On Hamilton cycles and other spanning structures

DISSERTATION
zur Erlangung des akademischen Grades
doctor rerum naturalium
(Dr. rer. nat.)
im Fach Mathematik

eingereicht am
Fachbereich Mathematik und Informatik
der Freien Universität Berlin

von
Roman Glebov

Erstgutachter: Prof. Tibor Szabó,
Zweitgutachter: Prof. Yoshiharu Kohayakawa

Tag der Disputation: 16.01.2013

Erklärung

Ich erkläre hiermit, dass

- ich die vorliegende Dissertationsschrift “On Hamilton cycles and other spanning structures” selbstständig, ohne unerlaubte Hilfe und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe,
- ich mich nicht anderwärts um einen Doktorgrad beworben habe oder einen solchen besitze, und
- mir die Promotionsordnung des Fachbereichs Mathematik und Informatik der Freien Universität Berlinn bekannt ist, gemäß Amtl. Mitteilungsblatt Nr.2 /2007.

Berlin, den 16. Januar 2014,

Roman Glebov

Contents

1 Introduction

1.1 Historical background

In their highly influential papers [28] and [29], Erdős and Rényi defined two related notions of a *random graph*. In the model $G(n, M)$, a graph is drawn uniformly at random among all n-vertex graphs with exactly M edges, whereas $G(n, p)$ can be seen as a product probability space with every unordered pair of vertices becoming an edge with probability p uniformly at random. In the probabilistic graph theory, we are usually interested in events happening *asymptotically almost surely*, or for short *a.a.s.*, that is, with probability tending to one as n tends to infinity. We also say that this event happens in a *typical* graph $G \sim G(n, p)$. The close relation between the two models is given by the fact that for every monotone graph property Q, the graph $G \sim G(n, p)$ has this property a.a.s. if and only if the graph $G' \sim G\left(n, p\binom{n}{2}\right)$ has this property a.a.s. (see e.g. [14]).

Erdős and Rényi found out that for a number of fundamental structural properties Q, there exists a function $t = t(n)$ such that the random graph $G \sim G(n, p)$ has the property a.a.s. if $p \gg t$, and $G \notin Q$ a.a.s. if $p \ll t$. We call t the *threshold function* for Hamiltonicity. They also observed that some properties Q have a so-called *sharp threshold* t, such that for every $\varepsilon > 0$, we obtain $G \in Q$ a.a.s. for $p \geq (1 + \varepsilon)t$, and $G \notin Q$ a.a.s. for $p \ll t$. They showed that the property of being connected has the sharp threshold $\log n/n$ by proving that $G(n, (1+\varepsilon) \log n/n)$ is a.a.s. connected, whereas $G(n, (1-\varepsilon) \log n/n)$ a.a.s. contains an isolated vertex. (In fact, they provided even a more detailed view on the probability of the random graph being connected for p being close to $\log n/n$.) In a question in [29], they asked about the threshold probability for a Hamilton path to appear.

Clearly, since $G(n, (1 - \varepsilon) \log n/n)$ is a.a.s. not connected, it contains neither a Hamilton path nor a Hamilton cycle. Komlós and Szemerédi [63] showed that $G\left(n, c \exp\left(\sqrt{\log n}\right)/n\right)$ is hamiltonian a.a.s. for a sufficiently large constant c, before Pósa [78] improved their result by proving that $G(n, c \log n/n)$ is hamiltonian a.a.s. for a sufficiently large constant c. Notice that this already settles the threshold function for Hamiltonicity as well as for a Hamilton path to be of order $\log n/n$.

It is well known (see e.g. [14]) that for $p \leq \frac{\log n + \log\log n - \omega(1)}{n}$, a.a.s. $G(n, p)$ contains more than two vertices of degree at most one, and therefore $G(n, p)$ contains no Hamilton cycle and no Hamilton path in this range of p a.a.s. Similarly, replacing $\omega(1)$ by any function with a finite limit leads to a constant positive probability for the random graph to contain more than two vertices of degree at most one. Komlós and Szemerédi [64] and Korshunov [65] were the first to show that this bound is tight, i.e., $G(n, p)$ is a.a.s. hamiltonian for every $p \geq \frac{\log n + \log\log n + \omega(1)}{n}$.

We could interpret the whole question as a search for the one edge in the random graph process that makes the graph universal for some prescribed class of bounded degree spanning trees, or creates a copy of a given bounded degree spanning tree. In this context, we could reformulate the above results in terms of the random graph model $G(n, M)$, where the graph is chosen uniformly at random among all n-vertex graphs with exactly M edges. Using a known equivalence between the two random graph models (see e.g. [14]), the results from [64] and [65] on the threshold probability for Hamiltonicity can be seen as follows. Let $\pi = e_1, \dots, e_{\binom{n}{2}}$ be a random ordering of the edges of the complete graph on the vertex set $[n]$, and let $G_M = ([n], \{e_1, \dots, e_M\})$ be the graph on the vertex set $[n]$ with the edge set consisting of the initial M edges of π. Clearly, $G_M \sim G(n, M)$. Hence, [64] and [65] state that The Edge (the one that makes the random graph G_M hamiltonian) has index at most $(\log n + \log\log n + \omega(1))n$ a.a.s., whereas its index is more than $(\log n + \log\log n - \omega(1))n$ a.a.s. by the argument on the number of vertices of degree at most one. From this point of view, the achievement of the above result was to determine an interval of length around n in a random ordering of the edges such that The Edge lies in this interval a.a.s.

We have already seen that this length of the interval is best possible if we are aiming to determine The Edge only by its index. However, this is not necessary, and it is similarly interesting to find The Edge e.g. by the number of vertices of degree at most one in the graph G_M. Formally, for an $i < n$, we are searching for the smallest integer M such that $G^{(i)} := G_M$ contains at most i vertices of degree at most one, but G_{M-1} contains at least $i+1$ of them. Bollobás [13] and independently Ajtai, Komlós, and Szemerédi [1] proved the hitting time version of the statement from [64] and [65] for the Hamilton cycle, showing that in the random graph process, the very edge that increases the minimum degree to two also makes the graph hamiltonian a.a.s. Formally, they showed that $G^{(0)}$ is a.a.s. hamiltonian. In the light of the previous discussion, this decreases the length of the "interval of uncertainness" from something of order at least n to one! This is an big improvement of the previous results from [64] and [65].

1.2 Main results

The results from [64], [65], [13], and [1] can be seen as a common motivation for most of the results presented in this thesis. We know now perfectly well when the random graph becomes hamiltonian. An intuitive (and informal) question to ask now is: *how* hamiltonian is the random graph, once it is hamiltonian? Probably the best known way to formalize this question among those approaches that are not considered in this thesis is the concept of *resilience* (see e.g. [88], [39], [11] and [73] for results on resilience with respect to Hamiltonicity). We, however, concentrate on the following interpretations of the question. In Chapter 2, we estimate the number of Hamilton cycles in the random graph and observe that it is relatively close to the expected number of Hamilton cycles in the considered range of edge probability. The results of this section are based on joint work with Michael Krivelevich [44]. In Chapter 3, we are interested in the smallest number of Hamilton cycles needed to cover all edges of the random graph. According to a joint project with Michael Krivelevich and Tibor Szabó [45], we show that this number is a.a.s. close to being optimal in the sense that a.a.s. this number asymptotically equals the maximum degree divided by two. And in Chapter 4, we investigate the game theoretic approach. We consider a game between two players, Maker and Breaker, on the edge set of a graph G. In every move, first Breaker claims b previously unclaimed edges of G, and then Maker claims one. Maker wins, if at some moment there exists a Hamilton cycle in the graph induced by his edges, otherwise Breaker wins. The largest bias b such that Maker still has a winning strategy can be seen as a measure of Hamiltonicity of G. Jointly with Asaf Ferber, Michael Krivelevich, and Alon Naor, we show in [31] that for every $\varepsilon > 0$ and $p = \omega(\log n/n)$, in the game played on the edge set of $G \sim G(n,p)$, Maker a.a.s. has a winning strategy if $b \leq (1-\varepsilon)np/\log n$, and Breaker a.a.s. has a winning strategy if $b \geq (1+\varepsilon)np/\log n$.

An other direction to generalize the results from [64], [65], [13], and [1] is to consider other spanning structures. Similarly to these results, with a light increase of technicalities one can show that in the random graph process, the edge that leaves only two vertices of degree at most one also creates the first Hamilton path a.a.s., that is, $G^{(2)}$ contains a Hamilton path a.a.s. In Chapter 5 we generalize this and obtain hitting time thresholds for the appearance of bounded degree spanning trees with linearly long bare paths or linearly many leaves. The results are based on joint work in preparation with Daniel Johannsen and Michael Krivelevich [43].

Finally, in Chapter 6 we consider Hamiltonicity in hypergraphs. There exist several definitions of a Hamilton cycle in hypergraphs, we follow the definition established by Katona and Kierstead in [60]. A Hamilton cycle for us is a spanning subhypergraph whose vertices can be cyclically ordered in such a way that the edges are segments of that ordering and every two consecutive edges intersect in the same number of vertices. Although this concept recently became more and more popular in research, we are not only missing the understanding of the random hypergraph process concerning the Hamilton cycle (for the recent developments, see e.g. [36], [24], [?], and [25]), but even the equivalent statements of the most basic graph theoretic Hamiltonicity results such as the theorem of Ore [76] and the theorem of Dirac [22] are not known. Aiming to fill this gap a bit more, we obtain both Turán- and Dirac-type results. While the Turán-type result gives an exact threshold for the appearance of a Hamilton cycle in a hypergraph depending only on the extremal number of a certain path, the Dirac-type result yields a sufficient condition relying solely on the minimum vertex degree. The corresponding results were obtained jointly with Yury Person and Wilma Weps in [40].

Our goal is to make the result from each chapter understandable for the interested reader without referring to other chapters, except for the general introduction. For this sake, we may prove similar statements in different chapters, trying to minimize the number of cross-references. Our notation also varies slightly depending mostly on the convenience in the corresponding settings (e.g. because of the expansion properties used in Chapter 2 and Chapter 3, it is convenient to use the concept of external neighborhood of a set; however, for the technical details in Section 5.2, it is simply not sufficient, hence in Chapter 5, the set is not by definition excluded from its neighborhood).

1.2.1 On the number of Hamilton cycles in sparse random graphs

The goal of Chapter 2 is to estimate the number of Hamilton cycles in the random graph $G(n,p)$. To be more formal, we show that the number of Hamilton cycle is asymptotically almost surely, or a.a.s. for brevity, concentrated around the expectation up to a factor $(1+o(1))^n$, provided the minimum degree is at least 2.

There exists a rich literature about Hamiltonicity of $G(n,p)$. Recent results include packing and covering problems (see e.g. [38], [62], [61], [69], [45], and [54]), local resilience (see e.g. [88], [39], [11], and [73]), and Maker-Breaker games ([87], [51], [9], and [31]). In Chapter 2, we are interested in estimating the typical number of Hamilton cycles in a random graph when it is a.a.s. hamiltonian. Several recent results about Hamiltonicity ([62],[69], [88],[31]) can be used to show fairly easily that $G(n,p)$ with $p=p(n)$ above the threshold for Hamiltonicity contains typically many, or even exponentially many Hamilton cycles. Here we aim however for (relatively) accurate bounds.

Using linearity of expectation we immediately see that the expected value of the number of Hamilton cycles in $G(n,p)$ is $\frac{(n-1)!}{2}p^n$. As the common intuition for random graphs may suggest, we expect the random variable to be concentrated around its mean, perhaps after some normalization (it is easy to see that the above expressions for the expectation become exponentially large in n already for p inverse linear in n).

The reality appears to confirm this intuition – to a certain extent. Denoting by X the number of Hamilton cycles in $G(n,p)$, we immediately obtain $X<\left(\frac{np}{e}\right)^n$ a.a.s. by Markov's inequality. Janson [56] considered the distribution of X for $p=\Omega\left(1/\sqrt{n}\right)$ and proved that X is log-normal distributed, implying that $X=\left(\frac{np}{e}\right)^n(1+o(1))^n$ a.a.s. It is instructive to observe that assuming $p=o(1)$, the distribution of X is in fact concentrated way below its expectation, in particular implying that $X/\mathbb{E}(X)\xrightarrow{p}0$. For random graphs of density $p=o\left(n^{-1/2}\right)$ not much appears to be known about the asymptotic behavior of the number of Hamilton cycles in corresponding random graphs. We nevertheless mention the result of Cooper and Frieze [20], who proved that in the random graph process typically at the very moment the minimum degree becomes two, not only the graph is hamiltonian but it has $(\log n)^{(1-o(1))n}$ Hamilton cycles.

Our main result is the following theorem, which can be interpreted as an extension of Janson's results [56] to the full range of $p(n)$.

Theorem 1.2.1. Let $G \sim G(n,p)$ with $p \geq \frac{\log n + \log\log n + \omega(1)}{n}$. Then the number of Hamilton cycles is $n!p^n(1-o(1))^n$ a.a.s.

Improving the main result of [20], we also show the following statement.

Theorem 1.2.2. In the random graph process, at the very moment the minimum degree becomes two, the number of Hamilton cycles becomes $(\log n/e)^n(1-o(1))^n$ a.a.s.

We continue with a short overview of related results for other models of random and pseudorandom graphs. For the model $G(n,M)$ of random graphs with n vertices and M edges, notice the result of Janson [56] showing in particular that for the regime $n^{3/2} \ll M \leq 0.99\binom{n}{2}$, the number of Hamilton cycles is indeed concentrated around its expectation. The situation appears to change around $M = \Theta\left(n^{3/2}\right)$, where the asymptotic distribution becomes log-normal instead. Notice also that the number of Hamilton cycles is more concentrated in $G(n,M)$ compared to $G(n,p)$; this is not surprising as $G(n,M)$ is obtained from $G(n,p)$ by conditioning on the number of edges of G being exactly equal to M, resulting in reducing the variance.

For the probability space of random regular graphs, it is the opposite case of very sparse graphs that is relatively well understood. Janson [57], following the previous work of Robinson and Wormald [79], [80], described the asymptotic distribution of the number of Hamilton cycles in a random d-regular graph $G(n,d)$ for a *constant* $d \geq 3$. The expression obtained is quite complicated, and we will not reproduce it here. For the case of growing degree $d = d(n)$, the result of Krivelevich [68] on the number of Hamilton cycles in (n,d,λ)-graphs in addition to known eigenvalue results for $G_{n,d}$ imply an estimation on the number of Hamilton cycles in $G_{n,d}$ with a superpolylogarithmic lower bound on d.

For an overview of these results as well as of the corresponding results in pseudorandom settings, we refer the interested reader to [68].

1.2.2 On covering expander graphs by Hamilton cycles

For an r-uniform hypergraph G and a family $\mathcal{F}$ of its subgraphs, we call a family $\mathcal{F}' \subset \mathcal{F}$ an *$\mathcal{F}$-decomposition* of G if every edge of G is contained in *exactly* one of the hypergraphs from $\mathcal{F}'$. We call a family $\mathcal{F}' \subset \mathcal{F}$ an *$\mathcal{F}$-packing* of G, if every edge of G is contained in *at most* one of the hypergraphs from $\mathcal{F}'$. Naturally, one tries to *maximize* the size of an $\mathcal{F}$-packing of G. The dual concept is that of an $\mathcal{F}$-covering: a family $\mathcal{F}'$ is called an *$\mathcal{F}$-covering* of G, if every edge of G is contained in *at least* one of the hypergraphs from $\mathcal{F}'$. Here the *minimum* size of an $\mathcal{F}$-covering of G is sought.

Decompositions, packings and coverings are in the core of combinatorial research (see [41] for a survey). One of the most famous problems in this area was the conjecture of Erdős and Hanani [27], dealing with the case when G is the complete r-uniform hypergraph on n vertices and $\mathcal{F}$ is the family of all k-cliques in G for some $k \geq r$. Clearly, if there was an $\mathcal{F}$-decomposition of G, its size would be $\binom{n}{r}/\binom{k}{r}$. Hence $\binom{n}{r}/\binom{k}{r}$ is an upper bound on the size of a largest $\mathcal{F}$-packing of G and a lower bound on the size of a smallest $\mathcal{F}$-covering of G. Erdős and Hanani conjectured both inequalities to be asymptotically tight for constant r and k, i.e., that the size of a largest $\mathcal{F}$-packing of G and the size of a smallest $\mathcal{F}$-covering of G are asymptotically equal to each other. Rödl [81] verified the conjecture by one of the first applications of the nibble method. Observe that in this setting, the two parts of the conjecture are trivially equivalent. The reason for this is that the size of the elements of the family $\mathcal{F}$ does not grow with n: from a packing $\mathcal{F}'$ of G of size $(1-\varepsilon)\binom{n}{r}/\binom{k}{r}$, one obtains a covering $\mathcal{F}''$ of G of size $(1+\varepsilon\binom{k}{r})\binom{n}{r}/\binom{k}{r}$ by simply taking additionally one k-clique for every r-edge that was not contained in any clique from $\mathcal{F}'$.

In Chapter 3 we study a covering problem where the sets in $\mathcal{F}$ grow with n and the above equivalence is not entirely clear. Let $r = 2$, so our objects are usual graphs. For a graph G, we consider the family $\mathcal{H} = \mathcal{H}(G)$ of all Hamilton cycles of G. The corresponding concepts of decomposition, packing, and covering are called *Hamilton decomposition*, *Hamilton packing* and *Hamilton covering*, respectively.

The most well-known fact about Hamilton decompositions is the nearly folklore result of Walecki (see e.g. [5]), stating that for every odd n, the complete graph K_n has a Hamilton decomposition. In general, however, not many graphs are known to have a Hamilton decomposition; the interested reader is referred to [6].

Given that the minimum degree of a Hamilton cycle is 2, the maximum size of a Hamilton packing of a graph with minimum degree δ is $\lfloor \delta/2 \rfloor$. Interestingly, the random graph $G(n,p)$ seems to match this bound tightly. There has been an extensive research considering Hamilton packings of the random graph $G(n,p)$. A classic result of Bollobás [13] and Komlós and Szemerédi [64] states that as soon as the minimum degree of the random graph is 2, it contains a Hamilton cycle a.a.s. This result was extended by Bollobás and Frieze [16] who showed that we can replace 2 by $2k$ for any constant k and obtain a Hamilton packing of size k a.a.s. Frieze and Krivelevich [39] proved that for every constant and slightly subconstant p, $G \sim G(n,p)$ contains a packing of $(1+o(1))\delta(G)/2$ Hamilton cycles a.a.s. They also conjectured that for every $p = p(n)$ there exists a Hamilton packing of $G \sim G(n,p)$ of size $\lfloor \delta(G)/2 \rfloor$ (and, in the case where $\delta(G)$ is odd, an additional (disjoint to the Hamilton cycles of the packing) matching of size $\lfloor n/2 \rfloor$) a.a.s. Frieze and Krivelevich [38] proved their conjecture as long as $p = (1+o(1))\log n/n$, which was extended to the range of $p \leq 1.02 \log n/n$ by Ben-Shimon, Krivelevich and Sudakov [11]. Meanwhile, Knox, Kühn and Osthus [62] extended the result from [39] to the range of $p = \omega(\log n/n)$, and then proved the conjecture for $\log^{50} n/n < p < 1 - n^{-1/4}\log^9 n$ [61]. Very recently it was also proven by Krivelevich and Samotij [69] that there exists a positive constant $\varepsilon > 0$ such that for the range of $\log n/n \leq p \leq n^{\varepsilon-1}$, $G \sim G(n,p)$ contains a Hamilton packing of size $\lfloor \delta(G)/2 \rfloor$ a.a.s., implying the conjecture in this range of p up to the existence of the additional matching.

To the best of our knowledge the dual concept of Hamilton covering of $G(n,p)$ has not been studied. Obviously, the size of any Hamilton covering of graph G is at least $\lceil \Delta(G)/2 \rceil$, where $\Delta(G)$ denotes the maximum degree of G. Recall that for $p = p(n) \gg \log n/n$, $\Delta(G(n,p)) = (1+o(1))np = \delta(G(n,p))$ a.a.s., hence the minimum size of a Hamilton cover and the maximum size of a Hamilton packing have a chance to be asymptotically equal. We prove that this, in fact, is the case for the range $p > n^{\alpha-1}$ where $\alpha > 0$ is an arbitrary small constant.

Theorem 1.2.3. For any $\alpha > 0$, for $p \geq n^{\alpha-1}$ a.a.s. $G(n,p)$ can be covered by $(1+o(1))np/2$ Hamilton cycles.

1.2.3 Biased games on random boards

In Chapter 4 we consider Maker-Breaker games, played on the edge set of a random graph $G \sim G(n,p)$.

Let $\mathcal{F} \subseteq 2^X$ be any hypergraph. In an (a,b) Maker-Breaker game $\mathcal{F}$, the two players are called Maker and Breaker, alternately claim a and b previously unclaimed elements of *the board* X, respectively. Maker's goal is to claim all the elements of some *target set* $F \in \mathcal{F}$. If Maker does not fully claim any target set by the time all board elements are claimed, then Breaker wins the game. The most basic case, where $a = b = 1$, is called the *unbiased* game. Any other case is called a *biased* game. Since being the first player is never a disadvantage in a Maker-Breaker game, in order to prove that Maker wins a certain game, it is enough to prove that he can win as a second player. Hence, throughout Chapter 4 we assume that Maker is the *second* player to move. We may also assume that there are no $F_1, F_2 \in \mathcal{F}$ such that $F_1 \subset F_2$, since in this case Maker wins once he claims all the elements in F_1, and so the two (a,b) games $\mathcal{F}$ and $\mathcal{F} \setminus \{F_2\}$ are identical.

It is natural to play positional games on the edge set of a graph G. In this case, the board is $X = E(G)$ and the target sets are all the edge sets of subgraphs $H \subseteq G$ which possess some given graph property $\mathcal{P}$. In the *connectivity game*, Maker wins if and only if his edges contain a spanning tree. In the *perfect matching* game $\mathcal{M}_n(G)$ the winning sets are all sets of $\lfloor n/2 \rfloor$ independent edges of G. Note that if n is odd, then such a matching covers all vertices of G but one. In the Hamiltonicity game $\mathcal{H}_n(G)$ the winning sets are all edge sets of Hamilton cycles of G. Given a positive integer k, in the k-connectivity game $\mathcal{C}_n^k(G)$ the winning sets are all edge sets of k-vertex-connected spanning subgraphs of G.

Maker-Breaker games played on the edge set of the complete graph K_n are well studied. In this case, many natural unbiased games are drastically in a favor of Maker (see e.g. [32], [50], [55], and [74]). Hence, in order to even out the odds, it is natural to give Breaker more power by increasing his *bias* (that is, to play a $(1, b)$ game instead of a $(1, 1)$ game), and/or to play on different types of boards.

Maker-Breaker games are *bias monotone*. That means that if Maker wins some game with bias (a, b), he also wins this game with bias (a', b'), for every $a' \geq a$, $b' \leq b$. Similarly, if Breaker wins a game with bias (a, b), he also wins this game with bias (a', b'), for every $a' \leq a$, $b' \geq b$. This bias monotonicity allows us to define the *critical bias* (also referred to as the *threshold bias*): for a given game $\mathcal{F}$, the critical bias b^* is the value for which Maker wins the game $\mathcal{F}$ with bias $(1, b)$ for every $b < b^*$, and Breaker wins the game $\mathcal{F}$ with bias $(1, b)$ for every $b \geq b^*$.

In their seminal paper [18], Chvatál and Erdős proved that playing the $(1, b)$ connectivity game on the edge set of the complete graph K_n, for every $\varepsilon > 0$, Breaker wins for every $b \geq \frac{(1+\varepsilon)n}{\log n}$, and Maker wins for every $b \leq \frac{n}{(4+\varepsilon)\log n}$. They conjectured that $b = \frac{n}{\log n}$ is (asymptotically) the threshold bias for this game. Gebauer and Szabó proved in [42] that this is indeed the case. Later on, Krivelevich proved in [67] that $b = \frac{n}{\log n}$ is also the threshold bias for the Hamiltonicity game.

Stojaković and Szabó suggested in [87] to play Maker-Breaker games on the edge set of a random board $G \sim G(n, p)$. They examined some games on this board such as the connectivity game, the perfect matching game, the Hamiltonicity game and building a k-clique game. Since then, much progress has been made in understanding Maker-Breaker games played on $G \sim G(n, p)$. For example, it was proved in [9] that for $p = \frac{(1+o(1))\log n}{n}$, $G \sim G(n, p)$ is typically such that Maker wins the $(1, 1)$ games $\mathcal{M}(G)$, $\mathcal{H}(G)$ and $\mathcal{C}_k(G)$. Moreover, the proofs in [9] are of a "hitting time" type. That means that, in the random graph process (see e.g. [58]), typically at the moment the graph reaches the needed minimum degree for Maker to win the desired game, Maker indeed win this game. Later on, in [**?**] fast winning strategies for Maker in various games played on $G \sim G(n, p)$ were considered, and in [89] a hitting time result was established for the "building a triangle" game, and it was proved that the threshold probability for the (monotone) property "Maker can build a k-clique" game is $p = \Theta\left(n^{-2/(k+1)}\right)$.

In [87], Stojaković and Szabó conjectured the following:

Conjecture 1.2.4 (Conjecture 1 in [87]**).** There exists a constant C such that for every $p \geq \frac{C \log n}{n}$, $G \sim G(n, p)$ is typically such that the threshold bias for the game $\mathcal{H}(G)$ is $b = \Theta\left(\frac{np}{\log n}\right)$.

In Chapter 4 we prove Conjecture 1.2.4, and in fact, for $p = \omega\left(\frac{\log n}{n}\right)$ we prove the following stronger statement:

Theorem 1.2.5. *Let $p = \omega\left(\frac{\log n}{n}\right)$. Then $G \sim G(n, p)$ is typically such that $\frac{np}{\log n}$ is the asymptotic threshold bias for the games $\mathcal{M}(G)$, $\mathcal{H}(G)$ and $\mathcal{C}_k(G)$.*

In order to prove Theorem 1.2.5 we prove the following two theorems:

Theorem 1.2.6. *Let $0 \leq p \leq 1$, $\varepsilon > 0$ and $b \geq (1 + \varepsilon)\frac{np}{\log n}$. Then $G \sim G(n, p)$ is typically such that in the $(1, b)$ Maker-Breaker game played on $E(G)$, Breaker has a strategy to isolate a vertex in Maker's graph, as a first or a second player.*

Theorem 1.2.7. *Let $p = \omega\left(\frac{\log n}{n}\right)$, $\varepsilon > 0$ and $b = (1-\varepsilon)\frac{np}{\log n}$. Then $G \sim G(n,p)$ is typically such that Maker has a winning strategy in the $(1,b)$ games $\mathcal{M}(G)$, $\mathcal{H}(G)$ and $\mathcal{C}_k(G)$ for a fixed positive integer k.*

In the case $p = \Theta\left(\frac{\log n}{n}\right)$ we establish two non-trivial bounds for the critical bias b^*. This also settles Conjecture 1.2.4 for this case but does not determine the exact value of b^* (notice that in this case, b^* is a constant!).

Theorem 1.2.8. *Let $p = \frac{c \log n}{n}$, where $c > 1600$ and let $\varepsilon > 0$. Then $G \sim G(n,p)$ is typically such that the threshold bias for the games $\mathcal{M}(G)$, $\mathcal{H}(G)$ and $\mathcal{C}_k(G)$ lies between $c/10$ and $c+\varepsilon$.*

Remark: In the terms of Theorem 1.2.8, if $1 < c \leq 1600$, we get by Theorem 1.2.6 that $b^* \leq c + \varepsilon$, and by the main result of [10] that $b^* > 1$, so indeed $b^* = \Theta(\frac{np}{\log n})$ in this case as well.

1.2.4 Hitting time appearance of certain spanning trees in the random graph process

In Chapter 5 we approach the question about universality of the binomial random graph for a special class of bounded degree spanning trees. The trees we deal with have either a long (almost linear) bare path or almost linearly many leaves. We show a hitting time statement, generalizing the famous hitting time result on the appearance of the first Hamilton path in the random graph process.

Given thebreakthrough knowledge about the behavior of the random graph process with respect to the Hamilton path, one has every reason to wonder if it can be generalized to other spanning trees. Given that the maximum degree of $G(n,p)$ in the interesting range of $p = \Theta(\log n)$ is of order $\log n$ a.a.s., we naturally restrict the considered trees to have bounded maximum degree. (In fact, in our results, the maximum degree might grow slightly sub-logarithmic in n). In this setting, the problem of *nearly* spanning bounded degree trees (where by nearly spanning we mean trees on $(1-\varepsilon)n$ many vertices for some $\varepsilon > 0$) is very well studied, see, e.g., [33], [2], [35], [34], [49], and recently [7]. In particular, Theorem 4 in [7] states that for every $\varepsilon > 0$ and $d \geq 2$, there exists a $c > 0$ such that the random graph $G(n, c/n)$ is a.a.s. *universal* for the class of bounded degree $(1-\varepsilon)n$-vertex trees with maximum degree at most d, i.e., a.a.s. $G(n, c/n)$ contains all of them. Notice that their bound on c is close to being linear in d and inverse linear in ε, significantly improving the previous bounds from [2].

