edges in $N(v; U_2 \setminus Z)$. Indeed due to the fact that there is no set $S' \subseteq X_2$ with $|S'| \geq \left(\frac{1}{3} - 2\mu\right)n$ and $\Delta(G[S']) \leq 2dn$, we can find dn^2 edges by repeatedly removing high degree vertices from $N(v; U_2 \setminus Z)$ and taking the edges adjacent to them. Thus there are at least $\frac{d}{10}n^3$ triangles in $G[U_2 \setminus Z]$. It follows from Lemma 4.6.1 (i) that whp there are at least $\frac{d}{3}n$ vertex-disjoint triangles in $G_1[U_2]$. Let $\mathcal{T}_1'$ be a triangle matching consisting of exactly $\frac{n}{3} - |U_1|$ of these and let $U_i' = U_i \setminus V(\mathcal{T}_1')$ for $i = 1, 2$. By construction, we have $|U_2'| = 2|U_1'| = 2|U_1| \geq 2\left(\frac{1}{3} - \rho^5\right)n$.

The process of removing *bad* vertices v in U_2' such that $\deg(v; U_1') \leq |U_1'| - \frac{\rho}{2}n$ is analogous to (and simpler than) the disconnected case and an application of Lemma 4.6.3 gives a triangle matching $\mathcal{T}_1'' \subset K_3(G_2[U_1' \cup U_2'])$ containing all the bad vertices and such that defining $\mathcal{T}_1 = \mathcal{T}_1' \cup \mathcal{T}_1''$ and $X_i = U_i' \setminus V(\mathcal{T}_1'')$ for $i = 1, 2$, gives the required conditions for the claim. Here in order to verify condition (iii), we use that for any $X \subset X_2$, we have

$$\overline{e}(X, X_2 \setminus X) \leq \overline{e}(X, U_2 \setminus X) \leq \frac{|U_2|^2}{4} - 10dn^2 \leq \frac{|X_2|^2}{4} - 8dn^2,$$

using that $|U_2| - |X_2| \leq 3|V(\mathcal{T}_1)| \leq \rho n$. ■

The disconnected case now follows without much more work, as we show now. Let us first remove more atypical vertices of our near-cliques. For $j \in [2]$, let $Z_{2,j} := \{v \in X_{2,j} : \deg(v; X_{2,j}) \leq |X_{2,j}| - \sqrt{d}n\}$. Observe that, since $X_{2,j}$ is $200d$-close to complete, by counting non-edges in $X_{2,j}$ we have $|Z_{2,j}| \leq 10\sqrt{d}n$ for both $j \in [2]$. Note that any two vertices in X_2 have at least $\frac{n}{4}$ common neighbours in X_1 by Claim 4.7.9 (ii) and for $j \in [2]$, any vertex $v \in X_{2,j}$ has $\deg(v; X_{2,j} \setminus Z_{2,j}) \geq \frac{n}{50}$ by Claim 4.7.9 (iii) and our upper bound on $|Z_{2,j}|$. Hence it follows from Lemma 4.6.3 that whp (in G_3) there is a triangle matching $\mathcal{T}_2$ of size at most $20\sqrt{d}n$ in $G_3[X_1 \cup X_2]$ with each triangle having one vertex in X_1 and two vertices in X_2 (both of which are in the same $X_{2,j}$) covering $Z_{2,1} \cup Z_{2,2}$. Let $X_i' = X_i \setminus V(\mathcal{T}_2)$ and $X_{2,j}' = X_{2,j} \setminus V(\mathcal{T}_2)$ for each $i, j \in [2]$. Let $X_1' = X_{1,1}' \cup X_{1,2}'$ be a partition such that $|X_{1,j}'| = \frac{1}{2}|X_{2,j}'|$ for each $j \in [2]$ (note that here the parity of $|X_{2,j}'|$ is important). Now, for both $j \in [2]$, $X_{1,j}' \cup X_{2,j}'$ induces a $\left(d^{1/6}, \left(1 - d^{1/3}\right)^+\right)$-super-regular triple (after splitting $X_{2,j}'$ arbitrarily in two sets of equal sizes) by Lemma 2.2.5. Therefore, by Theorem II*, whp there are vertex-disjoint triangles in G_4 covering the remaining vertices.

Thus, we may assume that X_2 is $8d$-strongly connected. This case is very similar to the proof of Lemma 4.7.1. Let $n_i := |X_i|$ for both $i \in [2]$ and recall that $n_2 = 2n_1$. We apply Lemma 2.2.1

to $G[X_2]$ with input m_0, ε and fixing $\gamma := \frac{1}{2}-\varepsilon$ to get an ε-regular partition $X_2 = V_0 \cup V_1 \cup \ldots \cup V_m$ for some $m_0 \leq m \leq M_0$. Let R be the corresponding (ε, d)-reduced graph (seen as a graph on $[m]$) and observe that we have $\delta(R) \geq \left(\frac{1}{2} - 2d\right)m$ and, as in the proof of Lemma 4.7.1, we have $\alpha(R) < \left(\frac{1}{2} - \mu\right)m$. It is well-known that every graph H contains a matching of size $\min\{\delta(H), \lfloor\frac{v(H)}{2}\rfloor\}$. Indeed, if $v(H)$ is even this is the $k = 2$ case of Theorem 2.5.1, whilst if n is odd this can be derived from Theorem 2.5.1 by adding a vertex to H that is adjacent to all other vertices. We conclude that R contains a matching $\mathcal{M}^*$ of size $\left(\frac{1}{2} - 2d\right)m$; let R' be the subgraph of R induced by $M^* := V(\mathcal{M}^*)$. Note that $\delta(R') \geq \left(\frac{1}{2} - 6d\right)m$ and we claim that R' is connected. Indeed, if not, there is a set $B \subset V(R')$ such that $e(B, V(R') \setminus B) = 0$. Observe that $|B|, |V(R') \setminus B| \geq \delta(R') \geq \left(\frac{1}{2} - 6d\right)m$. Let now $X' := \bigcup_{h\in B} V_h$ and so that $|X'| = \left(\frac{1}{2} \pm 20d\right)|X_2|$. Furthermore, we have $e(X', X_2 \setminus X') \leq (d + 4d + 2\varepsilon)n^2$ and consequently

$$\overline{e}(X', X_2 \setminus X') \geq |X'||X_2 \setminus X'| - 6dn^2 \geq \left(\frac{|X_2|}{2} + 20d|X_2|\right)\cdot\left(\frac{|X_2|}{2} - 20d|X_2|\right) - 6dn^2 > \frac{|X_2|^2}{4} - 8dn^2,$$

contradicting the fact that X_2 is $8d$-strongly connected.

By Lemma 2.2.4, there are $V_h' \subset V_h$ for each $h \in M^*$ such that $|V_h'| = \lceil(1-2\varepsilon)|V_h|\rceil$ and, for every edge $h\ell \in \mathcal{M}^*$, the pair (V_h', V_ℓ') is $(2\varepsilon, (d-\varepsilon)^+, d-2\varepsilon)$-super-regular. Let $Y = X_2 \setminus \bigcup_{h\in M^*} V_h'$ be the set of vertices in X_2 which are not in a cluster V_h' corresponding to a vertex in an edge of $\mathcal{M}^*$. Observe that $|Y| \leq 2\varepsilon n + \varepsilon n + 4dn \leq 5dn$, where the terms in the upper bound come from bounding the number of vertices in sets $V_h \setminus V_h'$ for $h \in M^*$, the number of vertices in V_0 and number of vertices in a set V_h for $h \in [m] \setminus M^*$, respectively. Let $W \subset X_2 \setminus Y$ be a set such that

1. $|W \cap V_h'| = \left(\frac{1}{2} \pm \frac{1}{20}\right)\frac{n_2}{m}$ for each $h \in M^*$,
2. $\deg_G(v; W) \geq \frac{1}{3}|W|$ for each $v \in X_2$, and
3. we have that $\deg_G(v; V_h' \cap W) = \left(\frac{1}{2} \pm \frac{1}{4}\right)\deg_G(v; V_h')$ for each $h \in M^*$ and $v \in X_2$ with $\deg_G(v; V_h') \geq \varepsilon|V_h'|$.

Such a set W can be found by choosing each vertex of $X_2 \setminus Y$ independently with probability $\frac{1}{2}$ and applying Chernoff's inequality (Theorem 2.1.1) and a union bound.

We will start by covering Y. We will not touch vertices outside of W in order to maintain super-regularity properties.

Claim 4.7.10. *Whp in G_3, there is a triangle matching $\mathcal{T}_2 \subset K_3(G_1)$ of size $|Y|$ with each triangle having two vertices in $W \cup Y \subset X_2$ and one in X_1, so that $Y \subset V(\mathcal{T}_2)$ and $|V(\mathcal{T}_2) \cap V_h'| \leq 50\sqrt{d}|V_h'|$ for all $h \in M^*$.*

The proof is essentially identical to the proof of Claim 4.7.5 (appealing to Lemma 4.6.3) and we omit the details. Let now $X_i'' = X_i \setminus V(\mathcal{T}_2)$ for each $i \in [2]$ and let $V_h'' = V_h' \setminus V(\mathcal{T}_2)$ for each $h \in M^*$. We will now balance the sizes of the clusters V_h''.

Claim 4.7.11. *Whp in G_4, there is a triangle matching $\mathcal{T}_3 \subset K_3(G_4)$ with each triangle having one vertex in X_1'' and two vertices in W, so that $|V_h'' \setminus V(\mathcal{T}_3)| = \lfloor \frac{9}{10}\frac{n_2}{m} \rfloor$ for all $h \in M^*$.*

<u>Proof of Claim:</u> Let $\lambda : M^* \to \mathbb{N}$ be given by $\lambda(h) = |V_h''| - \lfloor \frac{9}{10}\frac{n}{m} \rfloor$. Note that we have $\left(\frac{1}{10} - 60\sqrt{d}\right)\frac{n_2}{m} \leq \lambda(h) \leq \lceil \frac{1}{10}\frac{n_2}{m} \rceil$, and that $\sum_{h \in M^*} \lambda(h) = n_2 - 2|\mathcal{T}_2| - 2|\mathcal{M}^*|\lfloor \frac{9}{10}\frac{n}{m} \rfloor$ is even. Note also that $\delta(R') \geq \left(\frac{1}{2} - 6d\right)m \geq \left(\frac{1}{2} - 6d\right)|R'|$ and $\alpha(R') \leq \alpha(R) \leq \left(\frac{1}{2} - \mu\right)m \leq \left(\frac{1}{2} - \frac{\mu}{2}\right)|R'|$. Hence, by applying Theorem 4.5.4 to the connected graph R', there is a weight function $\omega : E(R') \to \mathbb{N}$ such that for each $h \in M^*$ we have $\sum_{\ell \in N_{R'}(h)} \omega(h\ell) = \lambda(h)$. We claim that we can remove $\omega(h\ell)$ triangles from $G_4[X_1'', V_h'' \cap W, V_\ell'' \cap W]$ for each edge $h\ell$ of R', making sure that all our choices are vertex-disjoint. Indeed, let $Y_1, \ldots, Y_m \subset X_1''$ be disjoint sets of size at least $\frac{2}{5} \cdot \lceil \frac{n_2}{m} \rceil$ and observe that for $i = 1, 2$ we have that $\deg(v; X_{3-i}'') \geq |X_{3-i}''| - 4\rho n$ for each $v \in X_i''$ by Claim 4.7.9. Since $\rho \ll \frac{1}{m} \ll \varepsilon$, this implies that, for each $k \in [m]$ and $h \in M^*$, the pair $(Y_k, V_h'' \cap W)$ is $\left(\varepsilon, \left(1 - \varepsilon^2\right)^+\right)$-super-regular, appealing to Lemma 2.2.5. It further follows from the Slicing Lemma (Lemma 2.2.2) and the choice of W that $(V_h'' \cap W, V_\ell'' \cap W)$ is $(10\varepsilon, (d/10)^+)$-super-regular for each $h\ell \in E(R')$. Hence the triple $(Y_k, V_h'' \cap W, V_\ell'' \cap W)$ is $(10\varepsilon, (d/10)^+)$-super-regular for each $h\ell \in E(R')$ and $k \in [m]$. Furthermore, we have $|V_h'' \cap W| \geq \frac{2}{5} \cdot \frac{n_2}{m}$. Hence, an application of Lemma 2.2.7 and Lemma 4.6.1 (*ii*) implies that whp, there are $\frac{7}{20} \cdot \frac{n_2}{m}$ vertex-disjoint triangles in $G_4[Y_k, V_h'' \cap W, V_\ell'' \cap W]$ for each $h\ell \in E(R')$ and $k \in [m]$. Thus we can select the desired number of triangles for each $e \in E(R')$ one at a time greedily as follows. When we look to find a triangle corresponding to the edge $h\ell \in E(R')$ with $h < \ell$ (one of $\omega(h\ell)$ many), we take the triangle from $G_4[Y_h, V_h'' \cap W, V_\ell'' \cap W]$, ensuring that it is vertex-disjoint from previous choices. From above we have that there is a collection of at least $\frac{7}{20} \cdot \frac{n_2}{m}$ vertex-disjoint triangles in $G_4[Y_h, V_h'' \cap W, V_\ell'' \cap W]$ to choose from and at most $3\max\{\lambda(h), \lambda(\ell)\} \leq \frac{3}{10} \cdot \frac{n_2}{m} < \frac{7}{20} \cdot \frac{n_2}{m}$ are unavailable due to their vertices having already been used in triangles in our triangle matching. This shows that the greedy process will succeed in finding a triangle matching $\mathcal{T}_3$ in G_4 such that $\mathcal{T}_3$ contains $\omega(h\ell)$ triangles in $G_4[X_1'', V_h'' \cap W, V_\ell'' \cap W]$ for each edge $h\ell$ of R'. ■

Let now $X_i''' = X_i'' \setminus V(\mathcal{T}_3)$ for each $i \in [2]$ and $V_h''' = V_h'' \setminus V(\mathcal{T}_3)$ for all $h \in M^*$ and observe that we have covered all vertices except for those in $X_1''' \cup X_2'''$. Since $|X_1'''| = \frac{1}{2}|X_2'''|$, we can partition $X_1''' = \bigcup_{e \in \mathcal{M}^*} X_e'''$ into $|\mathcal{M}^*|$ sets of size exactly $\lfloor \frac{9}{10}\frac{n_2}{m} \rfloor$. Observe that $\deg(v; X_2''') \geq |X_2'''| - 4\rho n$ for each $v \in X_1'''$ and vice versa by Claim 4.7.9. Since $\rho \ll \frac{1}{m} \ll \varepsilon$, Lemma 2.2.5 implies that, for each $e \in \mathcal{M}^*$ and $h \in M^*$, the pair (X_e''', V_h''') is $\left(\varepsilon, \left(1 - \varepsilon^2\right)^+\right)$-super-regular. Furthermore, the pair (V_h''', V_ℓ''') is $(8\varepsilon, (d/8)^+)$-super-regular for each $h\ell \in \mathcal{M}^*$ by the Slicing Lemma (Lemma 2.2.2) and $\deg(v; V_\ell''') \geq \deg(v; V_\ell' \setminus W) \geq \frac{1}{4}\deg_G(v; V_\ell') \geq \frac{d}{8}|V_\ell'|$ for all $v \in V_h'''$ and vice versa. Therefore, $(X_{h\ell}''', V_h''', V_\ell''')$ is $(8\varepsilon, (d/8)^+)$-super-regular for all $h\ell \in \mathcal{M}^*$. Finally, we apply Theorem II* to each of these triples individually in G_5 to obtain whp a triangle matching $\mathcal{T}_4$ covering exactly $X_1''' \cup X_2'''$. So we have that whp all of the triangle matchings $\mathcal{T}_1, \ldots, \mathcal{T}_4$ exist and taking $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2 \cup \mathcal{T}_3 \cup \mathcal{T}_4$, we have that $\mathcal{T}$ is a triangle factor in G_p as required. $\square$

Chapter 5

Clique factors in randomly perturbed graphs

In this Chapter, we establish the perturbed thresholds for clique factors, proving Theorem III, which we restate here for convenience.

Theorem III. *(Restated) Let* $2 \leq k \leq r$ *be integers. Then given any* $1 - \frac{k}{r} < \tau < 1 - \frac{k-1}{r}$,

$$p(K_r, \tau) = n^{-2/k}.$$

Let us also recall the definition of a perturbed threshold for factors.

Definition 1.6.2. (Restated) Given some $0 \leq \tau \leq 1$, and a graph H with r vertices, the *perturbed threshold* $p(H, \tau)$ for an H-factor satisfies the following. There exists constants $C = C(H, \tau)$, $c = c(H, \tau) > 0$ such that:

(i) If $p = p(n) \geq Cp(H, \tau)$, then for *any* n-vertex graph G with $n \in r\mathbb{N}$ and $\delta(G) \geq \tau n$, whp $G \cup G(n, p)$ contains an H-factor.

(ii) If $p = p(n) \leq cp(H, \tau)$, then for all $n \in r\mathbb{N}$ there is *some* n-vertex graph G with $\delta(G) \geq \tau n$ such that whp $G \cup G(n, p)$ does not contain an H-factor.

If it is the case that for sufficiently large $n \in r\mathbb{N}$, every n-vertex graph with minimum degree at least τn contains an H-factor we define $p(H, \tau) := 0$.

The lower bounds on perturbed thresholds in Theorem III follow from a simple construction which we detail in Section 5.1. The main difficulty of Theorem III is to prove the upper bounds on the perturbed thresholds and the rest of this chapter is devoted to proving these. In

Section 5.2, we give an overview of the proof, discussing how we use the absorption method to split the problem into finding an absorbing structure and an almost factor in $G \cup G(n,p)$. We also discuss how we will split the edges of (copies of K_r in) our desired factor between the deterministic dense graph G and the random graph $G(n,p)$. In Section 5.3, we give a technical result which will be used to embed subgraphs in $G(n,p)$. In Section 5.4, we then provide the details of how to find an almost factor in $G \cup G(n,p)$. Sections 5.5 and 5.6 are then devoted to proving the existence of an absorbing structure in $G \cup G(n,p)$. There is a lot of work to be done here and slightly different approaches are necessary to deal with different values of k and r. In Section 5.5 we deal only with edges coming from our deterministic graph G, finding many copies of certain subgraphs of our desired absorbing structure. In Section 5.6, we then introduce random edges to 'fill in the gaps' and obtain an absorbing structure in $G \cup G(n,p)$. Finally, in Section 5.7 we bring everything together to prove Theorem III.

5.1 Lower bound construction

In this section we give a construction that provides the lower bound in the proof of Theorem III. Our construction is a generalisation of that used for the lower bound in Theorem 1.6.1 (see Section 2.1 of [19]). We will make use of the following result, which can be readily shown using the second moment method, see for example [95, Remark 3.7].

Theorem 5.1.1. *For every $k \geq 2$ and for every $0 < \varepsilon < 1$ there exists a constant $c > 0$ such that if $p \leq cn^{-2/k}$, then*

$$\lim_{n\to\infty} \mathbb{P}[G(n,p) \text{ contains a } K_k\text{-matching of size } \varepsilon n] = 0.$$

Let k and r be as in the statement of Theorem III. Consider any $1-\frac{k}{r} < \tau < 1-\frac{k-1}{r}$ and let $\gamma > 0$ such that $(1-\gamma)(1-\frac{k-1}{r}) = \tau$. Now to show that $p(K_r,\tau) \leq n^{-2/k}$ we choose $0 < c \ll \varepsilon \ll \frac{1}{r}, \gamma$, fix some $p = p(n) \leq cn^{-2/k}$ and let $n \in r\mathbb{N}$. We will show that there exists an n-vertex graph G with $\delta(G) \geq \tau n$, such that whp $G \cup G(n,p)$ does not contain a K_r-factor. Note that we can assume that n is sufficiently large, as our desired conclusion is a "whp" statement. We define G to have vertex classes A and B such that $|B| = \lceil \tau n \rceil + 1$ and $|A| = n - |B|$, with all possible edges in G except that A is an independent set. So $\delta(G) \geq |B| - 1 \geq \tau n$.

Before proving formally that whp $G \cup G(n,p)$ does not contain K_r-factor, let us sketch the idea. Suppose for a contradiction that there is K_r-factor $\mathcal{K}$ in $G \cup G(n,p)$ and consider the average intersection of a copy of K_r in $\mathcal{K}$ with the set A. As A covers more than a $\frac{k-1}{r}$ proportion of the vertices, we have that the average intersection size of a copy of K_r in $\mathcal{K}$ with A must be strictly larger than $k-1$. However the only way this can happen is if, when we look at the intersections of cliques in $\mathcal{K}$ with A, we see linearly many cliques of size at least k. This will contradict the conclusion of Theorem 5.1.1. Let us now formally prove this.

From our choice of $c > 0$, we have from Theorem 5.1.1, that whp $G(n,p)[A] \cong G(|A|,p)$ does not contain a K_k-matching of size $\frac{\varepsilon n}{r}$. Here, we used that $|A| \geq \frac{n}{r}$ is linear in n. Observe that any copy of K_r in $G \cup G(n,p)$ either contains a K_k in A, or uses at least $r-(k-1)$ vertices of B. Thus, whp, the largest K_r-matching in $G \cup G(n,p)$ has size less than

$$\frac{|B|}{r-k+1} + \frac{\varepsilon n}{r} \leq \frac{(1-\gamma)n}{r} + \frac{2}{r-k+1} + \frac{\varepsilon n}{r} \leq \left(1 - \frac{\gamma}{2}\right)\frac{n}{r} < \frac{n}{r},$$

using that n is sufficiently large here. This concludes the proof of the lower bounds in Theorem III.

We remark that when $\tau < \frac{1}{2}$, the construction is perhaps a bit surprising as there is no need to include the edges within B in order to meet the minimum degree condition. However in this case we have that $k > \frac{r}{2}$ and so, if p is close to $n^{-2/k}$, copies of K_{r-k} will be in abundance in $G(n,p)[B]$. Therefore one would easily be able to extend copies of K_k in A to full copies of K_r in $G \cup G(n,p)$ even if B was empty in G. This shows that the determining factor for all such constructions (with a complete bipartite graph between A and B and A being independent) is the presence of large K_k matchings in the set A, as in our proof above.

5.2 Proof overview

It remains to prove the upper bounds in Theorem III and in this section we sketch some of the ideas involved in the proof, which will be the subject of the remainder of the chapter. So fix some $k, r \in \mathbb{N}$ and $\tau > 0$ as in Theorem III, let $C = (r, k, \tau) > 0$ be some sufficiently large constant and fix some $p \geq Cn^{-2/k}$. Now our aim is to show that for an n-vertex graph G with $n \in r\mathbb{N}$ and $\delta(G) \geq \tau n$, we have that whp $G \cup G(n,p)$ contains a K_r-factor. As with the proof of Theorem I in Chapter 3, we will use the *absorption method* to find a K_r-factor in $G \cup G(n,p)$. As discussed in Section 2.8, this reduces the problem into finding an *absorbing structure* on some vertex subset A (which contains a small constant proportion of the vertices of G) and finding an almost K_r-factor that leaves a leftover set of $o(n)$ vertices uncovered.

Now before addressing these two subproblems, let us discuss a much easier task: finding a copy of K_r in $G \cup G(n,p)$. Note that it might be the case that both G and $G(n,p)$ are K_r-free (whp for $G(n,p)$). Thus, to build even a single copy of K_r, we may have to use both deterministic edges (from G) and random edges (from $G(n,p)$). In order to do this, we define the following subgraph of K_r.

Definition 5.2.1. For $r \in \mathbb{N}$ and $k \in \mathbb{N}$ such that $2 \leq k \leq r$, let $r^*, q \in \mathbb{N}$ be such that $k(r^*-1) + q = r$ and $0 < q \leq k$. Then $H_{\det} = H_{\det}(r,k)$ is the complete r^*-partite graph with parts $V_1, V_2, \ldots, V_{r^*}$ such that $|V_1| = |V_2| = \ldots = |V_{r^*-1}| = k$ and $|V_{r^*}| = q$, i.e. $H_{\det} := K^{r^*}_{k,\ldots,k,q}$. We also define $\overline{H_{\det}}$ to be $K_r \setminus H_{\det}$, i.e. the complement of $H_{\det}$ on the same vertex set.

Some examples are given in Figure 5.1. Note that when $k = r$, $H_{\det}$ is simply an independent set of size $k = r$ and $\overline{H_{\det}}$ is an r-clique. The motivation for this partition comes from the following observation. We can build a copy of K_r in $G \cup G(n,p)$ by taking $\Omega(n^r)$ copies of $H_{\det}$ in G and then applying Janson's inequality (Lemma 2.1.3) to conclude that we can 'fill up' the independent sets in some copy of $H_{\det}$ by K_ks and a K_q and obtain a copy of K_r. With a few more ideas, one can repeatedly apply this naïve idea to greedily obtain an almost K_r-factor (see Theorem 5.4.1).

The task of defining and building an absorbing structure in $G \cup G(n,p)$ is much more involved and will be done in several steps. The first step is to find small building blocks for the absorbing structure which we will call *reachable paths*. Given two vertices u, v, a reachable path will be a (constant size) subgraph P of G containing both u and v and such that both $P \setminus u$ and $P \setminus v$ host K_r-factors. The simplest example of this is finding a copy of K^-_{r+1} with u and v playing the rôle of the *non-edge*. We will also use *longer* paths which are defined by gluing copies of K^-_{r+1} together sequentially at vertices of degree $r-1$. We remark that these reachable paths are exactly the K_r-diamond trees as used in Chapter 3 (see Definition 3.1.1), restricting to diamond trees whose auxiliary trees are paths. However, our focus in this chapter is different and, in particular, the *end-points* of the paths (u and v above) will play a crucial rôle. Indeed, our argument starts by first proving that u and v are *reachable*, meaning that there are many reachable paths relative to u and v, for every pair of vertices $u, v \in V(G)$. Once this is established, we will be able to greedily piece together reachable paths into larger (but still constant size) *absorbing gadgets*. Finally, we will piece together absorbing gadgets to build a linear sized *absorbing structure*. At this point we will use the *template absorption method* of Montgomery [136, 137] (see Section 2.8) in order to define the absorbing structure. That is, we will use an auxiliary *template* (Definition 2.8.1) to dictate how we interweave our absorbing gadgets, which will ensure that the resulting absorbing structure has a strong absorbing property, in that it can contribute to a K_r-factor in many ways.

Note that for all of this to work, we need to rely on both edges in G and $G(n,p)$ to find our absorbing structure. Indeed, even for the most basic building blocks, the reachable paths, we will need to build certain subgraphs with deterministic edges and 'fill in the gaps' with random edges. For this, we will need much more involved arguments than finding singular copies of K_r as discussed above. Indeed, our reachable paths (crucially) contain copies of K^-_{r+1} and we will also be interested in finding reachable paths relative to *fixed* endpoints. This complicates matters and in order to guarantee that both the deterministic and random parts of the reachable paths can be found (with sufficient abundance) in $G \cup G(n,p)$, we have to adopt several different approaches.

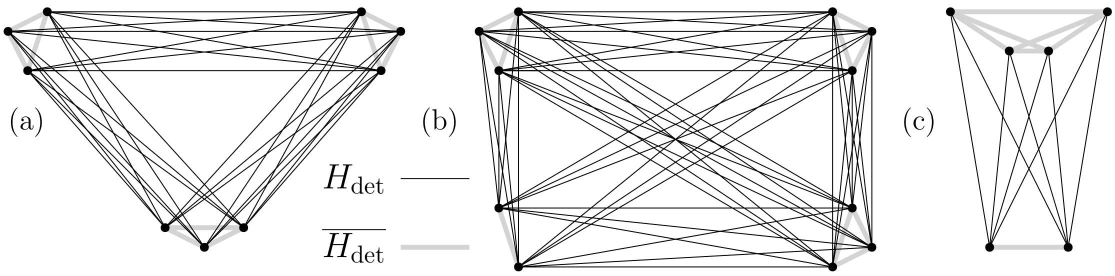


FIGURE 5.1: Some examples of the graphs H_{det} and $\overline{H_{\text{det}}}$ from Definition 5.2.1 for each case:

1. (a) $r = 9, k = q = r^* = 3$,
2. (b) $r = 11, k = 3, q = 2, r^* = 4$,
3. (c) $r = 6, k = 4, q = r^* = 2$.

Our analysis splits into three cases depending on the structure of H_{det} or equivalently, the values of r and k. The cases are as follows:

1. H_{det} is balanced i.e. $q = k$ and so $\frac{r}{k} \in \mathbb{N}$,
2. $\chi(H_{\text{det}}) = r^* \geq 3$ and H_{det} is not balanced i.e. $k < \frac{r}{2}$ and $\frac{r}{k} \notin \mathbb{N}$,
3. $\chi(H_{\text{det}}) = r^* = 2$ and H_{det} is not balanced i.e. $\frac{r}{2} < k < r$.

Examples of each case can be seen in Figure 5.1. By carefully partitioning edges between G and $G(n, p)$ depending on the case we are in, we will be able to show that there are reachable paths between pairs of vertices. When $k > \frac{r}{2}$ (case 3), there is an added complication as we cannot prove the reachability for *every* two vertices and have to pursue a weaker property, namely, building a partition of $V(G)$ such that the reachability can be established within each part. This stems from the fact that in this range, G may be disconnected.

Ultimately then, we can run a similar argument in case 3 and establish the existence of reachable paths in $G \cup G(n, p)$ for all cases. However, in order to go on and build absorbing gadgets and eventually, an absorbing structure, we need the existence of *many* reachable paths between a pair of vertices u and v. As soon as we introduce random edges, the number of reachable paths will be too small to run the greedy arguments used to build the absorbing structure. In order to handle this, we will introduce the random edges of $G(n, p)$ only in the *very last stage*, when proving the existence of the full absorbing structure in $G \cup G(n, p)$. Thus, we will first be occupied with finding many reachable paths and absorbing gadgets which use these reachable paths, restricting our attention *only* to the deterministic edges which will contribute to our eventual absorbing structure. Therefore all our notions of reachable paths and absorbing gadgets will be defined relative to certain subgraphs of K^-_{r+1} that we will find using deterministic edges. In order to motivate our definitions, we will keep track of the 'gaps' in reachable paths and absorbing gadgets, that we will eventually fill with random edges. The details of this process of defining and finding an absorbing structure, are given in Sections 5.5 and 5.6 but before embarking on this, we provide some probabilistic machinery in Section 5.3 and prove the existence of an almost K_r-factor in Section 5.4. Theorem III will then be proven in Section 5.7.

5.3 Embedding subgraphs in random graphs

In this section, we will discuss embeddings of subgraphs in $G(n, p)$. The main result is a technical lemma (Lemma 5.3.2) which will be crucial to our proof of Theorem III. This technical lemma will follow from Janson's inequality (Lemma 2.1.2) but provides us with much more than the immediate consequence (Lemma 2.1.3) for embedding subgraphs in $G(n, p)$. These strengthenings are necessary as we will need to embed subgraphs with certain vertices already fixed. We will also need to consider families of different subgraphs at the same time and we will need to have flexibility, being able to conclude the existence of many embeddings in $G(n, p)$, allowing us to find an embedding that avoids certain vertices and belongs to any large enough specified subset of the embeddings we are interested in. All of this is provided by Lemma 5.3.2, leading to its technical nature. Before embarking on Lemma 5.3.2, we define a graph parameter that will simplify our exposition.

Recall Janson's inequality (Lemma 2.1.2) and consider the random graph $G(n, p)$ on an n-vertex set V. Note that we can view $G(n, p)$ as Λ_p with $\Lambda := \binom{V}{2}$. Following [95], for a fixed graph F, we define $\Phi_F = \Phi_F(n, p) := \min\{n^{v_H} p^{e_H} : H \subseteq F, e_H > 0\}$. This parameter helps to simplify calculations of Δ_X in the context of counting the number of embeddings of the graph F in $G(n, p)$. As previously mentioned, we will also be interested in the appearance of graphs in $G(n, p)$ where we require some subset of vertices to be already fixed in place. Therefore, for a graph F, and some independent[1] subset of vertices $W \subset V(F)$, we define

$$\Phi_{F,W} = \Phi_{F,W}(n, p) := \min\{n^{v_H - v_{H[W]}} p^{e_H} : H \subseteq F, e_H > 0\}.$$

Recall that $F \setminus W$ denotes $F[V(F) \setminus W]$ and note that $\Phi_F = \Phi_{F,\emptyset}$ and $\Phi_{F\setminus W} \geq \Phi_{F,W}$ for any F and independent set $W \subset V(F)$. If $W = \{w\}$ for a single vertex $w \in V(F)$, we drop the set brackets and simply write $\Phi_{F,w}$ and $\Phi_{F\setminus w}$. Let us collect some more simple observations concerning Φ_F and $\Phi_{F,W}$ which will be useful later.

Lemma 5.3.1. *The following hold for all $n \in \mathbb{N}$:*

1. *Let $C > 1$ be some constant, $2 \leq k \in \mathbb{N}$ and $p = p(n) \geq Cn^{-2/k}$. Further, let $k' \in \mathbb{N}$ with $2 \leq k' \leq k$ and $F_1 := K_{k'}$, then we have that $\Phi_{F_1} \geq Cn$.*

[1] With respect to F i.e. $E(F[W]) = \emptyset$.

2. *As above, let $C > 1$ be some constant, $2 \leq k \in \mathbb{N}$ and $p = p(n) \geq Cn^{-2/k}$. Suppose now $k' \in \mathbb{N}$ with $3 \leq k' \leq k$ and let $F_2 := K^-_{k'}$ be the complete graph on k' vertices with one edge missing. Further, let $w \in V(F_2)$ be one of the endpoints of the missing edge. Then $\Phi_{F_2} \geq Cn$ and $\Phi_{F_2,w} \geq \min\{Cn^{1-2/k}, Cn^{2/k}\} \geq Cn^{1/k}$.*

3. *Let F_3, F_4 be graphs with vertex subsets $W_3 \subset V(F_3)$, $W_4 \subset V(F_4)$, let $\Phi_3 := \Phi_{F_3,W_3}$ and $\Phi_4 := \Phi_{F_4,W_4}$ and suppose that $\Phi_3, \Phi_4 \geq 1$. Let F_5 be the graph formed by the union of F_3 and F_4 meeting in exactly one vertex $x \in (V(F_3) \setminus W_3) \cap (V(F_4) \setminus W_4)$, and let F_6 be the graph obtained by taking a disjoint union of F_3 and F_4. Then letting $W_5 := W_3 \cup W_4$, we have that $\Phi_{F_5,W_5} \geq \min\{\Phi_3, \Phi_4, \Phi_3\Phi_4 n^{-1}\}$ and $\Phi_{F_6,W_5} = \min\{\Phi_3, \Phi_4\}$.*

Proof. For parts 1 and 2, it suffices to consider the case $k' = k$. For part 1, we have a simple calculation. Let H be a subgraph of K_k with v_H vertices and $e_H > 0$ edges. As $v_H \leq k$, we obtain

$$n^{v_H} p^{e_H} \geq n^{v_H} (Cn^{-2/k})^{\frac{v_H(v_H-1)}{2}} \geq Cn^{v_H-(v_H-1)} = Cn.$$

For part 2 first note that as $F_2 \subseteq K_k = F_1$ we have that $\Phi_{F_2} \geq \Phi_{F_1} \geq Cn$. Let H be a subgraph of K^-_k with $e_H > 0$. If $w \notin V(H)$, the calculation from part 1 gives that $n^{v_H} p^{e_H} \geq Cn$. So suppose $w \in V(H)$. Now let us distinguish two cases, depending on whether the vertex u is in $V(H)$, where u is the vertex in K^-_k such that uw is a *non*-edge. If $u \in V(H)$, we have that

$$n^{v_H-1} p^{e_H} \geq n^{v_H-1} Cn^{-2/k\left(\frac{v_H(v_H-1)}{2}-1\right)} \geq Cn^{v_H-1-(v_H-1-2/k)} \geq Cn^{2/k},$$

again using that $v_H \leq k$. Likewise, if $u \notin V(H)$, we have that

$$n^{v_H-1} p^{e_H} \geq n^{v_H-1} p^{\binom{v_H}{2}} \geq Cn^{(v_H-1)(1-v_H/k)} \geq Cn^{1-2/k},$$

where the last inequality follows as $(v_H - 1)\left(1 - \frac{v_H}{k}\right)$ is minimised in the range $2 \leq v_H \leq k-1$ at $v_H = 2$ and $v_H = k-1$. This shows that $\Phi_{F_2,w}$ is bounded as desired.

Part 3 also follows from the definition. Indeed, suppose a subgraph H of F_5 with $e_H > 0$ is a minimiser of the term in the definition of Φ_{F_5,W_5} and note that H can be expressed as $H = H_3 \cup H_4$ with H_3 and H_4 being edge-disjoint subgraphs of H and each H_i being a subgraph of F_i for $i = 3, 4$. Now for $i = 3, 4$, define $\Psi_i := n^{v_{H_i}-v_{H_i[W_i]}} p^{e_{H_i}}$ and note that $\Psi_i \geq \Phi_i$ if $e_{H_i} > 0$. If $e_{H_3} = 0$, then $e_{H_4} = e_H > 0$ and $\Phi_{F_5,W_5} = \Psi_4 \geq \Phi_4$. Similarly, if $e_{H_4} = 0$, then $\Phi_{F_5,W_5} = \Psi_3 \geq \Phi_3$. So we can suppose that $e_{H_3}, e_{H_4} > 0$. In this case we have that

$$\Phi_{F_5,W_5} = n^{v_H - v_{H[W]}} p^{e_H} \geq \Psi_3\Psi_4 n^{-1} \geq \Phi_3\Phi_4 n^{-1},$$

where the n^{-1} term comes from the fact that x might be counted as a vertex of both H_4 and H_5. Establishing that $\Phi_{F_6,W_5} = \min\{\Phi_3, \Phi_4\}$ is essentially identical, except that we do not have to worry about the vertex x and in the case that $e_{H_i} > 0$ for each $i = 3, 4$, we get that $\Phi_{F_6,W_6} \geq \Psi_3\Psi_4 \geq \min\{\Phi_3, \Phi_4\}$, using that $\Phi_3, \Phi_4 \geq 1$ here. $\square$

We now turn to a technical lemma for embedding constant-sized graphs in $G(n, p)$. Lemma 5.3.2 provides the basis for a greedy process in which we find some larger (linear size) graph in $G(n, p)$. We will require that the embedding of our larger graph has certain vertices already prescribed and repeated applications of Lemma 5.3.2 will then allow us to embed the remaining vertices of the graph in a greedy manner. So it is crucial that we can apply the lemma to *any* subset of s (remaining) indices while avoiding *any* small enough set of (previously used) vertices from being used.

Lemma 5.3.2. *Let $n, t(n), s(n) \in \mathbb{N}$, $0 < \beta < \frac{1}{2}$ and $L, v, w, e \in \mathbb{N}$ such that $Lt, sw \leq \frac{\beta n}{4v}$ and $\binom{t}{s} \leq 2^n$. Let $F_1, \dots, F_t$ be labelled graphs with distinguished vertex subsets $W_i \subset V(F_i)$ such that $|W_i| \leq w$, $|V(F_i \setminus W_i)| = v$, $e(F_i) = e$ and $e(F_i[W_i]) = 0$ for all $i \in [t]$. Now let V be an n-vertex set and let $U_1, \dots, U_t \subset V$ be labelled vertex subsets with $|U_i| = |W_i|$ for all $i \in [t]$. Finally, suppose there are families $\mathcal{F}_1, \dots, \mathcal{F}_t \subset \binom{V}{v}$ of labelled vertex sets such that for each $i \in [t]$, $|\mathcal{F}_i| \geq \beta n^v$.*

Now suppose that $1 \leq s(n) \leq t(n)$ and $p = p(n)$ are such that

$$s \cdot \Phi \geq \left(\frac{2^{v+7}v!}{\beta^2}\right) \min\{Lt \log n, n\} \text{ and } \Phi' \geq \left(\frac{2^{v+7}v!}{\beta^2}\right) n, \tag{5.3.1}$$

where $\Phi := \min\{\Phi_{F_i,W_i} : i \in [t]\}$ and $\Phi' := \min\{\Phi_{F_i \setminus W_i} : i \in [t]\}$ with respect to $p = p(n)$. Then, whp, for any $V' \subseteq V$, with $|V'| \geq n - Lt$ and any subset $S \subseteq [t]$ such that $|S| = s$ and $U_i \cap U_j = \emptyset$ for $i \neq j \in [s]$, there exists some *$i \in S$ such that there is an embedding (which respects labelling) of F_i in $G(n, p)$ on V which maps W_i to U_i and $V(F_i) \setminus W_i$ to a labelled set in $\mathcal{F}_i$ which lies in V'.*

Note by 'labelled' here we mean that for all j, the j^{th} vertex in W_i is mapped to the j^{th} vertex in U_i; moreover the j^{th} vertex in $V(F_i) \setminus W_i$ is mapped to the j^{th} vertex in some labelled set from $\mathcal{F}_i$.

Proof. Let us fix $S \subseteq [t]$ with $|S| = s$ and a vertex subset $V' \subset V$ as in the statement of the lemma. Let $U := \cup_{i \in S} U_i$ and fix $V'' := V' \setminus U$. Note that $(V \setminus V') \cup U$ intersects at most $\frac{\beta}{2} n^v$ of the elements of $\mathcal{F}_i$ for each i and so we can focus on a subset $\mathcal{F}_i'$ of each $\mathcal{F}_i$ of at least $\frac{\beta}{2} n^v$ sets which are all contained in V''. For each $i \in S$ and each labelled subset $X \in \mathcal{F}_i'$, let $I_{X,i}$ denote the indicator random variable that $X \cup U_i$ hosts a labelled copy of F_i with W_i mapped to U_i. To ease notation sometimes we write I_X instead of $I_{X,i}$. Note that $Z := \sum\{I_{X,i} : X \in \cup_{i \in S} \mathcal{F}_i'\}$ counts the number of suitable embeddings in $G(n, p)$. (So here if X is in a of the collections $\mathcal{F}_i'$, then there are a indicator random variables in this sum corresponding to X.)

An easy calculation (using the first part of (5.3.1)) gives that $\mathbb{E}[Z] > 2$ for large enough n. We will show that

$$\Delta_Z \leq \frac{\mathbb{E}[Z]^2}{16\min\{Lt\log n, n\}}, \tag{5.3.2}$$

and thus by Janson's inequality (2.1.1), $\mathbb{P}\big[Z \leq \frac{\mathbb{E}[Z]}{2}\big] \leq \exp(-2\min\{Lt\log n, n\})$. If $Lt\log n \leq n$, taking a union bound over the (at most 2^t) possible sets S and the $\binom{n}{Lt} \leq \exp(Lt(1+\log n))$ possible V', we have that whp, $Z \geq 1$ for all such S and V'; if $Lt\log n > n$, we instead bound both the number of V' and the number of S by 2^n (using that $\binom{t}{s} \leq 2^n$ here) and draw the same conclusion. So in both cases $Z \geq 1$ whp for all such S and V' and we are done.

Now it remains to verify (5.3.2). Firstly let $Z_i := \sum\{I_{X,i} : X \in \mathcal{F}_i'\}$. Then

$$\mathbb{E}[Z]^2 = \left(\sum_{i\in S}\mathbb{E}[Z_i]\right)^2 = \sum_{i,j\in S}\mathbb{E}[Z_i]\mathbb{E}[Z_j]. \tag{5.3.3}$$

To ease notation, let $\mathcal{F} := \cup_{i\in S}\mathcal{F}_i'$ and for $X, X' \in \mathcal{F}$, we write $X \sim X'$ if, assuming $X \in \mathcal{F}_i$, $X' \in \mathcal{F}_j$, the labelled copies of F_i on $X \cup U_i$ and F_j on $X' \cup U_j$ intersect in at least one edge. We split Δ_Z as follows:

$$\begin{aligned}\Delta_Z &= \sum_{\{(X,X')\in\mathcal{F}\times\mathcal{F}: X\sim X'\}}\mathbb{E}[I_X I_{X'}] \\ &= \sum_{i\in S}\Delta_{Z_i} + \sum_{\{(i,j)\in S^2:\, i\neq j\}}\ \sum_{\{(X,X')\in\mathcal{F}_i'\times\mathcal{F}_j': X\sim X'\}}\mathbb{E}[I_X I_{X'}],\end{aligned} \tag{5.3.4}$$

where Δ_{Z_i} is defined analogously to Δ_Z for the random variable Z_i.

Recall that for integers a and b, write $a!_b := a(a-1)\cdots(a-b+1)$. Fix $1 \leq k \leq v$. There are $\binom{v}{k}v!_k \leq \binom{v}{k}v!$ ways that two labelled v-sets share exactly k vertices. Fixing two such v-sets, there are at most $n!_{(2v-k)} \leq n^{2v-k}$ ways of mapping their $2v-k$ vertices into V. Let f_k denote the maximum number of edges of a k-vertex subgraph of $F_i \setminus W_i$, taken over all $i \in [t]$. Then we have that for $i \neq j \in S$,

$$\sum_{\{(X,X')\in\mathcal{F}_i'\times\mathcal{F}_j': X\sim X'\}}\mathbb{E}[I_X I_{X'}] \leq \sum_{k=1}^{v}\binom{v}{k}v!\,n^{2v-k}p^{2e-f_k} \leq \frac{2^v v!\,n^{2v}p^{2e}}{\Phi'} \leq \frac{2^{v+2}v!\mathbb{E}[Z_i]\mathbb{E}[Z_j]}{\beta^2\Phi'}.$$

Here, we crucially used that any copy of F_i on $X \in \mathcal{F}_i'$ does not have edges intersecting U_j for $j \neq i$. Note that the penultimate inequality follows by definition of Φ'. The last inequality follows as $\frac{\beta}{2}n^v p^e \leq \mathbb{E}[Z_i]$ for all $i \in S$.

Using the above calculation (and the second part of (5.3.1)) to compare (5.3.4) and (5.3.3), we see that the right-hand summand of (5.3.4) is less than $\frac{\mathbb{E}[Z]^2}{32n}$. We now estimate the left-hand summand of (5.3.4) in a similar fashion. For a fixed $i \in S$, let $1 \leq k \leq v$. We let g_k denote the maximum number of edges of a subgraph of F_i which has k vertices disjoint from W_i. We have, similarly to before, that

$$\Delta_{Z_i} \leq \sum_{k=1}^{v} \binom{v}{k} v!_k n!_{(2v-k)} p^{2e-g_k} \leq \sum_{k=1}^{v} \binom{v}{k} v!\, n^{2v-k} p^{2e-g_k} \leq \frac{2^v v!\, n^{2v} p^{2e}}{\Phi}.$$

Thus,

$$\sum_{i \in S} \Delta_{Z_i} \leq \frac{s 2^v v! n^{2v} p^{2e}}{\Phi} \overset{(5.3.1)}{\leq} \frac{\left(\frac{s\beta n^v p^e}{2}\right)^2}{32 \min\{Lt \log n, n\}}$$
$$\leq \frac{\left(\sum_{i \in S} \mathbb{E}[Z_i]\right)^2}{32 \min\{Lt \log n, n\}} = \frac{\mathbb{E}[Z]^2}{32 \min\{Lt \log n, n\}}.$$

So bringing both summands together, (5.3.2) holds and we are done. □

In its full generality, Lemma 5.3.2 will be a valuable tool in our proof. However, we will also have instances where we do not need to use the full power of the lemma. For instance, setting $s = 1$ and $W_i = U_i = \emptyset$ for all $i \in [t]$ (and $L = 0$), we recover a more standard application of Janson's inequality to subgraph containment which we state below for convenience.

Corollary 5.3.3. *Let $\beta > 0$, $1 \leq t \leq 2^n$ and F some fixed labelled graph on v vertices. Then there exists $C > 0$ such that the following holds. If V is a set of n vertices, $\mathcal{F}_1, \ldots, \mathcal{F}_t \subset \binom{V}{v}$ are families of labelled subsets such that $|\mathcal{F}_i| \geq \beta n^v$ and $p = p(n)$ is such that $\Phi_F \geq Cn$, then whp, for each $i \in [t]$, there is an embedding of F onto a set in $\mathcal{F}_i$, which respects labellings.*

5.4 An almost factor

In this section we study almost factors and prove Theorem 5.4.1 below. As is the case throughout, in this almost factor, the edges of G which contribute to the copies of K_r will be copies of H_{det} as defined in Definition 5.2.1. We will rely on $G(n, p)$ to then 'fill in the gaps', providing the missing edges i.e. $\overline{H_{\text{det}}}$, to guarantee that each copy of H_{det} is in fact part of a copy of K_r in $G \cup G(n, p)$. Note that $\chi(H_{\text{det}}) = \lceil \frac{r}{k} \rceil$ and recall the definition of χ_{cr} discussed in Section 2.5. When k divides r, we have $\chi_{cr}(H_{\text{det}}) = \frac{r}{k} = \chi(H_{\text{det}})$ and when k does not divide r, we have $\chi_{cr}(H_{\text{det}}) = \lfloor \frac{r}{k} \rfloor \frac{r}{r - \left(r - k \lfloor \frac{r}{k} \rfloor\right)} = \frac{r}{k}$.

Thus, the almost factor result of Komlós, Theorem 2.5.2, guarantees the existence of a partial H_{det}-factor in G which covers almost all the vertices. However, given such a partial factor we cannot guarantee that the correct edges appear in $G(n, p)$ in order to extend each copy of H_{det} in the partial factor to a copy of K_r. We aim instead to greedily build a partial K_r-factor and guarantee that at each step there are $\Omega(n^r)$ copies of H_{det}. To achieve this, we use the Regularity Lemma (Lemma 2.2.1) and apply Theorem 2.5.2 to the reduced graph of G. Then by the Counting

Lemma (Lemma 2.2.6), each copy of H_{det} in the reduced graph will provide many copies of H_{det} in G.

Theorem 5.4.1. *Let $2 \le k \le r \in \mathbb{N}$ and $\alpha, \gamma > 0$. Then there exists $C > 0$ such that if $p \ge Cn^{-2/k}$ and G is an n-vertex graph with $\delta(G) \ge \left(1 - \frac{k}{r} + \gamma\right)n$, then $G \cup G(n,p)$ whp contains a partial K_r-factor covering all but at most αn vertices.*

Proof. Choose $0 < \frac{1}{m_0} \ll \varepsilon \ll d \ll \gamma, \alpha, \frac{1}{r}$ and define $d_0 := \frac{(d-\varepsilon)^r}{(2+r)2^r} \ge \frac{4\varepsilon}{\alpha}$. Further, let $M_0 = M_0(m_0, \varepsilon)$ be returned by Lemma 2.2.1 and choose $0 \ll \frac{1}{C} \ll \varepsilon, \frac{1}{M_0}$. Now let $n \in \mathbb{N}$ be sufficiently large and G be a graph as in the statement of the theorem. Apply Lemma 2.2.1 to G with Note that by Lemma 2.2.1, the resulting (ε, d)-reduced graph R has $m \ge m_0$ vertices and satisfies $\delta(R) \ge \left(1 - \frac{k}{r} + \frac{\gamma}{2}\right)m$. Let the size of the clusters in the regularity partition be n' and note that $\frac{n}{M_0} \le n' \le \varepsilon n$. Now by Theorem 2.5.2, as $m \ge m_0$ is sufficiently large, there exists a partial H_{det}-factor $\mathcal{H}$ covering all but at most $\frac{\alpha m}{4}$ vertices of R. Let $\mathcal{U}_1, \ldots, \mathcal{U}_t \in \binom{V(R)}{r}$ such that the $\mathcal{U}_j$ span disjoint copies of H_{det} in $\mathcal{H}$.

Next, let $\mathcal{F}$ be the collection of subsets $W \subseteq V(G)$ such that there exists some $j \in [t]$ for which W intersects each $U \in \mathcal{U}_j$ in at least $\frac{\alpha n'}{2}$ elements (and W contains no vertices of clusters from outside of $\mathcal{U}_j$). Here we say that $\mathcal{U}_j$ *hosts* W. Moreover, we call a copy of K_r in W *crossing* if it contains precisely one vertex from each cluster in the class $\mathcal{U}_j$.

We claim that whp, *every* W in $\mathcal{F}$ contains a crossing copy of K_r in $G \cup G(n,p)$. Indeed, fix some $W \in \mathcal{F}$ and suppose $\mathcal{U}_j$ hosts W. Then there are subsets $W_1, \ldots, W_r \subset V(G)$ and clusters $\{U_1, \ldots, U_r\} = \mathcal{U}_j$ such that $W_i \subseteq U_i$,

$$|W_i| = \frac{\alpha n'}{2} \ge \frac{\alpha n}{2M},$$

$\cup_{i \in r} W_i \subseteq W$ and $U_1, U_2, \ldots, U_r$ form a copy H of H_{det} in R. By Lemma 2.2.2, for every $U_iU_j \in E(H)$, we have that $G[W_i, W_j]$ is a $\left(\frac{4\varepsilon}{\alpha}\right)$-regular pair with density at least $d - \varepsilon$. Thus by Lemma 2.2.6 $G[W]$ contains at least $\left(\frac{d_0 \alpha n'}{2}\right)^r = \Omega(n^r)$ copies of H_{det} where in each such copy of H_{det} precisely one vertex lies in each of $W_1, \ldots, W_r$; let C_W denote this collection. Now noting that $F := \overline{H_{\text{det}}} = K_r \setminus H_{\text{det}}$ is a collection of disjoint cliques of size at most k, Lemma 5.3.1 (part 1 and 3) implies that $\Phi_F \ge Cn$. Also, we have that $|\mathcal{F}| \le 2^n$. Thus for $C > 0$ sufficiently large, Corollary 5.3.3 gives that for every $W \in \mathcal{F}$ there is a copy of H_{det} from C_W which hosts a labelled copy of $\overline{H_{\text{det}}}$ in $G(n,p)$; thus the claim is satisfied.

One can now use the claim to greedily build the almost K_r-factor in $G \cup G(n,p)$. Indeed, initially set $\mathcal{K} := \emptyset$. At each step we will add a copy of K_r to $\mathcal{K}$ whilst ensuring $\mathcal{K}$ is a partial K_r-factor in $G \cup G(n,p)$. Further, at every step we only add a copy K of K_r if there is some $j \in [t]$ such that each vertex in K lies in a different cluster in $\mathcal{U}_j$ (recall each $\mathcal{U}_j$ consists of r clusters).

Suppose we are at a given step in this process such that there exists some cluster $U_i \in \mathcal{U}_j$ (for some j) that still has at least $\frac{\alpha n'}{2}$ vertices uncovered by $\mathcal{K}$. This in fact implies that every cluster in $\mathcal{U}_j$ contains at least $\frac{\alpha n'}{2}$ vertices uncovered by $\mathcal{K}$; these uncovered vertices correspond precisely to a set $W \in \mathcal{F}$ which is hosted by $\mathcal{U}_j$. Hence by the above claim there is a crossing copy of K_r in $(G \cup G(n,p))[W]$. Add this to $\mathcal{K}$. Thus, we can repeat this process, increasing the size of $\mathcal{K}$ at every step, until we find that for every $j \in [t]$, all the clusters in $\mathcal{U}_j$ have at least $\left(1 - \frac{\alpha}{2}\right)n'$ vertices covered by $\mathcal{K}$.

That is, whp there is a partial K_r-factor in $G \cup G(n,p)$ covering all but at most

$$\left(\frac{\alpha n'}{2} \times m\right) + \left(\frac{\alpha m}{4} \times n'\right) + |V_0| \leq \alpha n$$

vertices, as desired. Note that the first term in the above expression comes from the vertices in clusters from the classes $\mathcal{U}_j$; the second term comes from those vertices in clusters that were uncovered by $\mathcal{H}$. □

Note that one can in fact establish the case $k = r$ in a much simpler way because the copies of K_r that we look for can be completely provided by $G(n,p)$, see e.g. [95, Theorem 4.9].

5.5 The absorbing structure - deterministic edges

The aim of this section and the next is to prove the existence of an *absorbing structure* $\mathcal{A}$ in $G' := G \cup G(n,p)$. The main outcomes are Corollaries 5.6.6, 5.6.8 and 5.6.9, which will be used in the next section to prove Theorem III.

The key component of the absorbing structure will be some absorbing subgraph $F \subset G'$. We will define F so that it can contribute to a K_r-factor in many ways. In fact we will define F so that if we remove F from G' and we have a partial factor covering almost all of what remains (Theorem 5.4.1), then no matter which small set of vertices remains, the properties of F allow us to complete this partial factor to a factor in G'. There are some complications, and the absorbing structure will have different features depending on the exact values of minimum degree and the size of the cliques we want in our factor.

Our absorbing subgraph will be comprised of two sets of edges, namely the deterministic edges in G and the random edges in $G(n,p)$. Initially, we will be concerned with finding (parts of) the appropriate subgraph in G. In fact, we will need to prove the existence of many copies of the deterministic subgraphs we want, as we will rely on there being enough of these to guarantee that one of them will match up with random edges in $G(n,p)$ (Section 5.6) to give the desired subgraph. Therefore it is useful throughout to consider, with foresight, the random edges that we will be looking for to complete our desired structure, as this also motivates the form of our deterministic subgraphs.

The smallest building block in our absorbing graph will be K_{r+1}^-, the complete graph on $r+1$ vertices with one edge missing, say between w_1 and w_2. This is useful for the simple reason that it can contribute to a K_r-factor in two ways, namely $K_{r+1}^- \setminus \{w_i\}$ for $i = 1, 2$. We introduce the following notation to keep track of the partition of the edges between the deterministic graph and the random graph. Recall the definition of a blowup of a complete graph $K^t_{r_1,r_2,\ldots,r_t}$ from the Notation Section.

Definition 5.5.1. Suppose $t, r, r_1, r_2, \ldots, r_t \in \mathbb{N}$ such that $\sum_{i=1}^t r_i = r + 1$. We use the notation

$$H := (K^t_{r_1,r_2,\ldots,r_t}, i, j),$$

for not necessarily distinct $i, j \in [t]$, to denote the $(r+1)$-vertex graph $K^t_{r_1,r_2,\ldots,r_t}$ with two distinct distinguished vertices: w_1 in the i^{th} part (which has size r_i) and w_2 in the j^{th} part (which has size r_j).

Definition 5.5.2. Let $r \in \mathbb{N}$ and consider an $(r+1)$-vertex graph F with two distinguished vertices w_1 and w_2. (Typically we will take $F = H$ as in Definition 5.5.1.) We then write $\overline{F}$ to denote the graph on the same vertex set $V(K_{r+1}^-) = V(F)$ such that $E(\overline{F}) := E(K_{r+1}^-) \setminus E(F)$, where we take the non-edge of K_{r+1}^- to be w_1w_2. Thus $K_{r+1}^- \subseteq F \cup \overline{F}$.

Remark: Note that our use of the notation $\overline{F}$ is non-standard here, when referring to $(r+1)$-vertex graphs, we will always be interested in taking the complement of with respect to K_{r+1}^- and *not* K_{r+1}.

We think of $H, \overline{H}$ and K_{r+1}^- as all lying on the same vertex set throughout with the two distinguished vertices w_1, w_2 being defined for all three. The following graph gives the paradigm for how we split the edges of K_{r+1}^- between the deterministic and the random graph.

Definition 5.5.3. For $r \in \mathbb{N}$ and $k \in \mathbb{N}$ such that $2 \leq k \leq r$, let $r^*, q \in \mathbb{N}$ be such that $k(r^* - 1) + q = r$ and $0 < q \leq k$. Then $H_0 := (K^{r^*}_{k,\ldots,k,q+1}, r^*, r^*)$.

Some examples of H_0 and $\overline{H_0}$ can be seen in Figures 5.3 and 5.4. Note that if w_1 and w_2 are the distinguished vertices of H_0, then $H_0 \setminus w_i$ for $i = 1, 2$ are both copies of the graph H_{det} from Definition 5.2.1. Also note that $\overline{H_0}$ is a disjoint union of k-cliques as well as a disjoint copy of K_{q+1}^-. Thus, when $q \leq k-1$, it follows from Lemma 5.3.1 and Corollary 5.3.3 that the graph $\overline{H_0}$ is abundant[2] in $G(n, p)$ when $p \geq Cn^{-2/k}$ for some large enough C. Furthermore, as we will see, the minimum degree condition for G along with supersaturation arguments (Lemma 2.4.1) will imply that there are $\Omega(n^{r+1})$ copies of H_0 in G. This suggests the suitability

[2]Specifically, one can see that whp any linear sized set in $V(G(n, p))$ contains a copy of $\overline{H_0}$.

of this definition as a candidate for how to partition the edge set of K_{r+1}^- between deterministic and random edges. We remark that the case when $q = k$ is slightly more subtle and we have to adjust our decomposition accordingly. We will discuss this is more detail in the next section.

5.5.1 Reachability

In this subsection, we define reachable paths and show that we can find many of these in our deterministic graph G, when the graphs used to define such paths are chosen appropriately. The main results are Proposition 5.5.6, Proposition 5.5.7 and Proposition 5.5.12 which deal with Case 1, 2 and 3 respectively. We first define a reachable path which is a graph which connects together $(r+1)$-vertex graphs as follows.

Definition 5.5.4. Let $t, r \in \mathbb{N}$ and let $\boldsymbol{H} = (H^1, H^2, \ldots, H^t)$ be a vector of $(r+1)$-vertex graphs H^i such that each H^i has two distinguished vertices, w_1^i and w_2^i. Then an $\boldsymbol{H}$-path is the graph P obtained by taking one copy of each H^i and identifying w_2^i with w_1^{i+1}, for $i \in [t-1]$. We call w_1^1 and w_2^t the *endpoints* of P.

In the case where $H^1 = H^2 = \ldots = H^t = H$ for some $(r+1)$-vertex graph H, we use the notation $\boldsymbol{H} = (H, t)$ and thus refer to (H, t)-paths. For $\boldsymbol{H} = (H^1, H^2, \ldots, H^t)$, we also define $\overline{\boldsymbol{H}} := (\overline{H^1}, \overline{H^2}, \ldots, \overline{H^t})$ where $\overline{H^i}$ is as defined in Definition 5.5.2.

We give some explicit examples of $\boldsymbol{H}$-paths later in Figure 5.6. In the following, as we look to find embeddings of $\boldsymbol{H}$-paths and larger subgraphs in G and $G(n,p)$, we will always be considering *labelled* embeddings. Therefore, implicitly, when we define graphs such as the $\boldsymbol{H}$-paths above, we think of these graphs as having some fixed labelling of their vertices.

Again, the motivation for the definition of $\boldsymbol{H}$-paths comes from considering K_{r+1}^-, with vertices w_1, w_2 such that $w_1 w_2 \notin E(K_{r+1}^-)$. Indeed, then a (K_{r+1}^-, t)-path P has two K_r-matchings, each missing a single vertex; one on the vertices of $V(P) \setminus w_1^1$, and one on the vertices of $V(P) \setminus w_2^t$. Our first step is to find many $\boldsymbol{H}$-paths in the deterministic graph G, for an appropriately defined $\boldsymbol{H}$. In particular, we are interested in the images of the endpoints of the paths.

Definition 5.5.5. Let $\beta > 0$, $t, r \in \mathbb{N}$ and $\boldsymbol{H} = (H^1, \ldots, H^t)$ be a vector of $(r+1)$-vertex graphs (each of which is endowed with a tuple of distinguished vertices). We say that two vertices $x, y \in V(G)$ in an n-vertex graph G are *$(\boldsymbol{H}; \beta)$-reachable* (or $(H, t; \beta)$-reachable if $\boldsymbol{H} = (H, \ldots, H) = (H, t)$) if there are at least βn^{tr-1} distinct labelled embeddings of the $\boldsymbol{H}$-path P in G such that the endpoints of P are mapped to $\{x, y\}$.

As discussed before, the graph H_0 from Definition 5.5.3 will be used to provide deterministic edges for our absorbing structure. That is, we look for (H_0, t)-paths in G for some appropriate t.

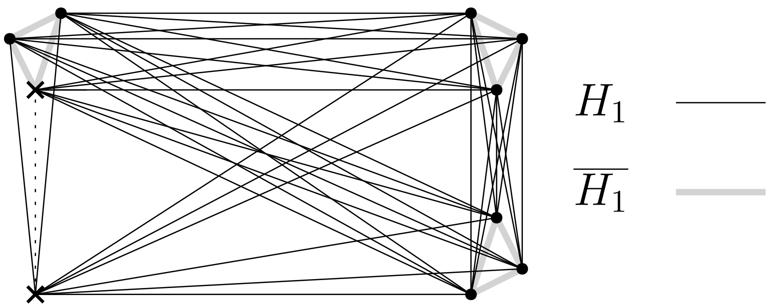


FIGURE 5.2: An example of H_1 and $\overline{H_1}$ for Case 1 (a) $r = 9, k = q = r^* = 3$ (see Definitions 5.5.1–5.5.2 and Proposition 5.5.6 for the definitions).