However, the picture changes drastically when we switch back to *spanning* trees. To our best knowledge, after the above mentioned well-known result on the Hamilton path, Alon, Krivelevich and Sudakov [2] were the first to consider a class of bounded degree spanning trees with respect to their appearance in the random graph process. Namely, they observed that if an n-vertex tree has a linear (in n) number of leaves, then a.a.s. it is contained in $G(n, C\log n/n)$ for some sufficiently large $C > 0$; the proof is not that hard and utilizes the embedding result for nearly spanning trees from the same paper.

A substantial step forward in solving this class of problems was made by the third author in [66]. He showed in Theorem 1 that for every $\varepsilon > 0$ and every n-vertex tree with maximum degree much smaller than $n^{\varepsilon}/\log n$, a.a.s. this tree is contained in the random graph $G(n, n^{\varepsilon-1})$.

However, since the number of bounded degree n-vertex trees grows with n (in fact, it grows exponentially), the above result does not imply the universality of $G(n, n^{\varepsilon-1})$ for the class of all bounded degree spanning trees. This question was addressed by the two last authors and Samotij in [59], who showed as a consequence of their Theorem 2.3 the existence of a constant $c > 0$ such that a.a.s. the random graph $G(n, cn^{-1/3}\log^2 n)$ is universal for the class of all bounded degree spanning trees. The probability $p = cn^{-1/3}\log^2 n$, however, in the light of the previous results and some intuition obtained from the strong result on the Hamilton path seems to be far from the correct threshold probability; for further speculations, the reader is referred to Section 5.5 of this paper.

To improve the estimations from [66] and [59] on the threshold for the probability p, Hefetz, Krivelevich, and Szabó [53] considered special classes of bounded degree spanning trees. They proved that for every $\varepsilon > 0$, the random graph $G(n,(1+\varepsilon)\log n/n)$ is a.a.s. universal for the class of bounded degree spanning trees with a linearly long bare path. Furthermore, they showed for every bounded degree spanning tree with linearly many leaves that it is a.a.s. contained in $G(n,(1+\varepsilon)\log n/n)$; notice that this statement is not universal. The probability $p = (1+\varepsilon)\log n/n$ is tight in the sense that, as we observed earlier, a.a.s. $G(n,(1-\varepsilon)\log n/n)$ is not connected, and thus contains no n-vertex tree as a subgraph.

We could interpret the whole question as a search for the one edge in the random graph process that makes the graph universal for some prescribed class of bounded degree spanning trees, or creates a copy of a given bounded degree spanning tree. In this context, we could reformulate the above results in terms of the random graph model $G(n,M)$, where the graph is chosen uniformly at random among all n-vertex graphs with exactly M edges. Using a known equivalence between the two random graph models (see, e.g., [14]), the results from [64] and [65] on the threshold probability for Hamiltonicity can be seen as follows. Let $\pi = e_1,\dots,e_{\binom{n}{2}}$ be a random ordering of the edges of the complete graph on the vertex set $[n]$, and let $G_M = ([n],\{e_1,\dots,e_M\})$ be the graph on the vertex set $[n]$ with the edge set consisting of the initial M edges of π. Clearly, $G_M \sim G(n,M)$. Hence, [64] and [65] state that The Edge (the one that makes the random graph G_M Hamiltonian) has index at most $(\log n + \log\log n + \omega(1))n$ a.a.s., whereas its index is more than $(\log n + \log\log n - \omega(1))n$ a.a.s. by the argument on the number of vertices of degree at most one. From this point of view, the achievement of the above result was to determine an interval of length around n in a random ordering of the edges such that The Edge lies in this interval a.a.s.

We have already seen that this length of the interval is best possible if we are aiming to determine The Edge only by its index. However, this is not necessary, and it is similarly interesting to find The Edge considering, e.g., the number of vertices of degree at most one in the graph G_M. Formally, for an $i < n$, we are searching for the smallest integer M such that $G^{(i)} := G_M$ contains at most i vertices of degree at most one, but G_{M-1} contains at least $i+1$ of them. Bollobás [13] and independently Ajtai, Komlós, and Szemerédi [1] proved the hitting time version of the statement from [64] and [65] for the Hamilton cycle, showing that in the random graph process, the very edge that increases the minimum degree to two also makes the graph Hamiltonian a.a.s. Formally, they showed that $G^{(0)}$ is a.a.s. Hamiltonian. Somewhat similarly with a light increase of technicalities one can show that in the random graph process, the edge that leaves only two vertices of degree at most one also creates the first Hamilton path a.a.s., that is, $G^{(2)}$ contains a Hamilton path a.a.s. In the light of the previous discussion, this decreases the length of the uncertainness interval from something of order at least n to one! This is an big improvement of the previous results from [64] and [65].

In general, one might wonder if this hitting time statement is true for arbitrary bounded degree spanning trees, i.e., if for an n-vertex tree T with exactly k leaves, the random graph $G^{(k)}$ a.a.s. contains a copy of T (or $G^{(k)}$ is not connected). In fact, as far as we are concerned with the degree sequences of T and $G^{(k)}$, it is not hard to see that the degree sequence of T is a.a.s. smaller or equal than the degree sequence of $G^{(k)}$ in every coordinate. However, a relatively simple example shows that the above question can be answered negatively. Let T be the n-vertex tree consisting of a path of length $n-1$ and one additional vertex joined to the neighbor of an endpoint of the path. This tree has three leaves. However, as we will see in Property $(P2)$ of Lemma 5.1.1, a.a.s. no two of the three vertices of degree one in $G^{(3)}$ have a common neighbor, making it impossible to embed T into $G^{(3)}$.

This example provides some intuition for the simple fact that for a "hardcore" hitting time statement, we need a set of leaves in the tree that are sufficiently far from each other. In the following theorem, this is only interpreted in terms of pairwise distance of leaves; a more technical Theorem 5.3.3 is presented in Section 5.3.

Our aim is to show the hitting time analogs of the statements from [53] as well as to make the statement on the trees with linearly many leaves universal. For the trees with a long bare path, we obtain the following result.

Theorem 1.2.9. *In the random graph process on n vertices, a.a.s. for every $k < n$, if $G^{(k)}$ is connected, then it contains every spanning tree of maximum degree at most $\frac{\log n}{2\log\log n \log\log\log n}$ containing a k-set of leaves of pairwise distances at least $(4.2\log n/\log\log n)$ and a bare path of length at least $\frac{23n}{\log\log\log n}$.*

Theorem 1.2.9 is tight in the sense of the following observation.

Observation 1.2.10. Let $k, \Delta \in \mathbb{N}$ and let T be an n-vertex tree with maximum degree at most Δ containing no k-set of leaves with pairwise distances at least $\frac{2\log n}{3\log\log n}$. Then a.a.s. in the random graph process on n vertices, for every $j > k$ the graph $G^{(j)}$ does not contain T.

For trees with linearly many leaves, we prove the following.

Theorem 1.2.11. *Let $\varepsilon > 0$ and $\Delta \in \mathbb{N}$, and let M^* be the random variable denoting the smallest integer M such that in the random graph process on n vertices, G_M is connected. Then, G_{M^*} a.a.s. contains a copy of every n-vertex tree of maximum degree at most Δ, provided the tree has at least εn leaves.*

1.2.5 On extremal hypergraphs for Hamilton cycles

For a fixed graph G and an integer n the *extremal number* $\mathrm{ex}\,(n, G)$ of G is the largest integer m such that there exists a graph on n vertices with m edges that does not contain a subgraph isomorphic to G. The corresponding graphs are called *extremal graphs*. Naturally, one can extend this definition to a forbidden spanning structure, e.g. a Hamilton cycle. In [76] Ore proved that a non-hamiltonian graph on n vertices has at most $\binom{n-1}{2} + 1$ edges, and further, that the unique extremal example is given by an $(n-1)$-clique and a vertex of degree one that is adjacent to one vertex of the clique.

A *k-uniform hypergraph H*, or k-graph for short, is a pair (V, E) with a vertex set $V = V(H)$ and an edge set $E = E(H) \subseteq \binom{V}{k}$. Since in Chapter 6 we always deal with k-graphs, and the usual 2-uniform graphs have no special meaning for us, we also might use the simplified term *graph* for k-graphs.

There are several definitions of Hamilton cycles in hypergraphs, e.g. Berge Hamilton cycles [12]. Chapter 6 yet follows the definition of Hamilton cycles established by Katona and Kierstead [60] as it has become more and more popular in research.

An *l-tight Hamilton cycle* in H, $0 \le l \le k-1$, $(k-l)\big||V(H)|$, is a spanning sub-k-graph whose vertices can be cyclically ordered in such a way that the edges are segments of that ordering and every two consecutive edges intersect in exactly l vertices. More formally, it is a graph isomorphic to $([n], E)$ with

$$E = \left\{\{i(k-l)+1, i(k-l)+2, \dots, i(k-l)+k\} :\ 0 \le i < \frac{n}{k-l}\right\},$$

where addition is made modulo n. We denote an l-tight Hamilton cycle in a k-graph H on n vertices by $C_n^{(k,l)}$, and call it *tight* if it is $(k-1)$-tight.

Working on her thesis [92] in coding theory, Woitas raised the question whether removing $\binom{n-1}{2} - 1$ edges from a complete 3-uniform hypergraph on n vertices leaves a hypergraph containing a 1-tight Hamilton cycle. A generalization of this problem is to estimate the extremal number of Hamilton cycles in k-graphs.

Katona and Kierstead were the first to study sufficient conditions for the appearance of a $C_n^{(k,k-1)}$ in k-graphs. In [60] they showed that for all integers k and n with $2 \le k$ and $2k-1 \le n$,

$$\operatorname{ex}(n, C_n^{(k,k-1)}) \ge \binom{n-1}{k} + \binom{n-2}{k-2}.$$

In the same paper Katona and Kierstead proved, that this bound is not tight for $k = 3$ by showing that for all integers n and q with $q \ge 2$ and $n = 3q+1$,

$$\operatorname{ex}(n, C_n^{(3,2)}) \ge \binom{n-1}{3} + n - 1.$$

In [90] Tuza gave a construction for general k and tight Hamilton cycles, improving the lower bound to

$$\operatorname{ex}(n, C_n^{(k,k-1)}) \ge \binom{n-1}{k} + \binom{n-1}{k-2},$$

if a Steiner system $S(k-2, 2k-3, n-1)$ exists. Also for all k, n and p such that a partial Steiner system $PS(k-2, 2k-3, n-1)$ of order $n-1$ with $p\binom{n-1}{k-2}/\binom{2k-3}{k-2}$ blocks exists, Tuza proved the bound

$$\operatorname{ex}(n, C_n^{(k,k-1)}) \ge \binom{n-1}{k} + p\binom{n-1}{k-2}.$$

An intuitive approach to forbid Hamilton cycles in hypergraphs is to prohibit certain structures in the *link* of one fixed vertex. For a vertex $v \in V$, we define the *link of v in H* to be the $(k-1)$-graph $H(v) = (V\backslash\{v\}, E_v)$ with $\{x_1, \ldots, x_{k-1}\} \in E_v$ iff $\{v, x_1, \ldots, x_{k-1}\} \in E(H)$.

The structure of interest in this case is a generalization of a path for hypergraphs.

An *l-tight k-uniform t-path*, denoted by $P_t^{(k,l)}$, is a k-graph on t vertices, $(k-l) \mid (t-l)$, such that there exists an ordering of the vertices, say $(x_1, \ldots, x_t)$, in such a way that the edges are segments of that ordering and every two consecutive edges intersect in exactly l vertices. Observe that a $P_t^{(k,l)}$ has $\frac{t-l}{k-l}$ edges. A k-uniform $(k-1)$-tight path is called *tight*, and whenever we consider a path we assume it to be tight unless stated otherwise.

For arbitrary k and l we give the exact extremal number and the extremal graphs of l-tight Hamilton cycles in Chapter 6. The extremal number and the extremal graphs rely on the extremal number of $P(k,l) := P^{(k-1,l-1)}_{\lfloor \frac{k}{k-l} \rfloor (k-l)+l-1}$, and its extremal graphs, respectively.

Theorem 1.2.12. For any $k \ge 2, l \in \{0, \ldots, k-1\}$ there exists an n_0 such that for any $n \ge n_0$ and $(k-l)|n$,

$$\operatorname{ex}\left(n, C_n^{(k,l)}\right) = \binom{n-1}{k} + \operatorname{ex}\left(n-1, P(k,l)\right)$$

holds. Furthermore, any extremal graph on n vertices contains an $(n-1)$-clique and a vertex whose link is $P(k,l)$-free.

Notice, that $P(k,l)$ contains $\left\lfloor \frac{k}{k-l} \right\rfloor$ hyperedges.

For $k = 3$ and $l = 1$ Theorem 1.2.12 answers the aforementioned question of Woitas [92] that indeed $\binom{n-1}{3} + 1$ hyperedges ensure an existence of a 1-tight Hamiltionian cycle $C_n^{(3,1)}$ for n large enough.

For $k = 3$ and $l = 2$ Theorem 1.2.12 states that there exists an n_0 such that for any $n \ge n_0$,

$$\begin{aligned}\operatorname{ex}\left(n, C_n^{(3,2)}\right) &= \binom{n-1}{3} + \operatorname{ex}\left(n-1, P_4^{(2,1)}\right)\\ &= \begin{cases} \binom{n-1}{3} + n - 1, & 3 \mid n-1 \\ \binom{n-1}{3} + n - 2, & \text{otherwise.} \end{cases}\end{aligned}$$

Note that this not only goes along with Katona and Kierstead's remark, but further specifies it for the special case $k = 3$.

Actually, in Chapter 6 we prove a stronger statement, namely that with one more edge we find a Hamilton cycle that is l-tight in the neighborhood of one vertex and is $(k-1)$-tight on the rest.

Using the result by Györi, Katona, and Lemons [46] stating that

$$(1+o(1))\binom{n-1}{k-2} \le \mathrm{ex}\left(n-1, P_{2k-2}^{(k-1,k-2)}\right) \le (k-1)\binom{n-1}{k-2},$$

we obtain lower and upper bounds for $l = k-1$:

$$\binom{n-1}{k} + (1+o(1))\binom{n-1}{k-2} \le \mathrm{ex}\left(n, C_n^{(k,l)}\right) \le \binom{n-1}{k} + (k-1)\binom{n-1}{k-2}.$$

Note that the upper bound also holds for $l \ne k-1$.

In our proof we make use of the *absorbing technique* that was originally developed by Rödl, Ruciński and Szemerédi.

Dirac-type Results

The problem of finding Hamilton cycles and perfect matchings in 2-graphs has been studied very intensively. There are plenty beautiful conditions guaranteeing the existence of such cycles, e.g. Dirac's condition [22].

Over the last couple of years several Dirac-type results in hypergraphs were shown, and along with them, different definitions of *degree* in a k-graph were introduced. They all can be captured by the following definition. The *degree* of $\{x_1, \dots, x_i\}$, $1 \le i \le k-1$, in a k-graph H is the number of edges the set is contained in and is denoted by $\deg(x_1, \dots, x_i)$. Let

$$\delta_d(H) := \min\{\deg(x_1, \dots, x_d) | \{x_1, \dots, x_d\} \subset V(H)\}$$

for $0 \le d \le k-1$. If the graph is clear from the context, we omit H and write for short δ_d. Note that $\delta_0 = e(H) := |E(H)|$ and δ_1 is the minimum vertex degree in H.

Following the definitions of Rödl and Ruciński in [82], denote for every d, k, l and n with $0 \le d \le k-1$ and $(k-l)|n$ the number $h_d^l(k,n)$ to be the smallest integer h such that every n-vertex k-graph H satisfying $\delta_d(H) \ge h$ contains an l-tight Hamilton cycle. Observe that $h_0^l(k,n) = \mathrm{ex}\left(n, C_n^{(k,l)}\right) + 1$.

In [60] Katona and Kierstead showed that $h_{k-1}^{k-1}(k,n) \ge \left\lfloor \frac{n-k+3}{2} \right\rfloor$ by giving an extremal construction. Their implicit conjecture that this bound is tight was confirmed for $k = 3$ by Rödl, Ruciński and Szemerédi in [83] asymptotically and in [86] exactly. For $k \ge 4$ the same authors showed in [85] that $h_{k-1}^{k-1}(k,n) \sim \frac{1}{2}n$. Generalizing the results to other tightnesses, Markström and Ruciński proved in [75] that $h_{k-1}^{l}(k,n) \sim \frac{1}{2}n$ if $(k-l)|k,n$. In [71] Kühn, Mycroft and Osthus proved that

$$h_{k-1}^l(k,n) \sim \frac{n}{\left\lceil \frac{k}{k-l} \right\rceil (k-l)}$$

if $k-l$ does not divide k and $(k-l)|n$, proving a conjecture by Hàn and Schacht [48]. For further information, an excellent survey of the recent results can be found in [82].

Rödl and Ruciński conjectured in [82] that for all $1 \le d \le k-1$, $k|n$,

$$h_d^{k-1}(k,n) \sim h_d^0(k,n).$$

Further notice that 0-tight Hamilton cycles $C_n^{(k,0)}$ are perfect matchings covering all vertices. A perfect matching may be considered the "simplest" spanning structure and there are several results about $h_d^0(k,n)$, see e.g. [84], [47], and [72].

Noting the fact that there are virtually no results on $h_d^l(k,n)$ for $d \leq k-2$, Rödl and Ruciński remarked in [82] that it does not even seem completely trivial to show $h_1^2(3,n) \leq c\binom{n-1}{2}$ for some constant $c < 1$. Further, they gave the following bounds

$$\left(\frac{5}{9}+o(1)\right)\binom{n-1}{2} \leq h_1^2(3,n) \leq \left(\frac{11}{12}+o(1)\right)\binom{n-1}{2}.$$

We show the following upper bound on $h_1^{k-1}(k,n)$.

Theorem 1.2.13. For any $k \in \mathbb{N}$ there exists an n_0 such that every k-graph H on $n \geq n_0$ vertices with $\delta_1 \geq \left(1 - \frac{1}{22(1280k^3)^{k-1}}\right)\binom{n-1}{k-1}$ contains a tight Hamilton cycle.

Note that Theorem 1.2.13 implies

$$h_d^l(k,n) \leq \left(1 - \frac{1}{22\left(1280k^3\right)^{k-1}}\right)\binom{n-d}{k-d}$$

for all $l \in \{0,\ldots,k-1\}$ and all $1 \leq d \leq k-1$. This shows that there exists a constant $c < 1$ such that for all l, d

$$h_d^l(k,n) \leq c\binom{n-d}{k-d}$$

holds, although this constant is clearly far from being optimal.

1.3 Chernoff bounds

In many estimations we will have to bound the probability for a random variable to deviate far from its expectation. For this aim, we extensively use Chernoff bounds. We decided to state them explicitly in the following lemma, see e.g. Appendix A of [3].

Lemma 1.3.1. *Let X be a binomially distributed random variable with parameters n and p. Then the following is true.*

- *For every $\varepsilon > 0$ we obtain $\mathbf{Pr}(X > (1+\varepsilon)np) < e^{-\frac{\varepsilon^2 np}{2}+\frac{\varepsilon^3 np}{2}}$.*
- *For every $\varepsilon > 0$ we obtain $\mathbf{Pr}(X < (1-\varepsilon)np) < e^{-\frac{\varepsilon^2 np}{2}}$.*
- *For every $\varepsilon > 0$ there exists a $c = c(\varepsilon) > 0$ such that $\mathbf{Pr}(|X - np| > \varepsilon np) < 2e^{-cnp}$.*
- *For $a > 2np$, $\mathbf{Pr}(X > a) < \left(\frac{enp}{a}\right)^a$.*

2 On the number of Hamilton cycles in sparse random graphs

2.1 Introduction

As we pointed out before, the results of this section are based on joint work with Michael Krivelevich [44].

2.1.1 Definitions and notation

The random oriented graph $\vec{G}(n,p)$ is obtained from $G(n,p)$ by randomly giving an orientation to every edge (every of the two possible directions with probability $1/2$). Notice that whenever we use the notation $\vec{G}$ for an oriented graph, there exists an underlying non-oriented graph obtained by omitting the orientations of the edges of $\vec{G}$; it is denoted by G. Making the notation consistent, when omitting the vector arrow above an oriented graph, we refer to the underlying non-oriented graph.

Given a graph G, we denote by $h(G)$ the number of Hamilton cycles in G. We call a spanning 2-regular subgraph of G a *2-factor*. Notice that every connected component of a 2-factor is a cycle. We denote by $f(G,s)$ the number of 2-factors in G with exactly s cycles. Similarly, a 1-factor of an oriented graph $\vec{G}$ is a spanning 1-regular subgraph, i.e., a spanning subgraph with all in- and outdegrees being exactly one. Analogously, the number of 1-factors in $\vec{G}$ with exactly s cycles is denoted by $f\left(\vec{G},s\right)$. For the purposes of our proofs, we relax the notion of a 2-factor and call a spanning subgraph $H \subseteq G$ an *almost 2-factor* of G if H is a collection of vertex-disjoint cycles and at most $|V(G)|/\log^2(|V(G)|)$ isolated vertices. We denote the number of almost 2-factors of G containing exactly s cycles by $f'(G,s)$. Similarly to the notation for non-oriented graphs, we call an oriented subgraph $\vec{H}$ of $\vec{G}$ an *almost 1-factor* of $\vec{G}$ if $\vec{H}$ is a 1-regular oriented graph on at least $|V(\vec{G})| - |V(\vec{G})|/\log^2(|V(\vec{G})|)$ vertices. The number of almost 1-factors of $\vec{G}$ with exactly s cycles is denoted by $f'\left(\vec{G},s\right)$.

As usual, in a graph G for a vertex $x \in V(G)$ we denote by $d_G(x) := |N_G(x)|$ its *degree*, i.e., the size of its neighborhood. We denote by $\delta(G)$ and respectively $\Delta(G)$ its minimum and maximum degrees. For a set $S \subseteq V(G)$, we denote by $N_G(S)$ the set of all vertices outside S having a neighbor in S. Whenever the underlying graph is clear from the context we might omit the graph from the index. Similarly, in an oriented graph $\vec{G}$ for a vertex $x \in V\left(\vec{G}\right)$ we call $d_{in,\vec{G}}(x) := \left|\left\{y \in V\left(\vec{G}\right) : yx \in E\left(\vec{G}\right)\right\}\right|$ the *indegree* of x and $d_{out,\vec{G}}(x) := \left|\left\{z \in V\left(\vec{G}\right) : xz \in E\left(\vec{G}\right)\right\}\right|$ the *outdegree* of x. We denote by $\delta_{in}\left(\vec{G}\right)$, $\Delta_{in}\left(\vec{G}\right)$, $\delta_{out}\left(\vec{G}\right)$, and $\Delta_{out}\left(\vec{G}\right)$ the minimum and maximum in- and outdegrees of $\vec{G}$.

In a graph G for two sets $A, B \subseteq V(G)$ we denote by $e_G(A,B)$ the number of edges incident with both sets. In an oriented graph $\vec{G}$, for two sets $A, B \subseteq V(G)$ the notation $e_{\vec{G}}(A,B)$ stands for the number of edges going from a vertex in A to a vertex in B. We write $e_G(A) := e_G(A,A)$ and $e_{\vec{G}}(A) := e_{\vec{G}}(A,A)$ for short. Similarly to the degrees, whenever the underlying graph is clear from the context we might omit the graph from the index.

To simplify the presentation, we omit all floor and ceiling signs whenever these are not crucial. Whenever we have a graph on n vertices, we suppose its vertex set to be $[n]$.

2.1.2 Outline of the proofs

In Section 2.2, the lower bounds for Theorems 1.2.1 and 1.2.2 are proven in the following steps.

- In Lemma 2.2.1 we show using the permanent of the incidence matrix that under certain pseudorandom conditions, an oriented graph contains sufficiently many oriented 1-factors.
- In Lemma 2.2.2 we prove that the random oriented graph $\vec{G}(n,p)$ a.a.s. contains a large subgraph with all in- and outdegrees being concentrated around the expected value. This subgraph then satisfies one of the conditions of Lemma 2.2.1.
- In Lemma 2.2.3 we show that the random graph $G(n,p)$ contains many almost 2-factors a.a.s. In the proof, we orient the edges of $G(n,p)$ randomly and apply Lemma 2.2.1 to the subgraph with almost equal degrees whose existence is guaranteed by Lemma 2.2.2 a.a.s.
- In Lemma 2.2.4 we prove that most of these almost 2-factors have few cycles a.a.s.
- We then call a graph p-expander if it satisfies certain expansion properties and show in Lemma 2.2.6 that in the random graph process, the graph $G(n,p)$ has these properties in a strong way.
- Lemma 2.2.7 shows that in any graph having the p-expander properties and minimum degree 2, for any path P_0 and its endpoint v_1 many other endpoints can be created by a small number of rotations with fixed endpoint v_1.
- Lemma 2.2.8 contains the main technical statement of this chapter. It states that in a graph satisfying certain pseudorandom conditions, for almost every almost 2-factor F with few components, there exists a Hamilton cycle with a small Hamming distance from F. The proof is a straightforward use of Lemma 2.2.7.
- The proofs of Theorems 1.2.1 and 1.2.2 are completed with a double counting argument. On the one hand, by Lemma 2.2.4 there exist many almost 2-factors with few cycles a.a.s. Furthermore, for each of these almost 2-factors there exists a Hamilton cycle with small Hamming distance from it a.a.s. by Lemma 2.2.8. On the other hand, for each Hamilton cycle, there are not many almost 2-factors with few cycles having a small Hamming distance from it. Hence, the number of Hamilton cycles is strongly related to the number of almost 2-factors with few cycles, finishing the proof.

2.2 The proofs

Let $G \sim G(n,p)$. Since $\mathbb{E}(h(G)) = (n-1)!p^n/2$, we obtain

$$h(G) < \log n(n-1)!p^n/2 < \left(\frac{np}{e}\right)^n$$

a.a.s., using just Markov's inequality. Thus, for the remainder of the section we are only interested in the lower bound on the typical number of Hamilton cycles in the random graph.

We know from [62] and using e.g. the results from [61] and [69] that in $G \sim G(n,p)$ there are at least $\left(\frac{\lfloor \delta(G)/2 \rfloor}{n}\right)^n n!$ 2-factors a.a.s. We now want to give an a.a.s. lower bound on the number of 2-factors in G, and we want to do it within a multiplicative error term of at most $2^{o(n)}$ from the "truth", basically deleting the 2 from the denominator in the above expression in the case $p \gg \log n/n$, and replacing the term $\lfloor \delta(G)/2 \rfloor$ by asymptotically np.

We first prove a pseudo-random technical statement that will give us the desired inequality once we show that G (or a large subgraph of it) satisfies the pseudo-random conditions. The proof is based on the permanent method as used in [37].

Lemma 2.2.1. *Let $r = r(n) = \omega(\log\log n)$, and let $\vec{G}$ be an oriented graph on n vertices satisfying the following (pseudo-random) conditions:*

- $\delta_{in}(\vec{G}), \delta_{out}(\vec{G}), \Delta_{in}(\vec{G}), \Delta_{out}(\vec{G}) \in (r - 4r/\log\log n, r + 4r/\log\log n)$
- *for any two sets $A, B \subset V(\vec{G})$ of size at most $|A|, |B| \leq 0.6n$, there are at most $0.8r\sqrt{|A||B|}$ edges going from A to B.*

Then $\vec{G}$ contains at least $\left(\frac{r-100r/\log\log n}{e}\right)^n$ oriented 1-factors, provided that n is sufficiently large.

Proof Create an auxiliary bipartite graph G' from $\vec{G}$ in the following way: take two copies X and Y of the vertex set $[n]$ by doubling each vertex $v \in [n]$ into $v_X \in X$ and $v_Y \in Y$. We put a (non-oriented) edge $uv \in E(G')$ between vertices $u_X \in X$ and $v_Y \in Y$ if $\vec{uv} \in E(\vec{G})$ is an edge oriented from u to v in $\vec{G}$. We observe a one-to-one correspondence between oriented 1-factors in $\vec{G}$ and perfect matchings in G'.

In order to use the permanent to obtain a lower bound on the number of perfect matchings of G', we need a (large) spanning regular subgraph of G'. Its existence is guaranteed by the following claim.