However, for various reasons there are complications with this approach. Sometimes using a slightly different graph H will allow more vertices to be reachable to each other. Also, as is the case below when $\frac{r}{k} \in \mathbb{N}$, it is possible that $\overline{H_0}$ is not sufficiently common in the random graph $G(n, p)$. Therefore, we have to tweak the graph H_0 in order to accommodate these subtleties. This is the reason for using a vector of graphs $\boldsymbol{H}$ as we will see. We will look first at Case 1, when $\frac{r}{k} \in \mathbb{N}$ and so $\overline{H_0}$ contains a copy of K^-_{k+1}. This is too dense to appear in the random graph $G(n, p)$ with the frequency that we require and thus we define H_1 as in the following proposition.

Proposition 5.5.6. *Let $\gamma > 0$, $n, r, k \in \mathbb{N}$ such that $\frac{r}{k} =: r^* \in \mathbb{N}$, $2 \le k \le r$ and n is sufficiently large. Let $H_1 := J_1 \cap K^-_{r+1}$ where $J_1 := (K^{(r^*+1)}_{k,k,\dots,k,1}, r^*+1, 1)$, as defined in Definition 5.5.1, and we consider K^-_{r+1} to be on the same vertex set as J_1 with a non-edge between the distinguished vertices of J_1. Likewise, let $H'_1 := (K^{(r^*+1)}_{k,k,\dots,k,1}, 1, r^*+1) \cap K^-_{r+1}$ be the same graph with the labels of the distinguished vertices switched. See Figure 5.2 for an example of H_1 (and H'_1 which is identical).*

Then there exists a $\beta_1 = \beta_1(r, k, \gamma) > 0$ such that for any n-vertex graph G of minimum degree $\delta(G) \ge \left(1 - \frac{k}{r} + \gamma\right)n$, any pair of distinct vertices in $V(G)$ are $(\boldsymbol{H_1}; \beta_1)$-reachable where $\boldsymbol{H_1} := (H_1, H'_1)$.

Proof. Fix some graph G as in the statement of the lemma. Let w_1, w_2 be the distinguished vertices of $H_1 = (K^{(r^*+1)}_{k,k,\dots,k,1}, r^*+1, 1) \cap K^-_{r+1}$ as defined in Definition 5.5.1. Fix a pair $x, y \in V(G)$.

We will show that for *any* $z \in V(G) \setminus \{x, y\}$, there are at least $\beta'_1 n^{r-1}$ labelled embeddings of H_1 which map w_1 to x and w_2 to z, for some $\beta'_1 = \beta'_1(r, k, \gamma) > 0$. Once we have established this property, this implies the proposition. Indeed, by symmetry, we can also find $\beta'_1 n^{r-1}$ embeddings of H_1 which map w_1 to y and w_2 to z. Set $\beta_1 := \frac{\beta'^2_1}{2}$. Thus there are at least

$$(n-2) \cdot \beta'_1 n^{r-1} \cdot (\beta'_1 n^{r-1} - r^2 n^{r-2}) \ge \beta_1 n^{2r-1}$$

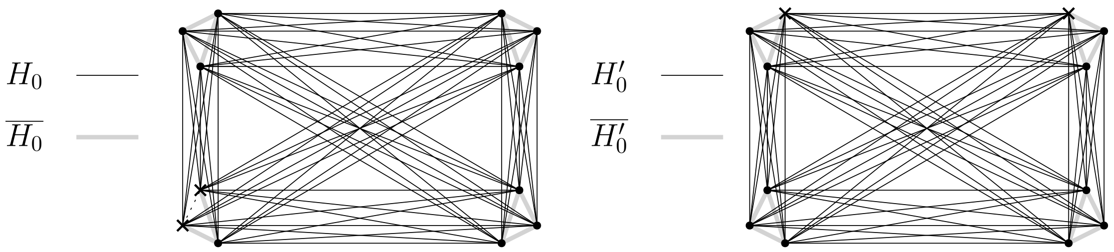


FIGURE 5.3: An example of H_0, $\overline{H_0}$, H'_0 and $\overline{H'_0}$ for Case 2 (b) $r = 11, k = 3, q = 2, r^* = 4$ (see Definitions 5.5.1–5.5.3 and Proposition 5.5.7 for the definitions).

distinct embeddings of the $\boldsymbol{H_1}$-path in G such that the endpoints are mapped to $\{x, y\}$, as desired. This follows as there are $n-2$ choices for z; at least $\beta_1' n^{r-1}$ choices for the copy of H_1 containing x and z; at least $(\beta_1' n^{r-1} - r^2 n^{r-2})$ choices for the copy of H_1' containing z and y that are *disjoint* from the choice of H_1 (except for the vertex z).

So let us fix $x, z \in V(G)$. The proof now follows easily from Lemma 2.4.1. As $kr^* = r$, we can express the minimum degree as $\delta(G) \geq \left(1 - \frac{1}{r^*} + \gamma\right)n$. Thus any set of at most r^* vertices has at least γn common neighbours. Therefore we have at least $(\gamma n)^{r^*}$ labelled copies K of K_{r^*} where $V(K) = \{x_1, \dots, x_{r^*}\} \subset N_G(x)$ and $\{x_2, \dots, x_{r^*}\} \subset N_G(x) \cap N_G(z)$. This follows by first choosing $\{x_2, \dots, x_{r^*}\}$ and then x_1 with the right adjacencies. Thus, by Lemma 2.4.1 we have $\beta_1' n^{r-1}$ labelled embeddings of the blow-up, $H_1 \setminus \{w_1, w_2\} = K^{r^*}_{k-1,k,\dots,k}$, of these cliques, crucially within the correct neighbourhoods ($N_G(x)$ and $N_G(x) \cap N_G(z)$) to ensure that together with $\{x, z\}$ they give us the required embeddings of H_1. □

Note that an $\overline{\boldsymbol{H_1}}$-path $\overline{P_1}$ has endpoints which are isolated. The other vertices of $\overline{P_1}$ lie in copies of K_k and these copies are disjoint from each other except for a single pair of K_ks that meet at a singular vertex. See Figure 5.6 for an example. We now turn to Case 2, as described in Section 5.2. Here we can use the graph H_0 from Definition 5.5.3. We also use a slight variant of H_0 where we redefine the distinguished vertices.

Proposition 5.5.7. *Suppose $\gamma > 0, n, r, k \in \mathbb{N}$, such that n is sufficiently large, $2 \leq k < \frac{r}{2}$ and $\frac{r}{k} \notin \mathbb{N}$. Further, let r^*, q and H_0 be as defined in Definition 5.5.3 and let $H_0' = (K^{r*}_{k,\dots,k,q+1}, 1, 2)$ be the same graph as H_0 with distinguished vertices in distinct*[3] *parts of size k (see Figure 5.3).*

Then there exists $\beta_2 = \beta_2(r, k, \gamma) > 0$ such that for any n-vertex G of minimum degree $\delta(G) \geq \left(1 - \frac{k}{r} + \gamma\right)n$, every pair of distinct vertices x, y in $V(G)$ are $(\boldsymbol{H_2}; \beta_2)$-reachable where $\boldsymbol{H_2} := (H_0, H_0', H_0', H_0)$.

Proof. Fix some graph G as in the statement of the lemma and let $x, y \in V(G)$ be arbitrary. We know that $\delta(G) \geq \left(1 - \frac{k}{r} + \gamma\right)n \geq \left(1 - \frac{1}{r^*-1} + \gamma\right)n$. Therefore every set of at most $r^* - 1$ vertices

[3]Note that this is possible as we are in the case where the number of parts, r^* of H_0 is at least 3.

has a common neighbourhood of size at least γn. We will appeal to Lemma 2.4.1 to give us the whole $\boldsymbol{H_2}$-path in one fell swoop. Let J be a graph with vertex set

$$V(J) = \{x_1, \dots, x_{r^*-1}, z_1, w_1, \dots, w_{r^*-2}, z_2, u_1, \dots, u_{r^*-2}, z_3, y_1, \dots, y_{r^*-1}\},$$

and $E(J)$ consisting of r^*-cliques on the vertex sets $\{x_1, \dots, x_{r^*-1}, z_1\}$, $\{z_1, w_1, \dots, w_{r^*-2}, z_2\}$, $\{z_2, u_1, \dots, u_{r^*-2}, z_3\}$ and $\{z_3, y_1, \dots, y_{r^*-1}\}$. We claim that if we can find $\frac{(\gamma n)^{4r^*-3}}{2}$ copies of J in G such that $x_i \in N_G(x)$ and $y_i \in N_G(y)$ for $i = 1, \dots, r^* - 1$, then we are done. Indeed, consider a blow-up J' of J with parts

$$\begin{aligned}\{X_1^{(k)}, \dots, X_{r^*-1}^{(k)}, Z_1^{(k+q-1)}, W_1^{(q+1)}, W_2^{(k)}, \dots, W_{r^*-2}^{(k)}, Z_2^{(2k-1)}, \\ U_1^{(q+1)}, U_2^{(k)}, \dots, U_{r^*-2}^{(k)}, Z_3^{(k+q-1)}, Y_1^{(k)}, \dots, Y_{r^*-1}^{(k)}\},\end{aligned}$$

where the parts correspond to the vertices of J in the obvious way and the size of each part is indicated by the superscript. Now if we have a copy of J' in $G_{\hat{x},\hat{y}} = G[V(G) \setminus \{x, y\}]$ with $X_i \subset N_G(x)$ and $Y_i \subset N_G(y)$ for all $i = 1, \dots, r^* - 1$, then this gives us an embedding of an $\boldsymbol{H_2}$-path. Indeed for $i = 1, 3$, arbitrarily partition $Z_i := \{z_i'\} \cup Z_i' \cup Z_i''$ with $|Z_i'| = q - 1$ and $|Z_i''| = k - 1$ and partition $Z_2 := \{z_2'\} \cup Z_2' \cup Z_2''$ with $|Z_2'| = |Z_2''| = k - 1$. Then $\{x, z_1'\} \cup Z_1' \cup_{i=1}^{r^*-1} X_i$ and $\{y, z_3'\} \cup Z_3' \cup_{i=1}^{r^*-1} Y_i$ both give copies of H_0 whilst $\{z_1', z_2'\} \cup Z_1'' \cup Z_2' \cup_{i=1}^{r^*-2} W_i$ and $\{z_2', z_3'\} \cup Z_3'' \cup Z_2'' \cup_{i=1}^{r^*-2} U_i$ both give copies of H_0', where in all cases the distinguished vertices appear in the first set of the union.

It suffices then, by Lemma 2.4.1, to find $\frac{(\gamma n)^{4r^*-3}}{2}$ embeddings of J in G with the $x_i \in N_G(x)$ and $y_i \in N_G(y)$. We can do this greedily. Indeed if we choose the x_i and y_i first, followed by z_1 and z_3, then $z_2 \in N_G(z_1) \cap N_G(z_3)$ and then the remaining vertices, we are always seeking to choose a vertex in G which has at most $r^* - 1$ neighbours which have already been chosen. Thus, by our degree condition, we have at least γn choices for each vertex with the right adjacencies. To ensure that these choices actually give an embedding of J we then discard any set of choices with repeated vertices, of which there are $O(n^{4r^*-4})$, and thus the conclusion holds as n is sufficiently large. $\square$

Consider an $\overline{\boldsymbol{H_2}}$-path which we denote $\overline{P_2}$ (see Figure 5.6 for an example). It is formed by copies of K_k and K_{q+1}^- which intersect in at most one vertex and such that the endpoints of $\overline{P_2}$ lie in copies of K_{q+1}^-. Furthermore, note that the endpoints of $\overline{P_2}$ are in distinct connected components. This will be an important feature when we start to address the random edges of our absorbing structure as it will allow us to use Lemma 5.3.1 to conclude certain statements about the likelihood of finding our desired random subgraph in $G(n, p)$. This motivated the introduction of H_0' in the previous proposition.

In Case 3, we cannot hope to prove reachability between every pair of vertices. Indeed our minimum degree in this case is $\delta(G) \geq \left(1-\frac{k}{r}+\gamma\right)n$ and $k > \frac{r}{2}$ and so it is possible that $\delta(G) < \frac{n}{2}$ and G is disconnected. Thus, as in [82, 89], we use a partition of the vertices into 'closed' parts, where we can guarantee that two vertices in the same part are reachable, with some set of parameters. We adopt the following notation which also allows us to consider different possibilities for what vectors we use for reachability.

Definition 5.5.8. Let $\mathcal{H}$ be a set of vectors, such that the entry of each vector in $\mathcal{H}$ is an $(r+1)$-vertex graph endowed with a tuple of distinguished vertices. We say that two vertices in G are $(\mathcal{H};\beta)$-reachable if they are $(\boldsymbol{H};\beta)$-reachable for some $\boldsymbol{H} \in \mathcal{H}$.

We say that a subset V of vertices in a graph G is $(\mathcal{H};\beta)$-closed if every pair of vertices[4] in V is $(\mathcal{H};\beta)$-reachable. We denote[5] by $N_{\mathcal{H},\beta}(v)$ the set of vertices in G that are $(\mathcal{H};\beta)$-reachable to v.

Thus, in this notation, the conclusion of Proposition 5.5.6 states that $V(G)$ is $(\boldsymbol{H_1};\beta_1)$-closed for all G satisfying the given hypothesis (and similarly for Proposition 5.5.7). Notice that if a set V is $(\mathcal{H};\beta)$-closed in a graph G it may be the case that two vertices $x, y \in V$ are $(\boldsymbol{H};\beta)$-reachable whilst two other vertices $z, w \in V$ are $(\boldsymbol{H}';\beta)$-reachable for some distinct $\boldsymbol{H}, \boldsymbol{H}' \in \mathcal{H}$ of different lengths.

It will be useful for us to consider the following notion.

Definition 5.5.9. Let $\mathcal{H}, \tilde{\mathcal{H}}$ be two sets of vectors as in Definition 5.5.8. Then

$$\mathcal{H} + \tilde{\mathcal{H}} := \mathcal{H} \cup \tilde{\mathcal{H}} \cup (\mathcal{H} \cdot \tilde{\mathcal{H}}),$$

where $\mathcal{H} \cdot \tilde{\mathcal{H}}$ is defined to be the set

$$\begin{aligned}\mathcal{H} \cdot \tilde{\mathcal{H}} := \{(\boldsymbol{H}, \tilde{\boldsymbol{H}}) :=& (H^1, \dots, H^t, \tilde{H}^1, \dots, \tilde{H}^{\tilde{t}}) : \\ & \boldsymbol{H} := (H^1, \dots, H^t) \in \mathcal{H}, \tilde{\boldsymbol{H}} := (\tilde{H}^1, \dots, \tilde{H}^{\tilde{t}}) \in \tilde{\mathcal{H}}\}.\end{aligned}$$

That is, $\mathcal{H}+\tilde{\mathcal{H}}$ comprises of all vectors that lie in $\mathcal{H}, \tilde{\mathcal{H}}$, or that can be obtained by a concatenation of a vector from $\mathcal{H}$ with a vector from $\tilde{\mathcal{H}}$.

As an important example, defining $\mathcal{H}(H, \leq t) := \{(H, s) : 1 \leq s \leq t\}$, we have that

$$\mathcal{H}(H, \leq t_1) + \mathcal{H}(H, \leq t_2) = \mathcal{H}(H, \leq t_1 + t_2).$$

In what follows we will apply the following simple lemma repeatedly.

[4]Note that we do not require the vertices of the $\boldsymbol{H}$-paths which give the reachability to lie in V.

[5]If $\mathcal{H}$ consists of just one vector $\boldsymbol{H}$, we simply refer to sets being $(\boldsymbol{H};\beta)$-closed and use $N_{\boldsymbol{H},\beta}(v)$ to denote the closed neighbourhood of a vertex.

Lemma 5.5.10. *Let $r \in \mathbb{N}$ and let $\mathcal{H}_x, \mathcal{H}_y$ be two sets of vectors of $(r+1)$-vertex graphs, each of which is endowed with a tuple of distinguished vertices and suppose that $t_x := |\mathcal{H}_x|$ and $t_y := |\mathcal{H}_y|$ are both finite. Suppose G is an n-vertex graph with n sufficiently large and let $x, y \in V(G)$. Suppose there exist $\beta_x, \beta_y, \varepsilon > 0$, and some subset $U \subseteq V(G)$ with $|U| \geq \varepsilon n$ such that for every $z \in U$, x and z are $(\mathcal{H}_x; \beta_x)$-reachable and z and y are $(\mathcal{H}_y; \beta_y)$-reachable. Then x and y are $(\mathcal{H}_x + \mathcal{H}_y; \beta)$-reachable for $\beta := \frac{\varepsilon \beta_x \beta_y}{2 t_x t_y} > 0$.*

Proof. By the pigeonhole principle, there exists some $U' \subseteq U$ such that $|U'| \geq \frac{\varepsilon n}{t_x t_y}$ and some $\boldsymbol{H}_x \in \mathcal{H}_x$, $\boldsymbol{H}_y \in \mathcal{H}_y$ such that for every $z \in U'$, z and x are $(\boldsymbol{H}_x; \beta_x)$-reachable and z and y are $(\boldsymbol{H}_y; \beta_y)$-reachable. Suppose $\boldsymbol{H}_x$ has length s_x and $\boldsymbol{H}_y$ has length s_y. Thus, fixing $z \in U'$, there are at least $\beta_x \beta_y n^{(s_x+s_y)r-2}$ pairs of labelled vertex sets S_x and S_y in G such that there is an embedding of an $\boldsymbol{H}_x$-path on $S_x \cup \{x, z\}$ mapping endpoints to $\{x, z\}$ and an embedding of a $\boldsymbol{H}_y$-path on the vertices $S_y \cup \{y, z\}$ which maps the endpoints to $\{y, z\}$. Of these pairs, at most

$$s_x s_y r^2 n^{(s_x+s_y)r-3}$$

are *not* vertex-disjoint or they intersect $\{x, y\}$. Hence, as n is sufficiently large we have at least $\frac{\beta_x \beta_y}{2} n^{(s_x+s_y)r-2}$ vertex-disjoint pairs which together form an embedding of an $(\boldsymbol{H}_x, \boldsymbol{H}_y)$-path. As we have at least $\frac{\varepsilon}{t_x t_y} n$ choices for z, this gives that x and y are $((\boldsymbol{H}_x, \boldsymbol{H}_y); \beta)$-reachable and $(\boldsymbol{H}_x, \boldsymbol{H}_y) \in \mathcal{H}_x + \mathcal{H}_y$. □

We now turn to proving reachability in Case 3. The following two lemmas together find the partition we will work on. Similar ideas have been used in [82, 89].

Lemma 5.5.11. *Suppose $\gamma > 0$ and $n, r, k, q \in \mathbb{N}$ such that $\frac{r}{2} < k \leq r-1$, $r = k + q$ and n is sufficiently large. Let $c := \lceil \frac{r}{q} \rceil$ and for $t \in \mathbb{N}$ define*

$$\mathcal{H}^t := \mathcal{H}(H_0, \leq 2^t) = \{(H_0, s) : 1 \leq s \leq 2^t\},$$

where $H_0 = (K^2_{k,q+1}, 2, 2)$ is as defined in Definition 5.5.3 with distinguished vertices w_1 and w_2.

Then there exists constants $0 < \beta_3' = \beta_3'(r, k, \gamma), \alpha = \alpha(r, k, \gamma)$ such that any n-vertex graph G of minimum degree $\delta(G) \geq \left(1 - \frac{k}{r} + \gamma\right)n$ can be partitioned into at most $c - 1$ parts, each of which is $(\mathcal{H}^c; \beta_3')$-closed and of size at least αn.

Proof. Fix some graph G as in the statement of the lemma. Firstly, observe that there is some $\eta = \eta(r, k, \gamma) > 0$ such that in every set of at least c vertices, there are two vertices which are $((H_0, 1); \eta)$-reachable. Indeed, fix some arbitrary set of vertices $S = \{v_1, \ldots, v_c\} \subset V(G)$, and for $v \in V = V(G)$, define $d_S(v) := |\{i \in [c] : vv_i \in E(G)\}|$. Let $\bar{d}_S := \frac{\sum_{v \in V} d_S(v)}{n}$ be the average. Then we have that

$$\sum_{v \in V} d_S(v) = \sum_{i \in [c]} \deg_G(v_i) \geq c\left(1 - \frac{k}{r} + \gamma\right) n \geq (1 + c\gamma)n.$$

Thus $\sum_{v\in V} \binom{d_S(v)}{2} \geq n\binom{\bar{d}_S}{2} \geq \frac{(c\gamma)^2}{2}n$ by Jensen's inequality. By averaging over all pairs we have that there exists a pair $i \neq j \in [c]$ so that both v_i and v_j are in the neighbourhood of at least $\gamma^2 n$ vertices. That is, $|N_G(v_i) \cap N_G(v_j)| \geq \gamma^2 n$.

Therefore there are at least $\gamma^3 n^2$ edges in G with one endpoint in $N_G(v_i) \cap N_G(v_j)$. Applying Lemma 2.4.1 this ensures that there is $\eta = \eta(r, k, \gamma) > 0$ so that there are ηn^{r-1} copies of $K^2_{k,q-1}$ where the first vertex class lies in $N_G(v_i) \cap N_G(v_j)$. Thus together they form copies of H_0 with distinguished vertices v_i and v_j; so v_i and v_j are $((H_0, 1); \eta)$-reachable in G.

Note also that there is some fixed $\alpha' = \alpha'(r, k, \gamma) > 0$ such that $|N_{(H_0,1),\alpha'}(v)| \geq \alpha' n$ for every $v \in V(G)$. Indeed, this follows as there are at least $\frac{(\gamma n)^2}{2}$ edges in G with one endpoint in $N_G(v)$. So, by Lemma 2.4.1, there is a fixed $\alpha'' = \alpha''(r, k, \gamma) > 0$ such that there are at least $\alpha'' n^r$ embeddings of H_0 which map w_1 to v. Setting $\alpha' := \frac{\alpha''}{3}$, this implies that there are at least $\alpha' n$ vertices which are $((H_0, 1); \alpha')$-reachable to v.

Now choose $0 < \varepsilon \ll \alpha', \frac{1}{c}$, $\eta =: \eta_0$ and $\eta_i := \frac{\varepsilon^4}{2^{2i+1}}\eta^2_{i-1}$ for $i = 1, \dots, c$. Set $\beta'_3 := \eta_c$. As in the statement of the lemma, define $\mathcal{H}^t := \mathcal{H}(H_0, \leq 2^t)$ for values of $t \leq c$ and note that $\mathcal{H}^t + \mathcal{H}^t = \mathcal{H}^{t+1}$. We will be interested in $(\mathcal{H}^t; \eta_t)$-reachability and so we will use the shorthand notation $\tilde{N}_t(v) := N_{\mathcal{H}^t,\eta_t}(v)$. Let ℓ be the maximal integer such that there exists a set of ℓ vertices, $v_1, \dots, v_\ell$ with v_j and $v_{j'}$ *not* $(\mathcal{H}^{c-\ell}; \eta_{c-\ell})$-reachable for any pair $j \neq j' \in [\ell]$.

Suppose $\ell = 1$. Then $V(G)$ is $(\mathcal{H}^{c-1}; \eta_{c-1})$-closed. As $\mathcal{H}^{c-1} \subseteq \mathcal{H}^c$ and $\eta_{c-1} > \beta'_3$, the lemma holds in this case.

We also have that $\ell \leq c - 1$ from our observations above, so we can assume $2 \leq \ell \leq c - 1$. Now fix such a set of ℓ vertices, $v_1, \dots, v_\ell$. We make the following two observations:

(i) Any $v \in V(G) \setminus \{v_1, \dots, v_\ell\}$ is in $\tilde{N}_{c-\ell-1}(v_j)$ for some $j \in [\ell]$ from our definition of ℓ, as otherwise v could be added to give a larger family contradicting the maximality of ℓ. Indeed, this follows because two vertices that are not $(\mathcal{H}^{c-\ell}; \eta_{c-\ell})$-reachable are certainly not $(\mathcal{H}^{c-\ell-1}; \eta_{c-\ell-1})$-reachable by definition.

(ii) $|\tilde{N}_{c-\ell-1}(v_j) \cap \tilde{N}_{c-\ell-1}(v_{j'})| \leq \varepsilon n$ for every pair $j \neq j' \in [\ell]$. This follows from Lemma 5.5.10 as otherwise we would have that v_j and $v_{j'}$ are $(\mathcal{H}^{c-\ell}; \eta_{c-\ell})$-reachable, a contradiction.

We define $U_j := \left(\tilde{N}_{c-\ell-1}(v_j) \cup \{v_j\}\right) \setminus \left(\bigcup_{j' \in [\ell]\setminus\{j\}} \tilde{N}_{c-\ell-1}(v_{j'})\right)$ for $j \in [\ell]$, and $U_0 := V(G) \setminus \cup_{j\in[\ell]} U_j$. Now for $j \in [\ell]$, we have that U_j is $(\mathcal{H}^{c-\ell-1}; \eta_{c-\ell-1})$-closed. Indeed, if there was a $j \in [\ell]$ and $u_1, u_2 \in U_j$ not reachable, then $\{u_1, u_2\} \cup \{v_1, \dots, v_\ell\} \setminus \{v_j\}$, is a larger family contradicting the definition of ℓ. Thus, the U_j almost form the partition we are looking for except that it remains to consider the vertices in U_0. For these, we greedily add them to the other U_j. We have that for each $u \in U_0$,

$$|N_{(H_0,1),\alpha'}(u) \setminus U_0| \geq \alpha' n - |U_0| \geq \alpha' n - \binom{\ell}{2}\varepsilon n \geq \ell \varepsilon n. \tag{5.5.1}$$

Here the second inequality holds due to (i), (ii) and the definition of the U_j; the final inequality holds by our choice of ε. Thus, there is a j such that $|N_{(H_0,1),\alpha'}(u) \cap U_j| \geq \varepsilon n$, and we add u to this U_j, arbitrarily choosing such a j if there are multiple choices. Let $V_1, \dots, V_\ell$ be the resulting partition.

Applications of Lemma 5.5.10 show that each V_j is $(\mathcal{H}^c; \eta_c)$-closed. Indeed suppose, for example, that w_1 and w_2 are two vertices that lie in U_0 and are added to U_j in the process of defining V_j. Then for each $i = 1, 2$, taking $W_i = N_{(H_0,1),\alpha'}(w_i) \cap U_j$, an application of Lemma 5.5.10 with $U = W_1$ gives that for any $x \in U_j$, w_1 and x are $(\mathcal{H}'; \eta')$-reachable where $\mathcal{H}' = \mathcal{H}(H_0, \leq 2^{c-\ell-1} + 1)$ and $\eta' = \frac{\varepsilon \alpha' \eta_{c-\ell-1}}{2^{c-\ell}}$. Another application of Lemma 5.5.10, this time with $U = W_2$ then gives that w_1 and w_2 are $(\mathcal{H}''; \eta'')$-reachable with $\mathcal{H}'' = \mathcal{H}(H_0, \leq 2^{c-\ell-1} + 2) \subseteq \mathcal{H}^c$ and $\eta'' = \frac{\varepsilon \eta' \alpha'}{2^{c-\ell+2}} > \eta_{c-\ell} > \beta_3'$. Showing other cases of reachability within each V_j are similar. We are now done since for each $j \in [\ell]$,

$$|V_j| \geq |U_j| \geq |N_{(H_0,1),\alpha'}(v_j) \setminus \big(N_{(H_0,1),\alpha'}(v_j) \cap U_0\big)| \overset{(ii)}{\geq} \alpha' n - \ell \varepsilon n = \alpha n,$$

where $\alpha := \alpha' - \ell\varepsilon > \eta_{c-\ell} \geq \eta_c = \beta_3'$. □

The rough idea for how to handle Case 3 is to run the same proof as in the other cases on each *part* of the partition given by Lemma 5.5.11. The point of Lemma 5.5.11 is that we recover the reachability within each part, albeit at the expense of allowing a family of possible paths used for reachability. However, in the process, we lose the minimum degree condition within each part. The purpose of the next proposition is to fix this, by adjusting parameters and making the partition coarser. Thus, we recover a minimum degree condition which is not quite as strong as what we had previously but good enough to work with in what follows.

Proposition 5.5.12. *Suppose $\gamma > 0$ and $n, r, k, q \in \mathbb{N}$ such that $\frac{r}{2} < k \leq r - 1$, $r = k + q$ and n is sufficiently large. Let $c := \lceil \frac{r}{q} \rceil$ and let $H_0 = (K^2_{k,q+1}, 2, 2)$ as defined in Definition 5.5.3 and $H_0' = (K^2_{k,q+1}, 1, 2)$ be the same graph with distinguished vertices in distinct parts of the bipartition*[6] *(see Figure 5.4). We define the following family of vectors of $(r+1)$-vertex graphs*

[6] This is analogous to the graph H_0' defined in Proposition 5.5.7.

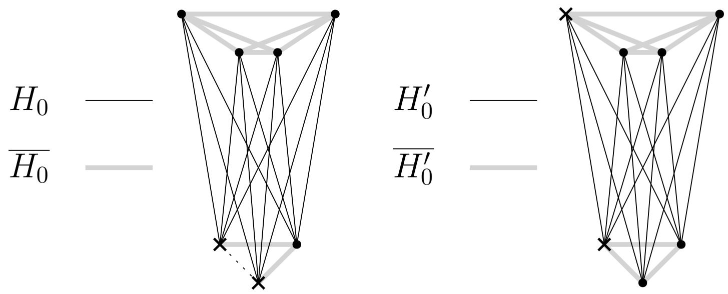


FIGURE 5.4: An example of H_0, $\overline{H_0}$, H_0' and $\overline{H_0'}$ for Case 3 (c) $r = 6, k = 4, q = r^* = 2$ (see Definitions 5.5.1–5.5.3 and Proposition 5.5.12 for the definitions).

(endowed with tuples of vertices):

$$\mathcal{H}_3 := \bigcup_{t=3}^{c(2^{c+1}+1)} \left\{ \boldsymbol{H} \in \{H_0, H_0'\}^t : \boldsymbol{H}[1] = \boldsymbol{H}[t] = H_0 \text{ and } \boldsymbol{H}[i] = H_0' \text{ for some } 2 \le i \le t-1 \right\},$$

where $\boldsymbol{H}[i]$ denotes the i^{th} entry of $\boldsymbol{H}$.

Then there exists $\alpha(r,k,\gamma) > 0$ such that for all $\varepsilon > 0$, the following holds. There exists $0 < \beta_3(r,k,\gamma,\varepsilon)$ such that for any n-vertex graph G with minimum degree $\delta(G) \ge \left(1 - \frac{k}{r} + \gamma\right)n$, there is a partition $\mathcal{P}$ of $V(G)$ into at most $c-1$ parts such that each part $U \in \mathcal{P}$ satisfies the following:

(i) $|U| \ge \alpha n$;

(ii) All but at most εn vertices $v \in U$ satisfy $\deg_G(v;U) \ge \left(1 - \frac{k}{r} + \frac{\gamma}{2}\right)|U|$;

(iii) U is $(\mathcal{H}_3;\beta_3)$-closed.

Proof. This is a simple case of adjusting the partition already obtained after applying the previous lemma, Lemma 5.5.11. Let α, β_3' be defined as in the outcome of Lemma 5.5.11 and let $\mathcal{P}'$ be the partition of $V(G)$ obtained, with vertex parts denoted $V_1, \ldots, V_s$. Fix $\mu := \frac{\varepsilon\gamma}{2c^3}$. We create an auxiliary graph J on vertex set $\{V_1, \ldots, V_s\}$ where for $i \ne j \in [s]$ we have an edge V_iV_j in J if and only if there are at least μn^2 edges in G with one endpoint in V_i and one in V_j. Then our new partition $\mathcal{P}$ in G will come from the connected components of J. That is, if $C_1, \ldots, C_t$ are the components of J, then for $i \in [t]$, we define $U_i := \cup_{j:V_j \in C_i} V_j$ and let $\mathcal{P}$ consist of the U_i with $i \in [t]$. Then certainly point (i) of the hypothesis is satisfied for all U_i. Also (ii) is satisfied. Indeed, suppose there exists $i \in [t]$, with $\deg_G(v;U_i) < \left(1 - \frac{k}{r} + \frac{\gamma}{2}\right)|U_i|$ for at least εn vertices of U_i. Thus, for such vertices $\deg_G(v;(V(G) \setminus U_i)) \ge \frac{\gamma}{2}n$ and by averaging there exists some $V_{j(v)} \in \mathcal{P}'$ such that $V_{j(v)} \cap U_i = \emptyset$ and $\deg_G(v;V_{j(v)}) \ge \frac{\gamma}{2c}n$. We average again to conclude that there is some $j, j' \in [s]$ such that $V_j \subset U_i$, $V_{j'} \cap U_i = \emptyset$ and V_j contains at least $\frac{\varepsilon}{c^2}n$ vertices v which have degree $\deg_G(v;V_{j'}) \ge \frac{\gamma}{2c}n$. This contradicts our definition of J as then $V_jV_{j'}$ should be an edge of J and thus in the same part of $\mathcal{P}$.