Claim 1. *G' contains a spanning regular subgraph G'' with regularity at least $d = r - 100r/\log\log n$.*

Proof Applying the Ore-Ryser theorem [77] we see that the statement of the claim is true provided that for every $Y' \subseteq Y$ we have

$$d|Y'| \leq \sum_{x \in X} \min\{d, e_{G'}(x, Y')\}.$$

Suppose to the contrary that this contrition does not hold, i.e., there exists a $Y' \subseteq Y$ s.t.

$$d|Y'| > \sum_{x \in X} \min\{d, e_{G'}(x, Y')\}.$$

We examine the number of edges incident to Y' that can be deleted from G' without disturbing the right hand side of the above inequality. Formally, we denote it by $c = \sum_{x \in X} \max\{0, e_{G'}(x, Y') - d\}$. Notice that

$$c = e_{G'}(X, Y') - \sum_{x \in X} \min\{d, e_{G'}(x, Y')\} > e_{G'}(X, Y') - d|Y'|$$

as supposed above.

Since $(r - 4r/\log\log n)|Y'| \leq \delta(G')|Y'| \leq e_{G'}(X, Y') < d|Y'| + c$, we obtain

$$c > \frac{96r}{\log\log n}|Y'|.$$

On the other hand, denoting by X' the set of vertices that have at least d neighbors in Y', and noticing that $\Delta(G') \leq r + 4r/\log\log n$, we obtain

$$c \leq \frac{104r}{\log\log n}|X'|.$$

Hence,

$$|X'| > 0.9|Y'|. \tag{2.1}$$

Notice that by the choices of Y' and X', we have

$$d|Y'| > \sum_{x \in X} \min\{d, e_{G'}(x, Y')\} = d|X'| + e_{G'}(X \setminus X', Y'). \tag{2.2}$$

For the number of edges between $Y \setminus Y'$ and $X \setminus X'$ we see that

$$\begin{aligned}(r + 4r/\log\log n)|Y \setminus Y'| &\geq e_{G'}(X \setminus X', Y \setminus Y') = e_{G'}(X \setminus X', Y) - e_{G'}(X \setminus X', Y') \\ &\overset{(2.2)}{>} \delta(G')|X \setminus X'| - d(|Y'| - |X'|) \\ &\geq \frac{96r}{\log\log n}|X \setminus X'| + (r - 100r/\log\log n)|Y \setminus Y'|, \end{aligned} \tag{2.3}$$

leading to

$$|X \setminus X'| < 1.1|Y \setminus Y'|. \tag{2.4}$$

Furthermore, notice that by (2.2) it holds that

$$|X'| < |Y'|. \tag{2.5}$$

We prove the claim by case analysis.

- If $|Y'| \leq n/2$, we obtain for the number of edges between X' and Y'

  $$e_{G'}(X', Y') \overset{\text{Choice of } X'}{\geq} d|X'| \overset{(2.1)}{>} 0.9d\sqrt{|X'||Y'|} > 0.8r\sqrt{|X'||Y'|},$$

  contradicting the second condition of the lemma.

- If $|Y'| > n/2$, then again by the definition of X' we obtain $e_{G'}(X', Y') \geq d|X'|$, leading to

  $$e_{G'}(X \setminus X', Y') \overset{(2.2)}{<} d(|Y'| - |X'|) = d(|X \setminus X'| - |Y \setminus Y'|) \overset{(2.4)}{<} 0.1d|Y \setminus Y'| \overset{(2.5)}{<} 0.1d|X \setminus X'|.$$

  Thus, using the fact that $\delta(G') > d$, we see that

  $$e_{G'}(X \setminus X', Y \setminus Y') \geq 0.9d|X \setminus X'| \overset{(2.5)}{>} 0.8r\sqrt{|X \setminus X'| \cdot |Y \setminus Y'|},$$

  again contradicting the same condition of the lemma, since now both $X \setminus X'$ and $Y \setminus Y'$ have size less than $0.6n$ by (2.4).

□

We observe that the number of perfect matchings in G'' equals the permanent of the incidence matrix of G''. Hence the result of Egorychev [26] and Falikman [30] on the conjecture of van der Waerden implies that the number of perfect matchings in G'' is at least $d^n n!/n^n > \left(\frac{d}{e}\right)^n$. □

In order to use Lemma 2.2.1, we first prove the a.a.s. existence of a large subgraph of $\vec{G}(n, p)$ satisfying the degree-conditions of Lemma 2.2.1 a.a.s.

Lemma 2.2.2. *Let $\vec{G} \sim \vec{G}(n, p)$ with $p \geq \log n/n$. Then there exists a set $V' \subseteq [n]$ of at least $n - n/\log^2 n$ vertices of $\vec{G}$ such that the graph $\vec{C} := \vec{G}[V']$ satisfies $\delta_{in}(\vec{C}), \delta_{out}(\vec{C}), \Delta_{in}(\vec{C}), \Delta_{out}(\vec{C}) \in \left(\frac{np - 3np/\log\log n}{2}, \frac{np + np/\log\log n}{2}\right)$ a.a.s.*

Proof We observe using Lemma 1.3.1 that for $p \gg \log n(\log\log n)^2/n$ the statement holds for $V' = [n]$ a.a.s. Hence, from now on we assume $np = O(\log n(\log\log n)^2)$.

Let L be the set of all vertices whose in- or outdegree is at most $\frac{np - np/\log\log n}{2} + 1$. For every $y \in [n]$, we can estimate using Lemma 1.3.1

$$\mathbf{Pr}(y \in L) = \exp\left(-\Omega\left(\log n/(\log\log n)^2\right)\right).$$

Thus, by Markov's inequality we obtain

$$\begin{aligned}|L| &\le \log n \cdot \mathbb{E}(|L|) = \log n \cdot (n-1)\exp\left(-\Omega\left(\log n/(\log\log n)^2\right)\right)\\ &= n\exp\left(-\Omega\left(\log n/(\log\log n)^2\right)\right)\end{aligned} \tag{2.6}$$

a.a.s.

Fix an arbitrary vertex $x \in [n]$. We denote

$$L_x = \left\{y \in [n]\setminus\{x\} : d_{in,\vec{G}-x}(y) \le \frac{np - np/\log\log n}{2} \text{ or } d_{out,\vec{G}-x}(y) \le \frac{np - np/\log\log n}{2}\right\}.$$

Notice that $L_x \subseteq L$, and thus (2.6) bounds $|L_x|$ as well.

Since for every $y \in [n]\setminus\{x\}$ the events "$xy \in E(G)$" and "$y \in L_x$" are independent, we obtain using Lemma 1.3.1 again

$$\begin{aligned}&\mathbf{Pr}\left[\left(|N_G(x)\cap L_x| \ge \frac{np}{2\log\log n}\right) \mid \left(|L_x| = n\exp\left(-\Omega\left(\log n/(\log\log n)^2\right)\right)\right)\right]\\ &\le \exp\left(-\frac{np}{2\log\log n}\Omega\left(\log n/(\log\log n)^2\right)\right) = o(1/n).\end{aligned} \tag{2.7}$$

Similarly, we let R be the set of all vertices whose in- or outdegree is at least $\frac{np+np/\log\log n}{2} - 1$ and obtain

$$|R| \le n\exp\left(-\Omega\left(\log n/(\log\log n)^2\right)\right) \tag{2.8}$$

a.a.s.

We define analogously

$$R_x = \left\{y \in [n]\setminus\{x\} : d_{in,\vec{G}-x}(y) \ge \frac{np + np/\log\log n}{2} - 1 \text{ or } d_{out,\vec{G}-x}(y) \ge \frac{np + np/\log\log n}{2} - 1\right\},$$

and observe analogously to (2.7) that $R_x \subseteq R$ and

$$\mathbf{Pr}\left[\left(|N_G(x)\cap R_x| \ge \frac{np}{2\log\log n}\right) \mid \left(|R_x| = n\exp\left(-\Omega\left(\log n/(\log\log n)^2\right)\right)\right)\right] = o(1/n). \tag{2.9}$$

We denote by V' the set of all vertices from $[n]$ whose in- an outdegrees in $\vec{G}$ lie in $\left(\frac{np-np/\log\log n}{2}, \frac{np+np/\log\log n}{2}\right)$. Notice that $[n]\setminus V' \subseteq L_x \cup R_x$ for every $x \in [n]$. Hence, we see that $|V'| > n - \frac{n}{\log^2 n}$ a.a.s. by (2.6) and (2.8). Furthermore, from (2.7) and (2.9) we obtain that all in- an outdegrees in $\vec{G}[V']$ lie in $\left(\frac{np-3np/\log\log n}{2}, \frac{np+np/\log\log n}{2}\right)$ a.a.s., completing the proof of the lemma. □

From now on, whenever we have n and p chosen, we denote

$$d = d(n,p) = np - 100np/\log\log n\,.$$

In the following lemma, we show that the random graph contains a.a.s. many 2-factors.

Lemma 2.2.3. *The random graph $G \sim G(n,p)$ with $p \ge \log n/n$ satisfies*

$$\sum_{s\in[n/3]} 2^s f'(G,s) \ge d^{-n/\log^2 n}(d/e)^n$$

a.a.s.

Proof In order to use Lemma 2.2.1, we orient G at random to obtain $\vec{G}$ (as always in this chapter, for every edge each of the two possible orientations gets probability $1/2$ independently of the choices of other edges).

First we show that the second condition of Lemma 2.2.1 holds a.a.s. for $\vec{G}$ with the intuitive choice $r = np/2$. Since the maximum degree of G is at most $3np$ a.a.s. (see e.g. [14]), we obtain that in G a.a.s. for any two sets A and B with $|A| > 100|B|$ the number of edges between them is at most $3np|B| < 0.4np\sqrt{|A||B|}$. Hence, $e_{\vec{G}}(A,B) \le 0.4np\sqrt{|A||B|}$ and $e_{\vec{G}}(B,A) \le 0.4np\sqrt{|A||B|}$. Thus, we are left with the case of sets A and B of sizes $|A| \le 100|B|$ and $|B| \le 100|A|$.

For small disjoint sets, we obtain using Lemma 1.3.1

$$\begin{aligned}
&\mathbf{Pr}\left(\exists A', B' \subset [n],\ A' \cap B' = \emptyset,\ |A'||B'| \le \frac{n^2}{\log\log n},\ |A| \le 100|B| \le 10^4|A| :\ e_{\vec{G}}(A',B') \ge 0.4np\sqrt{|A'||B'|}\right)\\
&\le \sum_{a,b=o(n),\, a=\Theta(b)} \binom{n}{a}\binom{n-a}{b} \exp\left(-\Omega\left(np\sqrt{ab}\log\left(\frac{np\sqrt{ab}}{pab}\right)\right)\right)\\
&\le \sum_{a,b=o(n)} \left(\frac{ne}{a}\right)^a \left(\frac{ne}{b}\right)^b \exp\left(-\Omega\left(a\log n\log\left(\Omega\left(\frac{n}{a}\right)\right)\right) - \Omega\left(b\log n\log\left(\Omega\left(\frac{n}{b}\right)\right)\right)\right)\\
&\le \sum_{a,b=o(n)} \exp\left(a\log\left(\frac{ne}{a}\right) + b\log\left(\frac{ne}{b}\right) - \Omega\left(a\log\left(\frac{n}{a}\right)\log n\right) - \Omega\left(b\log\left(\frac{n}{b}\right)\log n\right)\right)\\
&= o(1).
\end{aligned}$$

Similarly, for large disjoint sets we obtain using Lemma 1.3.1

$$\begin{aligned}
&\mathbf{Pr}\left(\exists A', B' \subset [n],\ A' \cap B' = \emptyset,\ |A'||B'| > \frac{n^2}{\log\log n},\ |A|, |B| \le 0.6n :\ e_{\vec{G}}(A',B') \ge 0.4np\sqrt{|A'||B'|}\right)\\
&\le \sum_{a,b\le n,\, ab > \frac{n^2}{\log\log n}} \binom{n}{a}\binom{n-a}{b} \exp\left(-\Omega\left(abp\right)\right)\\
&\le 4^n \exp\left(-\Omega\left(\frac{n\log n}{\log\log n}\right)\right) = o(1).
\end{aligned}$$

Hence, a.a.s. for every pair of disjoint sets A' and B', the number of edges going from A' to B' satisfies

$$e_{\vec{G}}(A',B') < 0.4np\sqrt{|A'||B'|}. \tag{2.10}$$

Analogously, we see that a.a.s. for every $M \subseteq [n]$ of size at most $0.6n$,

$$e_G(M) < 0.4np|M|. \tag{2.11}$$

Thus, a.a.s. for every $A, B \subset [n]$ of size $|A|, |B| \le 0.6n$, the number of edges going from A to B in $\vec{G}$ is bounded by

$$\begin{aligned}
e_{\vec{G}}(A,B) = e_{\vec{G}}(A\setminus B, B\setminus A) + e_G(A\cap B) &\overset{(2.10),\,(2.11)}{<} 0.4np\sqrt{|A\setminus B||B\setminus A|} + 0.4np|A\cap B|\\
&\le 0.4np\sqrt{|A||B|},
\end{aligned}$$

establishing that the second condition of Lemma 2.2.1 holds a.a.s. for every subgraph of $\vec{G}$.

Hence, by Lemma 2.2.2 the graph $G \sim G(n,p)$ a.a.s. is such that for a random orientation $\vec{G}$, there a.a.s. exists a vertex set $V' \subseteq [n]$ of size at least $n - n/\log^2 n$ such that the induced subgraph $\vec{G}[V']$ satisfies the conditions of Lemma 2.2.1 with $r = np/2$. Applying Lemma 2.2.1 to this induced subgraph, we obtain

$$\sum_{s\in[n/3]} f\left(\vec{G}[V'], s\right) \ge \left(\frac{d}{2e}\right)^{n-n/\log^2 n}$$

a.a.s. Thus, we obtain

$$\sum_{s\in[n/3]} \mathbb{E}\left(f'\left(\vec{G},s\right)\right) \geq (1-o(1))\left(\frac{d}{2e}\right)^{n-n/\log^2 n}$$

a.a.s., where the expectation is taken over the random choice of orienting the edges of G, the process creating $\vec{G}$ from G.

On the other hand, when we orient the edges, an almost 2-factor of G with exactly s cycles becomes an almost 1-factor of $\vec{G}$ with probability at most $2^{\frac{n}{\log^2 n}-n+s}$, implying

$$\sum_{s\in[n/3]} 2^s f'(G,s) \geq \sum_{s\in[n/3]} 2^{n-\frac{n}{\log^2 n}} \mathbb{E}\left(f'\left(\vec{G},s\right)\right).$$

Putting these two facts together, we obtain

$$\sum_{s\in[n/3]} 2^s f'(G,s) \geq (1+o(1))\left(\frac{d}{e}\right)^{n-n/\log^2 n} \geq d^{-n/\log^2 n}(d/e)^n$$

a.a.s., completing the proof of the lemma. □

We show now that there are typically many almost 2-factors in G with a small number of cycles. We denote

$$s^* = s^*(n) = \frac{n}{\log n\sqrt{\log\log n}}.$$

Lemma 2.2.4. *For every $p \geq \log n/n$, the random graph $G \sim G(n,p)$ satisfies*

$$\sum_{s=1}^{s^*} f'(G,s) \geq (np/e)^n\,(1-o(1))^n$$

a.a.s.

Proof By Lemma 2.2.3 we know that $\sum_{s\in[n/3]} 2^s f'(G,s) \geq d^{-n/\log^2 n}(d/e)^n$ a.a.s.

We show now that the contribution of almost 2-factors with too many cycles is negligible. We use the estimate (5) of [62]: in the random graph $H \sim G(n',p)$, for every $s \geq \log n'$,

$$\mathbb{E}(f(H,s)) \leq \frac{(n'-1)!\,(\log n')^{s-1}\,p^{n'}}{(s-1)!2^s}.$$

We obtain

$$\begin{aligned}
\sum_{s=s^*}^{n/3} \mathbb{E}(2^s f'(G,s)) &\leq \sum_{\ell\leq n/\log^2 n} \binom{n}{\ell} \sum_{s=s^*}^{n/3} \frac{n!\,(\log n)^s\,p^{n-\ell}}{s!} \\
&\leq n!p^n\left(\frac{n}{p}\right)^{n/\log^2 n} \sum_{s=s^*}^{n/3}\left(\frac{s}{e\log n}\right)^{-s} \\
&= (d/e)^n e^{O(n/\log\log n)}\left(\frac{s^*}{e\log n}\right)^{-s^*} \\
&= o\left(d^{-n/\log^2 n}\,(d/e)^n\right).
\end{aligned}$$

Hence, using this estimate together with Markov's inequality, we see that the number of almost 2-factors of G with at most s^* cycles is

$$\sum_{s=1}^{s^*} f'(G,s) \geq \frac{1}{2} 2^{-s^*} d^{-n/\log^2 n} (d/e)^n = (np/e)^n (1-o(1))^n$$

a.a.s. □

The next technicality we need to prove in order to be ready to prove the main theorem is the expansion of $G(n,p)$.

To collect all but one expansion properties that we need, we make the following definition.

Definition 2.2.5. We call a graph G with the vertex set $[n]$ a p-expander, if there exists a set $D \subset [n]$ such that G and D satisfy the following properties:

- $|D| \leq n^{0.09}$.
- The graph G does not contain a non-empty path of length at most $\frac{2\log n}{3\log\log n}$ such that both of its (possibly identical) endpoints lie in D.
- For every set $S \subset [n] \setminus D$ of size $|S| \leq \frac{1}{p}$, its external neighborhood satisfies $|N(S)| \geq \frac{np}{1000}|S|$.

The following lemma shows that these properties are pseudo-random.

Lemma 2.2.6. *Consider the two-round expansion of the random graph and fix $G \sim G(n,p)$ with $\log n/n \leq p \leq 1 - 2\log\log n/n$ and $G \subseteq \hat{G} \sim G(n,\hat{p})$ with $\hat{p} = p + 2\log\log n/n$. Then it is a.a.s. true that every graph G' satisfying $G \subseteq G' \subseteq \hat{G}$ is a p-expander.*

Proof We first expose G and fix $D = \{v \in [n] : d_G(v) < np/100\}$ to be the set of all vertices of G with degree less than $np/100$ in G. Since for a fixed set D the second property is decreasing and the third property is increasing, it suffices to prove the second statement for $\hat{G}$ and the third statement for G.

The first property is satisfied by Claim 4.3 from [11] a.a.s. The second property can be proven to hold in $\hat{G}$ a.a.s. similarly to Claim 4.4 from [11] (there it is proven to hold for G a.a.s.)

For the third property, suppose to the contrary that there exists a set $S \subset [n] \setminus D$ of size at most $|S| \leq \frac{1}{p}$ such that its external neighborhood in G satisfies $|N_G(S)| < \frac{np}{1000}|S|$. By the definition of D, the number of edges incident to S in G is

$$e_G(S, N_G(S) \cup S) \geq |S|np/200.$$

But Lemma 1.3.1 tells us that

$$\begin{aligned}
&\mathbf{Pr}\left(\exists A,B \subseteq [n],\, |A| \leq \frac{1}{p}, |B| < \frac{np}{1000}|A| :\, e_G(A, B\cup A) \geq |A|np/200\right) \\
&< \sum_{A,B\subset[n],\,|A|\leq 1/p,\,|B|<|A|np/1000} \left(e \cdot \frac{\mathbb{E}\left(|e_G(A,B\cup A)|\right)}{|A|np/200}\right)^{|A|np/200} \\
&< \sum_{A,B\subset[n],\,|A|\leq 1/p,\,|B|<|A|np/1000} \left(\frac{200e|A||A\cup B|p}{|A|np}\right)^{|A|np/200} \\
&< \sum_{a\leq 1/p} anp\binom{n}{a}\binom{n}{\frac{anp}{1000}}\left(\frac{3ap}{5}\right)^{anp/200} \\
&< \sum_{a\leq 1/p} anp\left(\frac{3ap}{5}\right)^{anp/400} = o(1),
\end{aligned}$$

providing that the third property holds in G a.a.s. $\square$

The proof of the next lemma is based on the ingenious rotation-extension technique, developed by Pósa [78], and applied later in a multitude of papers on Hamiltonicity, mostly of random or pseudorandom graphs (see for example [15], [37], and [64]).

Let G be a graph and let $P_0 = (v_1, v_2, \ldots, v_q)$ be a path in G. If $1 \le i \le q-2$ and (v_q, v_i) is an edge of G, then there exists a path $P' = (v_1 v_2 \ldots v_i v_q v_{q-1} \ldots v_{i+1})$ in G with the same set of vertices. The path P' is called a *rotation* of P_0 with *fixed endpoint* v_1 and *pivot* v_i. The edge (v_i, v_{i+1}) is called the *broken* edge of the rotation. We say that the segment $v_{i+1} \ldots v_q$ of P_0 is *reversed* in P'. In case the new endpoint v_{i+1} has a neighbor v_j such that $j \notin \{i, i+2\}$, then we can rotate P' further to obtain more paths of the same length. We will use rotations together with the expansion properties from Lemma 2.2.6 and the necessary minimum degree condition to find a path on the same vertex set as P_0 with large rotation endpoint sets.

The next lemma shows that in any graph having the p-expander property and minimum degree 2, for any path P_0 and its endpoint v_1, after a small number of rotations with fixed endpoint v_1, we either create many other endpoints or extend the path. Its proof has certain similarities to the proofs of Lemma 8 from [45] and of Claim 2.2 from [52].

Lemma 2.2.7. *Let n be a sufficiently large integer and G be an n-vertex p-expander with minimum degree $\delta(G) \ge 2$ and $np \ge \log n$. Let P_0 be a $v_1 w$-path in G. Denote by $B(v_1) \subset V(P_0)$ the set of all vertices $v \in V$ for which there is a $v_1 v$-path on the vertex set $V(P_0)$ which can be obtained from P_0 by at most $3\frac{\log n}{\log(np)}$ rotations with fixed endpoint v_1. Then $B(v_1)$ satisfies one of the following properties:*

- *there exists a vertex $v \in B(v_1)$ with a neighbor outside $V(P_0)$, or*
- *$|B(v_1)| \ge n/3000$.*

Proof Assume that $B(v_1)$ does not have the first property (i.e., for every $v \in B(v_1)$ it holds that $N(v) \subseteq V(P_0)$).

Let t_0 be the smallest integer such that $\left(\frac{np}{3000}\right)^{t_0-1} \ge \frac{1}{p}$; note that $t_0 \le 2\frac{\log n}{\log(np)}$. Since G is a p-expander, there is a corresponding vertex set D as in Definition 2.2.5.

At the first step, we find a neighbor $u \notin D \cup N(D)$ of w that is not a neighbor of w along P_0. Its existence is guaranteed since w has at least two neighbors along P_0, and by the second p-expansion property, at most one of them can have a neighbor in D. We rotate the initial path P_0 with pivot u and call the resulting path $P' = (v_1, \ldots, v_q)$. Notice that this way, v_q is guaranteed not to belong to D.

We construct a sequence of sets $S_0, \ldots, S_{t_0} \subseteq B(v_1) \setminus D \subseteq V(P_0) \setminus \{v_1\}$ of vertices, such that for every $0 \le t \le t_0$ and every $v \in S_t$, v is the endpoint of a path which can be obtained from P' by a sequence of t rotations with fixed endpoint v_1, such that for every $0 \le i < t$, the non-v_1-endpoint of the path after the ith rotation is contained in S_i. Moreover, $|S_t| = \left(\frac{np}{3000}\right)^t$ for every $t \le t_0 - 2$, $|S_{t_0-1}| = \frac{1}{p}$, and $|S_{t_0}| \ge \frac{n}{3000}$.

We construct these sets by induction on t. For $t = 0$, one can choose $S_0 = \{v_q\}$ and all requirements are trivially satisfied.

Let now t be an integer with $0 < t \le t_0 - 1$ and assume that the sets $S_0, \ldots, S_{t-1}$ with the appropriate properties have already been constructed. We will now construct S_t. Let

$$T = \left\{ v_i \in N(S_{t-1}) : v_{i-1}, v_i, v_{i+1} \notin \bigcup_{j=0}^{t-1} S_j \cup D \right\}$$

be the set of potential pivots for the tth rotation, and notice that $T \subset V(P_0)$ due to our assumption, since $T \subseteq N(S_{t-1})$ and $S_{t-1} \subseteq B(v_1)$. Assume now that $v_i \in T$, $y \in S_{t-1}$ and

$(v_i, y) \in E(G)$. Then, by the induction hypothesis, a v_1y-path Q can be obtained from P' by $t-1$ rotations such that after the jth rotation, the non-v_1-endpoint is in S_j for every $0 \le j \le t-1$. Each such rotation breaks an edge which is incident with the new endpoint, obtained in that rotation. Since v_{i-1}, v_i, v_{i+1} are not endpoints after any of these $t-1$ rotations, both edges (v_{i-1}, v_i) and (v_i, v_{i+1}) of the original path P' must be unbroken and thus must be present in Q.

Hence, rotating Q with pivot v_i will make either v_{i-1} or v_{i+1} an endpoint (which of the two, depends on whether the unbroken segment $v_{i-1}v_iv_{i+1}$ is reversed or not after the first $t-1$ rotations). Assume without loss of generality that the endpoint is v_{i-1}. We add v_{i-1} to the set $\hat{S}_t$ of new endpoints and say that v_i *placed* v_{i-1} *in* $\hat{S}_t$. The only other vertex that can place v_{i-1} in $\hat{S}_t$ is v_{i-2} (if it exists).

Observe now that if $t < 0.1\log n/\log\log n$, the distance between any vertex from S_{t-1} and v_q is at most $2t-2 < 0.2\log n/\log\log n$ by the way the sets were constructed. Hence, between any two vertices from $N(S_{t-1}) \cup N(N(S_{t-1}))$, there is a path of length at most $0.5\log n/\log\log n$. Thus at most one vertex from D can be in $N(S_{t-1}) \cup N(N(S_{t-1}))$. On the other hand, it $t \ge 0.1\log n/\log\log n$, then $|D| \le n^{0.09} = o(|S_{t-1}|) = o(|N(S_{t-1})|)$. Thus, in both cases $|D \cap (N(S_{t-1}) \cup N(N(S_{t-1})))| = o(|N(S_{t-1})|)$.

Combining all this information together, we obtain

$$\begin{aligned}
|\hat{S}_t| &\ge \frac{1}{2}|T| \\
&\ge \frac{1}{2}\left(|N(S_{t-1})| - 3(1 + |S_1| + \ldots + |S_{t-1}| + |D \cap (N(S_{t-1}) \cup N(N(S_{t-1})))|)\right) \\
&\ge \left(\frac{np}{3000}\right)^t.
\end{aligned}$$

Clearly we can delete arbitrary elements of $\hat{S}_t$ to obtain S_t of size $\left(\frac{np}{3000}\right)^t$ if $t \le t_0 - 2$ and of size $\frac{1}{p}$ if $t = t_0 - 1$. So the proof of the induction step is complete and we have constructed the sets $S_0, \ldots, S_{t_0-1}$.

To construct S_{t_0} we use the same technique as above, only the calculations are slightly different.

$$\begin{aligned}
|\hat{S}_{t_0}| &\ge \frac{1}{2}|T| \\
&\ge \frac{1}{2}\left(|N(S_{t_0-1})| - 3(1 + |S_1| + \ldots + |S_{t_0-2}| + |S_{t_0-1}| + |D \cap (N(S_{t-1}) \cup N(N(S_{t-1})))|)\right) \\
&\ge n/3000.
\end{aligned}$$

The set $S_{t_0} := \hat{S}_{t_0}$ is by construction a subset of $B(v_1)$, and the number of rotations needed to make any of its vertices an endpoint of the current path is at most $t_0 + 1$, concluding the proof of the lemma. $\square$

The proof of the following lemma relies on the final part of the proof of Theorem 1 from [68] and uses Lemma 2.2.7. It shows that under certain pseudorandom conditions in a graph G for every almost 2-factor, after adding few random edges, there exists a Hamilton cycle within a small Hamming distance from it a.a.s.