Thus it only remains to establish reachability. We begin by proving the following claim which is a slight variation of Lemma 5.5.10.

Claim 5.5.13. *Let $\mathcal{H}^c$ be as defined in Lemma 5.5.11. Suppose $x, y \in V(G)$ and that there exist (not necessarily disjoint) sets $S_x, S_y \subset V(G)$ such that for any $z_x \in S_x$, x and z_x are $(\mathcal{H}^c; \beta_3')$-reachable and for any $z_y \in S_y$, y and z_y are $(\mathcal{H}^c; \beta_3')$-reachable. If there exists at least μn^2 edges with one endpoint in S_x and one endpoint in S_y, then x and y are $(\boldsymbol{H}; \beta_3'')$-reachable for some $\beta_3'' = \beta_3''(\mu, \beta_3', c) > 0$ and $\boldsymbol{H} \in \mathcal{H}_3$ of length at most $2^{c+1} + 1$.*

<u>Proof of Claim:</u> Letting w_1', w_2' be the distinguished vertices of H_0', we have, by Lemma 2.4.1, that there are at least $\mu' n^{r+1}$ embeddings of H_0' into G which map w_1' to S_x and w_2' to S_y for some $\mu' = \mu'(\mu) > 0$. By averaging, there exists $\boldsymbol{H}_x, \boldsymbol{H}_y \in \mathcal{H}^c$ such that there are $\frac{\mu'}{2^{2c}} n^{r+1}$ embeddings of H_0' such that the image of w_1' and x are $(\boldsymbol{H}_x; \beta_3')$-reachable and the image of w_2' and y are $(\boldsymbol{H}_y; \beta_3')$-reachable. By considering the embeddings of $\boldsymbol{H}_x$, $\boldsymbol{H}_y$ and H_0' which join to give an embedding of an $(\boldsymbol{H}_x, H_0', \boldsymbol{H}_y)$-path (that is, ignoring choices of embeddings which are not vertex-disjoint), we see that x and y are $((\boldsymbol{H}_x, H_0', \boldsymbol{H}_y), \beta_3'')$-reachable with $\beta_3'' := \frac{\mu' \beta_3'^2}{2^{2c+1}}$. This completes the proof of the claim. ■

Recall the partition $\mathcal{P}' = \{V_1, V_2, \ldots, V_s\}$. Further consider any part $U \in \mathcal{P}$. First suppose $U = V_j$ for some j. Now given any $x, y \in U$, by Lemma 5.5.11, x and y are already (H_0, s)-reachable for some $s \leq 2^c$. However, (H_0, s) does not contain a copy of H_0' and so is not a valid vector in the family $\mathcal{H}_3$. We therefore apply Claim 5.5.13 with $S_x = S_y = V_j \setminus \{x, y\}$, to conclude that x and y are $(\boldsymbol{H}; \beta_3'')$-reachable for some $\boldsymbol{H} \in \mathcal{H}_3$ of length at most $2^{c+1} + 1$. Indeed since V_j sends fewer than μn^2 edges out to any other part V_i of $\mathcal{P}'$ and $|V_j| \geq \alpha n$, the minimum degree condition on G ensures that there are at least $2\mu n^2$ edges in $G[V_j]$ and hence μn^2 edges in $S_x = S_y$ allowing Claim 5.5.13 to be applied.

Next suppose U is the union of more than one part from $\mathcal{P}'$. If $x \in V_i \subseteq U$ and $y \in V_j \subseteq U$, for $i \neq j \in [s]$ and $V_i V_j \in E(J)$ as defined above, we can again apply Claim 5.5.13, this time with $S_x = V_i$ and $S_y = V_j$, to conclude x and y are $(\boldsymbol{H}; \beta_3'')$-reachable for some $\boldsymbol{H} \in \mathcal{H}_3$ of length at most $2^{c+1} + 1$. Therefore, we just need to establish reachability for vertices x, y such that $x \in V_i$, $y \in V_j$ with $V_i V_j \notin E(J)$ but such that V_i and V_j are in the same component of J. If $i \neq j$, there is a path of (at most c) edges from V_i to V_j in J; if $i = j$ there is a walk of length $2 \leq c$ in J that starts and ends at $V_i = V_j$ (i.e. traverse a single edge in J). In both cases we can repeatedly apply Lemma 5.5.10 to derive that x and y are $(\mathcal{H}_3; \beta_3)$-reachable with $\beta_3 := \left(\frac{\alpha \beta_3''}{2^{(2^{c+1}+2)}}\right)^c$. It is crucial here that we apply Claim 5.5.13 in all cases to establish the reachability here (even when $i = j$) in order to guarantee that the vectors witnessing the reachability contain a copy of H_0' and hence indeed lie in $\mathcal{H}_3$. □

We remark that the reason for the introduction of H_0' in Proposition 5.5.12 is two-fold. Firstly, it allows us to establish reachability between parts from Lemma 5.5.11 which have many edges between them. Moreover, as in Proposition 5.5.7, we have that for every $\boldsymbol{H} \in \mathcal{H}_3$, if $\overline{P}$ is an $\overline{\boldsymbol{H}}$-path, then the endpoints of $\overline{P}$ are in distinct connected components of $\overline{P}$ (see Figure 5.6 for an example), which is something that we will require later.

5.5.2 Absorbing gadgets

We now turn our focus to larger subgraphs which we look to embed in our graph and which will be used as part of an absorbing structure. These are formed by piecing together the $\boldsymbol{H}$-paths of the previous subsection and the aim will be to obtain subgraphs with even more flexibility, in that they will be able to contribute to a factor in many ways. The key definition is a graph which we call an *absorbing gadget*, which we define below.

Before this though, let us recall from Definition 5.5.4 that for a vector $\boldsymbol{H} = (H^1, H^2, \dots, H^t)$ of $(r+1)$-vertex graphs H^i such that each H^i has two distinguished vertices, w_1^i and w_2^i, an $\boldsymbol{H}$-path is the graph P obtained by sequentially gluing together the graphs H^i at their endpoints. That is, for all $2 \le i \le t$, H^i is glued to H^{i-1} by identifying w_1^i with w_2^{i-1}. The endpoints of the path P are then taken to be w_1^1 and w_2^t.

Definition 5.5.14. Let $r, s \in \mathbb{N}$, let H be an r-vertex graph and let $\underline{\boldsymbol{H}} := \{\boldsymbol{H}_{i,j} : i \in [r], j \in [s]\}$ be a labelled family of vectors of $(r+1)$-vertex graphs (with tuples of distinguished vertices). Then an $(\underline{\boldsymbol{H}}, H)$-absorbing gadget is a graph obtained by the following procedure. Take disjoint $\boldsymbol{H}_{i,j}$-paths for $1 \le i \le r$ and $1 \le j \le s$ and denote their endpoints by $u_{i,j}$ and $v_{i,j}$. Place a copy of H on $\{v_{i,j} : i \in [r]\}$ for each $j \in [s]$. For $2 \le i \le r$, identify all vertices $\{u_{i,j} : 1 \le j \le s\}$ and relabel this vertex u_i. Finally relabel $u_{1,j}$ as w_j for $j \in [s]$ and let $W := \{w_1, w_2, \dots, w_s\}$, which we refer to as the *base set* of vertices for the absorbing gadget.

An example of an absorbing gadget is given in Figure 5.5. Recall that we always consider K_{r+1}^- to have two distinguished vertices which form the only non-edge of the graph. In the previous section we commented on how a (K_{r+1}^-, t)-path P with endpoints x and y has two K_r-matchings covering all but one vertex; the first misses x, the other misses y. The point of the absorbing gadget is to generalise this property, giving a graph which can use any one of a number of vertices (the base set) in a K_r-matching. In more detail, suppose $s, t \in \mathbb{N}$ and for all $1 \le i \le r$ and $1 \le j \le s$, define $\boldsymbol{H}_{i,j} = (K_{r+1}^-, t)$ the vector composed of t copies of K_{r+1}^-. Furthermore fix the labelled family $\underline{\boldsymbol{H}} := \{\boldsymbol{H}_{i,j} : i \in [r], j \in [s]\}$ (which is simply rs copies the vector (K_{r+1}^-, t) with each copy indexed by an $i \in [r]$ and a $j \in [s]$) and fix $H = K_r$. Now consider the $(\underline{\boldsymbol{H}}, H)$-absorbing gadget F which is obtained by taking s copies of $H = K_r$ and rs copies of the (K_{r+1}^-, t)-path and gluing these graphs together at certain vertices according to Definition 5.5.14 (see Figure 5.5

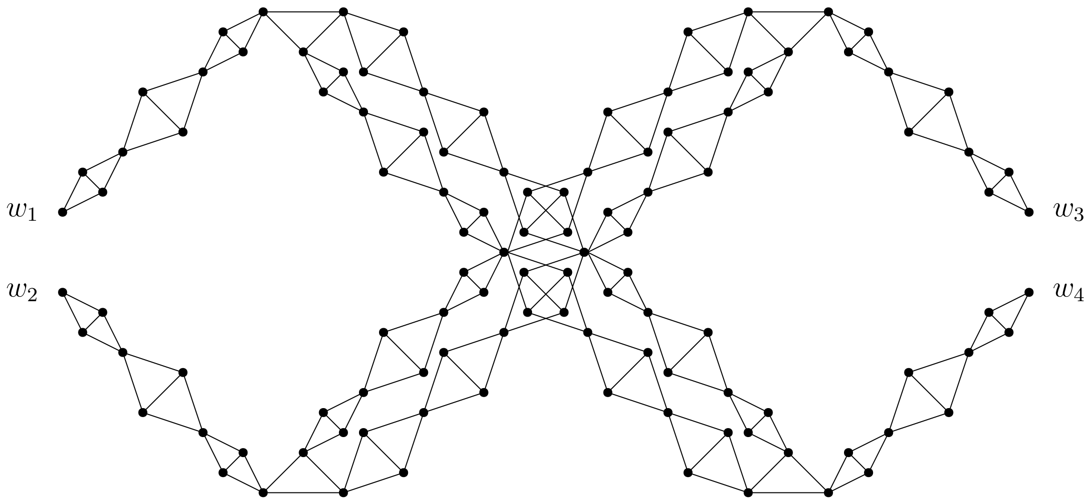


FIGURE 5.5: An $(\underline{\boldsymbol{H}}, K_3)$-absorbing gadget with $\underline{\boldsymbol{H}} := \{\boldsymbol{H}_{i,j} : i \in [3], j \in [4]\}$ such that each $\boldsymbol{H}_{i,j} = (K_4^-, 3)$. The base set of the absorbing gadget is $W := \{w_1, w_2, w_3, w_4\}$.

for an example with $r = 3$ and $s = 4$). Let $W = \{w_1, \ldots, w_s\}$ be the base set given by the definition of the absorbing gadget. We claim that F has the property that for *any* $j^* \in [s]$, there is a K_r-matching covering precisely $(V(F) \setminus W) \cup \{w_{j^*}\}$. Indeed, we have that for all $j \neq j^*$ and $i \in [r]$, there is a K_r-matching in the $\boldsymbol{H}_{i,j}$-path $P_{i,j}$ which uses[7] $v_{i,j}$ and not the other endpoint of $P_{i,j}$. Then there is a K_r-matching in the $\boldsymbol{H}_{1,j^*}$-path which uses w_{j^*}, a K_r-matching in the $\boldsymbol{H}_{i,j^*}$-path for $2 \leq i \leq r$ which uses u_i, and a copy of K_r on $\{v_{i,j^*} : i \in [r]\}$ which completes the desired K_r-matching.

More generally, for any $s, t^* \in \mathbb{N}$, letting $\mathcal{H} = \mathcal{H}(K_{r+1}^-, \leq t^*) = \{(K_{r+1}^-, t) : 1 \leq t \leq t^*\}$ be the set of vectors whose lengths are at most t^* and whose entries are all K_{r+1}^-, we have that for any labelled family of vectors $\underline{\boldsymbol{H}} := \{\boldsymbol{H}_{i,j} : i \in [r], j \in [s]\}$ where each $\boldsymbol{H}_{i,j}$ is an element from $\mathcal{H}$, an $(\underline{\boldsymbol{H}}, K_r)$-absorbing gadget F has the same property. That is, if $W = \{w_1, \ldots, w_s\}$ is the base set of F, then for *any* $j \in [s]$, there is a K_r-matching covering precisely $(V(F) \setminus W) \cup \{w_j\}$.

As in the previous subsection, we begin by showing that there are many absorbing gadgets in the deterministic graph. Again, although we are interested in $(\underline{\boldsymbol{H}}, K_r)$-absorbing gadgets for some $\underline{\boldsymbol{H}}$ consisting of vectors, all of whose entries are K_{r+1}^-, we split the edges of our absorbing gadget and rely on the deterministic graph to provide many copies of a subgraph of the gadget. In particular, we will use here our paradigm H_{det}, defined in Definition 5.2.1. The following general proposition allows us to show that we can find many absorbing gadgets if all the vertices which we hope to map the base set to, are reachable to each other.

Definition 5.5.15. Let $r, s \in \mathbb{N}$. Let $\mathcal{H}$ be a finite set of vectors, such that each entry of each vector in $\mathcal{H}$ is an $(r+1)$-vertex graph with a tuple of distinguished vertices. We write $\mathcal{H}(r \times s)$ for the collection of all ordered labelled sets $\underline{\boldsymbol{H}} := \{\boldsymbol{H}_{i,j} : i \in [r], j \in [s]\}$ where each $\boldsymbol{H}_{i,j}$ is

[7]We label all vertices in this discussion as in Definition 5.5.14.

an element from $\mathcal{H}$. If $\mathcal{H}$ consists of a single vector $\boldsymbol{H}$ we write $\boldsymbol{H}(r \times s) := \mathcal{H}(r \times s)$. That is, $\boldsymbol{H}(r \times s)$ is the ordered labelled (multi-)set with each element a copy of $\boldsymbol{H}$.

Proposition 5.5.16. *Let $\alpha, \gamma, \beta' > 0$, $s_0, k, r \in \mathbb{N}$ and let $\mathcal{H}$ be a finite set of vectors, such that each entry of each vector in $\mathcal{H}$ is an $(r+1)$-vertex graph with a tuple of distinguished vertices.*

Then there exists $\beta = \beta(\alpha, \gamma, \beta', s_0, k, r, \mathcal{H}) > 0$, such that for sufficiently large n, if G is an n-vertex graph with vertex subset $U \subseteq V(G)$ such that U is $(\mathcal{H}; \beta')$-closed, $|U| \geq \alpha n$ and $\delta(G[U]) \geq \left(1 - \frac{k}{r} + \gamma\right)|U|$, then for any set $X = \{x_1, \ldots, x_s\} \subset U$ with $|X| \leq s_0$, there exists some $\underline{\boldsymbol{H}} \in \mathcal{H}(r \times s)$ and some $(\underline{\boldsymbol{H}}, H_{det})$-absorbing gadget F with base set $W = \{w_1, \ldots, w_s\}$ such that there are at least $\beta n^{v(F)-s}$ embeddings of F in G which map w_i to x_i for $i \in [s]$.

Proof. Firstly notice that for a fixed $s \leq s_0$, there are a finite number (i.e. $|\mathcal{H}|^{rs}$) of $(\underline{\boldsymbol{H}}, H_{\det})$-absorbing gadgets F such that $\underline{\boldsymbol{H}} \in \mathcal{H}(r \times s)$ and F has a base set of size s. Let $\mathcal{F}_s$ be the set of all such absorbing gadgets, let $f := |\mathcal{F}_s|$ and set $Q := \max\{v_F - s : F \in \mathcal{F}_s\}$. We claim that there is some $\beta'' = \beta''(\alpha, \gamma, \beta', s_0, k, r, \mathcal{H}) > 0$ such that with G and U as in the statement of the proposition and $X \subset U$ of size s, there are at least $\beta'' n^Q$ subsets $S \subseteq V(G) \setminus X$ of Q ordered vertices such that there is an embedding of some $F \in \mathcal{F}_s$ in G which maps the base set of F to X and the other vertices to a subset[8] of S. Given this claim, the conclusion of the proposition follows easily. Indeed, by averaging we get that there is some $F \in \mathcal{F}_s$ and at least $\frac{\beta''}{f} n^Q$ ordered subsets S of Q vertices in $V(G)$ as above, that correspond to an embedding of F. Then setting $\beta := \frac{\beta''}{Q! f}$, we get that there must be at least $\beta n^{v_F - s}$ embeddings of F in G which map the base set to X. Indeed for each such embedding F' of F, the vertex set $V(F') \setminus X$ lies in at most $Q! n^{Q-(v_F-s)}$ different ordered sets of vertices $S \subseteq V(G)$.

So it remains to find these $\beta'' n^Q$ ordered subsets S. We will show that S can be generated in a series of steps so that every time we choose some a vertices, we have $\Omega(n^a)$ choices. We will use the notation of Definition 5.5.14. Firstly we select $r-1$ vertices $Y = \{y_2, y_3, \ldots, y_r\}$ in $U \setminus X$ which we can do in $\binom{|U \setminus X|}{r-1} = \Omega(n^{r-1})$ many ways. Now repeatedly find disjoint copies of $H_{\det}$ in $U \setminus (X \cup Y)$ and label these $\{z_{i,j} : 1 \leq i \leq r, 1 \leq j \leq s\}$ such that $\{z_{i,j} : 1 \leq i \leq r\}$ comprise a copy of $H_{\det}$ for each $j \in [s]$. In order to do this we repeatedly apply Lemma 2.4.1 and the degree condition which we can take to be $\delta(G[U]) \geq \left(1 - \frac{k}{r} + \frac{\gamma}{2}\right)|U|$ (ignoring any neighbours of vertices that have already been chosen in S). Hence there are $\Omega(n^{rs})$ choices for these copies of $H_{\det}$.

Now for $2 \leq i \leq r$ and $1 \leq j \leq s$, we have that y_i and $z_{i,j}$ are $(\boldsymbol{H}_{i,j}, \beta')$-reachable for some $\boldsymbol{H}_{i,j} \in \mathcal{H}$ of length $t_{i,j}$ say. Thus there are $\beta' n^{r t_{i,j} - 1}$ embeddings of an $\boldsymbol{H}_{i,j}$-path P in G which map the endpoints of P to $\{y_i, z_{i,j}\}$. We ignore those choices of embeddings of P which use previously chosen vertices of S, of which there are $O(n^{r t_{i,j} - 2})$. Similarly, for $1 \leq j \leq s$, x_j and $z_{1,j}$ are $(\boldsymbol{H}_{1,j}, \beta')$-reachable for some $\boldsymbol{H}_{1,j} \in \mathcal{H}$, so select an embedding of an $\boldsymbol{H}_{1,j}$-path

[8] In particular, if $v_F < Q$ then not all of the vertices of S are used in this embedding.

in G which maps the endpoints to $\{x_j, z_{1,j}\}$ and has all other vertices disjoint from previously chosen vertices. This gives an embedding of an $(\underline{\boldsymbol{H}}, H_{\det})$-absorbing gadget in G which maps the base set W to X, $v_{i,j}$ to $z_{i,j}$ for $2 \leq i \leq r$, $j \in [s]$ and maps u_i to y_i for $i \in [r]$. Choosing unused vertices arbitrarily until we have a set S of Q vertices, the claim and hence the proof of the proposition are settled. □

5.6 The absorbing structure - random edges

In this section, we will introduce the edges of $G(n, p)$ and show that $G \cup G(n, p)$ contains the absorbing structure we desire. The absorbing structure will be formed by choosing absorbing gadgets rooted on certain prescribed sets of vertices. The absorbing gadgets will be $(\underline{\boldsymbol{H}}, K_r)$-absorbing gadgets F^* for some $\underline{\boldsymbol{H}}$ consisting of vectors whose entries are all K_{r+1}^-. In order to obtain these absorbing gadgets, we consider the absorbing gadgets of just deterministic edges which we looked at in the previous section and show that with high probability, one of these matches up with random edges to get the required subgraph F^*. We begin by investigating the absorbing gadgets that we look for in the random graph.

5.6.1 Absorbing gadgets in the random graph

Recalling Definitions 5.2.1 and 5.5.14, let $\underline{\boldsymbol{H}} := \{\boldsymbol{H}_{i,j} : i \in [r], j \in [s]\}$ be a labelled family of vectors of $(r+1)$-vertex graphs and suppose that there is an embedding φ of an $(\underline{\boldsymbol{H}}, H_{\det})$-absorbing gadget F' in G which maps the base set of the gadget to some $U \subset V(G)$, with $|U| = s$. Recalling Definition 5.5.4, define $\overline{\underline{\boldsymbol{H}}} := \{\overline{\boldsymbol{H}_{i,j}} : i \in [r], j \in [s]\}$. Now in order to complete this absorbing gadget F' into one which has the form that we require, we have to find a labelled embedding of an $(\overline{\underline{\boldsymbol{H}}}, \overline{H_{\det}})$-absorbing gadget F onto the ordered vertex set $\varphi(V(F'))$ in $G(n, p)$. The following lemma will be used to show that there are sufficiently many embeddings in $G(n, p)$ of the necessary Fs defined as above. Indeed, the lemma lower bounds the parameters $\Phi_{F\setminus W}$ and $\Phi_{F,W}$ (where W is the base set of the absorber) which we will use to count embeddings of F in $G(n, p)$ via arguments using Janson's inequality. Indeed, we will soon appeal to Lemma 5.3.2 to prove the existence of absorbing gadgets $F \subset G(n, p)$ which match up with the deterministic absorbing gadgets $F' \subset G$, as discussed above. Lemma 5.3.2 requires lower bounds on the Φ values of the absorbing gadgets F, which can be seen as requiring that the gadgets F are sparse enough that we can expect to see many of them in $G(n, p)$. Note that, crucially, we will be interested in finding absorbing gadgets F in $G(n, p)$ with fixed base sets (hence our interest in the values $\Phi_{F,W}$). This will allow us to interweave our absorbing gadgets in a calculated way, leading to an absorbing structure with strong absorbing properties.

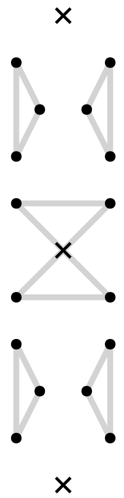
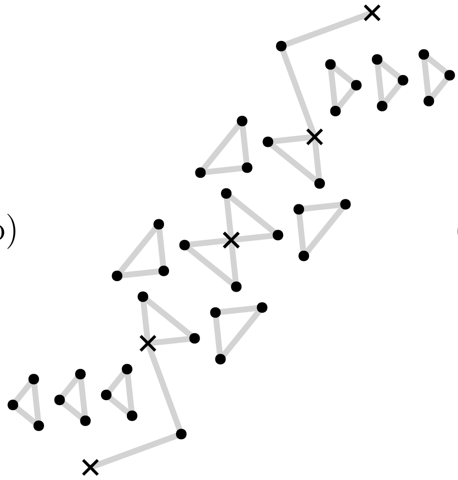
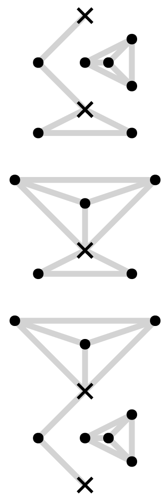
(a)  (b)  (c) 

FIGURE 5.6: Some examples of $\boldsymbol{H}$-paths for various $\boldsymbol{H}$ (see Definition 5.5.4 and Propositions 5.5.6, 5.5.7 and 5.5.12 for relevant definitions):

1. (a) An $\overline{\boldsymbol{H_1}}$ path with $r = 9$ and $k = q = r^* = 3$.
2. (b) An $\overline{\boldsymbol{H_2}}$ path with $r = 11, k = 3, q = 2$ and $r^* = 4$.
3. (c) An $\overline{\boldsymbol{H_3}}$ -path where $\boldsymbol{H_3} = (H_0, H_0', H_0', H_0) \in \boldsymbol{\mathcal{H}_3}$ and we have $r = 6, k = 4$ and $q = r^* = 2$.

It is worth noting that as F is uniquely defined by F', it is in fact the way that we chose our deterministic absorbing gadgets, that guarantees the following conclusions.

Lemma 5.6.1. *Let $k, r, s \in \mathbb{N}$ and $C > 1$, with $2 \le k \le r$ and suppose $p = p(n) \ge Cn^{-2/k}$. Suppose $\underline{\boldsymbol{H}}$ is such that:*

1. *$\underline{\boldsymbol{H}} \in \boldsymbol{H_1}(r \times s)$ if $\frac{r}{k} \in \mathbb{N}$, recalling the definition of $\boldsymbol{H_1}$ from Proposition 5.5.6;*
2. *$\underline{\boldsymbol{H}} \in \boldsymbol{H_2}(r \times s)$ if $\frac{r}{k} \notin \mathbb{N}$ and $k < \frac{r}{2}$, recalling the definition of $\boldsymbol{H_2}$ from Proposition 5.5.7;*
3. *$\underline{\boldsymbol{H}} \in \boldsymbol{\mathcal{H}_3}(r \times s)$ if $k > \frac{r}{2}$, recalling the definition of $\boldsymbol{\mathcal{H}_3}$ from Proposition 5.5.12.*

Then if F is an $(\underline{\boldsymbol{H}}, \overline{H_{det}})$-absorbing gadget with base set W such that $|W| = s$, we have that $\Phi_{F\setminus W} \ge Cn$ and $\Phi_{F,W} \ge Cn^{1/k}$.

Proof. We recommend that the reader refers to the examples in Figure 5.6 to help visualise some of the ideas in this proof. Note that as the endpoints of an $\overline{\boldsymbol{H_1}}$-path are isolated, we have that the base set of an $(\overline{\boldsymbol{H_1}}(r \times s), \overline{H_{\text{det}}})$ absorbing gadget F is also an isolated set of vertices and so $\Phi_{F,W} = \Phi_{F\setminus W}$. Defining $K_k + K_k$ as two copies of K_k which meet in a singular vertex, we have that $F \setminus W$ consists of disjoint copies of K_k and $r \times s$ disjoint copies of $K_k + K_k$, one for each $\overline{\boldsymbol{H_1}}$-path used in F. Therefore Lemma 5.3.1 (1) shows that $\Phi_{K_k} \ge Cn$, and repeated applications of Lemma 5.3.1 (3) show that $\Phi_{K_k+K_k} \ge Cn$ and in turn $\Phi_{F\setminus W} \ge Cn$ as required.

Case 2 is similar. Here we have that $q = r - k\lfloor \frac{r}{k} \rfloor < k$ and each of the base vertices w of F lie in a copy, say F_w, of the graph defined as follows. Take a copy of K_{q+1}^- and a copy of K_k that meet in exactly one vertex, which is one of the vertices of the nonedge in K^-_{q+1}. Furthermore, we have that the base vertex w is the other vertex in the nonedge of this copy of K^-_{q+1}. We have that each of the F_w is disconnected from the rest of F and an application of Lemma 5.3.1 (1), (2) and (3) gives that $\Phi_{F_w \setminus w} \geq Cn$ and $\Phi_{F_w,w} \geq Cn^{1/k}$ if $q \geq 2$. If $q = 1$, then F_w is an isolated vertex w and a copy of K_k so we have $\Phi_{F_w,w} = \Phi_{F_w \setminus w} \geq Cn$. Now note that $F \setminus (\cup_{w \in W} F_w)$ consists of copies of K_k, K^-_{q+1}, K_{q+1} and a copy of K_q (in the copy of $\overline{H_{\text{det}}}$ in F) which intersect each other in at most one vertex. Furthermore, one can view $F \setminus (\cup_{w \in W} F_w)$ as being 'built up' from these copies in the following way: there is an ordering (starting with $\overline{H_{\text{det}}}$) on these copies of K_k, K^-_{q+1}, K_{q+1} and K_q such that, starting with the empty graph and adding these copies in this order, each new copy shares at most one vertex with the previous copies already added, and at the end of the process we obtain $F \setminus (\cup_{w \in W} F_w)$. Each time we add a copy, we can apply Lemma 5.3.1 (3) and then again to add in the F_w (to obtain F). This leads us to conclude that $\Phi_{F \setminus W} \geq Cn$ and $\Phi_{F,W} \geq Cn^{1/k}$ as required.

In Case 3, let $q = r - k < k$ and let us fix some $\underline{\boldsymbol{H}} \in \mathcal{H}_3(r \times s)$ which then defines our F. For each $w \in W$, let F_w be the connected component of F which contains w. Due to the definition of $\mathcal{H}_3$, and in particular the fact that each $\boldsymbol{H} \in \mathcal{H}_3$ contains a copy of H'_0 as defined in Proposition 5.5.12, we have that $F_w \neq F_{w'}$ for all $w \neq w' \in W$. Also, for $q \geq 2$, it can be seen that F_w is a graph obtained by sequentially 'gluing' copies of K^-_{q+1} to vertices of degree $q - 1$. This gluing finishes with a copy of K_k being attached to a vertex of degree $q - 1$ (signifying that we have reached a copy of H'_0 in the vector $\boldsymbol{H}$) and with w being a vertex of degree $q - 1$ in the resulting graph. Similarly to the previous case, applications of Lemma 5.3.1 (2) and (3) imply that $\Phi_{F_w \setminus w} \geq Cn$ and $\Phi_{F_w,w} \geq Cn^{1/k}$ if $q \geq 2$ and if $q = 1$, we see that F_w is an isolated vertex, namely w itself. Also as before, we have that $F \setminus (\cup_{w \in W} F_w)$ consists of copies of K_k, K^-_{q+1}, K_{q+1} and a copy of K_q which intersect each other in at most one vertex. Thus, introducing the ordering of these copies as in Case 2, we can apply Lemma 5.3.1 repeatedly to obtain the desired conclusion. $\square$

We will use Lemma 5.6.1 to prove the existence of our desired absorbing gadgets in $G' = G \cup G(n, p)$. Before embarking on this however, we need to know how we wish our absorbing gadgets (in particular their base sets) to intersect in G'.

5.6.2 Defining an absorbing structure

Recall from Definition 2.8.1 that a *template* $\mathcal{T}$ with *flexibility* $m \in \mathbb{N}$ is a bipartite graph on $7m$ vertices with vertex classes I and $J_1 \cup J_2$, such that $|I| = 3m$, $|J_1| = |J_2| = 2m$, and for any $\overline{J} \subset J_2$, with $|\overline{J}| = m$, the induced graph $\mathcal{T}[V(\mathcal{T}) \setminus \overline{J}]$ has a perfect matching. Recall also that we call J_2 the *flexible* set of vertices and that for all m sufficiently large, templates of flexibility m

and maximum degree 40 exist (see Theorem 2.8.2). We will use a template here as an auxiliary graph in order to build an absorbing structure for our purposes.

Definition 5.6.2. Let $m, t^* \in \mathbb{N}$ and $\mathcal{T} = (I = \{1, \dots, 3m\}, J_1 \cup J_2 = \{1, \dots, 2m\} \cup \{2m + 1, \dots, 4m\}, E(\mathcal{T}))$ be a bipartite template with maximum degree $\Delta(\mathcal{T}) \leq 40$ and flexibility m as defined above. Further, let

$$\mathcal{H} := \mathcal{H}(K_{r+1}^-, \leq t^*) = \{(K_{r+1}^-, t) : 1 \leq t \leq t^*\}$$

be the set of vectors of length at most t^* whose entries are all K_{r+1}^-.