Lemma 2.2.8. *Let G be a connected n-vertex p-expander with minimum degree 2 and S be a set of vertices of G of size $|S| = o(n)$ such that there exist at least n non-edges in G not incident to S. Let F be an almost 2-factor of G with at most s^* cycles. Choose n non-edges $e_1, \ldots, e_n$ of G i.a.r. under the condition that none of them is incident to S and denote by G' the (random) graph obtained from G by turning them into edges. Then, if it is a.a.s. true that every graph $G \subseteq \hat{G} \subseteq G'$ is a p-expander, then G' a.a.s. contains a Hamilton cycle H with Hamming distance at most $17s^*\log n/\log(np)$ from F.*

Proof Fix an arbitrary component $C \subseteq F$. Since G is connected, there exists an edge in G connecting a vertex $v \in V(C)$ and $y \notin V(C)$ - unless of course C is already Hamiltonian. We denote by C' the component of y in F. Opening C up by deleting an edge of C incident to v (no need to do so if C is just one isolated vertex), we get a path P. We append the edge vy to P, go through it to C', and if C' is a cycle, then we open it up by deleting an edge of C' incident to y to get a longer path P' and repeat the argument. If at some point there are no edges between the endpoints of the current path P'' and other components from F, then we can fix one endpoint x of P'' and rotate P'' using Lemma 2.2.7 to extend it outside or to obtain a set $B(x)$ of size at least $|B(x)| \geq n/3000$ of potential other endpoints. For every vertex $z \in B(x)$, we can rotate the resulting path fixing z as one endpoint to obtain a set $A(z)$ of size at least $|A(z)| \geq n/3000$ of potential other endpoints or to extend the path outside. If the path still cannot be extended outside and we can still not close it to a cycle, we have a set E' of at least $10^{-8}n^2$ non-edges of G not incident to S, so that turning any of them into an edge would close the path to a cycle. We add pairs $e_1, e_2, \dots$ to $E(G)$, until one of them falls inside E'. Notice that for every $i \in [n]$, the pair e_i falls into E' with at least some constant positive probability. This means that considering events "$e_i \in E'$", every event has probability $\Theta(1)$ regardless of the previous events. Notice that in a successful round, the number of components gets reduced or a Hamilton cycle is created, since the edge that appeared in E' closed the path into a cycle or extended the path directly. To reduce the number of components by one, we do at most $\log n/\log(np)$ rotations by Lemma 2.2.7, therefore increasing the Hamming distance from F by at most $4 + 12\log n/\log(np)$. Since it is enough to have $s^* + n/\log^2 n$ successful events to obtain a Hamilton cycle, the expected number of needed turns of non-edges into edges is at most $O(1) \cdot (s^* + n/\log^2 n) = o(n)$. Hence the n additional edges suffice to create a Hamilton cycle H from F by Markov's inequality a.a.s., replacing at most $8\log n/\log(np)$ edges for every component of F. Thus, the Hamming distance between F and H is at most $2 \cdot 8\frac{\log n}{\log(np)}(s^* + n/\log^2 n) \leq 17s^* \log n/\log(np)$. □

We are now ready to prove Theorem 1.2.1.

Proof Notice that only the lower bound is of interest for us. We expose G in two rounds. We choose a function $p_1 = p_1(n)$ such that $\frac{\log n + \log\log n + \omega(1)}{n} \leq p_1 \leq p - \frac{\omega(1)}{n}$, $p_1 \leq 1 - 2\log\log n/n$, and $p_1 = (1-o(1))p$. In the first round, we expose $G_1 \sim G(n, p_1)$. We determine $D := \{v \in [n] : d_{G_1}(v) < np_1/100\}$.

In the second round, we expose the binomial random graph G_2 by including every edge from $K_n \setminus G_1$ into $E(G_2)$ with probability $p_2 := \frac{p-p_1}{1-p_1}$. Since $(1-p_1)(p_2) = 1-p$, we obtain a graph $G := G_1 \cup G_2 \sim G(n,p)$. We know by Lemma 2.2.6 that a.a.s. every graph between G_1 and G including them both is a p_1-expander. Furthermore, notice that the expected number of edges in G_2 is $\binom{n}{2}p_2 \geq \binom{n}{2}(1-p_1) = \omega(n)$, hence a.a.s. at least n additional random edges appeared in the second round of expansion by Markov's inequality. Since these edges were chosen i.a.r., and G_1 is a.a.s. connected with minimum degree at least 2 (see e.g. [14]), the conditions of Lemma 2.2.8 are satisfied for G_1 and the first n edges exposed in the second round with $S = \emptyset$.

Now, we put all we know together:

- By Lemma 2.2.4 we obtain $\sum_{s=1}^{s^*} f'(G_1, s) \geq (np_1/e)^n (1-o(1))^n$ a.a.s.
- For every almost 2-factor F of G_1 with at most s^* cycles, there a.a.s. exists a Hamilton cycle in G with Hamming distance at most $k := 17s^* \log n/\log(np_1) = \frac{17n}{\log(np_1)\sqrt{\log\log n}} = o(n)$ from F by Lemma 2.2.8.
- On the other hand, for every Hamilton cycle H in G, to obtain an almost 2-factor of G_1 of distance at most k from H, we can first delete at most k edges of H, thus obtaining a collection of at most k paths. These paths should then be tailored into an almost 2-factor, and the choices here are for each of the at most $2k$ endpoints of the paths to be

connected to one of its $\Delta(G_1)$ neighbors in G_1 or to stay isolated. Thus, there are at most $\binom{n}{k}(\Delta(G_1)+1)^{2k}$ almost 2-factors of G with Hamming distance at most k from H.

- Hence, by double counting almost 2-factors of G with at most s^* cycles, we obtain

$$h(G) \geq \frac{\sum_{s=1}^{s^*} f'(G',s)}{\binom{n}{k}(\Delta(G_1)+1)^{2k}} \geq (np_1/e)^n \frac{(1-o(1))^n}{2^{o(n)}(4\log(np_1))^{2k}} = (np/e)^n (1-o(1))^n$$

a.a.s. □

To strengthen the result of Cooper and Frieze [20], we now prove Theorem 1.2.2.

Proof The proof goes along the argument of Theorem 1.2.1, but now we expose the graph in three rounds. We first expose $G_1 \sim G(n, \log n/n)$ and fix the set D of vertices of degree at most $\log n/100$. Notice that similarly to the argument in the proof of Lemma 2.2.6, Claim 4.3 from [11] implies that $|D| \leq n^{0.09}$. In the second round of exposure, in addition to G_1 we expose those edges that are incident to D one by one, until the minimum degree becomes two; the resulting graph is called G'. In the third round of exposure we consider the binomial random graph G_2 by including every edge of $K_n \backslash G_1$ not incident to D with probability $p_2 := \frac{\log\log n}{2n}$.

Let us denote by G the graph obtained by stopping the random graph process at the moment the minimum degree becomes two. Notice first that since in the random graph process $\delta\left(G\left(n, \frac{\log n + 0.5\log\log n}{n}\right)\right) = 1$ a.a.s., we obtain $G' \cup G_2 \subseteq G$ a.a.s., where by the union of two graphs with vertex sets $[n]$ we denote the graph on the same vertex set where the union is taken over the edge sets. Furthermore, observe that since in the random graph process $\delta\left(G\left(n, \frac{\log n + 2\log\log n}{n}\right)\right) \geq 2$ a.a.s., we obtain $G' \cup G_2 \subseteq G\left(n, \frac{\log n + 2\log\log n}{n}\right)$ a.a.s. (The two statements above can be found e.g. in [14].) Finally, G' is connected a.a.s. because of the expansion properties and the fact that the edge set between two linearly large sets is not empty (see e.g. [52]).

Since $p_2 = \omega(1/n)$ and $|D| = o(n)$ a.a.s., we obtain $|E(G_2)| = \omega(n)$ a.a.s. Furthermore, these edges are random under the only conditions of being non-edges of G_1 and being not incident to D. Hence, the conditions of Lemma 2.2.8 are satisfied for G' and the first n edges exposed in the third round.

Thus, following the lines of the proof of Theorem 1.2.1, we obtain the desired estimate. □

2.3 Concluding remarks

In this chapter we have proven that for any value of the edge probability $p = p(n)$, for which the random graph G $G(n,p)$ is a.a.s. Hamiltonian, the number of Hamilton cycles in G is $n!p^n(1+o(1))^n$ a.a.s., thus being asymptotically equal to the expected value – up to smaller order exponential terms. Of course, it would be very nice to extend Janson's result [56] to smaller values of p and to understand more accurately the distribution of the number of Hamilton cycles in relatively sparse random graphs. However, given that the machinery used in [56] is rather involved, and the result (limiting distribution) is somewhat surprising, this will not necessarily be an easy task.

Our bound on the number of Hamilton cycles in $G(n,p)$ can be used to bound the number of perfect matching similarly to [68]. Let $m(G)$ denote the number of perfect matchings in the graph G. Since every Hamilton cycle is a union of two perfect matchings, we obtain $h(G) \leq \binom{m(G)}{2}$. Hence, for $G \sim G(n,p)$ the a.a.s. lower bound on $h(G)$ from Theorem 1.2.1 provides the a.a.s. lower bound $m(G) \geq (np/e)^{n/2}(1-o(1))^n$. Since the upper bound is easily obtained from the expected value by Markov's inequality similarly to the first paragraph of Section 2.2, we have $m(G) = (np/e)^{n/2}(1-o(1))^n$. The corresponding hitting time statement is obtained by a straightforward modification of the proof of Theorem 1.2.2: In the random graph process, the edge that makes the graph connected a.a.s. creates $(\log n/e)^{n/2}(1-o(1))^n$ perfect matchings.

3 On covering expander graphs by Hamilton cycles

3.1 Introduction

As we pointed out before, the results in this chapter are based on joint work with Michael Krivelevich and Tibor Szabó [45].

3.1.1 Pseudorandom setting

Our argument will proceed in an appropriately chosen pseudorandom setting.

By the *neighborhood* $N(A)$ of a set A, we mean all the vertices outside A having at least one neighbor in A. Note that we explicitly exclude A from $N(A)$.

The following definition contains the most important notions of the chapter.

Definition 3.1.1. We say that a graph G has the *small expansion property* $S(s, g)$ with *expansion factor* s and *boundary* g, if for any set $A \subset V(G)$ of size $|A| \leq g$, the neighborhood of A satisfies $|N(A)| \geq |A|s$.

We say that G has the *large expansion property* $L(l)$ with *frame* l, if there is an edge between any two disjoint sets $A, B \subset V(G)$ of size $|A|, |B| \geq l$.

We call a graph (s, g, l)-*expander*, if it satisfies properties $S(s, g)$ and $L(l)$.

We refer to an $\left(s, \frac{4n \log s}{s \log n}, \frac{n \log s}{3000 \log n}\right)$-expander on n vertices briefly as an s-*expander*.

Notice that the expander-property is monotone in all three parameters, meaning that every (s, g, l)-expander is also an $(s-1, g, l)$-, an $(s, g-1, l)$- and an $(s, g, l+1)$-expander.

The proof of Theorem 1.2.3 is based on the following result.

Theorem 3.1.2. For every constant $\alpha > 0$ there exists $n_0 = n_0(\alpha)$ such that for every $n \geq n_0$ and every $h > 0$, every n^{α}-expander graph G on n vertices with a Hamilton packing of size h has a Hamilton covering of size at most $h + \frac{28000(\Delta(G)-2h)}{\alpha^4}$.

3.1.2 Structure of the chapter and outline of the proofs

In Section 3.2, we prove Theorem 3.1.2 in the following main steps:

- in Lemma 3.2.2 we show that the small expansion property is "robust" in the sense that after deleting a small linear-size set of vertices from a graph satisfying S, we still have a large subgraph satisfying S with slightly worse parameters;

- in Lemma 3.2.4 we show that vertex-disjoint paths in an expander can be joined together without losing or gaining too many edges;

- in Lemma 3.2.5, Fact 3.2.6 and Lemma 3.2.7, we learn how to apply the rotation-extension technique, developed by Pósa [78], without losing too many important edges;

- Lemma 3.2.8 contains the main proof of the chapter. There we show applying the previous technical statements that a small matching in an expander can mostly be covered by a Hamilton cycle of the same graph.

- After having digested the statement of Lemma 3.2.8 in Corollary 3.2.9, we prove Theorem 3.1.2 in 6 easy lines.

Section 3.3 contains the proof of Theorem 1.2.3. There, we first prove in Lemma 3.3.1 that $G(n,p)$ is a $\sqrt[5]{np}$-expander a.a.s., and prove Theorem 1.2.3 using this fact and the result from [62].

And finally, in Section 3.4, we give some concluding remarks leading to further open questions.

In general, we may drop floor and ceiling signs to improve the readability when they do not influence the asymptotic statements.

3.2 Proof of Theorem 3.1.2

The maximum over all pairs of vertices $x, y \in V(G)$ of the length of a shortest xy-path in a graph G is called the *diameter* of G and is denoted by $\mathrm{diam}(G)$. We start with an observation showing that graphs with appropriate expander properties have small diameter.

Observation 3.2.1. Any n-vertex graph satisfying $S(s,g)$ and $L(l)$ for some s, g, l with $s > 1$ and $l \leq sg$ has diameter at most $2 \log n / \log s + 3$.

Proof Let G be a graph on n vertices. By the small expansion property of G, we know for every vertex $x \in V(G)$ that $|N(x)| \geq s$. In the following, for any $x \in V(G)$ and any $i \in \mathbb{N}$, let us denote by $B_i(x)$ the set of all vertices with distance at most i to x, i.e., $B_i(x)$ contains all those vertices $y \in V(G)$, for which there exists an xy-path of length at most i. Inductively, as long as for an $i \in \mathbb{N}$ and an $x \in V(G)$ it holds that $|B_i(x)| \leq g$, we obtain $|B_{i+1}(x)| \geq s^{i+1}$. Now, the last index i such that $|B_i(x)| \leq g$ is obviously at most $\lfloor \log n / \log s \rfloor - 1$, since $sg \leq n$. Hence, $|B_{\lfloor \log n / \log s \rfloor}(x)| \geq g$, and $|B_{\lfloor \log n / \log s \rfloor + 1}(x)| \geq sg \geq l$. Thus, for any two vertices $x, y \in V(G)$ we know that both sets $B_{\lfloor \log n / \log s \rfloor + 1}(x)$ and $B_{\lfloor \log n / \log s \rfloor + 1}(y)$ have at least l vertices. This guarantees that by property $L(l)$ of G, either these sets are not disjoint or there exists an edge between these sets. Each of these facts implies an x, y-path of length at most $2 \lfloor \log n / \log s \rfloor + 3$, and the observation follows. □

The following lemma shows that if we have a graph satisfying the small expansion property and we remove an arbitrary subset of small size from the vertex set, then with the additional removal of a (possibly) even smaller subset we can recover some of the small expansion property again.

Lemma 3.2.2 (Induced expander lemma)**.** For any s, g, every graph $G = (V, E)$ satisfying $S(s,g)$ has the following induced expansion property. For every $D \subset V$ of size $|D| \leq \frac{gs}{4}$, there exists a set $Z \subset V$ of size $|Z| \leq \frac{2|D|}{s}$, such that the graph $G[V \setminus (D \cup Z)]$ satisfies $S(\frac{s}{2}, \frac{g}{2})$.

Proof Let Z be a largest set in $V \setminus D$ among subsets of $V \setminus D$ of size at most g not satisfying the small expansion property with expansion factor $s/2$ in $G[V \setminus D]$, meaning that $|N(Z) \setminus D| < |Z| \frac{s}{2}$ (assuming there exists such a set; otherwise we are done by setting $Z = \emptyset$ in the statement of the lemma). We denote $U = V \setminus (D \cup Z)$ and remember that

$$|N(Z) \cap U| < \frac{|Z|s}{2}.$$

Thus, by the property $S(s,g)$ of G,

$$|D| + |N(Z) \cap U| \geq |N(Z)| \geq |Z|s,$$

implying that

$$|Z| \leq \frac{2|D|}{s} \leq \frac{g}{2}.$$

Assume now for the sake of contradiction that $G[U]$ does not satisfy $S(\frac{s}{2}, \frac{g}{2})$, i.e. there exists an $A \subset U$ with $|A| \leq \frac{g}{2}$ and $|N(A) \cap U| < |A|\frac{s}{2}$. Then the set $A \cup Z \subset V \setminus D$ satisfies both properties

$$|A \cup Z| \leq g$$

and

$$|N(A \cup Z) \cap U| \leq |N(A) \cap U| + |N(Z) \cap U| < \frac{|A \cup Z|s}{2},$$

contradicting the assumption that Z is a largest set in $V \setminus D$ with these properties. □

The following concept allows us to join together paths in an appropriate way.

By the *k-end* of a path we mean the at most $2k$ vertices of this path's vertex set with distance at most $k-1$ to one of the two endpoints, whereas the endpoints have distance 0 to themselves and so are part of any k-end with $k \geq 1$. We call a path *non-trivial* if its length is at least one. Given a family $\mathcal{M}$ of non-trivial vertex-disjoint paths, we call a family $\mathcal{M}'$ of non-trivial vertex-disjoint paths a *(d,k)-extension of* $\mathcal{M}$, if we have that

- $\mu := |\mathcal{M}| - |\mathcal{M}'| \geq 0$,
- $|\bigcup_{P \in \mathcal{M}} E(P) \setminus \bigcup_{P' \in \mathcal{M}'} E(P')| \leq 2(k-1)\mu$, and
- $|\bigcup_{P' \in \mathcal{M}'} E(P') \setminus \bigcup_{P \in \mathcal{M}} E(P)| \leq (d+2)\mu$.

Informally speaking, we neither gain nor lose too many edges relative to the decrease in the number of paths, while passing from $\mathcal{M}$ to $\mathcal{M}'$. Behind the definition, in our mind lies the iterative algorithm which in each step either deletes a path of length less than $2k-1$ from the family $\mathcal{M}$ or joins the k-ends of two paths in $\mathcal{M}$ by paths of length at most $d+2$ (and deletes the k-ends). At the end, when one cannot do either, we have a size-minimal (d,k)-extension, the main object we use in our proofs. This intuition is to be formalized in Lemma 3.2.3.

To clarify the notation, when speaking about the *size* $|\mathcal{M}|$ of a family of non-trivial vertex-disjoint paths $\mathcal{M}$, we mean the number of these paths. Note that any extension of $\mathcal{M}$ has at most as many paths as $\mathcal{M}$. Notice that the definition of extension is transitive, i.e., if for some pair (d,k), $\mathcal{M}'$ is a (d,k)-extension of $\mathcal{M}$ and $\mathcal{M}''$ is a (d,k)-extension of $\mathcal{M}'$, then $\mathcal{M}''$ is a (d,k)-extension of $\mathcal{M}$. Furthermore, the relation is also reflexive, i.e., $\mathcal{M}$ is a (d,k)-extension of itself for any $k \geq 1$ and every $d \geq 0$. We say that $\mathcal{M}'$ is a size-minimum (d,k)-extension of $\mathcal{M}$, if every (d,k)-extension $\mathcal{M}''$ of $\mathcal{M}$ has size at least $|\mathcal{M}''| \geq |\mathcal{M}'|$. Notice that the transitivity and reflexivity of extensions imply that then $\mathcal{M}'$ is a size-minimum (d,k)-extension of itself.

We use the notation $V(\mathcal{M}) := \bigcup_{P \in \mathcal{M}} V(P)$ for the set of all vertices appearing in one of the paths of $\mathcal{M}$ and $E(\mathcal{M}) := \bigcup_{P \in \mathcal{M}} E(P)$ for the corresponding edge set.

For our applications we will mostly be interested in size-minimum extensions. The following lemma provides us with two basic properties of size-minimum extensions: namely that they do not contain very short paths and that there are no short paths between the ends of two distinct paths from such an extension.

Lemma 3.2.3. *Let G be a graph, $k \geq 1$ and $d \geq 0$ arbitrary integers and $\mathcal{M}$ a family of non-trivial vertex-disjoint paths, which is a size-minimum (d,k)-extension of itself. Then the following holds:*

1. *there exists no path in $\mathcal{M}$ of length less than $2k-1$ and*
2. *for every two distinct paths $P_1, P_2 \in \mathcal{M}$, for every vertex x in the k-end of P_1 and every vertex y in the k-end of P_2, for every $a \in N(x) \setminus V(\mathcal{M})$ and $b \in N(y) \setminus V(\mathcal{M})$, there exists no ab-path of length at most d in $G - V(\mathcal{M})$.*

Proof 1. Suppose for the sake of contradiction that there exists a path P of length at most $2k-2$ in $\mathcal{M}$. Delete this path from $\mathcal{M}$ and call the resulting family of non-empty vertex-disjoint paths $\mathcal{M}'$. Then since $|\mathcal{M}|-|\mathcal{M}'|=1$, we obtain

$$|E(\mathcal{M})\setminus E(\mathcal{M}')| = |E(P)| \le 2k-2 = 2(k-1)\left(|\mathcal{M}|-|\mathcal{M}'|\right),$$

and

$$|E(\mathcal{M}')\setminus E(\mathcal{M})| = 0 \le (d+2)\left(|\mathcal{M}|-|\mathcal{M}'|\right).$$

Hence, $\mathcal{M}'$ is a (d,k)-extension of $\mathcal{M}$, contradicting the minimality of $\mathcal{M}$ as a (d,k)-extension of itself.

2. Suppose to the contrary that there exist two distinct paths $P_1, P_2 \in \mathcal{M}$ with a vertex x in the k-end of P_1, a vertex y in the k-end of P_2, and vertices $a \in N(x)\setminus V(\mathcal{M})$ and $b \in N(y)\setminus V(\mathcal{M})$ such that in $G - V(\mathcal{M})$, there is an ab-path of length at most d. Let us call this path P_3.

The vertex x splits P_1 into two subpaths, the shorter one has length at most $k-1$. Let us call the longer one P_x and construct P_y from P_2 analogously. By connecting the paths P_x, P_3 and P_y via the edges xa and by, we obtain a new path P'. Replace the two paths P_1 and P_2 in $\mathcal{M}$ by P' and call the resulting family of non-trivial vertex-disjoint paths $\mathcal{M}'$. Then $\mu = |\mathcal{M}|-|\mathcal{M}'| = 1$.

By construction $E(\mathcal{M})$ contains all but at most $k-1$ edges from each of P_1 and P_2, so

$$|E(\mathcal{M})\setminus E(\mathcal{M}')| = |E(P_1)\cup E(P_2)\setminus E(P')| \le 2(k-1) = 2(k-1)\mu,$$

Furthermore, $E(\mathcal{M}')$ contains at most $|E(P')\setminus (E(P_1)\cup E(P_2))| \le d+2$ edges that are not contained in $E(\mathcal{M})$, so

$$|E(\mathcal{M}')\setminus E(\mathcal{M})| \le d+2 = (d+2)\mu.$$

In conclusion, $\mathcal{M}'$ is a (d,k)-extension of $\mathcal{M}$, contradicting the minimality of $\mathcal{M}$ as a (d,k)-extension of itself. ☐

Lemma 3.2.4. For every $k \ge 1$, in every n-vertex graph G satisfying $S(s,g)$ and $L(l)$ for some s,g and l with $sg \ge 4l$, $s \ge 18$, and n sufficiently large the following holds. For every family of non-trivial vertex-disjoint paths $\mathcal{M}$ on at most $|V(\mathcal{M})| \le \alpha gs/20$ vertices in G, where $\alpha = \log s/\log n$, there exists a $(6/\alpha,k)$-extension $\mathcal{M}'$ of $\mathcal{M}$ of size $|\mathcal{M}'| \le \frac{5|V(\mathcal{M})|}{2\alpha ks}+1$.

Proof Let $\mathcal{M}'$ be a size-minimum $(6/\alpha,k)$-extension of $\mathcal{M}$. (One exists, since $\mathcal{M}$ is a $(6/\alpha,k)$-extension of itself.) Apply Lemma 3.2.2 with $D = V(\mathcal{M}')$ to find the corresponding sets Z and U (meaning that $|Z| \le 2|V(\mathcal{M}')|/s$, $U = V(G)\setminus(V(\mathcal{M}')\cup Z)$ and $G[U]$ satisfies $S(s/2,g/2)$). Lemma 3.2.2 can be applied since

$$\begin{aligned}
|D| &= |V(\mathcal{M}')| = |E(\mathcal{M}')| + |\mathcal{M}'| \\
&\le |E(\mathcal{M})| + (6/\alpha+2)(|\mathcal{M}|-|\mathcal{M}'|) + |\mathcal{M}'| \\
&= |V(\mathcal{M})| + (6/\alpha+1)(|\mathcal{M}|-|\mathcal{M}'|) \\
&< (1+(6/\alpha+1)/2)|V(\mathcal{M})| < 5|V(\mathcal{M})|/\alpha \le gs/4.
\end{aligned}$$

In the last line we used that the paths of $\mathcal{M}$ are non-trivial (implying that $|\mathcal{M}| \le |V(\mathcal{M})|/2$) and that $\alpha < 1$.

Let now x and y be two vertices from the k-ends of two distinct paths P_1 and $P_2 \in \mathcal{M}'$, respectively, and suppose each of them has a neighbor in U. Let $a \in U$ be a neighbor of x and let $b \in U$ be a neighbor of y. Since $G[U]$ satisfies both $S(s/2,g/2)$ and $L(l)$ with $l \le s/2 \cdot g/2$, Observation 3.2.1 implies that the diameter of $G[U]$ is at most

$$\operatorname{diam}(G[U]) \le 2\log n/\log(s/2) + 3 \le 2\frac{\log n}{\frac{2}{3}\log s} + 3 < \frac{6}{\alpha},$$

where the next to last inequality holds since $s \geq 8$. Hence, there exists an ab-path of length at most $\frac{6}{\alpha}$ in $G[U] \subseteq G[V(G) \setminus V(\mathcal{M})]$, contradicting the size-minimality of $\mathcal{M}'$ by Lemma 3.2.3.

Consequently, there can be at most one such path in $\mathcal{M}'$ that has a vertex in its k-end with a neighbor in U. Hence the k-ends of at least $|\mathcal{M}'| - 1$ paths have neighbors only in $V(\mathcal{M}') \cup Z$. By Lemma 3.2.3 each such path has length at least $2k - 1$ so we found a set S of $2k(|\mathcal{M}'| - 1)$ vertices such that S together with its neighborhood contains at most

$$\begin{aligned} |S| + |N(S)| &\leq |Z| + |V(\mathcal{M}')| \\ &< 2|V(\mathcal{M}')|/s + |V(\mathcal{M}')| \leq 10|V(\mathcal{M}')|/9 \\ &< \frac{10}{9}(|V(\mathcal{M})| + (6/\alpha + 1)|\mathcal{M}|) \\ &< 5|V(\mathcal{M})|/\alpha < gs \end{aligned}$$

vertices. Here in the next to last inequality we use the fact that $\mathcal{M}$ contains no path of length 0, implying that $|\mathcal{M}| \leq |V(\mathcal{M})|/2$. Now, if S contained at least g elements, then taking a subset $S' \subseteq S$ of size g would lead to a contradiction via the small expansion property: $gs = |S'|s \leq |N(S') \cup S'| \leq |N(S) \cup S| < gs$. Hence $|S| < g$ and then the small expansion property implies that

$$2sk(|\mathcal{M}'| - 1) = s|S| \leq |N(S)| < 5|V(\mathcal{M})|/\alpha.$$

$\square$

The proof of the main theorem is based on the ingenious rotation-extension technique, developed by Pósa [78], and applied later in a multitude of papers on Hamiltonicity, mostly of random or pseudorandom graphs (see for example [15], [37], and [64]).

Let G be a graph and let $P_0 = (v_1, v_2, \ldots, v_q)$ be a path in G. If $1 \leq i \leq q-2$ and (v_q, v_i) is an edge of G, then there exists a path $P' = (v_1 v_2 \ldots v_i v_q v_{q-1} \ldots v_{i+1})$ in G. P' is called a *rotation* of P_0 with *fixed endpoint* v_1 and *pivot* v_i. The edge (v_i, v_{i+1}) is called the *broken* edge of the rotation. We say that the segment $v_{i+1} \ldots v_q$ of P_0 is *reversed* in P'. In case the new endpoint v_{i+1} has a neighbor v_j such that $j \notin \{i, i+2\}$, then we can rotate P' further to obtain more paths of maximum length. We use rotations together with properties S and L to find a path on the same vertex set as P_0 with large rotation endpoint sets.

The next lemma is a slight strengthening of Claim 2.2 from [52] with a similar proof. It shows that in any graph having the small and large expansion properties for any path P_0 and its endpoint v_1 many other endpoints can be created by a small number of rotations with fixed endpoint v_1. In our setting we must also care about not breaking any of the edges from a small "forbidden" set F.

Lemma 3.2.5. *Let $G = (V, E)$ be a graph on n vertices that satisfies $S(s, g)$ and $L(l)$ with $s \geq 21$, $sg/3 > l$, and $l \leq n/24$. Let $P_0 = (v_1, v_2, \ldots, v_q)$ be a path in G and $F \subseteq E(P_0)$ with $|F| \leq s/24 - 1/2$. Denote by $B(v_1) \subset V(P_0)$ the set of all vertices $v \in V$ for which there is a v_1v-path on the vertex set $V(P_0)$ which can be obtained from P_0 by at most $3\frac{\log n}{\log s}$ rotations with fixed endpoint v_1 not breaking any of the edges of F. Then $B(v_1)$ satisfies one of the following properties:*

- *there exists a vertex $v \in B(v_1)$ with a neighbor outside $V(P_0)$, or*
- *$|B(v_1)| \geq n/3$.*

Proof Assume $B(v_1)$ does not have the first property (i.e., for every $v \in B(v_1)$ it holds that $N(v) \subseteq V(P_0)$).

Let t_0 be the smallest integer such that $\left(\frac{s}{3}\right)^{t_0 - 2} \geq g$; note that $t_0 \leq \frac{\log g}{\log(s/3)} + 3 \leq 3\frac{\log n}{\log s}$, because $21 \leq s$.