A (t^*-bounded) *absorbing structure* $\mathcal{A} = (\Pi, Z, Z_2)$ of flexibility m in a graph G' consists of a vertex set $Z = Z_1 \cup Z_2 \subset V(G')$ which we label $Z_1 := \{z_1, \dots, z_{2m}\}$ and $Z_2 := \{z_{2m+1}, \dots, z_{4m}\}$ and a set $\Pi := \{\varphi_1, \dots, \varphi_{3m}\}$ of embeddings of absorbing gadgets into G'. We require the following properties:

- For $i \in [3m]$, setting $N(i) := \{j : (i, j) \in E(\mathcal{T}) \subset I \times J\}$ and $n(i) := |N(i)|$, we have that φ_i is an embedding of some $(\underline{\boldsymbol{H}}, K_r)$-absorbing gadget F_i such that $\underline{\boldsymbol{H}} \in \mathcal{H}(r \times n(i))$ and the base set of F_i, which we denote W_i, is mapped to $\{z_j : j \in N(i)\} \subseteq Z$ by φ_i.
- The embeddings of the absorbing gadgets are vertex-disjoint other than the images of the base sets. That is, for all $i \in [3m]$, $\varphi_i(V(F_i) \setminus W_i) \subseteq V(G') \setminus Z$ and $\varphi_i(V(F_i) \setminus W_i) \cap \varphi_{i'}(V(F_{i'}) \setminus W_{i'}) = \emptyset$ for all $i \neq i' \in [3m]$.

We call Z_2 the *flexible set* of the absorbing structure.

Thus the absorbing structure is an embedding of a larger graph which is formed of $3m$ disjoint absorbing gadgets whose base vertices are then identified according to a template of flexibility m. We will refer to the vertices of $\mathcal{A}$ which are the vertices which feature in the embedding of this larger graph. That is,

$$V(\mathcal{A}) := \bigcup_{i \in [3m]} \varphi(V(F_i) \setminus W_i) \bigcup Z.$$

Remark 5.6.3. If $\mathcal{A}$ is a t^*-bounded absorbing structure of flexibility m, then it has less than $125t^*r^2m$ vertices in total (using that we assume $\Delta(\mathcal{T}) \leq 40$ here).

In our proof, we will bound t^* by a constant and look for an absorbing structure on a small linear number of vertices. The key property of the absorbing structure is that it inherits the flexibility of the template that defines it, but in the context of K_r-matchings, as detailed in the following remark.

Remark 5.6.4. If G' contains an absorbing structure $\mathcal{A} = (\Pi, Z, Z_2)$ of flexibility m, then for any subset of vertices $\overline{Z} \subset Z_2$ such that $|\overline{Z}| = m$, there is a K_r-matching in G' covering precisely $V(\mathcal{A}) \setminus \overline{Z}$.

Indeed given such a $\overline{Z}$, letting $\overline{J}$ be the corresponding indices from J, we have that $\mathcal{T}[V(\mathcal{T} \setminus \overline{J})]$ has a perfect matching. The perfect matching then indicates, for each $i \in [3m]$, which vertex z_{j_i} of Z to use in a K_r-matching of the corresponding absorbing gadget. That is, for each i, if φ_i is 'matched' to z_{j_i} by the perfect matching, then we take the K_r-matching covering $\varphi_i(F_i \backslash W_i) \cup \{z_{j_i}\}$ (which exists by the key property of the absorbing gadget mentioned after Definition 5.5.14) and then take their union.

We remark that our use of templates to define our absorbing structure in this way is reminiscent of the original application of Montgomery [136, 137] who used a template to dictate how absorbers are connected. His interest was in spanning trees and thus his absorbers were used to give flexibility over which vertices are included in certain paths in the tree. In relation to (hyper-)graph factors, although the use of templates had been used in several instances before this work (see for example [131, 146]), we are not aware of work before this which uses the template method in the same way as we do here. Indeed, most applications of the template method use a template to define a fixed absorbing structure which is then found in the graph. This is similar to our application of the method in Section 3.7 for example. Here our approach comes from a different angle. Instead of fixing representatives for all the vertices in I and J of a template $\mathcal{T}$ and then defining structures that correspond to the edges of the template, we rather focus only on the neighbourhoods in $\mathcal{T}$ of vertices in I, and use these to define our placement of absorbing gadgets. As previously mentioned, this is more in line with the original application of the method by Montgomery [136, 137], who worked with certain 'absorbing paths' as absorbing gadgets.

5.6.3 The existence of an absorbing structure

In order to prove the existence of an absorbing structure, we must find embeddings of absorbing gadgets in our graph. In the previous section we found many embeddings of certain absorbing gadgets with deterministic edges and thus it remains to find embeddings of complementary absorbing gadgets, using only random edges. Therefore we will turn to Lemma 5.3.2, which is a general result regarding embeddings in random graphs. However, there is still some work to do in the application of this lemma and the following proposition shows how we can use Lemma 5.3.2 repeatedly in order to embed a larger graph. We state the proposition in a more general form than just for showing the existence of absorbing structures as we will also use the result at other points in the proof. As the statement of the proposition is somewhat technical, we recommend that the reader sees how it is applied in Corollaries 5.6.6, 5.6.8 and 5.6.9 to help with digesting it.

Proposition 5.6.5. *Let $\kappa_d, \kappa_w, \kappa_e, \kappa_v, k \in \mathbb{N}$ and $\beta > 0$. Then there exists $\eta_0 > 0$ and $C > 0$ such that the following holds for any $0 < \eta < \eta_0$, $n \in \mathbb{N}$ and $t = \eta n \in \mathbb{N}$.*

Suppose that $F_1, \dots, F_t$ are labelled graphs with distinguished base vertex sets $W_i \subset V(F_i)$ such that $|W_i| \le \kappa_w$, $v_i := |V(F_i) \setminus W_i| \le \kappa_v$, $e(F_i[W_i]) = 0$ and $e(F_i) \le \kappa_e$ for all $i \in [t]$. Suppose further that $p = p(n)$ such that $\Phi_{F_i \setminus W_i} = \Phi_{F_i \setminus W_i}(n,p) \ge Cn$ and $\Phi_{F_i, W_i} = \Phi_{F_i, W_i}(n,p) \ge Cn^{1/k}$ for all $i \in [t]$. Let V be an n-vertex set, and $U_1, \dots, U_t \subset V$ be subsets such that $|U_i| = |W_i|$ for each $i \in [t]$, and defining

$$d(i) := |\{j \in [t] : U_i \cap U_j \neq \emptyset\}|,$$

we have that $d(i) \le \kappa_d$. Finally, suppose that $\mathcal{F}_1, \mathcal{F}_2, \dots, \mathcal{F}_t$ are families of vertex sets such that each $\mathcal{F}_i$ contains βn^{v_i} ordered subsets of V of size v_i.

Then whp there is a set of embeddings $\varphi_1, \varphi_2, \dots, \varphi_t$ such that each φ_i embeds a copy of F_i into $G(n,p)$ on V with W_i being mapped to U_i and $V(F_i) \setminus W_i$ being mapped to a set in $\mathcal{F}_i$ which does not intersect $\cup_{i \in [t]} U_i$. Furthermore for $i \neq i'$, we have that $\varphi_i(V(F_i) \setminus W_i) \cap \varphi_{i'}(V(F_{i'}) \setminus W_{i'}) = \emptyset$.

Proof. Fix $\kappa_v' := \kappa_v + 2\kappa_e$ and let $\eta_0 < \beta^2 2^{-(\kappa_v' + 2\kappa_e + 9)} (\kappa_v'!(\kappa_v' + \kappa_e + \kappa_w))^{-1}$. The idea here is to greedily extract the desired embeddings, finding them one at a time in $G(n,p)$. To achieve this, we use the multi-round exposure trick, having a constant number of phases such that in each phase we find a collection of embeddings. At the beginning of each phase we 'reveal' another copy of $G(n,p)$ on the same vertex set and focus *only* on the indices for which we have not yet found a suitable embedding, showing that in any sufficiently large subset of these indices there is an index for which we can find a suitable embedding. At each phase, we will apply Lemma 5.3.2 and so we first need to slightly adjust the sets we are considering in order to be in the setting of that lemma.

Firstly let us adjust each F_i so that it has κ_v' non-base vertices and κ_e edges. To each F_i add $\kappa_e - e(F_i)$ isolated edges. Then add isolated vertices until the resulting graph has $\kappa_v' + |W_i|$ vertices and redefine F_i as the resulting graph. Note that if $p = p(n)$ is such that $\Phi_{F_i \setminus W_i}(n,p) \ge Cn$ and $\Phi_{F_i, W_i}(n,p) \ge Cn^{1/k}$ for the original F_i as in the statement of the proposition, then these conditions are preserved under the above changes to F_i for each i, by Lemma 5.3.1. We also arbitrarily extend each set in each $\mathcal{F}_i$ to get sets of size κ_v'. As we can extend with any vertices not already in the set, it can be seen that we can have families $\mathcal{F}_i$ of size at least $\beta' n^{\kappa_v'}$ for some $\beta' > \frac{\beta}{2^{\kappa_v'}}$ which we now fix. Clearly, a set of valid embeddings of these new F_i (where[9] the new vertices of F_i are mapped to the new vertices from a set in $\mathcal{F}_i$) will also yield a set of embeddings of the original graphs we were interested in.

Now let us turn to the phases of our algorithm. We will generate $G(n,p)$ in $k+1$ rounds so that $G(n,p) = \cup_{j=1}^{k+1} G_j$ with each G_j an independent copy of $G(n,p')$, where p' is such

[9]This will be guaranteed in applications of Lemma 5.3.2 as the lemma is concerned with *labelled* embeddings.

that $(1-p)=(1-p')^{k+1}$. Note that for any graph F, vertex subset $W\subset V(F)$, constant $c>0$ and probability p, one has that $\Phi_{F\setminus W}(n,cp)=c'\Phi_{F\setminus W}(n,p)$ for some constant c' between 1 and $c^{e(F)}$. Likewise, multiplication of the probability by some constant $c>0$ results in multiplication of $\Phi_{F,W}$ by some constant factor. Hence, choosing $C>0$ sufficiently large, we can guarantee that if $\Phi_{F_i\setminus W_i}(n,p)\geq Cn$ and $\Phi_{F_i,W_i}(n,p)\geq Cn^{1/k}$ as in the statement of the proposition, then $\Phi_{F_i\setminus W_i}(n,p')\geq C'n$ and $\Phi_{F_i,W_i}(n,p')\geq C'n^{1/k}$ with C' such that $C'\geq\frac{2^{\kappa_v'+9}\kappa_v'!\kappa_v'}{\beta'^2}$. We fix such a $C>0$ and for $j=1,\dots,k$, we define

$$t_j:=\eta n^{1-(j-1)/k}((\kappa_d+1)\log n)^{j-1}\quad\text{ and }\quad s_j:=t_jn^{-1/k}\log n=\eta n^{1-j/k}(\kappa_d+1)^{j-1}(\log n)^j.$$

We also define $t_{k+1}:=(\kappa_d+1)s_k=\eta((\kappa_d+1)\log n)^k$, $s_{k+1}:=1$ and $t_{k+2}:=0$.

Now, as discussed, we look to choose embeddings one by one in order to reach the desired conclusion. Therefore, for the sake of brevity, at any point in the argument let us say that an embedding φ_i of F_i is *valid* if it maps W_i to U_i and maps $V(F_i\setminus W_i)$ to a set in $\mathcal{F}_i$ which is disjoint from $U:=\cup_{i\in[t]}U_i$ and also disjoint from $\varphi_{i'}(V(F_{i'}\setminus W_{i'}))$ for all indices $i'\in[t]$ for which we have already chosen an embedding. Our claim is that whp (with respect to $G(n,p)$) we can repeatedly choose valid embeddings until we have found embeddings for all t indices in $T:=\{1,\dots,t\}$. We therefore need to show that we never get stuck and that this greedy algorithm always finds a valid embedding. In order to do this, we split the algorithm into $k+1$ phases and rely on the edges of G_j in the j^{th} phase where we will find t_j-t_{j+1} valid embeddings. We will show that for all $j\in[k+1]$, conditioned on the fact that the algorithm has succeeded so far, we have that whp (with respect to $G_j=G(n,p')$) the algorithm will succeed for a further phase. The conclusion then follows easily as there are constantly many phases.

So let us analyse the j^{th} phase and condition on the fact that the process has been successful so far and so there are t_j indices that remain for us to find embeddings for. Let us further fix a specific set[10] of t_j indices $T_j\subseteq T$ that remain and some set of already chosen valid embeddings $\{\varphi_i:i\in R_j\}$ where $R_j:=T\setminus T_j$. By the law of total probability, it suffices to condition on this fixed set of embeddings so far and show that whp (with respect to G_j) we can repeatedly find valid embeddings, each time removing the corresponding index from T_j, until there are t_{j+1} indices remaining. So let $V_j'':=\cup_{i\in R_j}\varphi_i(V(F_i))\cup U$ and for $i\notin R_j$, define $\mathcal{F}_i^{(j)}:=\{S\in\mathcal{F}_i:S\cap V_j''=\emptyset\}$. We have that $|\mathcal{F}_i^{(j)}|\geq\frac{\beta'}{2}n^{\kappa_v'}$ as $|V_j''|<\frac{\beta'}{2}n$ due to our condition on η_0. We then apply Lemma 5.3.2 to the sets $\mathcal{F}_i^{(j)}$ such that $i\in T_j$, and where $t_j,s_j,\frac{\beta'}{2},\kappa_v',\kappa_v',\kappa_w,\kappa_e$ and p' play the rôles of t,s,β,L,v,w,e and p respectively. Let us check that the conditions needed for the lemma are satisfied. Indeed, we certainly have

[10]Note that when $j=1$ we must have that $T_1=T$.

that $\kappa'_v t_j \leq \kappa'_v t \leq \frac{\beta' n}{8\kappa'_v}$, $s_j \kappa_w \leq \frac{\beta' n}{8\kappa'_v}$ and $\binom{t_j}{s_j} \leq \binom{n}{s_j} \leq 2^n$. Moreover, when $1 \leq j \leq k$,

$$C' s_j n^{1/k} = C' t_j \log n \geq \left(\frac{2^{k'_v+9} \kappa'_v ! \kappa'_v}{\beta'^2} \right) t_j \log n \text{ and } C' n \geq \left(\frac{2^{\kappa'_v+9} \kappa'_v !}{\beta'^2} \right) n,$$

by our definition of C', whilst for $j = k+1$, $C' s_{k+1} n^{1/k} = \omega(t_{k+1} \log n)$. This verifies the conditions in (5.3.1) in all cases and so we conclude that whp, given any set B_j of at most $\kappa'_v t_j$ vertices and any set S_j of s_j indices in T_j such that the sets U_i with $i \in S_j$ are pairwise disjoint, there is an index $i^* \in S_j$ and a valid embedding of F_{i^*} in G_j which avoids B_j. This then implies that the greedy process will succeed throughout this phase. Indeed, we can now initiate with $B_j = \emptyset$ and repeatedly find indices $i \in T_j$ for which we have a valid embedding φ_i. We add this embedding to our chosen embeddings, add the vertices of it to B_j and delete the index i from T_j. The conclusion that we drew from Lemma 5.3.2 above asserts that we continue this process until we have t_{j+1} indices remaining in T_j, which is precisely what we need. Indeed, for $1 \leq j \leq k$, if we have more than t_{j+1} indices in T_j left then by the upper bound on $d(i)$ for i in T_j taking a maximal set $S \subset T_j$ such that U_i are all pairwise disjoint for $i \in S$, we have that $|S| \geq \frac{t_{j+1}}{\kappa_d+1} \geq s_j$. In the final phase when $j = k+1$ we can simply find embeddings one at a time as $s_{k+1} = 1$. This concludes the proof. □

As corollaries, we can conclude the existence of absorbing structures in $G \cup G(n,p)$. We split the cases here as Case 1 and 2 are much simpler.

Corollary 5.6.6. *Let $k, r \in \mathbb{N}$ such that either $2 \leq k \leq \frac{r}{2}$ or $k = r$ and let $\gamma > 0$. There exists $\eta_0 > 0$ and $C > 0$ such that if $p \geq Cn^{-2/k}$ and G is an n-vertex graph with minimum degree $\delta(G) \geq \left(1 - \frac{k}{r} + \gamma\right)n$, then for any $0 < \eta < \eta_0$ and any set of $2\eta n$ vertices $X \subseteq V(G)$, whp there exists a 4-bounded absorbing structure $\mathcal{A} = (\Pi, Z, Z_2)$ in $G' := G \cup G(n,p)$ of flexibility $m := \eta n$, which has flexible set $Z_2 = X$.*

Proof. We look to apply Proposition 5.6.5 and simply need to establish the hypothesis of the proposition. Consider a bipartite template $\mathcal{T} = (I = \{1, \ldots, 3m\}, J_1 \cup J_2 = \{1, \ldots, 2m\} \cup \{2m + 1, \ldots, 4m\}, E(\mathcal{T}))$ as in Definition 5.6.2; recall such a template exists (see Theorem 2.8.2). Fix $Z_2 = X = \{z_{2m+1}, \ldots, z_{4m}\}$ and choose an arbitrary set of $2m$ vertices $Z_1 \subset V(G) \setminus Z_2$ which we label $\{z_1, \ldots, z_{2m}\}$. Now towards applying Proposition 5.6.5, we set $t := 3m$ and for $i \in [t]$ we define the sets $U_i := \{z_j : j \in N(i)\}$ where $N(i)$ is as in Definition 5.6.2. Note that we can set $\kappa_d := 1600$ as we start with a template $\mathcal{T}$ with $\Delta(\mathcal{T}) \leq 40$, so for any set $N(i) \subset J$ (of at most 40 vertices), there are at most 1600 indices $i' \in I = [3m]$ such that $N(i') \cap N(i) \neq \emptyset$.

Now, fixing i, the collection $\mathcal{F}_i$, which we will use when applying Proposition 5.6.5, will be obtained from Proposition 5.5.16. Indeed, this proposition implies, along with Propositions 5.5.6 and 5.5.7, that there is some $\beta > 0$ such that the following holds with $a = 1$ if $\frac{r}{k} \in \mathbb{N}$ (Case 1) and $a = 2$ otherwise (Case 2).

Claim 5.6.7. *For any set U of at most 40 vertices, there is an $(\boldsymbol{H}_a(r \times |U|), H_{det})$-absorbing gadget*[11] *$F'$ such that there are at least $\beta n^{v_{F'}-|U|}$ embeddings of F' in G which map the base set of the absorbing gadget to U.*

For each i, apply Claim 5.6.7 with U_i playing the rôle of U to obtain a collection $\mathcal{F}_i$ of ordered vertex sets from $V(G)$ that combined with U_i each span such an absorbing gadget $F_i' = F'$. For each such embedding of F_i', if we have an ordered $(\overline{\boldsymbol{H}_a}(r \times |U_i|), \overline{H_{\text{det}}})$-absorbing gadget F_i (in $G(n,p)$), on the same vertex set, then we obtain the desired embedding φ_i of a $(\boldsymbol{K}_a(r \times |U_i|), K_r)$-absorbing gadget in $G \cup G(n,p)$, where $\boldsymbol{K}_a$ is a $(K_{r+1}^-, 2a)$-path. Applying Proposition 5.6.5 with small enough $\eta > 0$ thus gives us the absorbing structure, upon noticing that the conditions on Φ_{F_i,W_i} and $\Phi_{F_i \setminus W_i}$ are satisfied by Lemma 5.6.1. □

The third case, when $\frac{r}{2} < k \le r-1$, follows the exact same method of proof. The main difference comes from the fact that we do not have many absorbing gadgets for *all* small sets of vertices in the deterministic graph but only for sets which lie in one part of the partition dictated by Lemma 5.5.12. Therefore we look to find an absorbing structure in each part of the partition. Thus when we apply Proposition 5.6.5, we do so to find all these absorbing structures at once, in order to guarantee that these absorbing structures are disjoint. The conclusion is as follows.

Corollary 5.6.8. *Let $\frac{r}{2} < k \le r-1$ be integers, and define $q := r-k$, $c := \lceil \frac{r}{q} \rceil$ and $\gamma > 0$. Then there exists $\alpha > 0$ such that the following holds for all $0 < \varepsilon < \frac{\alpha\gamma}{4}$. There exists $C > 0$ and $\eta_0 > 0$ such that if $p \ge Cn^{-2/k}$ and G is an n-vertex graph with minimum degree $\delta(G) \ge \left(1 - \frac{k}{r} + \gamma\right)n$, then for any $0 < \eta < \eta_0$ there is a partition $\mathcal{P} = \{V_1, V_2, \ldots, V_\rho, W\}$ of $V(G)$ into at most c parts with the following properties:*

- *$|V_i| \ge \alpha n$ for $i \in [\rho]$;*
- *$|W| \le \varepsilon n$;*
- *$\delta(G[V_i]) \ge \left(1 - \frac{k}{r} + \frac{\gamma}{4}\right)|V_i|$ for each $i \in [\rho]$;*
- *For any collection of subsets $X_i \subset V_i$ such that $1.8\eta|V_i| \le |X_i| \le 2\eta|V_i|$ and $|X_i|$ even for all $i \in [\rho]$, whp there exists a set of $c(2^{c+1}+1)$-bounded absorbing structures $\{\mathcal{A}_i = (\Pi_i, Z_i, Z_{i2}) : i \in [\rho]\}$ in $G' := G \cup G(n,p)$ such that each $\mathcal{A}_i$ has flexibility $m_i := \frac{|X_i|}{2}$ and has flexible set $Z_{i2} = X_i$. Furthermore $V(\mathcal{A}_i) \cap V(\mathcal{A}_{i'}) = \emptyset$ for all $i \ne i' \in [\rho]$.*

Proof. We begin by applying Proposition 5.5.12 to get a vertex partition $\mathcal{P}$ with at most $c-1$ parts and in each part $U \in \mathcal{P}$ we remove any vertex v which has internal degree $\deg_G(v; U) < \left(1 - \frac{k}{r} + \frac{\gamma}{2}\right)|U|$, and add v to W. The resulting partition is the partition we will use. Choosing $\varepsilon_{5.5.12}$

[11] Recall here the definition of $\boldsymbol{H}_1$ from Proposition 5.5.6, of $\boldsymbol{H}_2$ from Proposition 5.5.7 and H_{det} from Definition 5.2.1. The notation $\boldsymbol{H}_a(r \times |U|)$ is also defined as in Definition 5.5.15.

in the application of Proposition 5.5.12 to be less than $\frac{\varepsilon}{c}$, we have that the first three bullet points are satisfied. Below we show the last bullet point, and to aid readability we temporarily fix $i = 1$.

Now given a set of $X_1 \subset V_1$ we choose a set $Z_1 \subset V_1 \setminus X_1$ such that $|Z_1| = 2m_1$. Further, according to some template $\mathcal{T} = (I = \{1, \dots, 3m_1\}, J_1 \cup J_2 = \{1, \dots, 2m_1\} \cup \{2m_1+1, \dots, 4m_1\}, E(\mathcal{T}))$ as in Definition 5.6.2, we label Z_1 according to J_1 and X_1 according to J_2 and identify sets $U_{i'} \subseteq X_1$ for each $i' \in [3m_1]$ according to the neighbourhood of i' in $\mathcal{T}$. As in Corollary 5.6.6, by Propositions 5.5.12 and 5.5.16 there exists some $\beta > 0$ such that for each $i' \in [3m_1]$, fixing $s_{i'} = |U_{i'}|$ the following holds. There is some $\underline{\boldsymbol{H}}_{i'} \in \mathcal{H}_3(r \times s_{i'})$ and some $(\underline{\boldsymbol{H}}_{i'}, H_{\mathrm{det}})$-absorbing gadget $F'_{i'}$ such that, defining $v_{i'} = v_{F'_{i'}}$ there are at least $\beta n^{v_{i'} - s_{i'}}$ embeddings of $F'_{i'}$ in G which map the base set of $F'_{i'}$ to $U_{i'}$. Each of these embeddings gives a candidate vertex set for which we could embed an $(\overline{\underline{\boldsymbol{H}}_{i'}}, \overline{H_{\mathrm{det}}})$-absorbing gadget, say $F_{i'}$ to get a copy of a $(\underline{\boldsymbol{K}}, K_r)$-absorbing gadget in G', with base set $U_{i'}$, where $\underline{\boldsymbol{K}} \in \mathcal{H}(r \times s_{i'})$ and $\mathcal{H} = \mathcal{H}(K^-_{r+1}, \leq c(2^{c+1} + 1))$. Using Lemma 5.6.1, we can now apply Proposition 5.6.5 (provided $\eta > 0$ is sufficiently small) to get the desired embeddings of all the $F_{i'}$ which gives an absorbing structure $\mathcal{A}_1$ as in the statement of the corollary. We in fact apply Proposition 5.6.5 for all $i \in [\rho]$ at once which gives the collection of absorbing structures as required. □

Before proving the upper bound in our main result, Theorem III, we give one last consequence of Proposition 5.6.5 which will be useful for us.

Corollary 5.6.9. *Suppose that $2 \leq k \leq r$ and $\gamma, \beta > 0$. Then there exists $\alpha > 0$ and $C > 0$ such that the following holds. Suppose G is an n-vertex graph with disjoint vertex sets U, W such that $|U| \leq \alpha n$, $|W| \geq \beta n$ and for all $v \in U \cup W$, $\deg_G(v; W) \geq \left(1 - \frac{k}{r} + \gamma\right)|W|$ and $p = p(n)$ is such that $p \geq Cn^{-2/k}$. Then whp in $G \cup G(n, p)$ there is a set of $|U|$ disjoint copies of K_r so that each copy of K_r contains a vertex of U and $r - 1$ vertices of W.*

Proof. Firstly, let $r^* := \lceil \frac{r}{k} \rceil$. By the fact that $\deg_G(v; W) \geq \left(1 - \frac{k}{r} + \gamma\right)|W|$ for all $v \in U \cup W$, we have that each vertex $u \in U$ is in at least $\left(\frac{\beta\gamma}{2} n\right)^{r^*}$ distinct copies of $K^-_{r^*+1}$ in G such that the other vertices of each copy lie in W, and u is contained in the nonedge of each $K^-_{r^*+1}$. Thus by Lemma 2.4.1, there exists some $\beta' > 0$ such that each $u \in U$ is in $\beta' n^{r-1}$ copies of H_{det} with the other vertices of each copy in W, and u in the part of size $q := r - (r^* - 1)k$ in H_{det}. Let $\mathcal{F}_u$ be the collection of $(r-1)$-sets of vertices in W that, together with u, give rise to these copies of H_{det} containing u. Set $F_u := \overline{H_{\mathrm{det}}} = K_r \setminus H_{\mathrm{det}}$ with an identified vertex w_u in the clique of size q in $\overline{H_{\mathrm{det}}}$. Thus an ordered embedding in $G(n, p)$ of F_u which maps w_u to u and $V(F_u) \setminus \{w_u\}$ to an ordered set in $\mathcal{F}_u$ will give an embedding of K_r in $G \cup G(n, p)$ containing u and vertices of W. By Lemma 5.3.1 we have that $\Phi_{F_u, w_u} \geq Cn^{1/k}$ and $\Phi_{F_u \setminus w_u} \geq Cn$. Thus, provided $\alpha > 0$ is sufficiently small, an application of Proposition 5.6.5 gives the desired set of embeddings of K_r in $G \cup G(n, p)$. □

5.7 Perturbed thresholds for clique factors

In this section we prove the upper bound of Theorem III. Fix some sufficiently large $n \in r\mathbb{N}$ and let G be an n-vertex graph with $\delta(G) \geq \left(1 - \frac{k}{r} + \gamma\right)n$. We will show that there exists $C = C(\gamma, k, r) > 0$ such that if $p \geq Cn^{-2/k}$, then $G' := G \cup G(n,p)$ whp contains a K_r-factor. Again, we split the proof according to the parameters. We first treat Cases 1 and 2 together (i.e. when $2 \leq k \leq \frac{r}{2}$ or $k = r$). Here we avoid many of the technicalities which occur in Case 3 and the main scheme of the proof is clear.

Theorem III upper bound: Proof of Cases 1 and 2. Suppose $2 \leq k \leq \frac{r}{2}$ or $k = r$, and let $C, C' > 0$ be chosen so that we can express $G(n,p) = \cup_{j=1}^{4} G_j$ with each G_j a copy of $G(n,p')$ where $p' \geq C'n^{-2/k}$ and $C' > 0$ is large enough to be able to draw the desired conclusions (whp) in what follows. In the following, we will reveal the graphs G_j for $j \in [4]$ at different points in the argument, making use of their independence. We will assume that each G_j satisfies some property that occurs whp in $G(n,p')$ and show that if this is the case, then we can find a K_r-factor. A simple union bound implies that whp all the G_j indeed satisfy their given properties and so whp we find a K_r-factor in $G \cup G(n,p) = G \cup (\cup_{j=1}^{4} G_j)$. Now fix $0 < \eta < \min\left\{\frac{\gamma}{2000r^2}, \eta_0\right\}$ where η_0 is as in Corollary 5.6.6 and consider $X' \subseteq V(G)$ to be the subset generated by taking every vertex in $V(G)$ in X' with probability 1.9η, independently of the other vertices. Whp, by Chernoff's theorem (Theorem 2.1.1), we have that $1.8\eta n \leq |X'| \leq 2\eta n$ and for every vertex $v \in V(G)$, $\deg_G(v; X') \geq \left(1 - \frac{k}{r} + \frac{3\gamma}{4}\right)|X'|$. Take an instance of X' where this is the case and let $X := X'$ if $|X'|$ is even and $X := X' \cup \{x\}$ for some arbitrary vertex $x \in V(G) \setminus X'$ if $|X'|$ is odd. Apply Corollary 5.6.6 to get a 4-bounded absorbing structure $\mathcal{A} = (\Pi, Z, Z_2)$ in $G \cup G_1$ with flexibility $\frac{|X|}{2}$ and flexible set $Z_2 = X$. Remark 5.6.3 implies $|V(\mathcal{A})| \leq 500r^2\eta n \leq \frac{\gamma n}{4}$.

Then letting $V' := V(G) \setminus V(\mathcal{A})$, we have that $\delta(G[V']) \geq \left(1 - \frac{k}{r} + \frac{\gamma}{2}\right)|V'|$. Choose $\alpha := \min\left\{\alpha_{5.6.9}, \frac{\gamma\eta}{4r}\right\}$, where $\alpha_{5.6.9}$ is the constant obtained when applying Corollary 5.6.9 with constants $r, k, \frac{\gamma}{2}, \eta$ playing the rôles of r, k, γ, β respectively.

Apply Theorem 5.4.1 to obtain a partial K_r-factor $\mathcal{K}_1$ in $(G \cup G_2)[V']$ covering all but at most αn vertices of V'. Let Y denote the set of those vertices in V' uncovered by $\mathcal{K}_1$. Apply Corollary 5.6.9 to obtain a partial K_r-factor $\mathcal{K}_2$ in $(G \cup G_3)[X \cup Y]$ which covers Y and covers precisely $(r-1)|Y| \leq \frac{\gamma\eta}{2}n \leq \frac{\gamma}{2}|X|$ vertices of X. Let $\tilde{X}$ be the set of those vertices in X not covered by $\mathcal{K}_2$. We have that $\delta(G[\tilde{X}]) \geq \left(1 - \frac{k}{r} + \frac{\gamma}{4}\right)|\tilde{X}|$ so we can apply Theorem 5.4.1 to obtain a K_r-matching $\mathcal{K}_3'$ in $(G \cup G_4)[\tilde{X}]$ which covers all but at most $\frac{|X|}{4}$ vertices of $\tilde{X}$. Here, we used that $|\tilde{X}| \geq \frac{\eta n}{2}$ is linear in size.

By Remark 5.6.4 we know that for any subset X'' of X of size $\frac{|X|}{2}$, there is a K_r-matching covering precisely $V(\mathcal{A}) \setminus X''$. Thus, $|V(\mathcal{A})| - \frac{|X|}{2}$ is divisible by r. Therefore, as the only vertices in $V(G)$ uncovered by $V(\mathcal{K}_1 \cup \mathcal{K}_2)$ are those from $(V(\mathcal{A}) \setminus X) \cup \tilde{X}$, there must be a (sub-)$K_r$-matching $\mathcal{K}_3 \subseteq \mathcal{K}_3'$ which covers all but exactly $\frac{|X|}{2}$ vertices of $\tilde{X}$.

Let $\overline{X}$ be the set of vertices of X that are covered by cliques in $\mathcal{K}_2 \cup \mathcal{K}_3$. Thus $|\overline{X}| = \frac{|X|}{2}$ and by Remark 5.6.4 there is a K_r-matching $\mathcal{K}_4$ in $G \cup G_1$ covering precisely $V(\mathcal{A}) \setminus \overline{X}$. Hence, $\mathcal{K} := \mathcal{K}_1 \cup \mathcal{K}_2 \cup \mathcal{K}_3 \cup \mathcal{K}_4$ gives a K_r-factor in $G \cup G(n, p)$ as required. □

If $\frac{r}{2} < k \leq r - 1$, we have to overcome a few technicalities. The idea is to apply Corollary 5.6.8 and to apply the same approach as above in each of the parts of the resulting partition to find a K_r-factor. Of course we also have to incorporate the vertices of the exceptional class W into copies of cliques in our factor; this is straightforward using Corollary 5.6.9. So we cover these vertices first before embarking on finding an almost factor.

More subtle is a problem that arises from divisibility. That is, when we cover each part with K_r-matchings according to the scheme above, we cannot guarantee that we are left with a subset of the flexible set of the right size to apply the key property of the absorbing structure. Therefore we embed 'crossing' copies of K_r in our flexible sets in order to resolve this divisibility hurdle at the end of our process. We find these copies in the following manner. Consider the graph $F := K_{\lceil \frac{r-1}{2} \rceil, \lfloor \frac{r-1}{2} \rfloor}$. Because of our minimum degree condition and Lemma 2.4.1, every part V_i contains at least $\gamma' n^{r-1}$ copies of F for some $\gamma' > 0$. Now let $\overline{F}$ be the graph consisting of a copy of $K_{\lceil \frac{r-1}{2} \rceil + 1}$ and a copy of $K_{\lfloor \frac{r-1}{2} \rfloor + 1}$ joined at a single vertex x, say. If we consider F and $\overline{F} \setminus x = \overline{F}[V(\overline{F}) \setminus \{x\}]$ to have the same vertex set so that $\overline{F} \setminus x = K_{r-1} \setminus F$, then $F \cup \overline{F}$ is a copy of K_r. Also note that it follows from Lemma 5.3.1 that $\Phi_{\overline{F}} \geq C'n$ for $p \geq C'n^{-2/k}$. We will look for embeddings of $K_r = F \cup \overline{F}$ in $G \cup G(n, p)$ such that the vertex x is mapped to one part of the partition and the $r - 1$ other vertices lie in another part of the partition.

Theorem III upper bound: Proof of Case 3. Suppose $\frac{r}{2} < k \leq r - 1$, $q := r - k$ and $c := \lceil \frac{r}{q} \rceil$. Now let $C, C' > 0$ be chosen so that we can express $G(n, p) = \cup_{j=1}^{4} G_j \cup_{i=1}^{c} (G_{i1} \cup G_{i2})$ with each G_j, G_{i1} and G_{i2} a copy of $G(n, p')$ where $p' \geq C'n^{-2/k}$ and $C' > 0$ is large enough to be able to draw the desired conclusions (whp) in what follows. As in the proof of cases 1 and 2, in each random graph we will use, we will assume that some high probability event occurs, and then prove the existence of a K_r-factor. As we have a constant number of independent random graphs, this will prove that whp a K_r-factor exists in $G \cup G(n, p)$, as required.

We use our first copy of $G(n, p')$ to find the crossing copies of K_r discussed above. Apply Corollary 5.6.8, letting $\alpha_1 > 0$ be the outcome of the corollary with input r, k, γ. Choose $0 < \varepsilon < \min\left\{\frac{\alpha_1 \gamma}{8r}, \alpha_{5.6.9}\right\}$, where $\alpha_{5.6.9}$ is the constant obtained when applying Corollary 5.6.9 with $r, k, \frac{\gamma}{4}, \frac{1}{2}$ playing the rôles of r, k, γ, β respectively. Thus, Corollary 5.6.8 yields a partition $V_1, \ldots, V_\rho, W$ of $V(G)$ where $\rho \leq c$ and $|W| \leq \varepsilon n$. Note $\delta(G[V_i]) \geq \left(1 - \frac{k}{r} + \frac{\gamma}{4}\right)|V_i|$ for each $i \in [\rho]$. Thus for each $i \in [\rho - 1]$ and every subset $V' \subseteq V(G)$ of at least $n - c(r - 1)r$

vertices, Lemma 2.4.1 implies that there exists some $\gamma' = \gamma'(r,k,\gamma) > 0$ such that there are at least $\gamma' n^r$ choices of pairs $(S,v) \in \binom{V'\cap V_{i+1}}{r-1} \times (V' \cap V_i)$ such that S hosts a copy of the graph $F := K_{\lceil \frac{r-1}{2} \rceil, \lfloor \frac{r-1}{2} \rfloor}$ discussed above.

Therefore, using that $\Phi_{\overline{F}} \geq C'n$ with the graph $\overline{F}$ as described above, we can apply Corollary 5.3.3 to conclude that for any subset V' of vertices of at least $n - c(r-1)r$ vertices and any $i \in [\rho - 1]$, there is a copy of K_r in $G \cup G_1$ which has $r-1$ vertices in V_{i+1} and one vertex in V_i. Therefore, we can greedily choose copies of K_r so that we have a set $\mathcal{R} := \cup_{i\in[\rho-1]} \mathcal{R}_i$ of disjoint copies of K_r in $G \cup G_1$ such that $\mathcal{R}_i$ contains $r-1$ copies of K_r with one vertex in V_i and $r-1$ vertices of V_{i+1}. Let $\mathcal{R}_\rho := \emptyset$ and $R_i := V(\mathcal{R}) \cap V_i$ for $i \in [\rho]$, where $V(\mathcal{R})$ denotes the vertices which feature in cliques in $\mathcal{R}$. Note that $|R_1| = r-1$, $|R_2| = |R_3| = \cdots = |R_{\rho-1}| = r(r-1)$ and $|R_\rho| = (r-1)^2$. We will incorporate these R_i into our flexible sets in order to use the copies of K_r that they define to fix divisibility issues that arise in the final stages of the argument.

Now fix $0 < \eta < \min\left\{\frac{\gamma}{4000c^2 2^c r^2}, \eta_0\right\}$ where η_0 is as in Corollary 5.6.8 and for each $i \in [\rho]$ consider $X_i' \subseteq V(G)$ to be a subset selected by taking every vertex in $V_i \setminus R_i$ with probability 1.9η, independently of the other vertices. Whp, by Chernoff's theorem (Theorem 2.1.1), we have that $1.8\eta|V_i| \leq |X_i'| \leq 2\eta|V_i| - r(r-1)$ and for every vertex $v \in V_i$, $\deg_G(v; X_i') \geq \left(1 - \frac{k}{r} + \frac{\gamma}{8}\right)|X_i'|$. Therefore, for each i, take an instance of X_i' where this is the case and let $X_i := X_i' \cup R_i$ if $|X_i'| + |R_i|$ is even and $X_i := X_i' \cup R_i \cup \{x\}$ for some arbitrary vertex $x \in V_i \setminus (X_i' \cup R_i)$ if $|X_i'| + |R_i|$ is odd. Apply Corollary 5.6.8 to get a collection $\{\mathcal{A}_i = (\Pi_i, Z_i, Z_{i2}) : i \in [\rho]\}$ of absorbing structures in $G \cup G_2$ such that each $\mathcal{A}_i$ has flexibility $\frac{|X_i|}{2}$ and flexible set $Z_{i2} = X_i$. By Remark 5.6.3 we have that $A := \cup_{i\in[\rho]} V(\mathcal{A}_i)$ is such that $|A| \leq 125c^2 2^{c+2} r^2 \eta n \leq \frac{\gamma}{8} n$.

Therefore, setting $V' := V(G) \setminus (W \cup A)$, we have that for every $w \in W \cup V'$, $\deg_G(w; V') \geq \left(1 - \frac{k}{r} + \frac{\gamma}{4}\right)|V'|$ and so an application of Corollary 5.6.9 yields a partial K_r-factor $\mathcal{K}_1$ in $G \cup G_3$ of $|W|$ cliques, each using one vertex of W and $r-1$ vertices of V'. Setting $V'' := V(G) \setminus (A \cup V(\mathcal{K}_1))$, we have that $\delta(G[V'']) \geq \left(1 - \frac{k}{r} + \frac{\gamma}{8}\right)|V''|$. So, as in the previous proof, we let $\alpha_2 := \min\left\{\alpha_{5.6.9}, \frac{\gamma\eta}{16r}\right\}$, where $\alpha_{5.6.9}$ is obtained from Corollary 5.6.9 (where $\frac{\gamma}{8}$ and η play the rôles of γ and β respectively), and we apply Theorem 5.4.1 to obtain a partial K_r-factor $\mathcal{K}_2$ in $(G \cup G_4)[V'']$ covering all but at most $\alpha_2 n$ vertices of V''. Let Y be the set of vertices from V'' uncovered by $\mathcal{K}_2$ and set $Y_i := Y \cap V_i$ for each $i \in [\rho]$.

Now for each $i \in [\rho]$ a simple application of Corollary 5.6.9 yields a K_r-matching $\mathcal{K}_{i1}$ in $G \cup G_{i1}$ which covers Y_i and uses precisely $(r-1)|Y_i| \leq \frac{\gamma\eta}{16} n \leq \frac{\gamma}{16}|X_i'|$ vertices of X_i'. Note that we do not use any vertices of $R = \cup_{i\in[\rho]} R_i$ in these cliques. For each $i \in [\rho]$ let $\tilde{X}_i$ be the vertices of $X_i \setminus R_i$ not involved in copies of K_r in $\mathcal{K}_{i1}$. As $\delta(G[\tilde{X}_i]) \geq \left(1 - \frac{k}{r} + \frac{\gamma}{16}\right)|\tilde{X}_i|$ and $|\tilde{X}_i|$ is linear, we can apply Theorem 5.4.1 to obtain a K_r-matching $\mathcal{K}_i'$ in $(G \cup G_{i2})[\tilde{X}_i]$ which covers all but at most $\frac{|X_i|}{4}$ vertices of $\tilde{X}_i$, for each $i \in [\rho]$.

Note that we will not use the full matchings $\mathcal{K}_i'$ in our final factor. So (ignoring for now the K_r-matchings $\mathcal{K}_i'$), it remains to cover the vertices in $(V(\mathcal{A}_i) \setminus X_i) \cup R_i \cup \tilde{X}_i$ for each $i \in [\rho]$. We do so by means of the following algorithm. We initiate with the $\mathcal{K}_i', \mathcal{R}_i$ as above and set $\overline{Z}_i := V(\mathcal{K}_{i1}) \cap X_i$ and $\mathcal{K}_{i2} := \emptyset$ for all $i \in [\rho]$, and set $i' = 1$. Now whilst $|\overline{Z}_{i'}| \leq \frac{|X_{i'}|}{2} - r + 1$, remove a clique from $\mathcal{K}_{i'}'$, add it to $\mathcal{K}_{i'2}$ and add its vertices to $\overline{Z}_{i'}$. Once this process stops, add $\frac{|X_{i'}|}{2} - |\overline{Z}_{i'}|$ copies of K_r in $\mathcal{R}_{i'}$ to $\mathcal{K}_{i'2}$, and add all their vertices in X_j to $\overline{Z}_j$ for $j = i', i' + 1$. If $i' \leq \rho - 1$, repeat this process, setting $i' = i' + 1$. Note that when $i' = \rho$, $\mathcal{R}_{i'} = \emptyset$ and there are no cliques which we could add in this process. However, setting $\mathcal{K}_0 = \mathcal{K}_1 \cup \mathcal{K}_2 \cup_{i \in [\rho]} (\mathcal{K}_{i1} \cup \mathcal{K}_{i2})$ we have that $|V(\mathcal{K}_0)|$, n, and $|V(\mathcal{A}_i)| - \frac{|X_i|}{2}$ are divisible by r for each i, so we can deduce that the algorithm takes no cliques from $\mathcal{R}_\rho$ and terminates with $|\overline{Z}_i| = \frac{|X_i|}{2}$ for all $i \in [\rho]$.

Finally, by the key property of the absorbing structure (Remark 5.6.4), we have that for each $i \in [\rho]$, there is a partial K_r-factor $\mathcal{K}_{i3}$ in $G \cup G_2$ covering $V(\mathcal{A}_i) \setminus \overline{Z}_i$ and thus $\cup_{i \in [\rho]} \mathcal{K}_{i3} \cup \mathcal{K}_0$ is the desired K_r-factor in $G \cup G(n, p)$. $\square$

Chapter 6

Related works and future directions

In this chapter, we conclude the thesis by discussing connections with other lines of research and the questions that remain open in the field. In Section 6.1, we discuss other related spanning structures that have been studied in the settings explored in this thesis. We discuss *universality* in Section 6.1.1, *powers of Hamilton cycles* in Section 6.1.2 and *H-factors* for different graphs H, in Section 6.1.3. In Section 6.2, we then introduce some further models and related questions about clique factors. In Section 6.2.1, we look at analogous questions to those studied in this thesis, in the context of *hypergraphs*. In Section 6.2.2, we consider clique factors in the so-called *Ramsey–Turán* setting and finally in Section 6.2.3, we address some further notions of robustness for clique factors. Before all of this, we discuss what remains to be proven for clique factors in our three principal settings of interest.

For pseudorandom graphs, Theorem I (and Theorem I*) give us a complete understanding for triangle factors due to the dense triangle-free pseudorandom construction of Alon [7] discussed in Section 1.4. Several further constructions [40, 115] of pseudorandom triangle-free graphs have also been given which are (near-)optimal. For $r \geq 4$, Theorem I gives that any n-vertex (p, β)-bijumbled graph G with $\beta = o(p^{r-1}n)$ contains a K_r-factor. However, it is not known whether this condition is tight, even for the existence of a singular copy of K_r. This represents a key challenge in the understanding of pseudorandom graphs. Various authors [41, 68, 126, 166] have stipulated that n-vertex K_r-free (p, β)-bijumbled graphs exist with $\beta = \Theta(p^{r-1}n)$. Such graphs would witness the tightness of Theorem I, as well as Conjecture 6.1.1 discussed below, for all values of $r \geq 4$. Focusing on optimally pseudorandom graphs (that is, fixing $\beta = \Theta(\sqrt{pn})$ in (p, β)-bijumbled graphs), we expect to be able to find K_r-free optimally pseudorandom graphs with $p = \Omega(n^{-1/(2r-3)})$. The best known construction comes from a recent improvement of Bishnoi, Ihringer and Pepe [22] who give K_r-free optimally pseudorandom graphs of density $p = \Theta(n^{-1/(r-1)})$. Further interest in finding denser such graphs comes from a recent remarkable connection discovered by Mubayi and Verstraëte [142] that shows that if, as we expect, the K_r-free optimally pseudorandom graphs with density $p = \Omega(n^{-1/(2r-3)})$ do exist, then it is possible to improve the lower bound on the off-diagonal Ramsey numbers to match the upper bound and thus determine the asymptotics of this extremal function. In detail, they show that if these pseudorandom graphs exist, then the off-diagonal Ramsey number is $R(r,t) = t^{r-1+o(1)}$ as t tends to infinity. In fact, even a construction with $p = \omega(n^{-1/r})$ would improve on the current best known lower bound on off-diagonal Ramsey numbers due to Bohman and Keevash [24].

In the robust setting, Theorem II only dealt with triangle factors. We believe that our methods can also be used to tackle factors for larger cliques and give a robust Hajnal–Szemerédi theorem (Theorem 1.1.2). That is, there exists a $C > 0$ such that for any with $n \in r\mathbb{N}$, $p \geq Cp_r^*(n)$ and n-vertex G with $\delta(G) \geq \left(1 - \frac{1}{r}\right)n$, the graph G_p whp contains a K_r-factor. Here $p_r^*(n)$ denotes the threshold for K_r-factors as in Theorem 1.2.1 (also given below (6.0.1)). We do not believe that significant new ideas would be needed for this, but that it would be technically much more involved, in particular in the analysis of the extremal cases which was done for triangle factors in Section 4.7. Consequently, we concentrated only on triangle factors here.

In the randomly perturbed setting, we have almost completely resolved the K_r-factor problem. The only cases that Theorem III does not resolve is when $\tau = \frac{1}{r}, \frac{2}{r}, \ldots, \frac{r-2}{r}$. Note however for $2 \leq k \leq r-1$ and $\tau := 1 - \frac{k}{r}$, Theorem III gives that

$$n^{-2/k} \leq p(K_r, \tau) \leq n^{-2/(k+1)}.$$

In fact, one can slightly improve the lower bound, giving that

$$p(K_r, \tau) \geq p(K_k, 0) = p_k^*(n) = n^{-2/k}(\log n)^{2/(k^2-k)}. \tag{6.0.1}$$

Indeed, as in our lower bound construction (Section 5.1), take G to be complete graph on n vertices with a clique of size $\frac{kn}{r}$ removed. Letting I be the resulting independent set of vertices, if $p = o(p_k^*)$ then whp we have that the number of copies of K_{k+1} in $G(n,p)$ is less than, say, $(\log n)^3$ by Markov's inequality, whilst the number of vertices in I which do not lie in copies of K_k is at least $n^{1-o(1)}$, as can be seen by a second moment calculation (see e.g. [95, Theorem 3.22]). This precludes the existence of a K_r-factor in $G \cup G(n,p)$, as the average intersection of a clique in such a factor with the vertex set I would be k and we cannot find a family of disjoint cliques in $G(n,p)[I]$, whose average size is k, given the restrictions above. This leaves a gap between the upper and lower bounds and Han, Treglown and the author [87] suggested to study these 'boundary' cases. In recent work, Böttcher, Parczyk, Sgueglia and Skokan [29] addressed this question. For triangle factors, they completed the picture by showing that $p(K_3, 1/3) = \frac{\log n}{n}$, confirming that the lower bound given here is in fact the correct value for the perturbed threshold. For larger cliques they showed, perhaps surprisingly, that the behaviour is more complicated and the lower bound given here is not always the correct answer. For example, they gave a construction giving a polynomial improvement, showing that there is some $\varepsilon > 0$ such that $p(K_4, 1/4) \geq n^{-2/3+\varepsilon}$. The full picture of the perturbed thresholds for cliques at these boundary values remains a mystery and provides an interesting challenge.

6.1 Related spanning structures

The study of clique factors is part of a wider research aim of understanding spanning structures in graphs. This is a large field with many interesting results and open problems. Here, we discuss some key themes in the area in relation to the settings studied in this thesis.

6.1.1 Universality

As a broad general aim, we want to understand conditions for the existence of *all* subgraphs in graphs and not just singular examples such as clique factors. In order to achieve this, we can categorise graphs according to their maximal degree, leading to the following notion. For $2 \leq k \in \mathbb{N}$, we say an n-vertex graph G is k-*universal* if for any graph F on at most n vertices, with maximum degree at most k, G contains a copy of F. Understanding universality in graphs seems to be a considerable challenge and many beautiful conjectures remain open.

A moment's thought may suggest that a K_{k+1}-factor is the '*hardest*' maximum degree k graph to find in a graph G, as a clique is the densest graph with maximum degree k and a clique factor maximises the number of cliques. This intuition appears to hold true and has manifested in various settings. For example, we know from the theorem of Hajnal and Szemerédi (Theorem 1.1.2) that any n-vertex graph G with $\delta(G) \geq \left(\frac{k}{k+1}\right)n$ contains a K_{k+1}-factor and that this is tight. Bollobás and Eldridge [25], and independently Catlin [33], conjectured that the same minimum degree condition actually guarantees k-universality. This has been proven for $k = 2, 3$ [1, 9, 49] (and large n when $k = 3$) but remains open in general. In the case of random graphs, we know from the theorem of Johansson, Kahn and Vu (Theorem 1.2.1) that the threshold for the appearance of a K_{k+1}-factor is $p^*_{k+1}(n)$. The recent breakthrough result of Frankston, Kahn, Narayanan and Park [69] on thresholds implies that for any n-vertex graph F with maximum degree k, the threshold for the appearance of F in $G(n, p)$ is at most $p^*_{k+1}(n)$. Note that this is *not* implying that $G(n, p)$ is k-universal whp when $p = \omega(p^*_{k+1}(n))$ as we can only guarantee that some fixed F appears whp. However, the stronger version that $p^*_{k+1}(n)$ is the threshold for k-universality is believed to be true but only verified for $k = 2$ [65]. We remark that in general the 2-universality question is considerably more assailable than the general case due to the fact that every maximum degree 2 graph is a union of disjoint cycles and paths whilst for larger k, no such simple structural classification exists.

We believe that the same phenomenon occurs in pseudorandom graphs and conjecture the following, which implies that the condition in Theorem I that forces a K_{k+1}-factor in fact forces all maximum degree k subgraphs.

Conjecture 6.1.1. For any $2 \leq k \in \mathbb{N}$ and $c > 0$ there exists an $\varepsilon > 0$ such that any n-vertex (p, β)-bijumbled graph with $\delta(G) \geq cpn$ and $\beta \leq \varepsilon p^k n$ is *k-universal.*

Note that Corollary 1.4.2 settles Conjecture 6.1.1 for the case $k = 2$. For $k \geq 3$, the best known result comes from the sparse blow-up lemma of Allen, Böttcher, Hàn, Kohayakawa and Person [2] which gives a condition of $\beta = o(p^{(3k+1)/2}n)$ guaranteeing k-universality in a (p, β)-bijumbled graph. One thing that sets aside the pseudorandom setting in stark contrast to the extremal and probabilistic settings is that, as discussed above, it might be possible to replace a K_{k+1}-factor as the benchmark for the 'hardest' graph to find in the host graph, by a single copy of K_{k+1}.

We also believe that there is robustness for universality and conjecture the following.

Conjecture 6.1.2. For any $k \geq 2$, there exists a $C > 0$ such that for all $n \in \mathbb{N}$ and $p \geq Cp^*_{k+1}$, the following holds. If G is a graph with $\delta(G) \geq (\frac{k}{k+1})n$ then whp G_p is k-universal.

Conjecture 6.1.2 is a common strengthening of the conjecture of Bollobás–Eldridge–Catlin [25, 33] and the threshold for universality and so a full solution to this conjecture at this point would be remarkable. However, establishing the case $k = 2$ seems attainable and would be very interesting.

Finally, in the randomly perturbed setting, we propose the following strengthening of Theorem III.

Conjecture 6.1.3. Let $2 \leq k' \leq k + 1$ be integers. Then given any $1 - \frac{k'}{k+1} < \tau < 1 - \frac{k'-1}{k+1}$, there exists a $C > 0$ such that the following holds for all

$$p \geq Cp(K_{k+1}, \tau) = Cn^{-2/k'}.$$

For any n-vertex graph G with $\delta(G) \geq \tau n$, whp $G \cup G(n, p)$ is k-universal.

Conjecture 6.1.3 would be best possible due to Theorem III. The case $k = 2$ and $k' = 3$ was proven by Parczyk [148] who also conjectured the same as Conjecture 6.1.3 for the subcase $k' = k + 1$ and all $k \geq 2$. Moreover, a solution to the case $k = 2$ and $k' = 2$ was announced in [29]. For $k \geq 5$ and $k' = k + 1$, Böttcher, Montgomery, Parczyk and Person proved a weaker version of Conjecture 6.1.3, showing that for any *fixed* maximum degree k graph F, adding $G(n, p)$ with $p = \omega(n^{-2/(k+1)})$ to a graph G with positive density, guarantees the existence of F in $G(n, p)$. Proving this weakening of Conjecture 6.1.3 for all values of $2 \leq k' \leq k + 1$ would be a natural first step and would provide compelling evidence towards the conjecture. Finally we remark that thresholds for universality with respect to all (bounded-degree) trees have also been studied in the randomly perturbed setting [27, 98, 121].

6.1.2 Powers of Hamilton cycles

Aside from studying all subgraphs through universality, one can also look at other natural examples of spanning structures. For $1 \leq k \in \mathbb{N}$, we say an n-vertex graph G contains *the k^{th} power of a Hamilton cycle* if it contains a copy of the graph obtained by taking a cycle C_n of length n and adding an edge between any pair of vertices that have distance at most k in C_n. When $k = 1$, this just corresponds to G being Hamiltonian. For $k = 2$, we say G contains the *square of a Hamilton cycle*. Powers of Hamilton cycles are a natural generalisation of Hamilton cycles and are well-studied. Note that for $k \geq 2$, if G has $n \in (k+1)\mathbb{N}$ vertices then the existence of the k^{th} power of a Hamilton cycle in G implies the existence of a K_{k+1}-factor in G. Therefore any threshold for containing the k^{th} power of a Hamilton cycle must be at least as large as the threshold for a K_{k+1}-factor.

In the extremal setting, perhaps surprisingly, it turns out that the minimum degree thresholds coincide. Indeed, Komlós, Sárközy and Szemerédi [111, 112] proved a strengthening of Theorem 1.1.2 for large $n \in \mathbb{N}$ by showing that any n-vertex graph with $\delta(G) \geq \left(\frac{k}{k+1}\right)n$ contains the k^{th} power of a Hamilton cycle. This confirmed conjectures (for large n) of Pósa (see [58]) for squares of Hamilton cycles and Seymour [157] for higher powers. In the probabilistic setting, the situation is different and we see a separation between the thresholds for K_{k+1}-factors, $p^*_{k+1} = n^{-2/(k+1)}(\log n)^{2/(k^2+k)}$ (see Theorem 1.2.1), and the thresholds for k^{th} powers of Hamilton cycles, which has been shown to be $n^{-1/k}$. For $k \geq 3$, this threshold follows from a general result of Riordan [151] using an argument based on the second moment method. For squares of Hamilton cycles, the problem of establishing the threshold took much longer and was only recently proven by Kahn, Narayanan and Park [101] by sharpening their general method from [69] for giving bounds on thresholds.

In the pseudorandom setting, in contrast to the other settings studied in this thesis, Theorem I* is, in some sense, the first result of its kind, giving a tight condition on pseudorandomness to guarantee the existence of a spanning structure. Indeed, even the case of Hamilton cycles remains an intriguing open problem. Krivelevich and Sudakov [125] conjectured that a condition of $\lambda = o(d)$ is sufficient in (n, d, λ)-graphs and proved the currently best known bound of

$$\lambda = o\left(\frac{(\log\log n)^2 d}{\log n(\log\log\log n)}\right).$$

For conditions forcing larger powers of Hamilton cycles in (n, d, λ)-graphs, it is known [3] that a condition of $\lambda = o(d^{5/2}n^{-3/2})$ suffices for squares and $\lambda = o(d^{3k/2}n^{1-3k/2})$ for k^{th} powers with $k \geq 3$. It is unclear how close these conditions are to being optimal and the only obstructions known come from the pseudorandom K_{k+1}-free graphs already discussed at the beginning of this chapter. In particular, we currently have no idea whether the optimal pseudorandom conditions for k^{th} powers of Hamilton cycles and K_{k+1}-factors coincide, mirroring the extremal case, or whether they are separated, as may be suggested by probabilistic intuition.

In the robust setting, a corollary of the sparse blow-up lemma of Allen, Böttcher, Hàn, Kohayakawa and Person [2] gives that for all $\varepsilon > 0$ and n-vertex graphs G with $\delta(G) \geq \left(\frac{k}{k+1} + \varepsilon\right)n$, if $p = \omega\left(\frac{\log n}{n}\right)^{1/2k}$, then G_p whp contains the k^{th} power of a Hamilton cycle. For squares of Hamilton cycles, this was improved by Fischer [67] giving a condition of $p \geq n^{-1/2+\varepsilon}$. It is believable that for all $k \geq 2$, an analogue of Theorem II holds in this setting and that the conclusions of the above results remain true without the ε in the minimum degree condition and with probability values all the way down to the threshold in random graphs, $n^{-1/k}$.

In randomly perturbed graphs, there has been a lot of interest in powers of Hamilton cycles, although the full picture remains elusive. Indeed, current research has focused on either the range of small positive densities [20, 28] or densities near the extremal threshold [21, 56] where several tight results [15, 147] have been established for certain powers and values of minimum degree. In particular, a result of Nenadov and Trujić [147] implies, among other things, the case when $k = 2$ and r is even in Theorem III and that adding $\omega(n)$ random edges to a graph with minimum degree τn for $\tau > \frac{1}{2}$, guarantees the square of a Hamilton cycle. At the other extreme, concentrating on arbitrary n-vertex dense graphs G with $\delta(G) \geq \tau n$ for some $\tau > 0$, Böttcher, Montgomery, Parczyk and Person [28] showed that there exists an $\eta = \eta(\tau, k)$ such that $G \cup G(n, p)$ whp contains the k^{th} power of a Hamilton cycle whenever $p \geq n^{1/k-\eta}$. This deviates from the observed behaviour of the randomly perturbed model with respect to other spanning structures, for example Hamilton cycles [23] and general H-factors [19], where only a logarithmic factor in probability is saved when adding an arbitrarily dense graph, compared to the purely random threshold. In the spirit of Theorem III, it would be very interesting to establish optimal perturbed threshold probabilities for the square of the Hamilton cycle, in the full range of minimum degrees $\delta(G) = \tau n$ for some $0 < \tau < \frac{2}{3}$, to bridge the gap between the extremal and probabilistic settings. Note that lower bounds follow from Theorem III for triangle factors and the result of Nenadov and Trujić [147] mentioned above gives a matching upper bound when $\tau > \frac{1}{2}$, showing that the perturbed threshold for squares of Hamilton cycles and triangle factors match in this range of minimum degree. The range $0 < \tau \leq \frac{1}{2}$ remains wide open and it would be very interesting to determine at what point the thresholds for triangle factors and the squares of Hamilton cycles separate. Moreover, for small $\tau > 0$ we can see that the picture is quite different to the 'jumping' phenomenon we saw for clique factors. Indeed, the best known upper bound is $p^{1/2-\eta}$ as mentioned above and for $0 < \tau < \frac{1}{8}$, Bennett, Dudek and Frieze [21] improved on the lower bound of $n^{-2/3}$ coming from Theorem III, showing that $p = \omega(n^{-1/(2-4\tau)})$ is necessary. As τ approaches 0, this gets close to the threshold for the square of the Hamilton cycle in the random graph. Therefore, it can not be the case that we have some window $0 < \tau < \tau_0$, for which the perturbed threshold is fixed. It would therefore be very interesting to determine the exact dependence of the perturbed threshold on τ here.

6.1.3 H-factors

Here we briefly discuss general H-*factors*, where one interesting in covering the vertex set of a graph G with $n \in r\mathbb{N}$ vertices with vertex-disjoint copies of some r-vertex graph H. In this thesis, we studied the case where $H = K_r$ and an eventual aim is to understand all H-factors. Towards this, a common tactic is to study other example cases, for instance the case where H is a cycle.

The extremal and probabilistic settings for H-factors are now both reasonably well understood. Indeed, concluding a large body of work (see [129]), Kühn and Osthus [130] characterised, up to an additive constant, the minimum degree which ensures that a graph G contains an H-factor for an arbitrary graph H. For H-factors in random graphs (and indeed random hypergraphs), the threshold depends on the so-called 1-*density* of H defined as

$$d^*(H) := \max\left\{\frac{e_F}{v_F - 1} : F \subseteq H,\ v_F \geq 2\right\}. \tag{6.1.1}$$

At the same time as solving the problem for clique factors (Theorem 1.2.1), Johansson, Kahn and Vu [96] made huge progress on this problem for general H. They conjectured that the threshold for the appearance of an H-factor in a binomial random (k-)graph is

$$\ell(n; H)n^{-1/d^*(H)}, \tag{6.1.2}$$

where $\ell(n; H)$ is an explicit polylogarithmic factor which depends on the structure of H; see [96] for details. Furthermore, they proved that the conjecture is true when we replace the $\ell(n; F)$ term by some function which is $n^{o(1)}$ and they determined the exact threshold for all so-called *strictly balanced* (k-)graphs H, in which case one has $\ell(n; H) = (\log n)^{1/e_H}$. The conjecture has now also been proven for the so-called *non-vertex-balanced* graphs H, by Gerke and McDowell [73]. In this case, one has that $\ell(n; H) = 1$.

For pseudorandom graphs, to our knowledge, very little is known for H-factors beyond the results of this thesis although some initial upper bounds on thresholds follow from the sparse blow-up lemma [2] (see the discussion on universality in pseudorandom graphs in Section 6.1.1). A bold conjecture would be that the situation observed for K_3 (Theorem I*) holds in general and the same optimal pseudorandom condition that forces a singular copy of H in fact forces an H-factor. It would be very interesting to explore whether this phenomenon occurs for general odd cycles. Indeed, the case $H = C_{2k+1}$ is essentially the only case where we understand the optimal pseudorandom condition [11, 126] forcing a singular copy of H and so this provides a benchmark for studying C_{2k+1}-factors.

In the robust setting, the situation is similar and it is only triangle factors (Theorem II) for which we have a full understanding. Again, the sparse blow-up lemma [2] gives a general result, showing that if for an n-vertex graph G we have $\delta(G) \geq \left(\frac{\chi(H)-1}{\chi(H)} + \varepsilon\right) n$ for some $\varepsilon > 0$ and $p = \omega\left(\frac{\log n}{n}\right)^{1/\Delta(H)}$, then G_p whp has an H-factor. Whilst the minimum degree condition here is asymptotically optimal for some H, the probability is much larger than the threshold for H-factors for all H on at least 3 vertices. In the same vein as Theorem II it would be interesting to explore whether for all H-factors one can get a strong robustness theorem which captures both the extremal and random thresholds.