We construct a sequence of sets $S_0, \dots, S_{t_0} \subseteq B(v_1) \subseteq V(P_0) \setminus \{v_1\}$ of vertices, such that for every $0 \le t \le t_0$ and every $v \in S_t$, v is the endpoint of a path which can be obtained from P_0 by a sequence of t rotations with fixed endpoint v_1, such that for every $0 \le i < t$, the non-v_1-endpoint of the path after the ith rotation is contained in S_i. Moreover, $|S_t| = \left(\frac{s}{3}\right)^t$ for every $t \le t_0 - 3$, $|S_{t_0-2}| = g$, $|S_{t_0-1}| = l$, and $|S_{t_0}| \ge n/3$. Furthermore, for any $1 \le t \le t_0$, $V(F) \cap S_t = \emptyset$, and hence no edge from F got broken in any of the rotations.

We construct these sets by induction on t. For $t = 0$, one can choose $S_0 = \{v_q\}$ and all requirements are trivially satisfied.

Let now t be an integer with $0 < t \le t_0 - 2$ and assume that the sets $S_0, \dots, S_{t-1}$ with the appropriate properties have already been constructed. We will now construct S_t. Let

$$T = \{v_i \in N(S_{t-1}) : v_{i-1}, v_i, v_{i+1} \notin \bigcup_{j=0}^{t-1} S_j \cup V(F)\}$$

be the set of potential pivots for the tth rotation, and notice that $T \subset V(P_0)$ due to our assumption, since $T \subseteq N(S_{t-1})$ and $S_{t-1} \subseteq B(v_1)$. Assume now that $v_i \in T$, $y \in S_{t-1}$ and $(v_i, y) \in E$. Then, by the induction hypothesis, a $v_1 y$-path Q can be obtained from P_0 by $t-1$ rotations not breaking any edge from F such that after the jth rotation, the non-v_1-endpoint is in S_j for every $0 \le j \le t-1$. Each such rotation breaks an edge which is incident with the new endpoint, obtained in that rotation. Since v_{i-1}, v_i, v_{i+1} are not endpoints after any of these $t-1$ rotations and also not in $V(F)$, both edges (v_{i-1}, v_i) and (v_i, v_{i+1}) of the original path P_0 must be unbroken and thus must be present in $Q \setminus F$.

Hence, rotating Q with pivot v_i will make either v_{i-1} or v_{i+1} an endpoint (which of the two, depends on whether the unbroken segment $v_{i-1}v_iv_{i+1}$ is reversed or not after the first $t-1$ rotations). Assume without loss of generality that the endpoint is v_{i-1}. We add v_{i-1} to the set $\hat{S}_t$ of new endpoints and say that v_i *placed* v_{i-1} *in* $\hat{S}_t$. The only other vertex that can place v_{i-1} in $\hat{S}_t$ is v_{i-2} (if it exists). Thus we have

$$\begin{aligned} |\hat{S}_t| &\ge \frac{1}{2}|T| \ge \frac{1}{2}\left(|N(S_{t-1})| - 3(1 + |S_1| + \ldots + |S_{t-1}| + 2|F|)\right) \\ &\ge \frac{s}{2}\left(\frac{s}{3}\right)^{t-1} - \frac{3}{2}\frac{(s/3)^t - 1}{s/3 - 1} - (s/8 - 3/2) \ge \left(\frac{s}{3}\right)^t, \end{aligned}$$

where in the third inequality we use the small expansion property for the set S_{t-1} (where $|S_{t-1}| \le g$ by the definition of t_0) and the last inequality follows since $s \ge 21$. Clearly we can delete arbitrary elements of $\hat{S}_t$ to obtain S_t of size $\left(\frac{s}{3}\right)^t$ if $t \le t_0 - 3$ and of size g if $t = t_0 - 2$. So the proof of the induction step is complete and we have constructed the sets $S_0, \dots, S_{t_0-2}$.

To construct S_{t_0-1} and S_{t_0} we use the same technique as above, only the calculations are slightly different. If $g = 1$, then $t_0 - 1 = 1$, and analogously to the above calculation we obtain $\hat{S}_1$ with $|\hat{S}_1| \ge s/3 \ge l$. Otherwise, for $g \ge 2$, since $|N(S_{t_0-2})| \ge sg$, we have

$$\begin{aligned} |\hat{S}_{t_0-1}| &\ge \frac{1}{2}|T| \ge \frac{1}{2}\left(|N(S_{t_0-2})| - 3(1 + |S_1| + \ldots + |S_{t_0-4}| + |S_{t_0-3}| + |S_{t_0-2}| + 2|F|)\right) \\ &\ge gs/2 - \frac{3}{2}\frac{(s/3)^{t_0-2} - 1}{s/3 - 1} - 3g/2 - (s/8 - 3/2) \\ &\ge gs/2 - 2 \cdot \left(\frac{s}{3}\right)^{t_0-3} - \frac{3}{2}g - (s/8 - 3/2) \\ &\ge gs/2 - 2g - \frac{3}{2}g - (s/8 - 3/2) \ge gs/3 > l, \end{aligned}$$

where the inequality in the last but one line and the last but one inequality follow since $s \ge 21$ and $g \ge 2$. We delete arbitrary elements of $\hat{S}_{t_0-1}$ to obtain S_{t_0-1} of size l.

For S_{t_0} the difference in the calculation comes from using the expansion guaranteed by the property L, rather than the property S. That is, we use the fact that $|N(S_{t_0-1})| \geq n - 2l$. Hence, we obtain

$$\begin{aligned} |S_{t_0}| &\geq \frac{1}{2}|T| \geq \frac{1}{2}\left(|N(S_{t_0-1})| - 3(1 + |S_1| + \ldots + |S_{t_0-2}| + |S_{t_0-1}| + 2|F|)\right) \\ &\geq \frac{n}{2} - l - 4g - \frac{3}{2}l - (s/8 - 3/2) \\ &> \frac{n}{3}, \end{aligned}$$

where the last inequality follows since $4g \leq sg/3 \leq l$, $s \leq sg \leq 3l$ and $l \leq n/24$.

The set S_{t_0} is by construction a subset of $B(v_1)$, concluding the proof of the lemma. □

Let H be a graph with a spanning path $P = (v_1, \ldots, v_m)$. For $2 \leq i < m$, let us define the auxiliary graph $H_i^+ = H_{v_i}^+$ by adding a vertex and two edges to H as follows: $V(H_i^+) = V(H) \cup \{w\}$, $E(H_i^+) = E(H) \cup \{(v_m, w), (v_i, w)\}$. Let $P_i = P_{v_i}$ be the spanning path of H_i^+ which we obtain from the path $P \cup \{(v_m, w)\}$ by rotating with pivot v_i. Note that the endpoints of P_i are v_1 and v_{i+1}.

For a vertex $v_i \in V(H)$, let S^{v_i} be the set of those vertices of $V(P) \setminus \{v_1\}$, which are endpoints of a spanning path of H_i^+ obtained from P_i by a series of rotations with fixed endpoint v_1.

A vertex $v_i \in V(P)$ is called a *bad initial pivot* (or simply a *bad vertex*) if $|S^{v_i}| < \frac{m}{43}$ and is called a *good initial pivot* (or a *good vertex*) otherwise. We can rotate P_i and find a large number of endpoints, provided that v_i is a good initial pivot.

Hefetz et al. [52] showed that H has many good initial pivots provided that property L is satisfied.

Fact 3.2.6 ([52] Lemma 2.3)**.** Let H be a graph satisfying $L(m/43)$ with a spanning path $P = (v_1, \ldots, v_m)$. Then

$$|R| \leq 7m/43,$$

where $R = R(P) \subseteq V(P)$ is the set of bad vertices.

With these statements in our toolbox, we can prove the following important technical lemma. It states that we can rotate a path until it can be extended, and still do not break too many of the important edges.

Lemma 3.2.7. For every sufficiently large n and every $s = s(n)$ with $s \geq 21$, in every s-expander graph G on n vertices every path P_0 in G has the following property. For every pair of sets $F \subset F' \subseteq E(P_0)$ of at most $|F| \leq s/24 - 1/2$ and $|F'| \leq \frac{n \log s}{9200 \log n}$ edges of P_0, there exists a path P' in G between some $x, y \in V(P_0)$, such that $V(P') = V(P_0)$, $F \subset E(P')$, $|F' \setminus E(P')| \leq 6 \log n / \log s$, and G contains the edge $\{x, y\}$, or the set $\{x, y\}$ has neighbors outside P'.

Proof Assume for the sake of contradiction that the statement is not true. Let $P_0 = (v_1, v_2, \ldots, v_q)$, and let $A_0 = B(v_1) \subset V(P_0)$ be the set corresponding to P_0 and F as in Lemma 3.2.5, meaning that for every $v \in B(v_1)$ there is a v_1v-path of maximum length which can be obtained from P_0 by at most $t_0 = 3\frac{\log n}{\log s}$ rotations with fixed endpoint v_1 not breaking any of the edges of F. Clearly, at most $3\frac{\log n}{\log s}$ edges from F' were broken by the rotations, thus by our assumption every $v \in A_0$ has no neighbors outside P_0, hence by Lemma 3.2.5 we obtain $|A_0| \geq n/3$. For every $v \in A_0$ fix a v_1v-path $P^{(v)}$ with the above properties and, again using our assumption and Lemma 3.2.5, construct sets $B(v)$, $|B(v)| \geq n/3$, of endpoints of paths with fixed endpoint v, obtained from the path $P^{(v)}$ by at most t_0 rotations not breaking any edge from F. To summarize, for every $a \in A_0$ and $b \in B(a)$ there is a path $P(a,b)$ joining a and b on the vertex set $V(P_0)$, which is obtainable from P_0 by at most $\rho := 2t_0 = \frac{6\log n}{\log s}$ rotations not breaking any of the edges from F. Moreover, this clearly entails $|V(P_0)| \geq n/3$.

We consider P_0 to be directed from v_1 to v_q and divided into 2ρ consecutive undirected vertex disjoint segments $I_1, I_2, \dots, I_{2\rho}$ of length at least $\lfloor |V(P_0)|/2\rho - 1\rfloor$ each. As every $P(a,b)$ is obtained from P_0 by at most ρ rotations, and every rotation breaks at most one edge of P_0, the number of segments of P_0 which also occur as segments of $P(a,b)$, although perhaps reversed, is at least ρ. We say that such a segment is *unbroken*. Although the segments themselves are undirected, they have an absolute orientation given to them by P_0, and another, relative to this one, given to them by $P(a,b)$, which we consider to be directed from a to b. We consider tuples $\sigma = (I_i, o_i, I_j, o_j)$, where I_i and I_j are unbroken segments of P_0, which occur in this order on $P(a,b)$, and o_i and o_j denote their corresponding relative orientation. We call such a tuple σ *unbroken*, and say that $P(a,b)$ *contains* σ.

For a given unbroken tuple σ, we consider the set $C(\sigma)$ of ordered pairs (a,b), $a \in A_0$, $b \in B(a)$, such that $P(a,b)$ contains σ.

The total number of unbroken tuples is at most $2^2(2\rho)_2$. Any path $P(a,b)$ contains at least ρ unbroken segments, and thus at least $\binom{\rho}{2}$ unbroken tuples. The average, over unbroken tuples, of the number of pairs (a,b) such that $P(a,b)$ contains a given unbroken tuple is therefore at least

$$\frac{n^2}{9} \cdot \frac{\binom{\rho}{2}}{2^2(2\rho)_2} \geq 0.003n^2.$$

Thus, there is an unbroken tuple σ_0 and a set $C = C(\sigma_0)$, $|C| \geq 0.003n^2$ of pairs (a,b), such that for each $(a,b) \in C$, the path $P(a,b)$ contains σ_0. Let $\hat{A} = \{a \in A_0 : C \text{ contains at least } 0.003n/2$ pairs with a as first element$\}$. Since $|A_0|, |B(a)| \leq n$, we have $0.003n^2 \leq |C| \leq |\hat{A}|n + n \cdot \frac{0.003n}{2}$, entailing $|\hat{A}| \geq 0.003n/2$. For every $a \in \hat{A}$, let $\hat{B}(a) = \{b : (a,b) \in C\}$. Then, by the definition of $\hat{A}$, for every $a \in \hat{A}$ we have $|\hat{B}(a)| \geq 0.003n/2$.

For an unbroken tuple $\sigma_0 = (I_i, o_i, I_j, o_j)$, we divide it into two oriented segments, $\sigma_0^1 = \vec{I_i}$ and $\sigma_0^2 = \vec{I_j}$, both of them maintaining the orientation in σ_0. Notice that for every $a \in \hat{A}$ and $b \in \hat{B}(a)$, in the path $P(a,b)$ the segment σ_0^1 comes before σ_0^2. For $i = 1,2$, let us denote by $|\sigma_0^i|$ the number of vertices in the segment σ_0^i. Then for both segments σ_0^1 and σ_0^2, we have that $|\sigma_0^i| > n/(7\rho)$. Let s_1 be the first vertex of σ_0^1, x be the last vertex of σ_0^1, and let y be the first vertex of σ_0^2 and s_2 be the last vertex of σ_0^2.

We construct a graph H_1 with $V(\sigma_0^1)$ as vertex set. The edge set of H_1 is defined as follows. First, we add all edges of $G[V(\sigma_0^1)]$, except for those that are incident with s_1, x or a vertex in $V(F')$. Further, we add all the edges from $E(\sigma_0^1)$. Note that all the edges in H_1 are also edges of G. By its construction, σ_0^1 is a spanning path in H_1 starting at s_1 and ending at x. Let us denote the path reversed to σ_0^1 (spanning path in H_1, starting at x and ending at s_1) by P. We would like to apply Fact 3.2.6 to H_1 with $m = |V(\sigma_0^1)|$ and P as the corresponding spanning path. The condition of the Fact holds since G satisfies property $L(l)$. Indeed, $l = \frac{n\log s}{3000\log n}$, $m > n/(7\rho) = \frac{n\log s}{42\log n}$, the edges of H_1 differ from the edges of G only at $V(F')$ and at the endpoints of the segment σ_0^1, and $|V(F') \cup \{x, s_1\}| \leq \frac{n\log s}{4600\log n} + 2$, implying $l + |V(F') \cup \{x, s_1\}| < \frac{n\log s}{43\cdot 42\log n} < m/43$. Notice that this are the lines justifying the choices of the integers 3000 and 9200 in the bounds for l and $|F'|$. Hence, H_1 satisfies $L(m/43)$, thus by Fact 3.2.6 at least a $\frac{36}{43}$-fraction of the vertices of H_1 are good.

For σ_0^2 we act similarly: construct a graph H_2 from σ_0^2 by adding all edges of G with both endpoints in the interior of σ_0^2 but not in $V(F')$ and edges from $E(\sigma_0^2)$ to H_2. Then σ_0^2 forms an oriented spanning path in H_2, starting at y and ending at the last vertex s_2 of σ_0^2. Again, due to property L, Lemma 3.2.6 applies here, so at least a $\frac{36}{43}$-fraction of the vertices of H_2 are good.

Recall that s_1 is the first vertex of σ_0^1. Since $|\hat{A}| \geq 0.003n/2 > l + 1$ and H_1 has at least

$\frac{36}{43}m > l + |V(F') \cup \{x, s_1\}|$ good vertices, there is an edge of G between a vertex $\hat{a} \in \hat{A} \setminus \{s_1\}$ and a good vertex $g_1 \in V(\sigma_0^1) \setminus (V(F') \cup \{x, s_1\})$.

Similarly, as $|\hat{B}(\hat{a})| \geq 0.003n/2 > l + 1$ and there are more than $l + |V(F') \cup \{y, s_2\}|$ good vertices in H_2, there is an edge from some $\hat{b} \in \hat{B}(\hat{a}) \setminus \{s_2\}$ to a good vertex $g_2 \in V(\sigma_0^2) \setminus (V(F') \cup \{y, s_2\})$.

Consider the path $P(\hat{a}, \hat{b})$ on the vertex set of P_0 connecting $\hat{a}$ and $\hat{b}$ and containing σ_0. The vertices x and y split this path into three sub-paths: R_1 from $\hat{a}$ to x, R_2 from y to $\hat{b}$ and R_3 from x to y. We will rotate R_1 with x as a fixed endpoint and R_2 with y as a fixed endpoint, making sure that no edge from F' gets broken. We will show that the obtained endpoint sets V_1 and V_2 are sufficiently large (clearly, they are disjoint). Then by property L there will be an edge of G between V_1 and V_2. Since we did not touch R_3, this edge closes the path into a cycle, contradicting the assumption from the beginning of the proof.

First we construct the endpoint set V_1, the endpoint set V_2 can be constructed analogously. Recall the notation from Fact 3.2.6: Let $H_{g_1}^+$ denote the graph we obtain from H_1 by adding the extra vertex w and the edges (w, g_1) and (w, s_1). The spanning path of $H_{g_1}^+$ obtained by rotating $P \cup \{(w, s_1)\}$ with fixed endpoint x at pivot g_1 is denoted by P_{g_1}. By the definition of a good vertex, the set S^{g_1} of vertices which are endpoints of a spanning path of $H_{g_1}^+$ that can be obtained from P_{g_1} by a sequence of rotations with fixed endpoint x, has at least $|\sigma_0^1|/43 > l$ vertices.

We claim that also in G, any vertex in S^{g_1} can be obtained as an endpoint by a sequence of rotations of R_1 with fixed endpoint x without breaking any edge from F'. The role of the vertex w will be played by $\hat{a}$ in G (note that we made sure that $\hat{a} \neq s_1$, so $\hat{a}$ is not contained in $V(\sigma_0^1)$). Hence, the edge $(\hat{a}, g_1)$ is present in G, while we will consider the edge $(\hat{a}, s_1)$ *artificial*.

For any endpoint $z \in S^{g_1}$ there is a sequence of pivots, such that performing the sequence of rotations with fixed endpoint x at these pivots results in an xz-path spanning $H_{g_1}^+$. We claim that in $G[V(R_1)]$ it is also possible to perform a sequence of rotations with the exact same pivot sequence and eventually to end up in an xz-path spanning $V(R_1)$. When performing these rotations, the subpath of R_1 that links $\hat{a}$ to s_1 corresponds to the artificial edge (w, s_1) in $H_{g_1}^+$.

Problems in performing these rotations in G could arise if a rotation is called for where (1) the pivot is connected to the endpoint of the current spanning path via an artificial edge of $H_{g_1}^+$: this rotation might not be possible in G as this edge might not exist in G, or (2) the broken edge is artificial: after such a rotation in G the endpoint of the new spanning path might be different from the one we have after performing the same rotation in $H_{g_1}^+$, or (3) the broken edge is in F'. However, the construction of $H_{g_1}^+$ ensures that these problems will never occur. Indeed, in all three cases (1), (2) and (3) the pivot vertex has an artificial edge or an edge from F' incident with it, while having degree at least 3 (as all pivots). However, both endpoints of an artificial edge and both endpoints of edges from $F' \cap H_{g_1}^+$ have degree 2 in $H_{g_1}^+$ (for this last assertion we use the fact that $g_1 \notin \{x, s_1\} \cup V(F')$; this is important as g_1 is the first pivot.)

Hence we have ensured that there is indeed a spanning path of $G[V(R_1)]$ from x to every vertex of $V_1 = S^{g_1}$ containing all edges from $E(R_1) \cap F'$.

Similarly, since there is an edge from $\hat{b}$ to a good vertex g_2 in H_2, $g_2 \notin V(F')$, we can rotate R_2, starting from this edge to get a set $V_2 = S^{g_2}$ of at least l endpoints not breaking any more edges from F'. In other words we have a spanning path of $G[V(R_2)]$ from y to every vertex of $V_2 = S^{g_2}$ containing the edges from $E(R_2) \cap F'$.

As we noted earlier, since by Fact 3.2.6 $|V_1|, |V_2| \geq \frac{m}{43} > l$, property $L(l)$ ensures that there is an edge between V_1 and V_2 in G, say $a'b' \in E(G)$ with $a' \in V_1$ and $b' \in V_2$. This contradicts our assumption, since the rotations we did to obtain $P(a', b')$ from P_0 did not break any edges from F and also all but at most $\rho \leq 6 \log n / \log s$ edges from F' are on the path $P(a', b')$. $\square$

We are now able to prove the main lemma, stating that for every matching there is a Hamilton cycle almost covering it. Notice that we use the same calligraphic letter $\mathcal{M}$ to denote a matching as for families of paths, since we are going to apply extension on the matching and hence we see it as a family of paths of length 1 each.

Lemma 3.2.8. For every constant $\alpha \in (0,1]$ and for every sufficiently large n the following holds. Let G be an n^α-expander graph on n vertices. For every matching $\mathcal{M}$ in G of size at most $|\mathcal{M}| \le \alpha^3 n/9200$ there exists a Hamilton cycle C in G with

$$|E(\mathcal{M}) \setminus E(C)| \le \left\lfloor \frac{1036|\mathcal{M}|}{\alpha^3 n^{\alpha/2}} \right\rfloor$$

.

Proof First we proceed inductively to construct a single path via (d,k)-extensions that contains most of the matching edges.

Using Lemma 3.2.4 with $s = n^\alpha$, $g = 4\alpha n^{1-\alpha}$, $l = \alpha n/3000$, $k = 1$, and setting $d = 6/\alpha$, we find a $(d,1)$-expansion $\mathcal{M}_2$ of $\mathcal{M}_1 = \mathcal{M}$ of size at most $\left\lfloor \frac{5|\mathcal{M}|}{\alpha n^\alpha} + 1 \right\rfloor$ containing all edges of $\mathcal{M}$.

For $i \ge 2$, given a family of vertex-disjoint non-trivial paths $\mathcal{M}_i$ of size

$$2 \le |\mathcal{M}_i| \le \left\lfloor \frac{45|\mathcal{M}|}{2\alpha^2 n^{i\alpha/2}} + 1 \right\rfloor$$

on at most

$$|V(\mathcal{M}_i)| \le (d+3)|\mathcal{M}|$$

vertices containing all but at most

$$(i-2)\frac{45|\mathcal{M}|}{\alpha^2 n^{\alpha/2}}$$

edges from $\mathcal{M}$, we construct a size-minimum $(d, n^{(i-1)\alpha/2})$-extension $\mathcal{M}_{i+1}$ of $\mathcal{M}_i$. Then $\mathcal{M}_{i+1}$ satisfies the above properties by construction: it contains all but at most

$$\begin{aligned} |E(\mathcal{M}) \setminus E(\mathcal{M}_{i+1})| &\le |E(\mathcal{M}_i) \setminus E(\mathcal{M}_{i+1})| + |E(\mathcal{M}) \setminus E(\mathcal{M}_i)| \\ &\le 2n^{(i-1)\alpha/2}(|\mathcal{M}_i| - 1) + (i-2)\frac{45|\mathcal{M}|}{\alpha^2 n^{\alpha/2}} \\ &\le (i-1)\frac{45|\mathcal{M}|}{\alpha^2 n^{\alpha/2}} \end{aligned}$$

edges from $\mathcal{M}$. Furthermore, since $\mathcal{M}_{i+1}$ was constructed from $\mathcal{M}$ by a series of (d,k)-extensions with varying k but fixed d, at most $d+1$ vertices on average were added for every path removed, thus $\mathcal{M}_{i+1}$ has at most $(d+3)|\mathcal{M}|$ vertices.

Finally, by Lemma 3.2.4 $\mathcal{M}_{i+1}$ has size at most

$$|\mathcal{M}_{i+1}| \le \left\lfloor \frac{5|V(\mathcal{M}_i)|}{2\alpha n^{(i-1)\alpha/2} n^\alpha} + 1 \right\rfloor \le \left\lfloor \frac{5(d+3)|\mathcal{M}|}{2\alpha n^{(i+1)\alpha/2}} + 1 \right\rfloor \le \left\lfloor \frac{45|\mathcal{M}|}{2\alpha^2 n^{(i+1)\alpha/2}} + 1 \right\rfloor.$$

The family $\mathcal{M}_{last}$, where $last \le 2/\alpha + 1 < 3/\alpha$ is the index we stop the induction with, contains only one path P. Let us apply Lemma 3.2.2 with $D = V(P)$. This is possible, since P contains at most $(d+3)|\mathcal{M}| < \alpha n = gs/4$ vertices. We obtain the corresponding sets Z and $U = V \setminus (D \cup Z)$ and conclude that the induced graph $G[U]$ satisfies the small expansion property $S(s/2, g/2)$. Theorem 2.5 from [52] states that for every choice of the expansion parameter r with $12 \le r \le \sqrt{n}$, every n-vertex graph G satisfying $S\left(r, \frac{n \log r}{r \log n}\right)$ and $L\left(\frac{n \log r}{1035 \log n}\right)$ is Hamiltonian. Hence, applying this statement to $G[U]$ with $s = n^\alpha$, $g = 4\alpha n/n^\alpha$ and $r = n^\alpha/2$, we see that $G[U]$ is Hamiltonian.

Furthermore, by Lemma 3.2.2 we know that

$$|Z| \le 2|V(P)|/n^\alpha. \tag{3.1}$$

Using the small expansion property of G, we obtain an edge between a vertex x in the $\lceil |\mathcal{M}|/n^{\alpha/2} \rceil$-end of P and a vertex y in U in the following way. Take a subset S of size $\min\{2\lceil |\mathcal{M}|/n^{\alpha/2} \rceil, 4\alpha n^{1-\alpha}\}$ of the $\lceil |\mathcal{M}|/n^{\alpha/2} \rceil$-end of P. If $2\lceil |\mathcal{M}|/n^{\alpha/2} \rceil > 4\alpha n^{1-\alpha} = g$, then by the small expansion property $|N(S)| \geq sg > |V(P) \cup Z|$. Otherwise, $|S| = 2\lceil |\mathcal{M}|/n^{\alpha/2} \rceil \leq 4\alpha n^{1-\alpha} = g$, and hence $|N(S)| \geq 2n^{\alpha}\lceil |\mathcal{M}|/n^{\alpha/2} \rceil > |V(P) \cup Z|$.

The vertex x breaks the path P into two subpaths, one of which contains all but at most $|\mathcal{M}|/n^{\alpha/2}$ edges from P. Connecting this path via the edge xy with a Hamilton path in $G[U]$, we create a path R containing all but at most

$$|\mathcal{M} \setminus E(R)| \leq \left\lfloor \frac{3}{\alpha} \cdot \frac{45|\mathcal{M}|}{\alpha^2 n^{\alpha/2}} \right\rfloor + \left\lfloor |\mathcal{M}|/n^{\alpha/2} \right\rfloor \leq \left\lfloor \frac{136|\mathcal{M}|}{\alpha^3 n^{\alpha/2}} \right\rfloor \tag{3.2}$$

edges from $\mathcal{M}$ and all vertices from U. Note that if $|\mathcal{M}| < \alpha^3 n^{\alpha/2}/136$ then $\left\lfloor \frac{136|\mathcal{M}|}{\alpha^3 n^{\alpha/2}} \right\rfloor = 0$, so that no edges of $\mathcal{M}$ are lost in forming R. Furthermore, notice that from (3.1) we have that R is missing only

$$n - |E(R)| \leq |Z| + |\mathcal{M}|/n^{\alpha/2} + 1 < 2|\mathcal{M}|/n^{\alpha/2} + 1. \tag{3.3}$$

vertices to be Hamiltonian.

We aim to use Lemma 3.2.7 to rotate/extend R into a Hamilton cycle without losing many edges of $\mathcal{M}$ that are already on it.

For this we set $P_0 = R$ and $F' = F = \mathcal{M}$, in case $|\mathcal{M}| < \alpha^3 n^{\alpha/2}/136$ (and thus $\mathcal{M}$ is contained in R by (3.2)). Otherwise, if $|\mathcal{M}| \geq \alpha^3 n^{\alpha/2}/136$, let $F = \emptyset$ and $F' = \mathcal{M} \cap E(R)$.

We will now use Lemma 3.2.7 iteratively, in each step rotating/extending our current path with an edge until it is spanning and then closing it into a Hamilton cycle. Notice that Lemma 3.2.7 can be applied throughout the process since $|F'| \leq |\mathcal{M}| < \alpha n/9200$ and $|F| = o(n^{\alpha})$.

Consider the x,y-path P' arising from the application of Lemma 3.2.7 to F, F' and P_0. If one of the two vertices x or y has neighbors outside P', we can extend P' with one more edge to obtain a longer path $\hat{P}$ containing P'. We update for our iteration $P_0 := \hat{P}$ and $F' := \mathcal{M} \cap E(\hat{P})$ and start the iteration step again. Notice that in this step, the size of F' decreased by at most $6/\alpha$. If x and y have no neighbors outside P', then there is a cycle C containing P'. If C is Hamilton, we stop the procedure since this is what we are aiming at. Otherwise, by the connectivity of G (guaranteed by properties S and L and stated implicitly in Observation 3.2.1), there is a vertex $w \in V(C)$ with a neighbor outside C, say $aw \in E(G)$, $a \in V(G) \setminus V(C)$. Notice that only one of the edges incident with w in C can be in F', since $F' \subset \mathcal{M}$ is a matching. Removing an edge incident with w in C which is not in F' and adding the edge aw, we obtain a path $\hat{P}$ of length $|\hat{P}| \geq |P'| + 1$ containing all edges from $F' \cap E(P')$. We update for our iteration $P_0 := \hat{P}$ and $F' := \mathcal{M} \cap E(\hat{P})$ and start the iteration step again. Notice that again, in this step, the size of F' decreased by at most $6/\alpha$.