Balogh, Treglown and Wagner [19] gave a general theorem for H-factors in randomly perturbed graphs showing that by starting with an arbitrarily dense graph G, the amount of random edges needed from $G(n,p)$ decreases by the log-factor $\ell(n;H)$ in (6.1.2) and adding $G(n,p)$ with $p = \omega(n^{-1/d^*(H)})$ suffices to find an H-factor. Note that, as mentioned above, for certain graphs H, we have $\ell(n;H) = 1$ and so we see that the randomly perturbed threshold actually matches the purely random threshold. That is, adding a graph of small linear minimum degree to $G(n,p)$ does not help substantially to create an H-factor for these cases. The result of Balogh, Treglown and Wagner [19] is tight for minimum degrees of the form τn with $0 < \tau < v_H$ and establishes $p(H,\tau)$ in this range (recall Definition 1.6.2 for the definition of a perturbed threshold for factors). The problem is still wide open for larger values of τ. The methods used here to prove Theorem III are likely to be useful for the general problem, although we suspect how $p(H,\tau)$ 'jumps' as τ increases will depend heavily on the structure of H. Thus we believe it would be a significant challenge to prove such a general result. Very recently, Böttcher, Parczyk, Sgueglia and Skokan [30] established the full picture for cycle factors, that is, when H is a cycle.

6.2 Related topics

In this section, we conclude by discussing related topics to those studied in this thesis. This includes some further results of the author whose proofs share some aspects with the proofs given in this thesis.

6.2.1 Hypergraphs

The study of clique factors transfers naturally to the setting of k-uniform hypergraphs (k-graphs for short) when $k \geq 3$. We denote by $\mathcal{K}_r^{(k)}$ the k-uniform clique on r vertices and $\mathcal{H}^{(k)}(n,p)$ denotes the binomial random hypergraph with each edge of $\mathcal{K}_n^{(k)}$ sampled independently with probability $p = p(n)$. There has been significant interest in factor problems in hypergraphs and establishing tight results has been a challenge, even in the simplest case of perfect matchings. Indeed, in the extremal setting, the problems seem to be far more difficult than the corresponding questions in graphs and the picture is complicated by different notions of minimum degree. We refer to the excellent surveys of Kühn and Osthus [129] and Zhao [174] for a detailed account of what is known. For random hypergraphs, establishing the threshold for perfect matchings was known as Shamir's problem (see [59]) and, as with the clique factor thresholds in graphs, this was a famous open problem which attracted a lot of attention. The methods of Johansson, Kahn and Vu [96] for graph factors (see Theorem 1.2.1 and Section 6.1.3), also transferred to the hypergraph setting and they were able to resolve the thresholds for many $\mathscr{F}$-factors, including perfect matchings and indeed all $\mathscr{K}_r^{(k)}$-factors (see (6.1.2) above).

In sparse hypergraphs, it is well-known that simple pseudorandom conditions requiring that the hypergraph is roughly regular, guarantee the existence of an almost perfect matching (see Section 2.7). However, in order to guarantee perfect matchings (and indeed other spanning structures), stronger notions of pseudorandomness are required. Here we focus on a natural generalisation of the notion of bijumbledness in Definition 1.4.1, defining a k-graph $\mathscr{H} = (V, E)$ to be (p, β)*-jumbled* if for all (not necessarily disjoint) $A_1, \ldots, A_k \subseteq V$ we have

$$e(A_1, \ldots, A_k) = p \prod_{i \in [k]} |A_i| \pm \beta \prod_{i \in [k]} |A_i|^{1/2}. \tag{6.2.1}$$

where $e(A_1, \ldots, A_k)$ denotes the number of tuples $(a_1, \ldots, a_k)$, $a_i \in A_i$, which form an edge in $\mathscr{H}$. In jumbled hypergraphs the picture becomes considerably more complex than in graphs. Indeed it turns out that the only subgraphs that one can guarantee by imposing conditions on jumbledness are *linear* subgraphs, those in which pairs of hyperedges intersect in at most one vertex. Building on previous work [42, 109, 132, 133] mainly concerned with dense hypergraphs (the so-called quasirandom regime), Hiệp Hàn, Jie Han and the author [79, 80] recently gave the best-known conditions on pseudorandomness that guarantee different linear subgraphs of hypergraphs. These include all fixed sized linear subgraphs as well as $\mathscr{F}$-factors for linear $\mathscr{F}$ and loose Hamilton cycles. In particular, we could show that n-vertex (p, β)-jumbled k-graphs with a mild minimum degree condition contain a perfect matching when $\beta = o(p^{k/2+1} n^{k/2})$. For $k = 3$, this gives $\beta = o(p^{5/2} n^{3/2})$ which improved on the previous best-known bound of Lenz and Mubayi [132] who worked in a more restrictive setting of hypergraph eigenvalues and whose condition corresponds to $\beta = o(p^{16} n^{3/2})$ in jumbled hypergraphs. Our proofs in [79, 80] rely on the template absorption method (see Section 2.8) and for $\mathscr{F}$-factors we build absorbing structures in a similar fashion to our proof of Theorem III (see Section 5.6). The tightness of our results in pseudorandom hypergraphs is unclear as no good constructions are known for $\mathscr{F}$-free pseudorandom hypergraphs.

In randomly perturbed hypergraphs, Yulin Chang, Jie Han, Yoshiharu Kohayakawa, Guilherme Mota and the author [34] determined, up to a multiplicative constant, the optimal number of random edges that need to be added to a k-graph $\mathscr{H}$ with minimum vertex degree $\Omega(n^{k-1})$ to ensure an $\mathscr{F}$-factor with high probability, for any $\mathscr{F}$ that belongs to a certain class of k-graphs, which includes all k-partite k-graphs. As with the results of Balogh, Treglown and Wagner, we observed that in the setting of hypergraphs, we lose the logarithmic factor $\ell(n;\mathscr{F})$ in the random threshold (6.1.2), when starting with a dense hypergraph. In particular, for perfect matchings, this settled a problem of Krivelevich, Kwan and Sudakov [120] who looked at perfect matchings in randomly perturbed dense k-graphs satisfying a more restrictive co-degree condition that every $(k-1)$-set of vertices lies in $\Omega(n)$ edges. Our proofs in [34] share many features with the proof of Theorem III given here. Establishing a corresponding perturbed result for general hypergraph clique factors poses an interesting open problem and we expect the techniques developed here in the proof of Theorem III to be useful for this.

6.2.2 The Ramsey–Turán perspective

In this section we discuss a recent trend of studying clique factors in the *Ramsey–Turán* setting. As demonstrated in Section 1.4, the extremal constructions that force the minimum degree thresholds for clique factors (Theorem 1.1.2), are atypical. We saw (Theorem I) that by additionally imposing pseudorandom conditions on the host graph, one can capture much sparser graphs that contain clique factors. However, one may argue that the pseudorandom condition given by the notion of bijumbledness (Definition 1.4.1) is quite strong and weaker pseudorandom notions may suffice to preclude the atypical behaviour of the extremal constructions. Perhaps the weakest pseudorandom condition one can impose on the host graph is to simply block the existence of large independent sets. This was proposed by Balogh, Molla and Sharifzadeh [18] who showed that for any $\tau > \frac{1}{2}$, an n-vertex graph G with $n \in 3\mathbb{N}$, $\delta(G) \geq \tau n$ and $\alpha(G) = o(n)$, contains a triangle factor. One can also consider stronger independence conditions which force the existence of cliques in linearly sized sets. Here, we let $\alpha_k(G)$ denote the *k-independence number* of a graph G, that is, the largest size of a K_k-free set of vertices in G. Nenadov and Pehova [146] proposed the study of what minimum degree and k-independence conditions force the existence of K_r-factors and provided some initial results.

These questions were inspired by the analogous question for Turán problems where one is interested in the density needed to force the existence of a fixed sized subgraph H. Again, the extremal examples are far from typical and contain large independent sets. Imposing an upper bound on independence numbers then leads to improvements on the density needed. This field, known as Ramsey–Turán theory, was initiated by Erdős and Sós [62] and led to a wealth of results, see e.g. [160] for an overview. In more detail, for a fixed graph H, define[1] the function $\mathbf{RT}_k(H)$ to be the maximum $\pi_0 \geq 0$ such that for any $\pi < \pi_0$ and $\alpha > 0$ there exist H-free graphs G with n vertices, $e(G) \geq \pi n^2$ and $\alpha_k(G) \leq \alpha n$ for all sufficiently large n. Much research [16, 26, 61, 167] has focused on establishing the value of $\mathbf{RT}_k(H)$ for various $k \in \mathbb{N}$ and graphs H, with a particular emphasis on establishing whether $\mathbf{RT}_k(H)$ is non-zero.

[1] This function is usually denoted by $\Theta_r(H)$ and defined (equivalently) in terms of a limit of Ramsey–Turán numbers, see e.g. [16].

In the setting of clique factors, it turns out that the sublinear independence conditions are not enough to force factors in sparser graphs with density approaching 0. Indeed, already for any $\tau < \frac{1}{2}$ and n sufficiently large, one can take G to be the union of two disjoint cliques whose sizes add to n. Such a G has the property that $\alpha_k(G) < 2k$ for all fixed $k \in \mathbb{N}$ whilst choosing the sizes of the cliques in G appropriately, we can guarantee that $\delta(G) \geq \tau n$ and that any K_r-matching will leave uncovered vertices. Therefore, in order to study sparser graphs we need to relax our expectations, moving away from studying K_r-factors and instead focusing on almost factors which are as large as possible, in that they cover all but a constant number of vertices.

Definition 6.2.1. Given $\tau \in (0, 1]$, $r \in \mathbb{N}$ and an n-vertex graph G, we say a K_r-matching in G is *τ-quasiperfect* if it covers all but $\ell(r-1)$ vertices of G where $\ell := \lfloor \frac{1}{\tau} \rfloor$. When τ is clear from context, we will simply call the matching quasiperfect.

Notice that when $\tau > \frac{1}{2}$, a K_r-matching is τ-quasiperfect if it leaves at most $r-1$ vertices uncovered. Therefore, with an additional condition that $n \in r\mathbb{N}$, a quasiperfect K_r-matching is in fact a K_r-factor. For (almost) all values of τ the definition of quasiperfect matchings captures the largest possible size of a matching we can hope for when looking at graphs G with $\delta(G) \geq \tau n$ and some bound on independence numbers. Indeed, generalising the construction above, if $\frac{\delta(G)}{n} = \tau$ and $\frac{1}{\tau} \notin \mathbb{N}$, we can take G to be $\ell = \lfloor \frac{1}{\tau} \rfloor$ disjoint cliques of equal size $\frac{n}{\ell}$ and choose n such that $\frac{n}{\ell}$ is equal to $(r-1) \mod r$. Then there are no copies of K_r using more than one of the large cliques in G and in each clique any K_r-matching must leave $r-1$ vertices uncovered due to divisibility constraints.

It turns out that the divisibility constraints given by the construction outlined above are the 'worst-case scenario' when we impose an appropriate independence condition. Indeed, in recent work, Jie Han, Guanghui Wang, Donglei Yang and the author [88] showed that with certain minimum degree independence conditions, we can always guarantee a quasiperfect matching. The first result of this kind was predicted by Alon and proven by Balogh, McDowell, Molla and Mycroft [17] who showed that for every $\tau > \frac{1}{3}$ there exists an $\alpha > 0$ such that every graph G on n vertices with $\delta(G) \geq \tau n$ and $\alpha(G) \leq \alpha n$ contains a K_3-matching covering all but 4 vertices. In order to systematically study optimal conditions guaranteeing the existence of quasiperfect K_r-matchings and K_r-factors in the Ramsey–Turán setting, we [88] introduced the following extremal function. For $r \in \mathbb{N}$ and an integer $1 \leq k \leq r$, define the *Ramsey–Turán factor threshold*, denoted by $\mathbf{RTF}_k(K_r)$, as the largest $\tau_0 \geq 0$ such that for all $0 < \tau < \tau_0$ and $\alpha > 0$, there exists n-vertex graphs G with $\delta(G) \geq \tau n$ and $\alpha_k(G) \leq \alpha n$ such that G does not contain an η-quasiperfect K_r-matching, for all sufficiently large n.

Before discussing the known results for Ramsey–Turán factor thresholds, we remark on a connection with the randomly perturbed setting. Indeed, given that the conditions $\alpha_k(G) = o(n)$ are typical in sparse graphs of a certain density, results in the Ramsey–Turán model have implications for the randomly perturbed model. These corollaries are often best possible (apart from the fact they leave a constant number of vertices uncovered) as one needs the random graph to provide small independence numbers in order to give the existence of factors in the perturbed model. Moreover, as noticed by Nenadov and Pehova [146], one in fact obtains something stronger. Indeed, in the perturbed model, one fixes an arbitrary graph G (which satisfies a dense minimum degree condition) and asks for p such that $G \cup G(n,p)$ contains a given factor whp. From Ramsey–Turán results we can conclude that whp, $G(n,p)$ has the property that no matter how an adversary places a graph G (satisfying a minimum degree condition), the resulting graph will have a given factor (or quasiperfect matching).

Moreover, constructions for the randomly perturbed setting provide lower bounds on Ramsey–Turán factor thresholds. In particular, our constructions given here for Theorem III when $\delta(G) \geq \tau n$ for some $1 - \frac{k+1}{r} < \tau < 1 - \frac{k}{r}$ (see Section 5.1) can be shown to have sublinear k-independence numbers and have the property that any K_r-matching leaves a linear number of vertices uncovered (see [88] for more details). This implies that $\mathbf{RTF}_k(K_r) \geq 1 - \frac{k}{r}$ for all $1 \leq k \leq r$. In [88] we conjecture that this is in fact the true value and the picture follows that of Theorem III.

Conjecture 6.2.2 (Han–Morris–Wang–Yang[88])**.** For $1 \leq k \leq r \in \mathbb{N}$ with $r \geq 3$,

$$\mathbf{RTF}_k(K_r) = 1 - \frac{k}{r}.$$

Several cases of Conjecture 6.2.2 have been established. Indeed, the case $k = 1$ is essentially the Hajnal–Szemerédi Theorem (Theorem 1.1.2). The case $k = 2$ (and $r \geq 4$) was shown by Knierim and Su [108] and, due to the connection above, this reproved Theorem III for these values of k, r. The case $k = 2$ and $r = 3$ follows from the aforementioned result of Balogh, McDowell, Molla and Mycroft [17] and a result of Nenadov and Pehova [146]. In [88], Han, Wang, Yang and the author established Conjecture 6.2.2 for $k = r$ and $k = r - 1$, thus completing the picture for $r = 3, 4$. Our proofs use similar techniques to our proof of Theorem III in Chapter 5 but we also needed to adopt the *lattice-based absorption* method developed in [81, 82, 103], in order to get the optimal number of vertices uncovered by our K_r-matchings. For larger values of r, the conjecture is open for intermediate values for k (namely, $3 \leq k \leq r - 2$). It would be very interesting to close this gap.

6.2.3 Further notions of robustness

We close this thesis by discussing some further notions of robustness for clique factors. Many of these have been studied in the setting of Hamiltonicity and we refer to the excellent survey of Sudakov [165] for a detailed account of this. In the clique factor setting, it seems very little is known. For this discussion, we say an n-vertex graph G is *r-full* for $r \geq 3$ if $n \in r\mathbb{N}$ and $\delta(G) \geq \left(1 - \frac{1}{r}\right)n$. We will also refer to n-vertex graphs G with $\delta(G) \geq \frac{n}{2}$ as *Dirac graphs*. Thus for $r \geq 3$ the Hajnal–Szemerédi theorem (Theorem 1.1.2) gives that every r-full graph contains a K_r-factor and Dirac's theorem (Theorem 1.3.1) gives that every Dirac graph is Hamiltonian. As we discuss at the beginning of this chapter, and prove for $r = 3$ in Theorem II, we expect that random sparsifications of r-full graphs contain K_r-factors whp, with probabilities right down to the threshold for K_r-factors. The only other result towards robustness for clique factors, that we are aware of, is a result of Coulson, Keevash, Perarnau and Yepremyan [46] that establishes the existence of rainbow K_r-factors in suitably bounded colourings of r-full graphs[2]. It is interesting to consider other notions of robustness for K_r-factors in r-full graphs.

Firstly, as mentioned for triangle factors in Section 1.5, it would be interesting to establish how many K_r-factors are necessarily contained in an r-full graph. In particular, it would be interesting to establish that there is some constant $c = c(r)$ such that in any n-vertex r-full graph the number of distinct K_r-factors is at least $(cn)^{n(1-1/r)}$. This would be tight up to the value of c and is established for triangle factors in Corollary 1.5.2 with an extra log factor. One can also consider edge-disjoint K_r-factors. By considering a random partition of edges, Theorem II implies that any n-vertex 3-full graph contains a family of at least $\Omega(n^{2/3}(\log n)^{-1/3})$ edge-disjoint triangle factors. In terms of upper bounds, by considering triangles at a fixed vertex v with $\deg(v) = \frac{2n}{3}$, it is clear that one cannot hope for more than $\frac{n}{3}$ edge-disjoint triangle factors. In fact one can do slightly better than this by considering a construction similar to that of Nash-Williams [143] for the number of edge-disjoint Hamilton cycles in Dirac graphs. Indeed, let $n \in 3\mathbb{N}$ and $m := \frac{n}{3}$. Consider the n-vertex complete tripartite graph on vertex parts $X \cup Y \cup Z$ such that $|X| = m + 2$ and $|Y| = |Z| = m - 1$. Let G be the graph obtained from this tripartite graph by adding the edges of some cycle C of length $m + 2$ on the vertices of X. It is easy to check that G is 3-full. Moreover, any triangle factor in G must contain at least 2 edges of C. Hence G contains at most $\lfloor \frac{m+2}{2} \rfloor = \lfloor \frac{n}{6} \rfloor + 1$ edge-disjoint triangle factors. This leaves a big gap and it would be very interesting to bring these bounds closer together.

Finally we mention the study of *Maker-Breaker games*. We only give a brief account here and refer the reader to [91] for a comprehensive treatment of the area. Given an r-full graph G, the *Maker-Breaker K_r-factor game* on G is played by two players, *Maker* and *Breaker*. The players take it in turns to claim edges with the aim of Maker being to create a K_r-factor and the aim of Breaker being to stop Maker. By convention, Breaker has the first move. Moreover, in order to make life harder for Maker, we can give Breaker extra moves. We say the game is *b-biased* if for every one move of Maker, Breaker gets b moves. The question is then to determine for what b Maker is able to win the game. When $G = K_n$, this was recently studied and Liebenau and Nenadov [134], building on results in [4], proved that for $r \geq 4$, if $b = o\left(n^{2/(r+2)}\right)$ then Maker has a winning strategy for the K_r-factor game on $G = K_n$ whilst if $b = \omega\left(n^{2/(r+2)}\right)$ then Breaker

[2] In fact they need the slightly stronger minimum degree condition of $\delta(G) \geq \left(1 - \frac{1}{r} + \varepsilon\right)n$ for some $\varepsilon > 0$.

has a winning strategy. For $r = 3$, Allen, Böttcher, Kohayakawa, Naves and Person [4] showed that $b = o\left(\left(\frac{n}{\log n}\right)^{1/2}\right)$ leads to Maker's win whilst $b \geq 4n^{1/2}$ is Breaker's win, leaving a window for the so-called *bias threshold* that remains open. The proof of Liebenau and Nenadov [134] for Maker's strategy works by showing that Maker can claim a pseudorandom graph that has certain properties and showing that such a pseudorandom graph is guaranteed to contain a K_r-factor. In this way the ideas given here in Chapter 3 are pertinent and indeed Liebenau and Nenadov use the *cascading absorption* idea used here (see Section 4.1) and in [145] to study K_r-factors in pseudorandom graphs. The obvious open problem here is to close the gap for the triangle factors and it is believed [4] that one can remove the log factor in the condition for Maker's win. Beyond this it would be interesting to explore K_r-factor games in r-full graphs G other than K_n. Indeed, for the Hamilton cycle Maker-Breaker game defined analogously, Krivelevich, Lee and Sudakov [122] proved that Maker can win in any Dirac graph if the bias is $b = o\left(\frac{\log n}{n}\right)$. This is tight as $\frac{\log n}{n}$ is the threshold bias for the Maker-Breaker game for Hamilton cycles in the complete graph, and provides another notion of robustness in Dirac graphs.

Glossary

Here we provide a glossary of notation used, for the convenience of the reader. The definitions here are provided to serve as a quick reminder for the reader and are *not* supposed to be formal definitions. The precise definitions can be found in the Notation section or within the text (in which case we direct the reader to the relevant section).

$\mathbb{Z}, \mathbb{N}, \mathbb{R}$	the integers, the non-negative integers, the real numbers
$n!_t$	the number of ways to choose t distinct numbers from $[n]$
$x = y \pm z$	implies that $x \leq y + z$ and $x \geq y - z$
whp	with high probability i.e. with probability tending to 1 as n tends to ∞
A^k	the set of ordered k-tuples of elements in a set A
$\binom{A}{k}$	the set of all k-element subsets of a set A
$X \setminus Y$	the set $X \setminus (X \cap Y)$ when X and Y are sets
v_G	the number of vertices of a graph G
e_G	the number of edges of a graph G
$\Delta(G)$	the maximum degree of a graph G
$\delta(G)$	the minimum degree of a graph G
$\alpha(G)$	the independence number of a graph G
log	the natural logarithm function
$N_G(v)$	the set of neighbours of a vertex v in a graph G
$N_G(S;U)$	the set $(\cap_{v \in S} N_G(v)) \cap U$ for vertex sets $S, U \subseteq V(G)$
$\deg_G(S;U)$	the size of the neighbourhood $\lvert N_G(S;U)\rvert$ for vertex sets $S, U \subseteq V(G)$
$\deg_{E'}(v)$	the number of edges in a subset $E' \subseteq E(G)$ containing a vertex v
$\delta(E')$	the minimum value of $\deg_{E'}(v)$ over all vertices that lie in some edge in E'
$E'[A]$	the set of edges in $E' \subset E(G)$ that lie in $A \subset V(G)$
$K_r(G)$	the set of copies of K_r in a graph G
$K_r(G,u)$	the set of copies of K_r that contain a vertex u in a graph G
$\mathrm{Tr}_v(G)$	the set of edges that form triangles with a vertex v in a graph G
V^i	a vertex part of a tripartite graph with vertex partition $V^1 \cup V^2 \cup V^3$
$\mathcal{V}$	the set of ordered tuples of vertices of a tripartite graph on $V^1 \cup V^2 \cup V^3$
$\ell(\underline{u})$	the length of a tuple $\underline{u} \in \mathcal{V}$

$W_{\hat{u}_1,\dots,\hat{u}_\ell}$	the set of vertices $W \setminus \{u_1, \dots, u_\ell\}$
$W_{\underline{\hat{u}}}$	the set of vertices $W \setminus \{u : u \in \underline{u}\}$
D_t	the graph consisting of t labelled vertex-disjoint triangles
$\Psi^t(G)$	the set of embeddings of D_t in a tripartite graph G (that respect labellings)
$\Psi^t_{\underline{\hat{u}}}(G)$	the set of embeddings of D_t that avoid vertices in $\underline{u} \in \mathcal{V}$
$\Psi^t_v(G)$	the set of embeddings ψ of D_t into G such that $\psi((1,1)) = v$
$G[U]$	the graph induced by G on a vertex subset U
$G[U_1, \dots, U_k]$	the k-partite graph induced by G between vertex subsets $U_1, \dots, U_k$
$G_{\underline{\hat{u}}}$	the graph induced by G on $V_{\underline{\hat{u}}} = V(G)_{\underline{\hat{u}}}$
$F \setminus W$	the graph induced by F on $V(F) \setminus W$
$F \setminus w$	the graph induced by F on $V(F) \setminus \{w\}$
$\deg^{\mathscr{H}}(v)$	the number of edges containing a vertex v in a hypergraph $\mathscr{H}$
$\operatorname{codeg}^{\mathscr{H}}(u,v)$	the number of edges containing two vertices u and v in a hypergraph $\mathscr{H}$
$\mathscr{H}_J$	the subhypergraph of $\mathscr{H}$ defined by all the edges of $\mathscr{H}$ that contain an edge of J, where J is a 2-uniform graph with $V(J) = V(\mathscr{H})$
$G \cup \tilde{G}$	the graph with vertex set $V(G) = V(\tilde{G})$ and edge set $E(G) \cup E(\tilde{G})$
$G \setminus \tilde{G}$	the graph with vertex set $V(G) = V(\tilde{G})$ and edge set $E(G) \setminus E(\tilde{G})$
$\mathscr{H} \setminus \mathscr{H}'$	the hypergraph with vertex set $V(\mathscr{H}) = V(\mathscr{H}')$ and edge set $E(\mathscr{H}) \setminus E(\mathscr{H}')$
$K^r_{m_1,m_2,\dots,m_r}$	the complete r-partite graph with parts of sizes $m_1, m_2, \dots, m_r$
$J_{m_1,m_2,\dots,m_r}$	the blowup of a graph J with the i^{th} vertex of J blown up to a set of size m_i
$G(n,p)$	the binomial random graph (see Section 1.2)
$p^*_r(n)$	the threshold for containing K_r-factors (see Theorem 1.2.1)
$e(A,B)$	the number of edges between vertex sets A and B (see Section 1.4)
G_p	the random sparsification of a graph G with probability p (see Section 1.5)
$p(H,\alpha)$	the perturbed threshold for an H-factor (see Definition 1.6.2)
Δ_X	a parameter for a random variable X in applications of Janson's inequality (see Lemma 2.1.2)
$\bar{\Delta}$	a parameter for a random variable X in applications of Janson's inequality for subgraphs (see Lemma 2.1.3)
$d_G(X,Y)$	the density of G between vertex sets X and Y (see Section 2.2)
$(\Omega, \mathbb{P})$	a finite probability space (see Section 2.3)
$p(x)$	probability that a random variable X equals x (see Section 2.3)
$h(X)$	entropy of a random variable X (see Section 2.3)
$\operatorname{rg}(X)$	range of a random variable X (see Section 2.3)
$h(X_1, \dots, X_n)$	entropy of the random vector $(X_1, \dots, X_n)$ (see Section 2.3)
$h(X\|E)$	conditional entropy of X given an event E (see Section 2.3)
$p(x\|E)$	probability that X equals x given an event E (see Section 2.3)
$h(X\|Y)$	conditional entropy of X given Y (see Section 2.3)

$\chi(H)$	chromatic number of a graph H (see Section 2.5)
$\chi_{cr}(H)$	critical chromatic number of a graph H (see Section 2.5)
$\sigma(H)$	the smallest size of colour class over all proper colourings of H with $\chi(H)$ colours (see Section 2.5)
$\lvert f\rvert$	the value of a fractional matching f in a hypergraph $\mathscr{H}$ (see Section 2.6)
$\vartheta^*(\mathscr{H})$	the fractional matching number of $\mathscr{H}$ (see Section 2.6)
$\vartheta(\mathscr{H})$	the matching number of $\mathscr{H}$ (see Section 2.6)
$\lvert g\rvert$	the value of a fractional cover g in $\mathscr{H}$ (see Section 2.6)
$\tau^*(\mathscr{H})$	the fractional cover number of $\mathscr{H}$ (see Section 2.6)
$\mathcal{T}$	a template graph (see Definition 2.8.1)
$\mathcal{B}, \mathcal{C}, \mathcal{D}$	a diamond tree (see Definition 3.1.1)
$\Sigma, \Pi, \Upsilon, \Xi$	K_{r-1}-matchings (see e.g. Definition 3.1.1)
$\boldsymbol{\mathcal{O}}, \boldsymbol{\mathcal{P}}, \boldsymbol{\mathcal{R}}, \boldsymbol{\mathcal{Q}}, \boldsymbol{\mathcal{J}}$	an orchard (see Definition 3.1.3)
$\mathscr{H}(\boldsymbol{\mathcal{O}})$	the K_r-hypergraph defined by an orchard $\boldsymbol{\mathcal{O}}$ (see Definition 3.1.5)
$\mathbf{\Lambda}, \mathbf{\Gamma}, \mathbf{\Phi}, \mathbf{\Theta}$	a system of pairwise disjoint vertex sets (see Definition 3.5.1)
$\mathscr{H}(\mathbf{\Lambda}; r)$	the K_r-hypergraph generated by $\mathbf{\Lambda}$ (see Definition 3.5.1)
$<_L$	the lexicographic order on pairs $(i, j) \in [t] \times [\ell]$ (see Section 3.6.4)
$I_{<ij}$	the indices coming before (i, j) according to $<_L$ (see Section 3.6.4)
$\mathbb{A}$	a K_r-absorbing structure (see Definition 3.7.1)
$\mathbb{J}, \mathbb{K}, \mathbb{L}$	index sets appearing in the proof of Claim 3.7.6
Γ	a tripartite graph with n vertices in each part (see e.g. Section 4.1)
$I^\ell(\psi)$	the vertices in V^ℓ that are isolated in an embedding $\psi \in \Psi^t$ (see Section 4.4.2)
ψ_v	the pair of vertices that lie in a triangle with v in the embedding $\psi(D_t)$ (see Section 4.4.2)
$Y_v(\psi)$	the indicator function for the event that v is *not* isolated under ψ (see Section 4.4.2)
ψ^*	a uniformly random embedding $\psi \in \Psi^t(\Gamma_p)$ (see Section 4.4.2)
$H(n, p, d)$	a benchmark function for maximum entropy (see Definition 4.4.2)
σ	a linear ordering on V^ℓ (see Section 4.4.2)
$w <_\sigma v$	w appears before v in the ordering σ (see Section 4.4.2)
$A_v^j(\psi, \sigma, \underline{u})$	the set of alive vertices in $V_{\hat{\underline{u}}}^j$ at the point of revealing ψ at vertex v according to a linear ordering σ (see Section 4.4.2)
$t_v(\psi)$	the number of vertices before v in the order σ, that are not isolated under ψ (see proof of Lemma 4.4.4)
$\zeta_v(w_1, w_2)$	a weight function that encodes the number of embeddings $\psi \in \Psi^{(t-1)}(G)$ that avoid v, w_1 and w_2 (see (4.4.16))
$\overline{e}(X, Y)$	the number of edges missing between sets X and Y (see proof of

	Lemma 4.7.2)
H_{det}	a partite subgraph of K_r (see Definition 5.2.1)
$\overline{H_{\text{det}}}$	the complement of H_{det} (see Definition 5.2.1)
Φ_F	a graph parameter capturing the expected number of the densest subgraph of F in $G(n,p)$ (see Section 5.3)
$\Phi_{F,W}$	as above but with embeddings that have some independent set W of vertices of F already fixed (see Section 5.3)
K_k^-	the clique on k vertices with one edge removed (see Section 5.3)
$(K^t_{r_1,r_2,\ldots,r_t}, i, j)$	a copy of $K^t_{r_1,r_2,\ldots,r_t}$ with distinguished vertices in the i^{th} part and the j^{th} part (see Definition 5.5.1)
$\overline{H}$	the complement of an $(r+1)$-vertex graph H with respect to K^-_{r+1} (see Definition 5.5.2)
H_0	a specific graph with two distinguished vertices (see Definition 5.5.3)
$\boldsymbol{H}$	a vector of $(r+1)$-vertex graphs, each with two distinguished vertices (see Definition 5.5.4)
(H,t)	a vector with t copies of H (see Definition 5.5.4)
$\overline{\boldsymbol{H}}$	the complement of a vector of $(r+1)$-vertex graphs $\boldsymbol{H}$ (see Definition 5.5.4)
H_1	a specific $(r+1)$-vertex graph with two distinguished vertices (see Proposition 5.5.6)
H_1'	as H_1 with the two distinguished vertices switched (see Proposition 5.5.6)
$\boldsymbol{H_1}$	the vector (H_1, H_1') (see Proposition 5.5.6)
H_0'	as H_0 with different distinguished vertices (see Proposition 5.5.7)
$\boldsymbol{H_2}$	the vector (H_0, H_0', H_0', H_0) (see Proposition 5.5.7)
$\boldsymbol{\mathcal{H}}$	a set of vectors $\boldsymbol{H}$ (see Definition 5.5.8)
$N_{\boldsymbol{\mathcal{H}},\beta}(v)$	the set of vertices that are $(\boldsymbol{\mathcal{H}},\beta)$-reachable to v (see Definition 5.5.8)
$N_{\boldsymbol{H},\beta}(v)$	the set of vertices that are $(\boldsymbol{H},\beta)$-reachable to v (see Definition 5.5.8)
$\boldsymbol{\mathcal{H}} \cdot \tilde{\boldsymbol{\mathcal{H}}}$	the set of concatenations of vectors in $\boldsymbol{\mathcal{H}}$ with vectors in $\tilde{\boldsymbol{\mathcal{H}}}$ (see Definition 5.5.9)
$\boldsymbol{\mathcal{H}} + \tilde{\boldsymbol{\mathcal{H}}}$	the set of vectors that lie in one of $\boldsymbol{\mathcal{H}}$ or $\tilde{\boldsymbol{\mathcal{H}}}$ or are a concatenation of a vector in $\boldsymbol{\mathcal{H}}$ with a vector in $\tilde{\boldsymbol{\mathcal{H}}}$ (see Definition 5.5.9)
$\boldsymbol{\mathcal{H}}(H, \leq t)$	the set of vectors of length at most t whose entries are all H (see Section 5.5)
$\boldsymbol{\mathcal{H}}^t$	the set of vectors $\boldsymbol{\mathcal{H}}(H_0, \leq 2^t)$ (see Lemma 5.5.11)
$\boldsymbol{\mathcal{H}_3}$	a set of specific vectors $\boldsymbol{H}$ (see Proposition 5.5.12)
$\boldsymbol{H}[i]$	the i^{th} entry of a vector $\boldsymbol{H}$ (see Proposition 5.5.12)
$\underline{\boldsymbol{H}}$	a labelled family of vectors of $(r+1)$-vertex graphs (see Definition 5.5.14)
$\boldsymbol{\mathcal{H}}(r \times s)$	collection of ordered labelled sets $\underline{\boldsymbol{H}}$ whose entries lie in $\boldsymbol{\mathcal{H}}$ (see Definition 5.5.15)
$\boldsymbol{H}(r \times s)$	ordered labelled multiset $\underline{\boldsymbol{H}}$ whose entries are all $\boldsymbol{H}$ (see Definition 5.5.15)
$\overline{\underline{\boldsymbol{H}}}$	the labelled family of vectors obtained by taking the complement of each

	entry of $\underline{\boldsymbol{H}}$ (see Section 5.6)
$K_k + K_k$	two copies of K_k meeting at a singular vertex (see proof of Lemma 5.6.1)
$\mathcal{A}$	an absorbing structure (see Definition 5.6.2)
$d^*(H)$	the 1-density of H (see (6.1.1))
$\ell(n; H)$	the logarithmic factor in the threshold for an H-factor (see (6.1.2))
$\mathcal{K}_r^{(k)}$	the k-uniform clique with r vertices (see Section 6.2.1)
$\mathcal{H}^{(k)}(n, p)$	the binomial random k-graph (see Section 6.2.1)
$\alpha_k(G)$	the largest size of a K_k-free subset of vertices in G (see Section 6.2.2)
$\mathbf{RT}_k(H)$	the Ramsey–Turán threshold for H (see Section 6.2.2)
$\mathbf{RTF}_k(K_r)$	the Ramsey–Turán factor threshold for K_r (see Section 6.2.2)

Bibliography

[1] M. Aigner and S. Brandt. Embedding arbitrary graphs of maximum degree two. *Journal of the London Mathematical Society*, 2(1):39–51, 1993.