Using (3.3), we see that after at most $n - |R| \leq 2|\mathcal{M}|/n^{\alpha/2} + 1$ steps the iteration ends and we obtain a Hamilton cycle C. If $|\mathcal{M}| < \alpha^3 n^{\alpha/2}/136$, then C contains all edges from $\mathcal{M}$. Otherwise, C contains all but at most

$$\begin{aligned} |\mathcal{M} \setminus E(C)| \leq |E(R) \setminus E(C)| + |\mathcal{M} \setminus E(R)| &\leq \left(2|\mathcal{M}|/n^{\alpha/2} + 1\right) \cdot \frac{6}{\alpha} + \frac{136|\mathcal{M}|}{\alpha^3 n^{\alpha/2}} \\ &= \frac{6}{\alpha} + \frac{12|\mathcal{M}|}{\alpha n^{\alpha/2}} + \frac{136|\mathcal{M}|}{\alpha^3 n^{\alpha/2}} < \frac{7}{\alpha} \cdot \frac{148|\mathcal{M}|}{\alpha^3 n^{\alpha/2}} \end{aligned}$$

edges from $\mathcal{M}$, completing the proof of the lemma. □

The following corollary condenses all the previous technical work. It states that every matching of an n^{α}-expander graph can be covered with a constant-size collection of Hamilton cycles.

Corollary 3.2.9. For every constant $\alpha, 0 < \alpha \le 1$, there exists n_0, such that for every $n \ge n_0$ the following holds. In every n^α-expander graph G on n vertices, for every matching M of G there exist at most $14000/\alpha^4$ Hamilton cycles such that M is contained in their union.

Proof We start by splitting M into at most $\lceil 4600/\alpha^3 \rceil < 4601/\alpha^3$ matchings of size at most $\alpha^3 n/9200$ each. For every such matching M_1 we set $i = 1$ and perform the following iterative procedure:

- take a Hamilton cycle C_i covering as many of the edges of M_i as possible. If $M_i \subset E(C_i)$, then we found our covering and finish the procedure.
- otherwise, we set $M_{i+1} := M_i \setminus E(C_i)$ and remark that by Lemma 3.2.8

$$|M_{i+1}| \le \frac{1036|M_i|}{\alpha^3 n^{\alpha/2}} \le \left(\frac{1036}{\alpha^3 n^{\alpha/2}}\right)^{i+1} |M_1|. \tag{3.4}$$

  Update $i := i + 1$ and start again from the first iteration step.

From (3.4) we see that after at most $\lfloor 2/\alpha + 1 \rfloor < 3/\alpha$ iteration steps, we get a collection of at most $3/\alpha$ Hamilton cycles $C_1, C_2, \ldots$ covering M_1. Hence, there exist a collection of at most $\frac{3}{\alpha} \cdot \frac{4601}{\alpha^3}$ Hamilton cycles covering M, implying the statement of the lemma. $\square$

We are now able to prove Theorem 3.1.2.

Proof We start by taking h disjoint Hamilton cycles into our covering. Removing the union of these cycles from G, we are left with a graph H of maximum degree exactly $\Delta(H) = \Delta(G) - 2h$. Using at most $2\Delta(H)$ colors we color the edges of H greedily, partitioning them into at most $2\Delta(H)$ matchings. By Corollary 3.2.9 for every of these matchings there exist $14000/\alpha^4$ Hamilton cycles covering it, completing the proof of the theorem.

$\square$

3.3 Proof of Theorem 1.2.3

In this section we derive the proof of Theorem 1.2.3 from Theorem 3.1.2 by checking that $G(n,p)$ is an s-expander a.a.s. with the appropriate choice of p and s. Notice that this choice of s is clearly not optimal, but suffices for our purposes.

Lemma 3.3.1. For every constant α with $0 < \alpha < 1$ and every function $p = p(n) \ge n^{\alpha-1}$, $G(n,p)$ is a $\sqrt[5]{np}$-expander a.a.s.

Proof Let $s = \sqrt[5]{np}$.

First we prove that $G(n,p)$ has the small expansion property $S(s, \frac{4n\log s}{s\log n})$.

Let $A \subset V(G(n,p))$ be an arbitrary subset of size at most $|A| \le \frac{4n\log s}{s\log n} \le \frac{4n}{5s}$. The random variable $|N(A)|$ is the sum of the $n - |A|$ characteristic variables of the events $v \in N(A)$ for $v \in V \setminus A$. Hence for the expectation we obtain

$$\begin{aligned}\mathbb{E}[|N(A)|] &= \sum_{v\in V\setminus A} \mathbf{Pr}[v \in N(A)] = (n - |A|)\left(1 - (1-p)^{|A|}\right) > (n - |A|)\frac{|A|p}{1+|A|p} \\ &\ge (1+o(1))n\frac{|A|p}{1+|A|p} \ge (1+o(1))|A|s\frac{np/s}{1+\frac{4}{5}np/s} = |A|s\left(\frac{5}{4} + o(1)\right).\end{aligned}$$

Here we first used the simple fact that $(1-p)^{|A|} < \frac{1}{1+|A|p}$, then a couple of times that $|A| \le \frac{4n}{5s}$. Since the elementary events $v \in N(A)$ that make up $|N(A)|$ are mutually independent the

Chernoff bound can be applied to estimate the probability that A is not expanding. We use the above estimate on $\mathbb{E}[|N(A)|]$ several times.

$$\mathbf{Pr}\left[|N(A)| \le s|A|\right] < \exp\left[-\frac{(\mathbb{E}[|N(A)|] - |A|s)^2}{2\mathbb{E}[|N(A)|]}\right] < \exp\left(-\frac{\left(\left(\frac{1}{5}+o(1)\right)\mathbb{E}[|N(A)|]\right)^2}{2\mathbb{E}[|N(A)|]}\right)$$
$$= \exp\left(-\left(\frac{1}{50}+o(1)\right)\mathbb{E}[|N(A)|]\right) < \exp\left(-\left(\frac{1}{40}+o(1)\right)|A|s\right).$$

By the union bound the probability that $G(n,p)$ does not satisfy property $S\left(s, \frac{4n\log s}{s\log n}\right)$ is bounded by

$$\mathbf{Pr}\left[\exists A \subset V, |A| \le \frac{4n}{5s} : \ |N(A)| \le s|A|\right] < \sum_{a=1}^{n}\binom{n}{a}\exp\left[-\left(\frac{1}{40}+o(1)\right)as\right]$$
$$< \sum_{a=1}^{\infty}\left(n\exp\left[-0.01n^{\alpha/5}\right]\right)^a = o(1).$$

To complete the proof we show that $G(n,p)$ has the large expansion property $L\left(\frac{\alpha n}{15000}\right)$.

Let $A, B \subseteq V(G(n,p))$ be fixed subsets of size $|A|, |B| \ge \frac{\alpha n}{15000}$ with $A \cap B = \emptyset$. Then we have that

$$\mathbf{Pr}[\text{there are no edges between } A \text{ and } B] = (1-p)^{|A||B|} < \exp\left(-\frac{\alpha^2 n^2 p}{15000^2}\right).$$

Using union bound over all pairs of such disjoint sets $A, B \subseteq V(G(n,p))$, we get the desired probability

$$\mathbf{Pr}\left[G(n,p) \text{ satisfies } L\left(\frac{\alpha n}{15000}\right)\right] \ge 1 - 4^n\exp\left(-\frac{\alpha^2 n^2 p}{15000^2}\right) = 1 - o(1),$$

proving the lemma. □

We are now able to prove Theorem 1.2.3 using Theorem 3.1.2.

Proof Let p be in the range of the theorem. By Lemma 3.3.1 $G(n,p)$ is an $n^{\alpha/5}$-expander a.a.s. We know from [62] that there exists a packing of $(1-o(1))np/2$ Hamilton cycles into $G(n,p)$. Finally, the maximum degree a.a.s. satisfies $\Delta(G(n,p)) = (1+o(1))np$. Hence, by Theorem 3.1.2, we obtain a covering of $G(n,p)$ by $(1-o(1))np/2 + 28000((1+o(1))np - 2(1-o(1))np/2)/(\alpha/5)^4 = (1+o(1))np/2$ Hamilton cycles, finishing the proof of the theorem. □

3.4 Concluding remarks and open questions

In this chapter we verified that the size of a largest Hamilton cycle packing and the size of a smallest Hamilton cycle covering are asymptotically equal a.a.s. in the random graph $G(n,p)$, provided $p \ge n^{\alpha-1}$ for an arbitrary constant $\alpha > 0$. Our result calls for at least two natural directions of possible improvement.

First of all, the only explanation why the lower bound on the edge probability needs to be at least $n^{\alpha-1}$ and the corresponding expansion factor s needs to be at least $n^{\alpha/5}$ is that for lower values of p and s most of our technical arguments would break down. We think these bounds on p and s are only artifacts of our proof. Since the minimum and maximum degrees of the random graph have to be asymptotically equal in order for the minimum Hamilton covering and the maximum Hamilton packing to be asymptotically of the same size, we need to assume that $p = \omega(\log n)/n$. We strongly believe though that Theorem 1.2.3 holds already whenever $p = \omega(\log n)/n$.

Conjecture 3.4.1. For any $p = \omega(\log n)/n$ the random graph $G(n,p)$ admits a covering of its edges with at most $(1+o(1))np/2$ Hamilton cycles a.a.s.

Even though [62] provides the corresponding packing result for this range, we were not able to extend our techniques to prove e.g. the analog of Lemma 3.2.8 for $\alpha = o(1)$, not to mention for $\alpha = (\log\log n + \omega(1))/\log n$.

Another direction in which Theorem 1.2.3 could be tightened is to make the statement exact instead of approximate. The trivial lower bound on the size of a Hamilton covering in terms of the maximum degree is $\lceil \Delta(G) \rceil$. In what range of p will this be tight.

Question 3.4.2. In what range of p does there exist a Hamilton covering of $G \sim G(n,p)$ of size $\lceil \Delta(G)/2 \rceil$ a.a.s.?

Recall that the analogous precise statement in terms of the minimum degree *is* true for Hamilton packings [61], [69]. To have a positive answer for the question, we clearly need to be above the Hamiltonicity threshold, but it is plausible that the statement is true immediately after that. Very recently, Hefetz, Kühn, Lapinskas and Osthus [54] answered Question 3.4.2 positively for $log^{117}n/n \leq p \leq 1 - n^{1/8}$.

The question of covering the edges of a graph by Hamilton cycles can also be considered in the *pseudorandom* setup. A graph G is called an (n,d,λ)-graph if it is d-regular on n vertices and the second largest absolute value of its eigenvalues is λ. The concept of (n,d,λ)-graphs is a common way to formally express pseudorandomness, as (n,d,λ)-graphs with $\lambda = o(d)$ behave in many ways as random graphs are expected to do. (See, e.g., [70] for a general discussion on pseudorandom graphs and (n,d,λ)-graphs.) Theorem 2 from [38] implies that for (n,d,λ)-graph with $d = \Theta(n)$ and $\lambda = o(d)$ there exists a Hamilton packing of size $d/2 - 3\sqrt{\lambda n} = d/2 - o(d)$. The expansion properties of (n,d,λ)-graphs are also well-known. For example, it is stated in Section 3.1 of [68], it is stated for example that every (n,d,λ)-graph is an $\left(\frac{(d-2\lambda)^2}{3\lambda^2}, \frac{\lambda^2 n}{d^2}\right)$-expander (translated into our notation). This implies that for every $\alpha > 0$, every (n,d,λ)-graph with $\lambda \leq \frac{d}{2n^{\alpha/2}}$ is an n^{α}-expander. Hence, Theorem 3.1.2 implies that any such (n,d,λ)-graph has a Hamilton covering of size $d/2 + o(d)$, which is of course asymptotically best possible. It would be interesting to decide whether a similar statement holds for sparser pseudorandom graphs, maybe as sparse as $d = n^{\epsilon}$, for arbitrarily small $\epsilon > 0$.

A Hamilton cycle is a particular spanning structure of the complete graph, which can be used to decompose its edges. A further group of problems related to our result is to determine the typical sizes of a largest packing and of a smallest covering of various other spanning structures in the random graph. Here often the corresponding decomposition result for the complete graph is not known or is just conjectured. Still asymptotic packing and covering results would be of interest, for example for trees of bounded maximum degree.

4 Biased games on random boards

4.1 Introduction

As we already mentioned, this chapter is based on joined work with Asaf Ferber, Michael Krivelevich, and Alon Naor [31].

4.1.1 Notation and terminology

Our graph-theoretic notation is standard and follows that of [91]. In particular, we use the following:

For a graph G, let $V = V(G)$ and $E = E(G)$ denote its sets of vertices and edges, respectively. For subsets $U, W \subseteq V$, and for a vertex $v \in V$, we denote by $E_G(U)$ all the edges with both endpoints in U, by $E_G(U, W)$ all the edges e with both endpoints in $U \cup W$, for which $e \cap U \neq \emptyset$ and $e \cap W \neq \emptyset$, and by $E_G(v, U)$ all the edges with one endpoint being v and the other endpoint in U. We further denote $e_G(U) := |E(U)|$, $e_G(U, W) := |E(U, W)|$ and $e_G(v, U) := |E(v, U)|$. For a subset $U \subseteq V(G)$ we denote by $N_G(U)$ the *external* neighborhood of U, that is: $N_G(U) := \{v \in V \setminus U : \exists u \in U \text{ s.t. } uv \in E\}$. For simplicity of notation, whenever the underlying graph is clear from the context we omit the graph from the index.

Assume that some Maker-Breaker game, played on the edge set of some graph G, is in progress. At any given moment during the game, we denote the graph formed by Maker's edges by M, the graph formed by Breaker's edges by B, and the edges of $G \setminus (M \cup B)$ by F. For any vertex $v \in V$, $d_M(v)$ and $d_B(v)$ denote the degree of v in M and in B, respectively. The edges of $G \setminus (M \cup B)$ are called *free edges*, and $d_F(v)$ denotes the number of free edges incident to v, for any $v \in V$.

Whenever we say that $G \sim G(n, p)$ *typically* has some property, we mean that G has that property with probability tending to 1 as n tends to infinity.

We use the following notation throughout this chapter:

$$f(n) := \frac{np}{\ln n}.$$

For the sake of simplicity and clarity of presentation, and in order to shorten some of our proofs, no real effort has been made here to optimize the constants appearing in our results. We also omit floor and ceiling signs whenever these are not crucial. Most of our results are asymptotic in nature and whenever necessary we assume that n is sufficiently large.

4.2 Auxiliary results

In this section we present some auxiliary results that will be used throughout the chapter.

4.2.1 Basic positional games results

The following fundamental theorem, due to Beck [8], is a useful sufficient condition for Breaker's win in the (a, b) game $(X, \mathcal{F})$. It will be used in the proof of Theorem 1.2.7.

Theorem 4.2.1 (**[8], Theorem 20.1**)**.** *Let X be a finite set and let $\mathcal{F} \subseteq 2^X$. Breaker, as a first or a second player, has a winning strategy in the (a,b) game $(X,\mathcal{F})$, provided that:*

$$\sum_{F\in\mathcal{F}} (1+b)^{-|F|/a} < \frac{1}{1+b}.$$

While Theorem 4.2.1 simply shows that Breaker can win certain games, the following lemma shows that Maker can win certain games quickly (see [8]):

Lemma 4.2.2 (Trick of fake moves)**.** *Let X be a finite set and let $\mathcal{F} \subseteq 2^X$. Let $b' < b$ be positive integers. If Maker has a winning strategy for the $(1,b)$ game $(X,\mathcal{F})$, then he has a strategy to win the $(1,b')$ game $(X,\mathcal{F})$ within $\left\lceil \frac{|X|}{b+1} \right\rceil$ moves.*

The main idea of the proof of Lemma 4.2.2 is that, in every move of the $(1,b')$ game $(X,\mathcal{F})$, Maker (in his mind) gives Breaker $b-b'$ additional board elements. The straightforward details can be found in [8].

Recall the classic *box game* which was first introduced by Chvátal and Erdős in [18]. In the Box Game $Box(m,\ell,b)$ there are m pairwise disjoint boxes $A_1,\dots,A_m$, each of size ℓ. In every round, the first player, called *BoxMaker*, claims b elements of $\bigcup_{i=1}^m A_i$ and then the second player, called *BoxBreaker*, destroys one box. BoxMaker wins the game $Box(m,\ell,b)$ if and only if he is able to claim all elements of some box before it is destroyed. We use the following theorem which was proved in [18]:

Theorem 4.2.3. *Let m,ℓ be two integers. Then, BoxMaker wins the game $Box(m,\ell,b)$ for every $b > \frac{\ell}{\ln m} + 1$.*

4.2.2 (R,c)-Expanders

Definition 4.2.4. For every $c > 0$ and every positive integer R we say that a graph $G = (V,E)$ is an (R,c)-*expander* if $|N(U)| \geq c|U|$ for every subset of vertices $U \subseteq V$ such that $|U| \leq R$.

In the proof of Theorem 1.2.7 Maker builds an expander and then he turns it into a Hamiltonian graph. In order to describe the relevant connection between Hamiltonicity and (R,c)-expanders, we need the notion of *boosters*.

Given a graph G, we denote by $\ell(G)$ the maximum length of a path in G.

Definition 4.2.5. For every non-Hamiltonian graph G, we say that a non-edge $uv \notin E(G)$ is a *booster* with respect to G, if either $G \cup \{uv\}$ is Hamiltonian or $\ell(G\cup\{uv\}) > \ell(G)$. We denote by $\mathcal{B}_G$ the set of boosters with respect to G.

The following is a well-known property of $(R,2)$-expanders (see e.g. [39]).

Lemma 4.2.6. *If G is a connected non-Hamiltonian $(R,2)$-expander, then $|\mathcal{B}_G| \geq R^2/2$.*

Our goal is to show that during a game on an appropriate graph G, assuming Maker can build a subgraph of G which is an $(R,2)$-expander, he can also claim sufficiently many such boosters, so that his $(R,2)$-expander becomes Hamiltonian. In order to do so, we need the following lemma:

Lemma 4.2.7. *Let $a > 0$ and $p > \frac{50000a\ln n}{n}$. Then $G \sim G(n,p)$ is typically such that every subgraph $\Gamma \subseteq G$ which is a non-Hamiltonian $(n/5,2)$-expander with $\frac{an\ln n}{2\ln\ln n} \leq |E(\Gamma)| \leq \frac{100an\ln n}{\ln\ln n}$ satisfies $|E(G)\cap\mathcal{B}_\Gamma| > \frac{n^2p}{100}$.*

Proof. First, notice that any $(n/5,2)$-expander is connected. Indeed, let C be a connected component of G. If $|C| \le n/5$ then clearly C has neighbors outside, a contradiction. Otherwise, since G is an $(n/5,2)$-expander, C must be of size at least $3n/5 > n/2$. Hence there is exactly one such component and G is connected. Now, fix a non-Hamiltonian $(n/5,2)$-expander Γ in the complete graph K_n. Then clearly $\mathbf{Pr}\,(\Gamma \subseteq G) = p^{|E(\Gamma)|}$. By definition, the set of boosters of Γ, $\mathcal{B}_\Gamma$, is a subset of the potential edges of G. Therefore, $|E(G)\cap \mathcal{B}_\Gamma| \sim \mathrm{Bin}\left(|\mathcal{B}_\Gamma|, p\right)$ and the expected number of boosters is $|\mathcal{B}_\Gamma|p \ge \frac{n^2p}{50}$ by Lemma 4.2.6. Now, by Lemma 1.3.1 we get that $\mathbf{Pr}\left(|E(G)\cap\mathcal{B}_\Gamma| \le \frac{n^2p}{100}\right) \le \exp\left(-\frac{n^2p}{8}\right)$. Running over all choices of Γ with $\frac{an\ln n}{2\ln\ln n} \le |E(\Gamma)| \le \frac{100an\ln n}{\ln\ln n}$ and using the union bound we get

$$
\begin{aligned}
&\mathbf{Pr}\left(\exists\, \Gamma \text{ such that } \Gamma \subseteq G,\ \frac{an\ln n}{2\ln\ln n} \le |E(\Gamma)| \le \frac{100an\ln n}{\ln\ln n}, \text{ and } |E(G)\cap\mathcal{B}_\Gamma| \le \frac{n^2p}{100}\right)\\
&\le \sum_{m=\frac{an\ln n}{2\ln\ln n}}^{\frac{100an\ln n}{\ln\ln n}} \binom{\binom{n}{2}}{m} p^m \exp\left(-\frac{n^2p}{400}\right)\\
&\le \sum_{m=\frac{an\ln n}{2\ln\ln n}}^{\frac{100an\ln n}{\ln\ln n}} \left(\frac{en^2p}{2m}\right)^m \exp\left(-\frac{n^2p}{400}\right)\\
&\le \sum_{m=\frac{an\ln n}{2\ln\ln n}}^{\frac{100an\ln n}{\ln\ln n}} \exp\left(m\ln\left(\frac{en^2p}{2m}\right) - \frac{n^2p}{400}\right) = \heartsuit
\end{aligned}
$$

To complete the proof we should show that $\heartsuit = o(1)$. For that goal we consider each of the cases $np = \omega\left(\ln^2 n\right)$ and $np = O\left(\ln^2 n\right)$ separately. For the former we have that

$$
\heartsuit \le \sum_{m=\frac{an\ln n}{2\ln\ln n}}^{\frac{100an\ln n}{\ln\ln n}} \exp\left(n\ln^2 n - \frac{n^2p}{400}\right) = o(1);
$$

and for the latter, recalling that $p > \frac{50000a\ln n}{n}$, we have

$$
\begin{aligned}
\heartsuit &\le \sum_{m=\frac{an\ln n}{\ln\ln n}}^{\frac{100an\ln n}{\ln\ln n}} \exp\left(\frac{100an\ln n}{\ln\ln n}\ln\left(\frac{enp\ln\ln n}{a\ln n}\right) - \frac{n^2p}{400}\right)\\
&\le \sum_{m=\frac{an\ln n}{2\ln\ln n}}^{\frac{100an\ln n}{\ln\ln n}} \exp\left(\frac{100an\ln n}{\ln\ln n}\ln\left(C\ln n\ln\ln n\right) - \frac{n^2p}{400}\right)\\
&= \sum_{m=\frac{an\ln n}{2\ln\ln n}}^{\frac{100an\ln n}{\ln\ln n}} \exp\left((1+o(1))\,100an\ln n - \frac{n^2p}{400}\right)\\
&< \frac{100an\ln n}{\ln\ln n}\exp\left(-24an\ln n\right) = o(1).
\end{aligned}
$$

This completes the proof. □

The following lemma shows that an (R,c)-expander with the appropriate parameters is also k-vertex-connected.

Lemma 4.2.8 ([10], Lemma 5.1)**.** *For every positive integer k, if $G=(V,E)$ is an (R,c)-expander such that $c \ge k$, and $Rc \ge \frac{1}{2}(|V|+k)$, then G is k-vertex-connected.*

4.2.3 Properties of $G \sim G(n,p)$

Throughout this chapter we use the following properties of $G \sim G(n,p)$:

Theorem 4.2.9. *Let $p \geq \frac{\ln n}{n}$ and recall our notation $f(n) := \frac{np}{\ln n}$. A random graph $G \sim G(n,p)$ is typically such that the following properties hold:*

(P1) For every $v \in V$, $d(v) \leq 4np$. For every $\alpha > 0$ there are only $o(n)$ vertices with degree at least $(1+\alpha)np$.
If $f(n) = \omega(1)$ then for every $0 < \alpha < 1$ and for every $v \in V$,

$$(1-\alpha)np \leq d(v) \leq (1+\alpha)np.$$

(P2) For every subset $U \subseteq V$, $e(U) \leq \max\left\{3|U|\ln n, 3|U|^2 p\right\}$.

(P3) For every subset $U \subseteq V$ of size $|U| \leq \frac{3n\ln\ln n}{\ln n}$, $e(U) \leq 100|U|f(n)\ln\ln n$.

(P4) Let $\varepsilon > 0$. For every constant $\alpha > 0$ and for every subset $U \subseteq V$ where $1 \leq |U| \leq \frac{\alpha}{p}$, $|N(U)| \geq \beta|U|np$, for $\beta = \frac{1-\sqrt{\frac{(2+\varepsilon)(\alpha+1)}{f(n)}}}{\alpha+1}$.

(P5) For every $U \subseteq V$, $\frac{1}{p} \leq |U| \leq \frac{n}{\ln n}$, $|N(U)| \geq n/4$.

(P6) Let $\varepsilon > 0$. For every $\alpha \geq \sqrt{\frac{4}{f(n)}} + \varepsilon$ and for every set $U \subseteq V$, the number of edges between the set and its complement U^c satisfies:

$$e(U,U^c) \geq (1-\alpha)|U|(n-|U|)p.$$

(P7) Let ε, α be two positive constants which satisfy $\alpha^2\varepsilon f(n) > 4$, and denote $m := \frac{\varepsilon n \ln\ln n}{\ln n}$. For every two disjoint subsets $A, B \subseteq V$ with $|A| = |B| = m$, $e(A,B) \geq (1-\alpha)m^2 p$.

(P8) $e(A,B) \geq (1-\alpha)|A||B|p$ for every two disjoint subsets $A, B \subseteq V$ with $|A| = \frac{10000n}{\ln\ln n}$, $|B| = n/10$ and for every $\alpha > 0$.

(P9) For every subset $U \subseteq V$ such that $1 \leq |U| \leq \frac{n}{\ln^2 n}$, and for every $\varepsilon > 0$, $\left|\left\{v \in V \setminus U : d(v,U) \leq \frac{\varepsilon np}{\ln n}\right\}\right| = (1-o(1))n$.

Proof. For the proofs of *(P4)*,*(P5)* below we will use the following:

Let $U \subseteq V$. For every vertex $v \in V \setminus U$ we have that $\mathbf{Pr}(v \in N(U)) = 1-(1-p)^{|U|}$ independently of all other vertices. Therefore $|N(U)| \sim Bin\left(n-|U|, 1-(1-p)^{|U|}\right)$. Notice that for any $0 < p < 1$ (all the properties above trivially hold for $p = 1$) and for any positive integer k we have the following variation of Bernoulli's inequality: $(1-p)^{-k} \geq 1+kp$. Therefore, $\left(1-(1-p)^{|U|}\right) \geq \left(1-\frac{1}{1+|U|p}\right) = \frac{|U|p}{1+|U|p}$. It follows that:

$$\mathbb{E}\left(|N(U)|\right) = (n-|U|)\left(1-(1-p)^{|U|}\right) \geq \frac{(n-|U|)\,|U|p}{1+|U|p}. \tag{4.1}$$

(P1) For every $v \in V$, since $d(v) \sim Bin(n-1,p)$, it follows by Lemma 1.3.1 that

$$\mathbf{Pr}\left(d(v) \geq 4np\right) \leq \left(\frac{enp}{4np}\right)^{4np} < e^{-1.2np} \leq e^{-1.2\ln n} = n^{-1.2}.$$

Applying the union bound we get that

$$\mathbf{Pr}\left(\exists v \in V \text{ with } d(v) \geq 4np\right) \leq n \cdot n^{-1.2} = o(1).$$

Now let $\alpha > 0$. By Lemma 1.3.1 we get that for every $v \in V$:

$$\mathbf{Pr}\left(d(v) > (1+\alpha)np\right) \leq \exp\left(-\alpha' np\right) \leq n^{-\alpha'},$$

for some constant α'. Denote by S the set of all vertices with such degree. $\mathbb{E}(|S|) \leq n^{1-\alpha'}$. $|S|$ is a nonnegative random variable, so by Markov's inequality we get that:

$$\mathbf{Pr}\left(|S| > n^{1-\frac{\alpha'}{2}}\right) \leq \frac{n^{1-\alpha'}}{n^{1-\frac{\alpha'}{2}}} = n^{-\frac{\alpha'}{2}} = o(1).$$

Therefore, w.h.p. $|S| \leq n^{1-\frac{\alpha'}{2}} = o(n)$.

Assume now that $f(n) = \omega(1)$, and let $0 < \alpha < 1$ be a constant. By Lemma 1.3.1 and the union bound we get that

$$\mathbf{Pr}\left(\exists v \in V \text{ with } d(v) \geq (1+\alpha)np\right) \leq n \exp\left(-\frac{\alpha^2}{3}np\right)$$

$$= n \exp\left(-\frac{\alpha^2}{3} f(n) \ln n\right) = n^{-\omega(1)} = o(1).$$

The lower bound is achieved in a similar way.

(P2) Since $e(U) \sim Bin\left(\binom{|U|}{2}, p\right)$, using Lemma 1.3.1 and the union bound we get that:

$$\begin{aligned}
&\mathbf{Pr}\left(\exists U \subseteq V \text{ with } e(U) > \max\left\{3|U|\ln n, 3|U|^2 p\right\}\right) \\
&\leq \sum_{t=1}^{\frac{\ln n}{p}} \binom{n}{t} \left(\frac{e\binom{t}{2}p}{3t\ln n}\right)^{3t\ln n} + \sum_{t=\frac{\ln n}{p}}^{n} \binom{n}{t} \left(\frac{e\binom{t}{2}p}{3t^2 p}\right)^{3t^2 p} \\
&\leq \sum_{t=1}^{\frac{\ln n}{p}} \left[n\left(\frac{tp}{2\ln n}\right)^{3\ln n}\right]^t + \sum_{t=\frac{\ln n}{p}}^{n} \left[n\left(\frac{1}{2}\right)^{3tp}\right]^t \\
&\leq \sum_{t=1}^{n} \left(\frac{e}{8}\right)^{t\ln n} \leq \sum_{t=1}^{n} n^{-t} = o(1).
\end{aligned}$$

(P3) Let $U \subset V$ be a subset of size at most $\frac{3n\ln\ln n}{\ln n}$. Since $e(U) \sim \mathrm{Bin}\left(\binom{|U|}{2}, p\right)$, by Lemma 1.3.1 we get that

$$\mathbf{Pr}\left(e\left(U\right) \geq 10|U| f\left(n\right) \ln\ln n\right) \leq \left(\frac{e|U|^2 p}{20|U| f\left(n\right) \ln\ln n}\right)^{10|U| f(n) \ln\ln n}.$$

Applying the union bound we get that

$$\mathbf{Pr}\left(\exists\, U \text{ such that } |U| \leq \frac{3n\ln\ln n}{\ln n} \text{ with } e\left(U\right) \geq 10|U| f\left(n\right) \ln\ln n\right)$$

$$
\begin{aligned}
&\leq \sum_{k=1}^{\frac{3n\ln\ln n}{\ln n}} \binom{n}{k} \left(\frac{ek^2p}{20kf(n)\ln\ln n}\right)^{10kf(n)\ln\ln n} \\
&\leq \sum_{k=1}^{\frac{3n\ln\ln n}{\ln n}} \left[\frac{en}{k}\left(\frac{ekp}{20f(n)\ln\ln n}\right)^{10f(n)\ln\ln n}\right]^k \\
&= \sum_{k=1}^{\frac{3n\ln\ln n}{\ln n}} \left[\frac{e^2np}{20f(n)\ln\ln n}\left(\frac{ekp}{20f(n)\ln\ln n}\right)^{10f(n)\ln\ln n-1}\right]^k \\
&\leq \sum_{k=1}^{\frac{3n\ln\ln n}{\ln n}} \left[\frac{e^2\ln n}{20\ln\ln n}\left(\frac{3enp\ln\ln n}{20f(n)\ln n\ln\ln n}\right)^{10f(n)\ln\ln n-1}\right]^k \\
&\leq \sum_{k=1}^{\frac{3n\ln\ln n}{\ln n}} \left[\frac{e^2\ln n}{20\ln\ln n}\left(\frac{3e}{20}\right)^{10f(n)\ln\ln n-1}\right]^k \\
&= o(1).
\end{aligned}
$$

(P4) Since $n-|U|=(1-o(1))n$ in this range, by (4.1) we have that:

$$\mathbb{E}(|N(U)|) \geq \frac{(n-|U|)|U|p}{1+|U|p} \geq (1-o(1))\frac{|U|np}{\alpha+1}.$$

By Lemma 1.3.1 we have that for any $\delta>0$:

$$\mathbf{Pr}(|N(U)|<(1-\delta)\mathbb{E}(|N(U)|)) \leq e^{-\frac{\delta^2}{2}\mathbb{E}(|N(U)|)} \leq e^{-\alpha'|U|np},$$

where $\alpha'=\frac{\delta^2}{(2+o(1))(\alpha+1)}$. Now, by taking $\delta=\sqrt{\frac{(2+\varepsilon)(\alpha+1)}{f(n)}}$ (for some $\varepsilon>0$) we get that $\alpha' f(n)>1+\frac{\varepsilon}{3}$, and so by applying the union bound we get that:

$$\mathbf{Pr}\left(\exists U\subseteq V,\ |U|\leq\frac{\alpha}{p},\ |N(U)|<(1-\delta)\mathbb{E}(|N(U)|)\right) \leq \sum_{k=1}^{\alpha/p}\binom{n}{k}e^{-\alpha' knp} \leq \sum_{k=1}^{\alpha/p}\left[ne^{-\alpha' f(n)\ln n}\right]^k = o(1).$$

Therefore, w.h.p. for every U of size at most α/p, we obtain $|N(U)|\geq(1-\delta)\mathbb{E}(|N(U)|)\geq \beta|U|np$ with $\beta=\frac{1-\sqrt{\frac{(2+\varepsilon)(\alpha+1)}{f(n)}}}{\alpha+1}$.

(P5) Let $\frac{1}{p}\leq|U|\leq\frac{n}{\ln n}$. By (4.1), $\mathbb{E}(|N(U)|)\geq\frac{(n-|U|)|U|p}{1+|U|p}\geq n/3$.
By Lemma 1.3.1 we have that $\mathbf{Pr}(|N(U)|\leq n/4)\leq e^{-0.01n}$.
Applying the union bound we get that

$$\mathbf{Pr}(\exists\text{ such }U)\leq\sum_{k=1/p}^{n/\ln n}\binom{n}{k}e^{-0.01n}\leq n\binom{n}{\frac{n}{\ln n}}e^{-0.01n}$$

$$\leq n(e\ln n)^{\frac{n}{\ln n}}e^{-0.01n}=n\exp\left(\frac{n}{\ln n}\ln(e\ln n)-0.01n\right)=o(1).$$

(P6) Assume first that $|U|\leq n/2$, otherwise switch the roles of U and U^c. Since every edge between U and U^c is chosen independently, $e(U,U^c)\sim Bin(|U||U^c|,p)$. By Lemma 1.3.1 we have that for given $\alpha>0$ and $U\subseteq V$:

$$\mathbf{Pr}(e(U,U^c)<(1-\alpha)|U|(n-|U|)p)\leq\exp\left(-\frac{\alpha^2}{2}|U|(n-|U|)p\right)$$

$$\leq \exp\left(-\frac{\alpha^2}{4}|U|np\right) \leq \exp\left(-\left(\frac{1}{f(n)}+\delta\right)|U|np\right)$$
$$= \exp\left(-|U|\left(\ln n + \delta np\right)\right),$$

for some $\delta = \delta(\varepsilon) > 0$. By the union bound we get that:

$$\mathbf{Pr}\left(\exists \text{ such } U\right) \leq \sum_{k=1}^{n/2} \binom{n}{k} \exp\left(-k\left(\ln n + \delta np\right)\right) \leq \sum_{k=1}^{n/2} \left[n \exp\left(-\ln n - \delta np\right)\right]^k$$

$$= \sum_{k=1}^{n/2} \left(n^{-\delta f(n)}\right)^k = o(1).$$

(P7) Similarly to *(P6)*, given $A, B \subset V$, $|A| = |B| = m$, $e(A,B) \sim Bin\left(m^2, p\right)$. Therefore, by Lemma 1.3.1 we have that:

$$\mathbf{Pr}\left(e(A,B) \leq (1-\alpha)\, m^2 p\right) \leq \exp\left(-\frac{\alpha^2}{2} m^2 p\right).$$

Applying the union bound we get that:

$$\mathbf{Pr}\left(\exists \text{ such } A, B\right) \leq \binom{n}{m}^2 \exp\left(-\frac{\alpha^2}{2} m^2 p\right) \leq \left[\left(\frac{en}{m}\right)^2 \exp\left(-\frac{\alpha^2}{2} mp\right)\right]^m$$

$$= \left[\left(\frac{e \ln n}{\varepsilon \ln\ln n}\right)^2 \exp\left(-\frac{\alpha^2}{2} \varepsilon f(n) \ln\ln n\right)\right]^m < \left(\frac{e}{\varepsilon \ln\ln n}\right)^{2m} = o(1).$$

(P8) Given subsets $A, B \subseteq V$ as described, since $e(A,B) \sim \mathrm{Bin}\left(|A||B|, p\right)$, by Lemma 1.3.1 we get that

$$\mathbf{Pr}\left(e(A,B) \leq (1-\alpha)\, |A||B| p\right) \leq \exp\left(-\frac{\alpha^2}{2} |A||B| p\right) = \exp\left(-\frac{\alpha' n^2 p}{\ln\ln n}\right),$$

for some constant α'. Applying the union bound we get that:

$$\mathbf{Pr}\left(\exists \text{ such } A, B\right) \leq \binom{n}{\frac{10000n}{\ln\ln n}} \binom{n}{n/10} \exp\left(-\frac{\alpha' n^2 p}{\ln\ln n}\right) \leq 4^n \exp\left(-\omega(n)\right) = o(1).$$

(P9) Suppose towards a contradiction that there exists a subset $U \subseteq V$ such that $1 \leq |U| \leq \frac{n}{\ln^2 n}$ and that there are $\Theta(n)$ vertices $v \in V \setminus U$ with $d(v, U) \geq \frac{\varepsilon np}{\ln n}$. Therefore, the average degree of the vertices in U is at least $\Omega\left(n \cdot \frac{np}{\ln n} \cdot \frac{1}{|U|}\right) = \Omega\left(np \ln n\right)$. But by *(P1)*, $d(v) \leq 4np$ for every $v \in V$ — a contradiction. Hence, $\left|\left\{v \in V \setminus U : d(v,U) \leq \frac{\varepsilon np}{\ln n}\right\}\right| = o(n)$. $\square$

The following lemma will be a key ingredient in the main proof of the next subsection.

Lemma 4.2.10. *Let $p = \omega\left(\frac{\ln n}{n}\right)$. Then $G \sim G(n,p)$ is typically such that the following holds:*

For every subset $J_N = \{v_1, \ldots, v_N\} \subseteq V$ we have that

$$\sum_{j=1}^{N} \frac{e(v_j, J_j)}{j} = o(np)$$

where $N = \frac{n}{\ln^3 n}$ and $J_j = \{v_1, \ldots, v_j\}$, $1 \leq j \leq N$.

Proof. Let $t = \lfloor \log_2 (N+1) \rfloor$. We have that:

$$\sum_{j=1}^{N} \frac{e(v_j, J_j)}{j} \leq \sum_{i=0}^{t} \sum_{j=2^i}^{\min\left(N, 2^{i+1}-1\right)} \frac{e(v_j, J_j)}{2^i} \leq \sum_{i=1}^{t} \frac{e\left(J_{2^{i+1}}\right)}{2^i} = \spadesuit.$$

Now, note that $|J_j| = j$ for every $1 \leq j \leq N$ and distinguish between the following two cases:

(i) $pn = \omega\left(\ln^2 n\right)$. In this case, using Property $(P2)$ of Theorem 4.2.9 we have that:

$$\spadesuit \leq \sum_{i=1}^{\log_2\left(\frac{\ln n}{p}\right)} \frac{3|J_{2^{i+1}}| \ln n}{2^i} + \sum_{i=\log_2\left(\frac{\ln n}{p}\right)}^{t} \frac{3|J_{2^{i+1}}|^2 p}{2^i}$$

$$\leq \sum_{i=1}^{\log_2\left(\frac{\ln n}{p}\right)} \frac{3 \cdot 2^{i+1} \ln n}{2^i} + \sum_{i=\log_2\left(\frac{\ln n}{p}\right)}^{t} \frac{3 \cdot 2^{2i+2} p}{2^i}$$

$$\leq \log_2\left(\frac{\ln n}{p}\right) 6 \ln n + 12p \cdot \frac{2^{t+1}-1}{2-1} \leq c_1 \ln^2 n + c_2 N p,$$

for some positive constants c_1 and c_2. This is clearly $o(np)$ as desired.

(ii) $pn = O\left(\ln^2 n\right)$. In this case we need a more careful calculation. First, we prove the following claim:

Claim 2. *If $np = O\left(\ln^2 n\right)$, then for every $c > 3$, $G \sim G(n,p)$ is typically such that $e(X) \leq c|X|$ for every subset $X \subseteq V$ of size $|X| \leq \frac{n}{\ln^3 n}$.*

Proof of Claim 2. Let $X \subset V$ be a subset of size at most $\frac{n}{\ln^3 n}$. Since $e(X) \sim \mathrm{Bin}\left(\binom{|X|}{2}, p\right)$, by Lemma 1.3.1 we get that $\mathbf{Pr}(e(X) \geq c|X|) \leq \left(\frac{e|X|^2 p}{c|X|}\right)^{c|X|}$. Applying the union bound we get that

$$\mathbf{Pr}\left(\exists X \text{ such that } |X| \leq \frac{n}{\ln^3 n} \text{ and } e(X) \geq c|X|\right)$$

$$\leq \sum_{k=1}^{\frac{n}{\ln^3 n}} \binom{n}{k} \left(\frac{ek^2p}{ck}\right)^{ck} \leq \sum_{k=1}^{\frac{n}{\ln^3 n}} \left[\frac{en}{k}\left(\frac{ekp}{c}\right)^c\right]^k$$

$$= \sum_{k=1}^{\frac{n}{\ln^3 n}} \left[\frac{e^2 np}{c}\left(\frac{ekp}{c}\right)^{c-1}\right]^k \leq \sum_{k=1}^{\frac{n}{\ln^3 n}} \left[O\left(\ln^2 n\right) O\left(\ln^{-1} n\right)^{c-1}\right]^k$$

$$= \sum_{k=1}^{\frac{n}{\ln^3 n}} \left[O(\ln n)^{3-c}\right]^k = o(1).$$

□

Now, applying Claim 2 with $c = 4$, we get that

$$\spadesuit \leq \sum_{i=1}^{t} \frac{4|J_{2^{i+1}}|}{2^i} \leq \sum_{i=1}^{t} 8 = O(\ln n) = o(np).$$

This completes the proof of Lemma 4.2.10. □

4.2.4 The minimum degree game

In the proof of Theorem 1.2.7, Maker has to build a suitable expander which possesses some relevant properties. The first step towards creating a good expander is to create a spanning subgraph with a large enough minimum degree. The following theorem was proved in [42]:

Theorem 4.2.11 (**[42], Theorem 1.3**). *Let $\varepsilon > 0$ be a constant. Maker has a strategy to build a graph with minimum degree at least $\frac{\varepsilon}{3(1-\varepsilon)} \ln n$ while playing against Breaker's bias of $(1-\varepsilon)\frac{n}{\ln n}$ on $E(K_n)$.*

In fact, the following theorem can be derived immediately from the proof of Theorem 4.2.11:

Theorem 4.2.12. *Let $\varepsilon > 0$ be a constant. Maker has a strategy to build a graph with minimum degree at least $c = c(n) = \frac{\varepsilon}{3(1-\varepsilon)} \ln n$ while playing against a Breaker's bias of $(1-\varepsilon)\frac{n}{\ln n}$ on $E(K_n)$. Moreover, Maker can do so within cn moves and in such a way that for every vertex $v \in V(K_n)$, at the same moment v becomes of degree c in Maker's graph, there are still $\Theta(n)$ free edges incident with v.*

Using Theorem 4.2.12, Krivelevich proved in [67] that Maker has a strategy to build a good expander. Here, we wish to prove an analog of Theorem 4.2.12 for $G(n,p)$:

Theorem 4.2.13. *Let $p = \omega\left(\frac{\ln n}{n}\right)$, $\varepsilon > 0$ and let $b = (1-\varepsilon)\frac{np}{\ln n}$. Then $G \sim G(n,p)$ is typically such that in the $(1,b)$ Maker-Breaker game played on $E(G)$, Maker has a strategy to build a graph with minimum degree $c = c(n) \le \frac{\varepsilon}{6}\ln n$. Moreover, Maker can do so within cn moves and in such a way that for every vertex $v \in V(G)$, at the same moment that v becomes of degree c in Maker's graph, at least $\varepsilon np/3$ edges incident with v are free.*

Proof of Theorem 4.2.13. The proof is very similar to the proof of Theorem 4.2.11 so we omit some of the calculations (for more details, the reader is referred to [42]). Since claiming an extra edge is never a disadvantage for any of the players, we can assume that Breaker is the first player to move. A vertex $v \in V$ is called *dangerous* if $d_M(v) < c$. The game ends at the first moment in which either none of the vertices is dangerous (and Maker won), or there exists a dangerous vertex $v \in V$ with less than $\varepsilon np/3$ free edges incident to it (and Breaker won). For every vertex $v \in V$ let $\mathsf{dang}(v) := d_B(v) - 2b \cdot d_M(v)$ be the *danger value* of v. For a subset $X \subseteq V$, we define $\overline{\mathsf{dang}}(X) = \frac{\sum_{v\in X}\mathsf{dang}(v)}{|X|}$ (the average danger of vertices in X).

The strategy proposed to Maker is the following one:

Maker's strategy S_M: As long as there is a vertex of degree less than c in Maker's graph, Maker claims a free edge vu for some v which satisfies $\mathsf{dang}(v) = \max\{\mathsf{dang}(u) : u \in V\}$ (ties are broken arbitrarily).

Suppose towards a contradiction that Breaker has a strategy S_B to win against Maker who plays according to the strategy S_M as suggested above. Let g be the length of this game and let $I = \{v_1, \dots, v_g\}$ be the multi-set which defines Maker's game, i.e, in his ith move, Maker plays at v_i (in fact, according to the assumption Maker does not make his gth move, so let v_g be the vertex which made him lose). For every $0 \le i \le g-1$, let $I_i = \{v_{g-i}, \dots, v_g\}$. Following the notation of [42], let $\mathsf{dang}_{B_i}(v)$ and $\mathsf{dang}_{M_i}(v)$ denote the danger value of a vertex $v \in V$ *directly before* Breaker's and Maker's ith move, respectively. Notice that in his last move, Breaker claims b edges to decrease the minimum degree of the free graph to at most $\varepsilon np/3$. In order to be able to do that, directly before Breaker's last move B_g, there must be a dangerous vertex v_g with $d_M(v_g) \le c-1$ and $d_F(v_g) \le \frac{\varepsilon np}{3} + b$. By $(P1)$ of Theorem 4.2.9 we can assume that $\delta(G) \ge \left(1-\frac{\varepsilon}{12}\right)np$. Therefore we have that $\mathsf{dang}_{B_g}(v_g) \ge \left(1-\frac{\varepsilon}{12}-\frac{\varepsilon}{3}\right)np - (c-1) - b - 2b(c-1) = \left(1-\frac{5}{12}\varepsilon\right)np - (c-1) - b(2c-1) > \left(1-\frac{3}{4}\varepsilon\right)np$.

Analogously to the proof of Theorem 4.2.11 in [42], we state the following lemmas which estimate the change of the average danger after each move of any of the players. In the first lemma we estimate the change after Maker's move:

Lemma 4.2.14. *Let i be an integer, $1 \leq i \leq g-1$. Then*
(i) if $I_i \neq I_{i-1}$, then $\overline{\mathsf{dang}}_{M_{g-i}}(I_i) - \overline{\mathsf{dang}}_{B_{g-i+1}}(I_{i-1}) \geq 0$, and
(ii) if $I_i = I_{i-1}$, then $\overline{\mathsf{dang}}_{M_{g-i}}(I_i) - \overline{\mathsf{dang}}_{B_{g-i+1}}(I_{i-1}) \geq \frac{2b}{|I_i|}$.

In the second lemma we estimate the change of the average danger during Breaker's moves:

Lemma 4.2.15. *Let i be an integer, $1 \leq i \leq g-1$. Then*
(i) $\overline{\mathsf{dang}}_{M_{g-i}}(I_i) - \overline{\mathsf{dang}}_{B_{g-i}}(I_i) \leq \frac{2b}{|I_i|}$, and
(ii) $\overline{\mathsf{dang}}_{M_{g-i}}(I_i) - \overline{\mathsf{dang}}_{B_{g-i}}(I_i) \leq \frac{b+e(v_{g-i},I_i)+a(i-1)-a(i)}{|I_i|}$, where $a(i)$ denotes the number of edges spanned by I_i which Breaker claimed in the first $g-i-1$ rounds.

Combining Lemmas 4.2.14 and 4.2.15, we get the following corollary which estimates the change of the average danger after a full round:

Corollary 4.2.16. *Let i be an integer, $1 \leq i \leq g-1$. Then*
(i) if $I_i = I_{i-1}$, then $\overline{\mathsf{dang}}_{B_{g-i}}(I_i) - \overline{\mathsf{dang}}_{B_{g-i+1}}(I_{i-1}) \geq 0$,
(ii) if $I_i \neq I_{i-1}$, then $\overline{\mathsf{dang}}_{B_{g-i}}(I_i) - \overline{\mathsf{dang}}_{B_{g-i+1}}(I_{i-1}) \geq -\frac{2b}{|I_i|}$, and
(iii) if $I_i \neq I_{i-1}$, then $\overline{\mathsf{dang}}_{B_{g-i}}(I_i) - \overline{\mathsf{dang}}_{B_{g-i+1}}(I_{i-1}) \geq -\frac{b+e(v_{g-i},I_i)+a(i-1)-a(i)}{|I_i|}$, where $a(i)$ denotes the number of edges spanned by I_i which Breaker took in the first $g-i-1$ rounds.

In order to complete the proof, we prove that before Breaker's first move, $\overline{\mathsf{dang}}_{B_1}(I_{g-1}) > 0$, thus obtaining a contradiction.

Let $N := \frac{n}{\ln^3 n}$. For the analysis, we split the game into two parts: the main game, and the end game which starts when $|I_i| \leq N$.

Let $|I_{g-1}| = r$ and let $i_1 < \ldots < i_{r-1}$ be those indices for which $I_{i_j} \neq I_{i_j-1}$. Note that $|I_{i_j}| = j+1$. Note also that since $I_{i_j-1} = I_{i_{j-1}}$ and $i_{j-1} \leq i_j - 1$, $a(i_j-1) \leq a(i_{j-1})$.

Recall that the danger value of v_g directly before B_g is at least

$$\mathsf{dang}_{B_g}(v_g) > \left(1 - \frac{3}{4}\varepsilon\right) np. \tag{4.2}$$

Assume first that $r < N$.

$$\begin{aligned}
\overline{\mathsf{dang}}_{B_1}(I_{g-1}) &= \overline{\mathsf{dang}}_{B_g}(I_0) + \sum_{i=1}^{g-1}\left(\overline{\mathsf{dang}}_{B_{g-i}}(I_i) - \overline{\mathsf{dang}}_{B_{g-i+1}}(I_{i-1})\right) \\
&\geq \overline{\mathsf{dang}}_{B_g}(I_0) + \sum_{j=1}^{r-1}\left(\overline{\mathsf{dang}}_{B_{g-i_j}}(I_{i_j}) - \overline{\mathsf{dang}}_{B_{g-i_j+1}}(I_{i_j-1})\right) \quad \text{[by Corollary 4.2.16}(i)] \\
&\geq \overline{\mathsf{dang}}_{B_g}(I_0) - \sum_{j=1}^{r-1}\frac{b + e\left(v_{g-i_j}, I_{i_j}\right) + a(i_j-1) - a(i_j)}{j+1} \quad \text{[by Corollary 4.2.16}(iii)] \\
&\geq \overline{\mathsf{dang}}_{B_g}(I_0) - b\ln r - \sum_{j=1}^{r-1}\frac{e\left(v_{g-i_j}, I_{i_j}\right)}{j+1} - \frac{a(0)}{2} + \sum_{j=2}^{r-1}\frac{a(i_{j-1})}{(j+1)j} + \frac{a(i_{r-1})}{r} \\
&\geq \overline{\mathsf{dang}}_{B_g}(I_0) - b\ln r - o(np) \quad \text{[by Lemma 4.2.10]} \\
&> \left(1 - \frac{3}{4}\varepsilon\right) np - (1 - \varepsilon + o(1))\, np \quad \text{[by (4.2)]} \\
&> 0.
\end{aligned} \tag{4.3}$$

Assume now that $r \geq N$.

$$
\begin{aligned}
\overline{\mathsf{dang}}_{B_1}(I_{g-1}) &= \overline{\mathsf{dang}}_{B_g}(I_0) + \sum_{i=1}^{g-1}\left(\overline{\mathsf{dang}}_{B_{g-i}}(I_i) - \overline{\mathsf{dang}}_{B_{g-i+1}}(I_{i-1})\right) \\
&\geq \overline{\mathsf{dang}}_{B_g}(I_0) + \sum_{j=1}^{r-1}\left(\overline{\mathsf{dang}}_{B_{g-i_j}}(I_{i_j}) - \overline{\mathsf{dang}}_{B_{g-i_j+1}}(I_{i_j-1})\right) \quad \text{[by Corollary 4.2.16}(i)\text{]} \\
&= \overline{\mathsf{dang}}_{B_g}(I_0) + \sum_{j=1}^{N-1}\left(\overline{\mathsf{dang}}_{B_{g-i_j}}(I_{i_j}) - \overline{\mathsf{dang}}_{B_{g-i_j+1}}(I_{i_j-1})\right) + \\
&\quad \sum_{j=N}^{r-1}\left(\overline{\mathsf{dang}}_{B_{g-i_j}}(I_{i_j}) - \overline{\mathsf{dang}}_{B_{g-i_j+1}}(I_{i_j-1})\right) \\
&\geq \mathsf{dang}_{B_g}(v_g) - \sum_{j=1}^{N-1}\frac{b}{j+1} - o(np) - \sum_{j=N}^{r-1}\frac{2b}{j+1} \quad \text{[by Corollary 4.2.16}(ii)\text{ and (4.3)]} \\
&\geq \left(1-\frac{3}{4}\varepsilon\right)np - b\ln n - o(np) - 2b\left(\ln n - \ln\frac{n}{\ln^3 n}\right) \quad \text{[by (4.2)]} \\
&= \left(1-\frac{3}{4}\varepsilon\right)np - (1-\varepsilon)np - o(np) - 6b\ln\ln n \\
&= \frac{\varepsilon}{4}np - o(np) \\
&> 0.
\end{aligned}
$$

This completes the proof. □

4.3 Maker-Breaker games on $G(n,p)$

4.3.1 Breaker's win

In this subsection we prove Theorem 1.2.6.

Chvátal and Erdős proved in [18] that playing on the edge set of the complete graph K_n, if Breaker's bias is $b = (1+\varepsilon)\frac{n}{\ln n}$, then Breaker is able to isolate a vertex in his graph and thus to win a lot of natural games such as the perfect matching game, the Hamiltonicity game and the k-connectivity game.

In their proof, Breaker wins by creating a large clique which is disjoint of Maker's graph and then playing the box game on the stars centered in this clique. Our proof is based on the same idea.

Proof of Theorem 1.2.6: First, we may assume that $p \geq \frac{\ln n}{n}$, since otherwise $G \sim G(n,p)$ typically contains isolated vertices and Breaker wins no matter how he plays. Now we introduce a strategy for Breaker and then we prove it is a winning strategy. At any point during the game, if Breaker cannot follow the proposed strategy then he forfeits the game. Breaker's strategy is divided into the following two stages:

Stage I: Throughout this stage Breaker maintains a subset $C \subseteq V$ which satisfies the following properties:

(*i*) $E_G(C) = E_B(C)$.

(*ii*) $d_M(v) = 0$ for every $v \in C$.

(*iii*) $d_G(v) \leq (1+\varepsilon/2)np$ for every $v \in C$.

Initially, $C = \emptyset$. In every move, Breaker increases the size of C by at least one. This stage ends after the first move in which $|C| \geq \frac{n}{\ln^2 n}$.

Stage II: For every $v \in C$, let $A_v = \{vu \in E(G) : vu \notin E(B)\}$. In this stage, Breaker claims all the elements of one of these sets.

It is evident that if Breaker can follow the proposed strategy, then he isolates a vertex in Maker's graph and wins the game. It thus remains to prove that Breaker can follow the proposed strategy. We consider each stage separately.

Stage I: Notice that in every move Maker can decrease the size of C by at most one. Hence, it is enough to prove that in every move Breaker is able to find at least two vertices which are isolated in Maker's graph and have bounded degree as required, and to claim all the free edges between them and C, as well as the edge between the two vertices, if it exists in G. For this it is enough to prove that Breaker can always find two vertices $u, v \in V \setminus C$ which have the proper degree in G and are isolated in Maker's graph, and such that $e(u, C)$, $e(v, C) \leq \frac{b-1}{2}$. Since this stage lasts $o(n)$ moves, and the number of vertices with too high degree in G is $o(n)$ by property $(P1)$ of Theorem 4.2.9, the existence of such vertices is trivial by property $(P9)$ of Theorem 4.2.9.