[2] P. Allen, J. Böttcher, H. Hàn, Y. Kohayakawa, and Y. Person. Blow-up lemmas for sparse graphs. *arXiv preprint arXiv:1612.00622*, 2016.

[3] P. Allen, J. Böttcher, H. Hàn, Y. Kohayakawa, and Y. Person. Powers of Hamilton cycles in pseudorandom graphs. *Combinatorica*, 37(4):573–616, 2017.

[4] P. Allen, J. Böttcher, Y. Kohayakawa, H. Naves, and Y. Person. Making spanning graphs. *arXiv preprint arXiv:1711.05311*, 2017.

[5] P. Allen, J. Böttcher, J. Corsten, E. Davies, M. Jenssen, P. Morris, B. Roberts, and J. Skokan. A robust Corrádi–Hajnal Theorem. In preparation.

[6] P. Allen, J. Böttcher, E. Davies, Y. Kohayakawa, M. Jenssen, and B. Roberts. Shamir's problem revisited. Unpublished.

[7] N. Alon. Explicit Ramsey graphs and orthonormal labelings. *Electronic Journal of Combinatorics*, 1:R12, 8pp, 1994.

[8] N. Alon and J. Bourgain. Additive patterns in multiplicative subgroups. *Geometric and Functional Analysis*, 24(3):721–739, 2014.

[9] N. Alon and E. Fischer. 2-factors in dense graphs. *Discrete Mathematics*, 152(1-3):13–23, 1996.

[10] N. Alon, P. Frankl, H. Huang, V. Rödl, A. Ruciński, and B. Sudakov. Large matchings in uniform hypergraphs and the conjectures of Erdős and Samuels. *Journal of Combinatorial Theory, Series A*, 119(6):1200–1215, 2012.

[11] N. Alon and N. Kahale. Approximating the independence number via the ϑ-function. *Mathematical Programming*, 80(3):253–264, 1998.

[12] N. Alon and J. Spencer. *The probabilistic method*. Wiley Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, fourth edition, 2016.

[13] N. Alon and R. Yuster. Threshold functions for H-factors. *Combinatorics, Probability and Computing*, 2(2):137–144, 1993.

[14] Y. Alon and M. Krivelevich. Hitting time of edge disjoint Hamilton cycles in random subgraph processes on dense base graphs. *arXiv preprint arXiv:1912.01251*, 2019.

[15] S. Antoniuk, A. Dudek, C. Reiher, A. Ruciński, and M. Schacht. High powers of Hamiltonian cycles in randomly augmented graphs. *Journal of Graph Theory*, 2021. To appear.

[16] J. Balogh and J. Lenz. On the Ramsey–Turán numbers of graphs and hypergraphs. *Israel Journal of mathematics*, 194(1):45–68, 2013.

[17] J. Balogh, A. McDowell, T. Molla, and R. Mycroft. Triangle-tilings in graphs without large independent sets. *Combinatorics, Probability and Computing*, 27(4):449–474, 2018.

[18] J. Balogh, T. Molla, and M. Sharifzadeh. Triangle factors of graphs without large independent sets and of weighted graphs. *Random Structures & Algorithms*, 49(4):669–693, 2016.

[19] J. Balogh, A. Treglown, and A. Z. Wagner. Tilings in randomly perturbed dense graphs. *Combinatorics, Probability and Computing*, 28:159–176, 2019.

[20] W. Bedenknecht, J. Han, Y. Kohayakawa, and G. O. Mota. Powers of tight Hamilton cycles in randomly perturbed hypergraphs. *Random Structures & Algorithms*, 55(4):795–807, 2019.

[21] P. Bennett, A. Dudek, and A. Frieze. Adding random edges to create the square of a Hamilton cycle. *arXiv preprint arXiv:1710.02716*, 2017.

[22] A. Bishnoi, F. Ihringer, and V. Pepe. A construction for clique-free pseudorandom graphs. *Combinatorica*, 40(3):307–314, 2020.

[23] T. Bohman, A. Frieze, and R. Martin. How many random edges make a dense graph Hamiltonian? *Random Structures & Algorithms*, 22(1):33–42, 2003.

[24] T. Bohman and P. Keevash. The early evolution of the H-free process. *Inventiones mathematicae*, 181(2):291–336, 2010.

[25] B. Bollobás and S. E. Eldridge. Packings of graphs and applications to computational complexity. *Journal of Combinatorial Theory, Series B*, 25(2):105–124, 1978.

[26] B. Bollobás and P. Erdős. On a Ramsey–Turán type problem. *Journal of Combinatorial Theory, Series B*, 21(2):166–168, 1976.

[27] J. Böttcher, J. Han, Y. Kohayakawa, R. Montgomery, O. Parczyk, and Y. Person. Universality for bounded degree spanning trees in randomly perturbed graphs. *Random Structures & Algorithms*, 55(4):854–864, 2019.

[28] J. Böttcher, R. Montgomery, O. Parczyk, and Y. Person. Embedding spanning bounded degree graphs in randomly perturbed graphs. *Mathematika*, 66(2):422–447, 2020.

[29] J. Böttcher, O. Parczyk, A. Sgueglia, and J. Skokan. Triangles in randomly perturbed graphs. *arXiv preprint arXiv:2011.07612*, 2020.

[30] J. Böttcher, O. Parczyk, A. Sgueglia, and J. Skokan. Cycle factors in randomly perturbed graphs. *arXiv preprint arXiv:2103.06136*, 2021.

[31] A. Z. Broder, A. M. Frieze, S. Suen, and E. Upfal. Optimal construction of edge-disjoint paths in random graphs. *SIAM Journal on Computing*, 28(2):541–573, 1998.

[32] A. E. Brouwer and W. H. Haemers. Eigenvalues and perfect matchings. *Linear Algebra and its Applications*, 395:155–162, 2005.

[33] P. A. Catlin. *Embeddings subgraphs and coloring graphs under extremal degree conditions*. PhD thesis, Ohio State University, 1976.

[34] Y. Chang, J. Han, Y. Kohayakawa, P. Morris, and G. O. Mota. Factors in randomly perturbed hypergraphs. *Random Structures & Algorithms*, 2021. To appear.

[35] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23:493–507, 1952.

[36] V. Chvátal. The tail of the hypergeometric distribution. *Discrete Mathematics*, 25(3):285–287, 1979.

[37] S. M. Cioabă. Perfect matchings, eigenvalues and expansion. *Mathematical Reports of the Academy of Sciences*, 27(4):101–104, 2005.

[38] P. Condon, A. E. Díaz, A. Girão, D. Kühn, and D. Osthus. Hamiltonicity of random subgraphs of the hypercube. *arXiv preprint arXiv:2007.02891*, 2020.

[39] P. Condon, A. E. Díaz, A. Girão, D. Kühn, and D. Osthus. Hamiltonicity of random subgraphs of the hypercube. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 889–898. SIAM, 2021.

[40] D. Conlon. A sequence of triangle-free pseudorandom graphs. *Combinatorics, Probability and Computing*, 26(2):195–200, 2017.

[41] D. Conlon, J. Fox, and Y. Zhao. Extremal results in sparse pseudorandom graphs. *Advances in Mathematics*, 256:206–290, 2014.

[42] D. Conlon, H. Hàn, Y. Person, and M. Schacht. Weak quasi-randomness for uniform hypergraphs. *Random Structures & Algorithms*, 40(1):1–38, 2012.

[43] N. Cook, L. Goldstein, and T. Johnson. Size biased couplings and the spectral gap for random regular graphs. *The Annals of Probability*, 46(1):72–125, 2018.

[44] K. Corrádi and A. Hajnal. On the maximal number of independent circuits in a graph. *Acta Mathematica Academiae Scientiarum Hungarica*, 14(3-4):423–439, 1963.

[45] J. Corsten. *Extremal graph colouring and tiling problems*. PhD thesis, The London School of Economics and Political Science (LSE), 2020.

[46] M. Coulson, P. Keevash, G. Perarnau, and L. Yepremyan. Rainbow factors in hypergraphs. *Journal of Combinatorial Theory, Series A*, 172:105184, 2020.

[47] M. Coulson and G. Perarnau. A rainbow Dirac's theorem. *SIAM Journal on Discrete Mathematics*, 34(3):1670–1692, 2020.

[48] B. Csaba, D. Kühn, A. Lo, D. Osthus, and A. Treglown. Proof of the 1-factorization and Hamilton decomposition conjectures. *Memoirs of the American Mathematical Society*, 244(1154), 2016.

[49] B. Csaba, A. Shokoufandeh, and E. Szemerédi. Proof of a conjecture of Bollobás and Eldridge for graphs of maximum degree three. *Combinatorica*, 23(1):35–72, 2003.

[50] B. Cuckler and J. Kahn. Hamiltonian cycles in Dirac graphs. *Combinatorica*, 29(3):299–326, 2009.

[51] A. Czygrinow, L. DeBiasio, H. A. Kierstead, and T. Molla. An extension of the Hajnal–Szemerédi theorem to directed graphs. *Combinatorics, Probability and Computing*, 24(5):754–773, 2015.

[52] A. Czygrinow, L. DeBiasio, T. Molla, and A. Treglown. Tiling directed graphs with tournaments. *Forum of Mathematics, Sigma*, 6:e2, 2018.

[53] S. Das, P. Morris, and A. Treglown. Vertex Ramsey properties of randomly perturbed graphs. *Random Structures & Algorithms*, 57(4):983–1006, 2020.

[54] S. Das and A. Treglown. Ramsey properties of randomly perturbed graphs: cliques and cycles. *Combinatorics, Probability and Computing*, 29(6):830–867, 2020.

[55] G. A. Dirac. Some theorems on abstract graphs. *Proceedings of the London Mathematical Society*, 3(1):69–81, 1952.

[56] A. Dudek, C. Reiher, A. Ruciński, and M. Schacht. Powers of Hamiltonian cycles in randomly augmented graphs. *Random Structures & Algorithms*, 56(1):122–141, 2020.

[57] S. Ehard and F. Joos. Paths and cycles in random subgraphs of graphs with large minimum degree. *The Electronic Journal of Combinatorics*, 25(2):2–31, 2018.

[58] P. Erdős. Problem 9. In *Theory of Graphs and its Applications*. Czechoslovak Academy of Sciences, 1964.

[59] P. Erdős. On the combinatorial problems which I would most like to see solved. *Combinatorica*, 1(1):25–42, 1981.

[60] P. Erdős, A. Gyárfás, and L. Pyber. Vertex coverings by monochromatic cycles and trees. *Journal of Combinatorial Theory. Series B*, 51(1):90–95, 1991.

[61] P. Erdős, A. Hajnal, V. T. Sós, and E. Szemerédi. More results on Ramsey–Turán type problems. *Combinatorica*, 3(1):69–81, 1983.

[62] P. Erdős and V. T. Sós. Some remarks on Ramsey's and Turán's theorem. In *Combinatorial theory and its applications, II (Proceedings of Colloquium, Balatonfüred, 1969)*, pages 395–404, 1970.

[63] P. Erdős. Appendix B: Some open problems. In *The Probabilistic Method, First Edition*, pages 234–237. John Wiley & Sons, Inc., Hoboken, NJ, 1992. Book authors: Alon, N. and Spencer, J.

[64] P. Erdős and M. Simonovits. Supersaturated graphs and hypergraphs. *Combinatorica*, 3(2):181–192, 1983.

[65] A. Ferber, G. Kronenberg, and K. Luh. Optimal threshold for a random graph to be 2-universal. *Transactions of the American Mathematical Society*, 372(6):4239–4262, 2019.

[66] A. Ferber and M. Kwan. Almost all Steiner triple systems are almost resolvable. *Forum of Mathematics, Sigma*, 8:e39, 2020.

[67] M. Fischer. Robustness of Pósa's conjecture. Master's thesis, Eidgenössische Technische Hochschule (ETH) Zürich, 2016.

[68] J. Fox, C. Lee, and B. Sudakov. Chromatic number, clique subdivisions, and the conjectures of Hajós and Erdős–Fajtlowicz. *Combinatorica*, 33(2):181–197, 2013.

[69] K. Frankston, J. Kahn, B. Narayanan, and J. Park. Thresholds versus fractional expectation-thresholds. *Annals of Mathematics*, 2021. To appear.

[70] A. Frieze. Hamilton cycles in random graphs: a bibliography. *arXiv preprint arXiv:1901.07139*, 2019.

[71] A. Frieze and M. Krivelevich. Hamilton cycles in random subgraphs of pseudo-random graphs. *Discrete Mathematics*, 256(1-2):137–150, 2002.

[72] D. Galvin. Three tutorial lectures on entropy and counting. arXiv preprint arXiv:1406.7872, 2014.

[73] S. Gerke and A. McDowell. Nonvertex-balanced factors in random graphs. *Journal of Graph Theory*, 78(4):269–286, 2015.

[74] R. Glebov, Z. Luria, and M. Simkin. Perfect matchings in random subgraphs of regular bipartite graphs. *Journal of Graph Theory*, 97(2):208–231, 2021.

[75] R. Glebov, H. Naves, and B. Sudakov. The threshold probability for long cycles. *Combinatorics, Probability and Computing*, 26(2):208–247, 2017.

[76] S. Glock, D. Kühn, R. Montgomery, and D. Osthus. Decompositions into isomorphic rainbow spanning trees. *Journal of Combinatorial Theory, Series B*, 146:439–484, 2021.

[77] A. Goel, M. Kapralov, and S. Khanna. Perfect matchings via uniform sampling in regular bipartite graphs. *ACM Transactions on Algorithms (TALG)*, 6(2):1–13, 2010.

[78] A. Hajnal and E. Szemerédi. Proof of a conjecture of P. Erdős. In *Combinatorial theory and its applications, II (Proceedings of Colloquium, Balatonfüred, 1969)*, pages 601–623, 1970.

[79] H. Hàn, J. Han, and P. Morris. Factors and loose Hamilton cycles in sparse pseudo-random hypergraphs. In *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 702–717. SIAM, 2020.

[80] H. Hàn, J. Han, and P. Morris. Factors and loose Hamilton cycles in sparse pseudo-random hypergraphs. *Random Structures & Algorithms*, 2021. To appear.

[81] J. Han. Near perfect matchings in k-uniform hypergraphs II. *SIAM J. Discrete Math.*, 30(3):1453–1469, 2016.

[82] J. Han. Decision problem for perfect matchings in dense k-uniform hypergraphs. *Transactions of the American Mathematical Society*, 369(7):5197–5218, 2017.

[83] J. Han, Y. Kohayakawa, P. Morris, and Y. Person. Clique-factors in sparse pseudorandom graphs. *European Journal of Combinatorics*, 82:102999, 2019.

[84] J. Han, Y. Kohayakawa, P. Morris, and Y. Person. Finding any given 2-factor in sparse pseudorandom graphs efficiently. *Journal of Graph Theory*, 96:87–108, 2021.

[85] J. Han, Y. Kohayakawa, and Y. Person. Near-perfect clique-factors in sparse pseudorandom graphs. In *Proceedings of the 11th "Jornadas de matemática discreta y algorítmica" (JMDA), Sevilla, Spain, June 27–29*, pages 221–226. Amsterdam: Elsevier, 2018.

[86] J. Han, Y. Kohayakawa, and Y. Person. Near-perfect clique-factors in sparse pseudorandom graphs. *Combinatorics, Probability & Computing*, 30(4):570–590, 2021.

[87] J. Han, P. Morris, and A. Treglown. Tilings in randomly perturbed graphs: bridging the gap between Hajnal–Szemerédi and Johansson–Kahn–Vu. *Random Structures & Algorithms*, 58(3):480–516, 2021.

[88] J. Han, P. Morris, G. Wang, and D. Yang. A Ramsey–Turán theory for tilings in graphs. *arXiv preprint arXiv:2106.09688*, 2021.

[89] J. Han and A. Treglown. The complexity of perfect matchings and packings in dense hypergraphs. *Journal of Combinatorial Theory, Series B*, 141:72–104, 2020.

[90] A. Heckel. Random triangles in random graphs. *Random Structures & Algorithms*, 2018. To appear.

[91] D. Hefetz, M. Krivelevich, M. Stojaković, and T. Szabó. *Positional games*, volume 44 of *Oberwolfach Seminars*. Birkhäuser Basel, 2014.

[92] P. Hell and D. G. Kirkpatrick. On the complexity of general graph factor problems. *SIAM Journal on Computing*, 12(3):601–609, 1983.

[93] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[94] S. Janson. Poisson approximation for large deviations. *Random Structures & Algorithms*, 1(2):221–229, 1990.

[95] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000.

[96] A. Johansson, J. Kahn, and V. Vu. Factors in random graphs. *Random Structures & Algorithms*, 33(1):1–28, 2008.

[97] T. Johansson. On Hamilton cycles in Erdős–Rényi subgraphs of large graphs. *Random Structures & Algorithms*, 57(1):132–149, 2020.

[98] F. Joos and J. Kim. Spanning trees in randomly perturbed graphs. *Random Structures & Algorithms*, 56(1):169–219, 2020.

[99] J. Kahn. Asymptotics for Shamir's problem. *arXiv preprint arXiv:1909.06834*, 2019.

[100] J. Kahn. Hitting times for Shamir's problem. *arXiv preprint arXiv:2008.01605*, 2020.

[101] J. Kahn, B. Narayanan, and J. Park. The threshold for the square of a Hamilton cycle. *Proceedings of the American Mathematical Society*, 149(08):3201–3208, 2021.

[102] R. M. Karp. On the computational complexity of combinatorial problems. *Networks*, 5(1):45–68, 1975.

[103] P. Keevash and R. Mycroft. A geometric theory for hypergraph matching. *Memoirs of the American Mathematical Society*, 233(1098), 2015.

[104] P. Keevash and R. Mycroft. A multipartite Hajnal–Szemerédi theorem. *Journal of Combinatorial Theory, Series B*, 114:187–236, 2015.

[105] H. A. Kierstead and A. V. Kostochka. A short proof of the Hajnal–Szemerédi theorem on equitable colouring. *Combinatorics, Probability and Computing*, 17(2):265–270, 2008.

[106] J. H. Kim. Perfect matchings in random uniform hypergraphs. *Random Structures & Algorithms*, 23(2):111–132, 2003.

[107] J. H. Kim and V. Vu. Concentration of multivariate polynomials and its applications. *Combinatorica*, 20(3):417–434, 2000.

[108] C. Knierim and P. Su. K_r-factors in graphs with low independence number. *Journal of Combinatorial Theory, Series B*, 148:60–83, 2021.

[109] Y. Kohayakawa, B. Nagle, V. Rödl, and M. Schacht. Weak hypergraph regularity and linear hypergraphs. *Journal of Combinatorial Theory, Series B*, 100(2):151–160, 2010.

[110] J. Komlós. Tiling Turán theorems. *Combinatorica*, 20(2):203–218, 2000.

[111] J. Komlós, G. N. Sárközy, and E. Szemerédi. On the square of a Hamiltonian cycle in dense graphs. *Random Structures & Algorithms*, 9(1-2):193–211, 1996.

[112] J. Komlós, G. N. Sárközy, and E. Szemerédi. On the Pósa-Seymour conjecture. *Journal of Graph Theory*, 29(3):167–176, 1998.

[113] J. Komlós and M. Simonovits. Szemerédi's regularity lemma and its applications in graph theory. In *Paul Erdős is eighty*, volume 2 of *Bolyai Society Mathematical Studies*, pages 295–352. Springer, Berlin, 1996.

[114] J. Komlós and E. Szemerédi. Limit distribution for the existence of Hamiltonian cycles in a random graph. *Discrete Mathematics*, 43(1):55–63, 1983.

[115] S. Kopparty. Cayley graphs (lecture notes), 2011. available: http://sites.math.rutgers.edu/~sk1233/courses/graphtheory-F11/cayley.pdf.

[116] A. D. Koršunov. Solution of a problem of P. Erdős and A. Rényi on Hamiltonian cycles in undirected graphs. *Doklady Akademii Nauk SSSR*, 228(3):529–532, 1976.

[117] A. V. Kostochka and V. Rödl. Partial Steiner systems and matchings in hypergraphs. *Random Structures & Algorithms*, 13(3-4):335–347, 1998.

[118] M. Krivelevich. Perfect fractional matchings in random hypergraphs. *Random Structures & Algorithms*, 9(3):317–334, 1996.

[119] M. Krivelevich. Triangle factors in random graphs. *Combinatorics, Probability and Computing*, 6(3):337–347, 1997.

[120] M. Krivelevich, M. Kwan, and B. Sudakov. Cycles and matchings in randomly perturbed digraphs and hypergraphs. *Combinatorics, Probability and Computing*, 25(6):909–927, 2016.

[121] M. Krivelevich, M. Kwan, and B. Sudakov. Bounded-degree spanning trees in randomly perturbed graphs. *SIAM Journal on Discrete Mathematics*, 31(1):155–171, 2017.

[122] M. Krivelevich, C. Lee, and B. Sudakov. Robust Hamiltonicity of Dirac graphs. *Transactions of the American Mathematical Society*, 366(6):3095–3130, 2014.

[123] M. Krivelevich, C. Lee, and B. Sudakov. Long paths and cycles in random subgraphs of graphs with large minimum degree. *Random Structures & Algorithms*, 46(2):320–345, 2015.

[124] M. Krivelevich, C. Lee, and B. Sudakov. Compatible Hamilton cycles in Dirac graphs. *Combinatorica*, 37(4):697–732, 2017.

[125] M. Krivelevich and B. Sudakov. Sparse pseudo-random graphs are Hamiltonian. *Journal of Graph Theory*, 42(1):17–33, 2003.

[126] M. Krivelevich and B. Sudakov. Pseudo-random graphs. In *More sets, graphs and numbers*, volume 15 of *Bolyai Society Mathematical Studies*, pages 199–262. Springer, Berlin, 2006.

[127] M. Krivelevich, B. Sudakov, and T. Szabó. Triangle factors in sparse pseudo-random graphs. *Combinatorica*, 24(3):403–426, 2004.

[128] M. Krivelevich, B. Sudakov, and P. Tetali. On smoothed analysis in dense graphs and formulas. *Random Structures & Algorithms*, 29:180–193, 2006.

[129] D. Kühn and D. Osthus. Embedding large subgraphs into dense graphs. In *Surveys in Combinatorics*, volume 365 of *London Mathematical Society Lecture Notes*, pages 137–167. Cambridge University Press, 2009.

[130] D. Kühn and D. Osthus. The minimum degree threshold for perfect graph packings. *Combinatorica*, 29(1):65–107, 2009.

[131] M. Kwan. Almost all Steiner triple systems have perfect matchings. *Proceedings of the London Mathematical Society*, 121(6):1468–1495, 2020.

[132] J. Lenz and D. Mubayi. Perfect packings in quasirandom hypergraphs I. *Journal of Combinatorial Theory, Series B*, 119:155–177, 2016.

[133] J. Lenz, D. Mubayi, and R. Mycroft. Hamilton cycles in quasirandom hypergraphs. *Random Structures & Algorithms*, 49(2):363–378, 2016.

[134] A. Liebenau and R. Nenadov. The threshold bias of the clique-factor game. *arXiv preprint arXiv:2002.02578*, 2020.

[135] A. Lo and K. Markström. f-factors in hypergraphs via absorption. *Graphs and Combinatorics*, 31(3):679–712, 2015.

[136] R. Montgomery. Embedding bounded degree spanning trees in random graphs. *arXiv preprint arXiv:1405.6559*, 2014.

[137] R. Montgomery. Spanning trees in random graphs. *Advances in Mathematics*, 356:106793, 2019.

[138] R. Montgomery, A. Pokrovskiy, and B. Sudakov. A proof of Ringel's conjecture. *arXiv preprint arXiv:2001.02665*, 2020.

[139] P. Morris. Random Steiner triple systems. Master's thesis, Freie Universität Berlin, 2017.

[140] P. Morris. A tight condition for triangle factors in pseudorandom graphs. In *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 899–918. SIAM, 2021.

[141] P. Morris. Clique factors in pseudorandom graphs. *arXiv preprint arXiv:2101.05092*, 2021.

[142] D. Mubayi and J. Verstraëte. A note on pseudorandom Ramsey graphs. *arXiv preprint arXiv:1909.01461*, 2019.

[143] C. S. J. Nash-Williams. Hamiltonian lines in graphs whose vertices have sufficiently large valencies. In *Combinatorial theory and its applications, III (Proceedings of Colloquium, Balatonfüred, 1969)*, pages 813–819, 1970.

[144] C. S. J. Nash-Williams. Edge-disjoint Hamiltonian circuits in graphs with vertices of large valency. In *Studies in Pure Mathematics (Presented to Richard Rado)*, pages 157–183. Academic Press London, 1971.

[145] R. Nenadov. Triangle-factors in pseudorandom graphs. *Bulletin of the London Mathematical Society*, 51(3):421–430, 2019.

[146] R. Nenadov and Y. Pehova. On a Ramsey–Turán variant of the Hajnal–Szemerédi theorem. *SIAM Journal on Discrete Mathematics*, 34(2):1001–1010, 2020.

[147] R. Nenadov and M. Trujić. Sprinkling a few random edges doubles the power. *SIAM Journal on Discrete Mathematics*, 35(2):988–1004, 2021.

[148] O. Parczyk. 2-universality in randomly perturbed graphs. *European Journal of Combinatorics*, 87:103118, 2020.

[149] L. Pósa. Hamiltonian circuits in random graphs. *Discrete Mathematics*, 14(4):359–364, 1976.

[150] E. Powierski. Ramsey properties of randomly perturbed dense graphs. *arXiv:1902.02197*, 2019.

[151] O. Riordan. Spanning subgraphs of random graphs. *Combinatorics, Probability & Computing*, 9(2):125–148, 2000.

[152] O. Riordan. Long cycles in random subgraphs of graphs with large minimum degree. *Random Structures & Algorithms*, 45(4):764–767, 2014.

[153] O. Riordan. Random cliques in random graphs. *arXiv preprint arXiv:1802.01948*, 2018.

[154] V. Rödl, A. Ruciński, and E. Szemerédi. A Dirac-type theorem for 3-uniform hypergraphs. *Combinatorics, Probability and Computing*, 15(1-2):229–251, 2006.

[155] A. Ruciński. Matching and covering the vertices of a random graph by copies of a given graph. *Discrete Mathematics*, 105(1-3):185–197, 1992.

[156] G. N. Sárközy, S. M. Selkow, and E. Szemerédi. On the number of Hamiltonian cycles in Dirac graphs. *Discrete Mathematics*, 265(1-3):237–250, 2003.

[157] P. Seymour. Problem section. In *Combinatorics: Proceedings of the British Combinatorial Conference 1973*, pages 201–202. Cambridge University Press, 1974.

[158] A. Shokoufandeh and Y. Zhao. Proof of a tiling conjecture of Komlós. *Random Structures & Algorithms*, 23(2):180–205, 2003.

[159] M. Šileikis and L. Warnke. Counting extensions revisited. *Random Structures & Algorithms*, 2021. To appear.

[160] M. Simonovits and V. T. Sós. Ramsey–Turán theory. *Discrete Mathematics*, 229(1-3):293–340, 2001.

[161] M. Skala. Hypergeometric tail inequalities: ending the insanity. *arXiv preprint arXiv:1311.5939*, 2013.

[162] J. Spencer. Counting extensions. *Journal of Combinatorial Theory, Series A*, 55(2):247–255, 1990.

[163] J. Spencer. Threshold functions for extension statements. *Journal of Combinatorial Theory, Series A*, 53(2):286–305, 1990.

[164] D. A. Spielman and S. H. Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *Journal of the ACM (JACM)*, 51(3):385–463, 2004.

[165] B. Sudakov. Robustness of graph properties. *Surveys in combinatorics*, 440:372, 2017.

[166] B. Sudakov, T. Szabó, and V. Vu. A generalization of Turán's theorem. *Journal of Graph Theory*, 49(3):187–195, 2005.

[167] E. Szemerédi. On graphs containing no complete subgraph with 4 vertices. *Matematikai Lapok*, 23:113–116, 1972.

[168] E. Szemerédi. Regular partitions of graphs. *Problémes Combinatoires et Théorie des Graphes Colloques Internationaux CNRS 260*, pages 399–401, 1978.

[169] A. Thomason. Pseudorandom graphs. In *Random graphs '85 (Poznań, 1985)*, volume 144 of *North-Holland Mathematics Studies*, pages 307–331. North-Holland, Amsterdam, 1987.

[170] A. Thomason. Random graphs, strongly regular graphs and pseudorandom graphs. In *Surveys in combinatorics*, volume 123 of *London Mathematical Society Lecture Note Series*, pages 173–195. Cambridge University Press, Cambridge, 1987.

[171] K. Tikhomirov and P. Youssef. The spectral gap of dense random regular graphs. *The Annals of Probability*, 47(1):362–419, 2019.

[172] A. Treglown. On directed versions of the Hajnal–Szemerédi theorem. *Combinatorics, Probability and Computing*, 24(6):873–928, 2015.

[173] A. Treglown. A degree sequence Hajnal–Szemerédi theorem. *Journal of Combinatorial Theory, Series B*, 118:13–43, 2016.

[174] Y. Zhao. Recent advances on Dirac-type problems for hypergraphs. In *Recent trends in combinatorics*, volume 159 of *IMA Volumes in Mathematics and its Applications*, pages 145–165. Springer, Cham, New York, 2015.