Stage II: Notice that $|C| \geq \frac{n}{\ln^2 n}$ and that $A_v \cap A_u = \emptyset$ for every two vertices $u \neq v$ in C. In addition, by the way Breaker chooses his vertices we have that $|A_v| \leq (1 + \varepsilon/2)\, np$ for every $v \in C$. Recall that $b = (1 + \varepsilon)\frac{np}{\ln n} > \frac{(1+\varepsilon/2)np}{\ln |C|} + 1$. Therefore, by Theorem 4.2.3 Breaker (as BoxMaker) wins the Box Game on these sets.

This completes the proof. □

4.3.2 Maker's win

In this subsection we prove Theorems 1.2.7 and 1.2.8. We start with providing Maker with a winning strategy in the Hamiltonicity game for each case (which implies the perfect matching game) and then we sketch the changes which need to be done in order to turn it into a winning strategy in the k-connectivity game as well.

Proof of Theorem 1.2.7. First we describe a strategy for Maker and then prove it is a winning strategy.

At any point during the game, if Maker is unable to follow the proposed strategy (including the time limits), then he forfeits the game. Maker's strategy is divided into the following three stages:

Stage I: Maker builds an $\left(\frac{10000n}{\ln\ln n}, 2\right)$-expander within $\frac{100n\ln n}{\ln\ln n}$ moves.

Stage II: Maker makes his graph an $(n/5, 2)$-expander within additional $\frac{300n\ln n}{\ln\ln n}$ moves.

Stage III: Maker makes his graph Hamiltonian by adding at most n boosters.

It is evident that if Maker can follow the proposed strategy without forfeiting the game he wins. It thus suffices to prove that indeed Maker can follow the proposed strategy. We consider each stage separately.

Stage I: In his first $\frac{100n\ln n}{\ln\ln n}$ moves, Maker creates a graph with minimum degree $c = c(n) = \frac{100\ln n}{\ln\ln n}$. Maker plays according to the strategy proposed in Theorem 4.2.13 except of the seemingly minor but crucial change that in every move, when Maker needs to claim an edge incident with a vertex v, Maker randomly chooses such a free edge. We prove that, with a positive probability, this non-deterministic strategy ensures that Maker's graph is an $\left(\frac{10000n}{\ln\ln n}, 2\right)$-expander and then, since our game is a perfect information game, we conclude that indeed there exists a deterministic such strategy for Maker. Recall that according to the strategy proposed in Theorem 4.2.13, at any move Maker claims a free edge vu with $\mathsf{dang}(v) = \max\{\mathsf{dang}(u) : u \in V\}$. In this case we say that the edge vu is *chosen* by v. We wish to show that the probability of having a subset $A \subseteq V$ with $|A| \leq \frac{10000n}{\ln\ln n}$ and $|N_M(A)| \leq 2|A| - 1$ is $o(1)$. To that end, we can assume that G satisfies all the properties listed in Theorem 4.2.9 and Theorem 4.2.13.

Assume that there exists a subset $A \subset V$ of size $|A| \le \frac{10000n}{\ln\ln n}$ such that after this stage $N_M(A)$ is contained in a set B of size at most $2|A|-1$. This implies that

$$|E_M(A, A\cup B)| \ge c|A|/2 = \frac{50|A|\ln n}{\ln\ln n}.$$

Recall that $f(n) := \frac{np}{\ln n}$. We distinguish between the following two cases:

Case I: At least $c|A|/4$ edges of Maker which are incident to A were chosen by vertices from A.

Notice that if $|A| \le \frac{n\ln\ln n}{\ln n}$, then there are at most $o(|A|)$ vertices $v \in A$ such that $e(v, A\cup B) = \Omega\left(f(n)(\ln\ln n)^2\right)$, since otherwise we have that $e(A\cup B) = \Omega\left(f(n)(\ln\ln n)^2|A|\right)$ which contradicts $(P3)$ of Theorem 4.2.9. Furthermore, if $\frac{n\ln\ln n}{\ln n} < |A| \le \frac{10000n}{\ln\ln n}$, then there are at most $o(|A|)$ vertices $v \in A$ such that $e(v, A\cup B) = \Omega(np)$ (follows from $(P2)$ of Theorem 4.2.9). Consider an edge $e = ab$ with $a \in A$ and $b \in A\cup B$ and assume that e has been chosen by a. Notice that by Theorem 4.2.13, when Maker chose e, the vertex a had at least $\varepsilon np/3$ free neighbors. Therefore, for at least $(1-o(1))|A|$ such vertices $a \in A$, the probability that Maker chose an edge with a second endpoint in $A\cup B$ is at most $\left(\frac{f(n)(\ln\ln n)^2}{\varepsilon np/3}\right) = \frac{3(\ln\ln n)^2}{\varepsilon \ln n}$ when $|A| \le \frac{n\ln\ln n}{\ln n}$ or an arbitrarily small constant $\delta > 0$ when $\frac{n\ln\ln n}{\ln n} < |A| \le \frac{10000n}{\ln\ln n}$. Therefore, the probability that all of Maker's edges incident to A were chosen in $A\cup B$ is at most $\left(\frac{3(\ln\ln n)^2}{\varepsilon\ln n}\right)^{(1-o(1))c|A|/4}$ for $|A| \le \frac{n\ln\ln n}{\ln n}$ and at most $\delta^{(1-o(1))c|A|/4}$ otherwise. Applying the union bound we get that the probability that there exists such A (with $|N_M(A)| \le 2|A|-1$ and at least $c|A|/4$ edges chosen by A) is at most

$$\begin{aligned}
&\sum_{|A|<\frac{n\ln\ln n}{\ln n}} \binom{n}{|A|}\binom{n}{2|A|-1}\left(\frac{3(\ln\ln n)^2}{\varepsilon\ln n}\right)^{(1-o(1))c|A|/4} + \sum_{|A|=\frac{n\ln\ln n}{\ln n}}^{\frac{10000n}{\ln\ln n}} \binom{n}{|A|}\binom{n}{2|A|-1}\delta^{(1-o(1))c|A|/4} \\
&\le \sum_{|A|<\frac{n\ln\ln n}{\ln n}} n^{3|A|}\left(\frac{3(\ln\ln n)^2}{\varepsilon\ln n}\right)^{\frac{24|A|\ln n}{\ln\ln n}} + \sum_{|A|=\frac{n\ln\ln n}{\ln n}}^{\frac{10000n}{\ln\ln n}} \left(\frac{e^3n^3}{4|A|^3}\right)^{|A|}\delta^{\frac{24|A|\ln n}{\ln\ln n}} \\
&\le \sum_{|A|<\frac{n\ln\ln n}{\ln n}} \left[n^3\exp\left(\frac{24\ln n}{\ln\ln n}\ln\left(\frac{3(\ln\ln n)^2}{\varepsilon\ln n}\right)\right)\right]^{|A|} + \sum_{|A|=\frac{n\ln\ln n}{\ln n}}^{\frac{10000n}{\ln\ln n}} \left(\alpha\frac{\ln^3 n}{(\ln\ln n)^3}\delta^{\frac{24\ln n}{\ln\ln n}}\right)^{|A|} \\
&\le \sum_{|A|<\frac{n\ln\ln n}{\ln n}} \left[n^3\exp\left(-(1-o(1))\,24\ln n\right)\right]^{|A|} + o(1) = o(1).
\end{aligned}$$

Case II: At least $c|A|/4$ edges of Maker which are incident to A were chosen by vertices from B.

As in Case I, notice that there are at most $o(|B|)$ vertices $v \in B$ such that $e(v, A) \ge f(n)(\ln\ln n)^2$ when $|B| \le \frac{2n\ln\ln n}{\ln n}$ and at most $o(|B|)$ vertices $v \in B$ such that $e(v, A) = \Omega(np)$ when $\frac{2n\ln\ln n}{\ln n} \le |B|\frac{20000n}{\ln\ln n}$. Similar to the previous case, with the only difference being that not all the edges which were chosen by vertices from B have to touch A, we get that the probability that all Maker's edges incident to A were chosen in $A\cup B$ is at most

$$\binom{c|B|}{c|A|/4}\left(\frac{3(\ln\ln n)^2}{\varepsilon\ln n}\right)^{(1-o(1))c|A|/4}$$

or

$$\binom{c|B|}{c|A|/4}\delta^{(1-o(1))c|A|/4},$$

for an arbitrarily small δ, for $|B| \leq \frac{2n \ln \ln n}{\ln n}$ or $\frac{2n \ln \ln n}{\ln n} \leq |B| \leq \frac{20000n}{\ln \ln n}$, respectively (the binomial coefficient corresponds to the number of possible choices of edges from $E_M(A, B)$ out of all edges chosen by vertices from B). Applying the union bound, similar to the computations in Case I, we get that the probability that there exists such A (with $|N_M(A)| \leq 2|A| - 1$ and at least $c|A|/4$ edges incident to A chosen by $N(A)$) is $o(1)$.

This completes the proof that Maker can build a $\left(\frac{10000n}{\ln \ln n}, 2\right)$-expander fast and thus is able to follow Stage I of the proposed strategy.

Stage II: It is enough to prove that Maker has a strategy to ensure that $E_M(A, B) \neq \emptyset$ for every two disjoint subsets $A, B \subseteq V$ of sizes $|A| = \frac{10000n}{\ln \ln n}$ and $|B| = n/10$. Indeed, if there exists a subset $X \subseteq V$ of size $\frac{10000n}{\ln \ln n} \leq |X| \leq n/5$ such that $|X \cup N(X)| < 3|X|$, then there exist two subsets $A \subseteq X$ and $B \subseteq V \setminus (X \cup N(X))$ with $|A| = \frac{10000n}{\ln \ln n}$ and $|B| = n/10$ such that $E_M(A, B) = \emptyset$.

Recall that by Property $(P8)$ of Theorem 4.2.9, $G \sim G(n, p)$ is typically such that for every two such subsets $A, B \subseteq V$ and for sufficiently small $\alpha > 0$ we have that

$$e_G(A, B) \geq (1 - \alpha)|A||B|p \geq \frac{999n^2p}{\ln \ln n}.$$

To achieve his goal for this stage, Maker can use the trick of fake moves and to play as $\mathcal{F}$-Breaker in the $\left(\frac{np \ln \ln n}{100 \ln n}, 1\right)$ Maker-Breaker game where the winning sets are

$$\mathcal{F} = \left\{E_F(A, B) : A, B \subseteq V,\ A \cap B = \emptyset,\ |A| = \frac{10000n}{\ln \ln n} \text{ and } |B| = n/10\right\}.$$

Notice that since so far Breaker has played at most $\frac{100n \ln n}{\ln \ln n}$ rounds, we get that $e_F(A, B) \geq \frac{899n^2p}{\ln \ln n}$ for every $A, B \subset V(G)$ of sizes $|A| = \frac{10000n}{\ln \ln n}$ and $|B| = n/10$. Finally, since the following inequality holds

$$\binom{n}{\frac{10000n}{\ln \ln n}}\binom{n}{n/10} 2^{-89900n \ln n/(\ln \ln n)^2} \leq 4^n 2^{-\omega(n)} = o(1)$$

it follows by Theorems 4.2.1 and 4.2.2 that indeed Maker (or, as we called him in this stage, $\mathcal{F}$-Breaker) can achieve his goals for this stage within $\frac{e(G)}{np \ln \ln n/100 \ln n} < \frac{300n \ln n}{\ln \ln n}$ moves (recall that $e(G) \leq 3n^2p$).

Stage III: So far Maker has played at most $\frac{400n \ln n}{\ln \ln n}$ moves (and at least $\frac{50n \ln n}{\ln \ln n}$ moves) and his graph is an $(n/5, 2)$-expander. Notice that for the choice $a = 2$ Lemma 4.2.7 holds. In addition, Maker and Breaker together claimed $o\left(n^2p\right)$ edges of G. Therefore, there are still $\Theta\left(n^2p\right)$ free boosters in G, so Maker can easily claim n boosters and to turn his graph into a Hamiltonian graph.

This completes the proof that Maker wins the game $\mathcal{H}(G)$ (and of course also the game $\mathcal{M}(G)$). □

Now, we briefly sketch the proof of Theorem 1.2.8.

Sketch of proof of Theorem 1.2.8. Let $K > 10^5$, $p = \frac{Kn}{\ln n}$ and $G \sim G(n, p)$. The upper bound on b^* is obtained immediately from Theorem 1.2.6. We wish to show that G is typically such that given $b \leq K/10$, Maker has a winning strategy in the $(1, b)$ game $\mathcal{H}(G)$. First, we make the following modifications to Theorem 4.2.13:

- In the statement of the theorem, $p = \frac{K \ln n}{n}$, $b \leq \frac{np}{10 \ln n} = \frac{K}{10}$, and ε is some positive constant.
- By similar calculations to those in $(P1)$ of Theorem 4.2.9, we can assume that $\delta(G) \geq \frac{1}{2}np$.

- We conclude that $\mathsf{dang}_{B_g}(v_g) \geq \left(\frac{1}{2} - \frac{\varepsilon}{3}\right) np - (c-1) - b(2c-1) > \left(\frac{1}{2} - \frac{\varepsilon}{3} - \frac{\varepsilon}{60}\right) np$.
- Finally, we use the following calculation:

$$\begin{aligned}
\overline{\mathsf{dang}}_{B_1}(I_{g-1}) &\geq \overline{\mathsf{dang}}_{B_g}(I_0) + \sum_{j=1}^{r-1}\left(\overline{\mathsf{dang}}_{B_{g-i_j}}(I_{i_j}) - \overline{\mathsf{dang}}_{B_{g-i_j+1}}(I_{i_j-1})\right) \\
&\geq \overline{\mathsf{dang}}_{B_g}(I_0) - \sum_{j=1}^{r-1}\frac{2b}{j+1} \\
&\geq \overline{\mathsf{dang}}_{B_g}(I_0) - 2b\ln n \\
&\geq \left(\frac{1}{2} - \frac{\varepsilon}{3} - \frac{\varepsilon}{60}\right) np - \frac{1}{5}np \\
&> 0
\end{aligned}$$

  to get a contradiction (for sufficiently small ε).

With this variant of Theorem 4.2.13, adapted to the case $p = \Theta\left(\frac{\ln n}{n}\right)$, the proof of Theorem 1.2.8 goes the same as the proof of Theorem 1.2.7, mutatis mutandis. □

Remark: To win the k-connectivity game, Maker follows Stages I and II of the proposed strategy S_M with the following parameter changes:

- In Stage I, Maker creates an $\left(\frac{10000n}{\ln\ln n}, k\right)$-expander by creating a graph with minimum degree at least $c = 100k \ln n / \ln\ln n$. The calculations are almost identical to these appear in the proof of Theorem 1.2.7, Stage I.
- In Stage II, Maker makes his graph an $\left(\frac{n+k}{2k}, k\right)$-expander by claiming an edge between any two disjoint subsets $A, B \subseteq V$ such that $|A| = \frac{10000n}{\ln\ln n}$, $|B| = \frac{n}{10k}$, in a similar way as in Stage II of Theorem 1.2.7.

Then, by Lemma 4.2.8, Maker's graph is k-connected and he wins the game. We omit the straightforward details and the calculations, which are almost identical to those of the Hamiltonicity game.

5 Hitting time appearence of certain spanning trees in the random graph process

As we mentioned earlier, the results of this chapter are based on joined work with D. Johannsen and Michael Krivelevich [43].

We use standard graph theoretic notation (see e.g. [21]) with the following slight modifications. In a graph G, we denote the degree of a vertex x by $\deg_G(x)$. The *neighborhood* $N_G(A)$ of a set $A \subseteq V(G)$ is the union of the neighborhoods of its vertices, i.e., a vertex $x \in A$ could also lie in $N_G(A)$. We denote by $e(X, Y)$ the number of ordered pairs (x, y) with $x \in X$ and $y \in Y$ (edges in $X \cap Y$ are therefore counted twice). For the sake of readability, for subgraphs $S, S' \subseteq G$ (not necessarily vertex- or edge-disjoint), we denote $S + S' = (V(S) \cup V(S'), E(S) \cup E(S'))$. For an event $\mathcal{E}$, we denote by $I[\mathcal{E}]$ the random variable that equals one if $\mathcal{E}$ occurs and zero otherwise.

In general, we may drop floor and ceiling signs to improve the readability when they do not influence the asymptotic statements. Whenever we have a graph on n vertices, we assume its vertex set to be $[n]$.

The chapter is organized as follows. Since we are aiming for universal statements, we have to prove our results for pseudo-random instead of truly random graphs. In Section 5.1, we collect all the pseudo-random properties we will need for the proofs of the later statements and show that the random graph $G(n, p)$ satisfies this properties a.a.s. for p in the considered range. In Section 5.3, we prove Observation 1.2.10 and Theorem 1.2.9 in the pseudo-random setting introduced in Section 5.1. In Section 5.4 we prove Theorem 1.2.11. Finally, in Section 5.5 we state a conjecture generalizing the intuition obtained from Theorem 1.2.9 and Theorem 1.2.11.

5.1 The Random Graph Process

In this section, we collect all properties that we need that are a.a.s. satisfied in the random graph process. Notice that our theorems are pretty demanding in terms of the pseudo-random properties the process has to satisfy: we need that a.a.s. all graphs in the interesting part of the random graph process behave pseudo-randomly. However, we deal with this explicitly in the following lemma, proving a statement that holds a.a.s. for all these graphs simultaneously.

Lemma 5.1.1. *Let $n \in \mathbb{N}$, $m = \frac{2n \log\log n}{\log n}$, $p_1 = \frac{\log n - \log\log n}{n}$, and $p_2 = \frac{\log n + 2\log\log n}{n}$. Consider the two-round exposure of the random graph: from $G_1 \sim G(n, p_1)$, we create $G_1 \subseteq G_2 \sim G(n, p_2)$ by turning every non-edge of G_1 into an edge of G_2 independently at random with probability $\frac{p_2 - p_1}{1 - p_1}$. Let G be an arbitrary graph such that $G_1 \subseteq G \subseteq G_2$. Then the vertex set of G can a.a.s. be partitioned into two parts U and D and contains disjoint sets $W, P \subseteq U$ such that all the following properties are satisfied.*

(P1) *The set D has size at most $n^{0.9}$.*

(P2) *G does not contain a non-trivial path between two (not necessarily distinct) vertices from D of length less than $\frac{2\log n}{3\log\log n}$.*

(P3) *Every vertex in D of degree one in G has its neighbor in P. Every vertex in D of degree at least two in G has exactly two neighbors in P. There are no other vertices in P, i.e., $P \subseteq N(D)$.*

(P4) *The set W has size at most $7m$.*

(P5) *The induced graph $G[U]$ satisfies $\left|N_{G[U]}(X) \setminus (W \cup P)\right| \geq \frac{\log n}{5 \log\log n}|X|$ for all $X \subseteq U$ with $|X| \leq 2m$.*

(P6) *The induced graph $G[U]$ satisfies $\left|N_{G[U]}(X) \cap W\right| > |X|$ for all $X \subseteq U \setminus (W \cup P)$ with $|X| \leq m$.*

(P7) *The induced graph $G[U]$ satisfies $e(X, Y) > 0$ for all $X, Y \subseteq U$ with $|X|, |Y| \geq m$.*

It is well-known (see, e.g., [14]) that G_1 is a.a.s. not connected, and G_2 has a.a.s. minimum degree at least two. Hence, the random graphs that are of interest for us a.a.s. lie between G_1 and G_2, making the lemma the key technical statement.

Proof. We fix $D = \{v \in [n] : d_{G_1}(v) < \log n/100\}$ to be the set of all vertices with degree less than $\log n/100$ in G_1. Correspondingly, U is the complement of D.

Proof of **(P1)**. Since the property is increasing, it suffices to prove it for G_1 instead of G. The proof is similar to the proof of Claim 4.3 in [11]. Setting $t_0 = \log n/100$, we bound the probability that a vertex $v \in [n]$ is in D as follows:

$$\begin{aligned}
\mathbf{Pr}\left(d_{G_1}(v) < t_0\right) &= \mathbf{Pr}\left(\textsc{Bin}\left(n-1, p_1\right) < t_0\right) \\
&= \sum_{i=0}^{t_0-1} \binom{n-1}{i} p_1^i (1-p_1)^{n-1-i} \\
&< t_0 \cdot \binom{n-1}{t_0} p_1^{t_0} (1-p_1)^{n-1-t_0} \\
&< t_0 \cdot \left(\frac{e(n-1)p_1}{t_0}\right)^{t_0} e^{-p_1(n-1-t_0)} \\
&= \exp\left(O(\log\log n) + \frac{\log n}{100}\left(1 + \log 100 + o(1)\right) - (1-o(1))\frac{\log n}{100}\right) \\
&< n^{-0.92}.
\end{aligned}$$

This implies that $\mathbb{E}(D) < n^{0.08}$, and the property follows from Markov's inequality. □

Proof of **(P2)**. Since the property is increasing, it suffices to prove it for G_2. The proof is similar to the proof of Claim 4.4 in [11]. We prove the claim for two distinct endpoints in D and for paths of length r with $2 \leq r \leq \frac{2\log n}{3\log\log n}$. The other cases (i.e. identical endpoints or $r = 1$) are similar and a little simpler. Fix two vertices $u, w \in [n]$ and let $(u = v_0, \ldots, v_r = w)$ be a sequence of vertices from $[n]$, where $2 \leq r \leq \frac{2\log n}{3\log\log n}$. Denote by $\mathcal{A}$ the event "$\{v_i, v_{i+1}\} \in E(G_2)$ for every $0 \leq i \leq r-1$", and by $\mathcal{B}$ the event that both u and w are elements of D. Clearly, $\mathbf{Pr}(\mathcal{A}) = p_2^r$, hence

$$\mathbf{Pr}\left(\mathcal{B} \wedge \mathcal{A}\right) = p_2^r \cdot \mathbf{Pr}\left(\mathcal{B} | \mathcal{A}\right).$$

Let X denote the random variable which counts the number of edges in G_2 incident with u or w disregarding the pairs $\{u, v_1\}$, $\{v_{r-1}, w\}$, and $\{u, w\}$. We can therefore bound $\mathbf{Pr}\left(\mathcal{B}|\mathcal{A}\right) \leq$

$\mathbf{Pr}(X < 2t_0 - 2)$ and as $X \sim \textsc{Bin}(2n-6, p_2)$, we derive setting $t_0 = \log n/100$ that

$$\begin{aligned}
\mathbf{Pr}(X < 2t_0 - 2) &\leq \sum_{i=0}^{2t_0-2} \binom{2n-6}{i} p_2^i (1-p_2)^{2n-6-i} \\
&< 2t_0 \binom{2n-6}{2t_0} p_2^{2t_0} (1-p_2)^{2n-6-2t_0} \\
&< 2t_0 \cdot \left(\frac{e(2n-6)p_2}{2t_0}\right)^{2t_0} e^{-p_2(2n-6-2t_0)} \\
&= \exp\left(O(\log\log n) + 2\frac{\log n}{100}(1 + \log 100 + o(1)) - (2 - o(1))\frac{\log n}{100}\right) \\
&< n^{-1.8}.
\end{aligned}$$

Fixing the two endpoints u, w, the number of such sequences is at most $(r-1)!\binom{n}{r-1} \leq n^{r-1}$. Applying a union bound argument over all such pairs of vertices and possible sequences connecting them we conclude that the probability of a path in G_2 of length $r \leq \frac{2\log n}{3\log\log n}$, connecting two vertices of D is at most

$$\begin{aligned}
\sum_{r=1}^{\frac{2\log n}{3\log\log n}} \binom{n}{2} \cdot n^{r-1} \cdot p_2^r \cdot n^{-1.8} &< \sum_{r=1}^{\frac{2\log n}{3\log\log n}} \frac{n^{r+1}}{2} \cdot \frac{(\log n + 2\log\log n)^r}{n^r} \cdot n^{-1.8} \\
&< \frac{2\log n}{3\log\log n} \cdot n^{-0.8} \cdot (\log n + 2\log\log n)^{\frac{2\log n}{3\log\log n}} \\
&= o(1).
\end{aligned}$$

This completes the proof of the property. □

Proof of **(P3)**. For every vertex $x \in D$, if its degree in G is one, we add its neighbor to P, and if it is at least two, we add arbitrary two of its neighbors to P. Since by Property ***(P2)***, the neighborhoods of vertices from P do not intersect a.a.s., the created set P satisfies Property ***(P3)*** a.a.s. □

We denote $W_1 = \left\{1, \ldots, \left\lceil \frac{13.8n\log\log n}{\log n} \right\rceil\right\}$ to be the set of the first $6.9m$ vertices. We denote $W_2 = \{x \in [n] \setminus W_1 : \ |N_{G_1}(x) \cap W_1| \leq 8\log\log n\}$. Finally, we set $W = W_1 \cup W_2 \setminus (D \cup P)$.

Proof of **(P4)**. Observe that for every vertex $x \in [n] \setminus W_1$, the expected number of neighbors of x in W_1 in the graph G_1 is

$$\mathbb{E}(|N_{G_1}(x) \cap W_1|) = |W_1|p_1 \sim 13.8\log\log n.$$

Thus, using Chernoff's inequality (Lemma 1.3.1), we obtain

$$\mathbf{Pr}(|N_{G_1}(x) \cap W_1| \leq 8\log\log n) \ll \frac{1}{\log n}. \tag{5.1}$$

By Markov's inequality, $|W_2| < \frac{n\log\log n}{5\log n}$ a.a.s. Hence, $|W| \leq |W_1 \cup W_2| < 7m$ a.a.s. □

The next property of a typical graph G_1 is not listed in the statement of the lemma, since it is used here only as a tool to prove Property ***(P5)***.

Claim 3. *The graph G_1 is a.a.s. such that the induced graph $G_1[U]$ satisfies $e_{G_1}(X,Y) > 0$ for all disjoint $X, Y \subseteq U$ with $|X| \cdot |Y| \geq \frac{2n^2}{\log n}$.*

Notice that by definition, the set U only depends on G_1, and so the statement of the claim makes sense without specifying any graph G.

Proof of Claim 3. For the probability of the complementary event, we obtain using the union bound

$$\begin{aligned}
&\mathbf{Pr}\left(\exists X, Y \subset [n], X \cap Y = \emptyset, |X| \cdot |Y| \geq \frac{2n^2}{\log n} : e_{G_1}(X,Y) = 0\right) \\
&\leq \sum_{X,Y\subset[n], X\cap Y=\emptyset, |X|\cdot|Y|\geq \frac{2n^2}{\log n}} (1-p_1)^{|X|\cdot|Y|} < 2^n 2^n e^{-(2-o(1))n} = o(1).
\end{aligned}$$

□

For the proof of the next property, the following claim bounds the degree in G_2, and thus also in G, of an arbitrary vertex into $W \cup P$ from above.

Claim 4. *The largest degree in G_2 of a vertex from $[n]$ into $W \cup P$ a.a.s. satisfies the relation* $\max\{e_{G_2}(x, W \cup P) : \ x \in [n]\} \leq \frac{3\log n}{\log\log n}$.

Proof of Claim 4. Recall that in G_2, all edges are drawn independently with probability p_2. Furthermore, for vertices $x \in [n]$ and $y, z \in [n] \setminus W_1$, the events "$xy \in E(G_2)$" and "$z \in W_2$" as well as the events "$y \in W_2$" and "$z \in W_2$" are independent. Finally, notice crucially that the events "$xy \in E(G_2)$" and "$y \in W_2$" are independent if $x \notin W_1$, and negatively correlated if $x \in W_1$. Thus, we can estimate

$$\begin{aligned}
\mathbb{E}\left(|N_{G_2}(x) \cap W|\right) &\leq \mathbb{E}\left(|N_{G_2}(x) \cap W_1|\right) + \mathbb{E}\left(|N_{G_2}(x) \cap W_2|\right) \\
&\leq \sum_{y \in W_1 \setminus \{x\}} \mathbf{Pr}\left(xy \in E(G_2)\right) + \sum_{y \in [n]\setminus(W_1 \cup \{x\})} \mathbf{Pr}\left(xy \in E(G_2)\right) \cdot \mathbf{Pr}(y \in W_2) \\
&\overset{(5.1)}{<} |W_1| p_2 + np_2 \cdot \frac{1}{\log n} < 14 \log\log n.
\end{aligned}$$

Now, since for every $x, y, z \in [n]$, the events "$xy \in E(G_2)$ and $y \in W$" and "$xz \in E(G_2)$ and $z \in W$" are mutually independent, we obtain using union bound and Chernoff's inequality (Lemma 1.3.1)

$$\mathbf{Pr}\left[\exists x \in [n] : \ e_{G_2}(x, W) > \frac{2\log n}{\log\log n}\right] < n\left(\frac{2\log n}{14e(\log\log n)^2}\right)^{-\frac{2\log n}{\log\log n}} = o(1).$$

Combining it with the fact that every vertex $x \in [n]$ a.a.s. has at most 2 neighbors in P by ***(P2)*** and ***(P3)*** we have the statement of the claim. □

Appropriate expansion properties of $G[U]$ are widely known. Similarly to Lemma 10 in [44], the following statement is true.

Claim 5. *The graph G_1 is a.a.s. such that for every set $X \subset U$ of size $|X| \leq n/\log n$, its neighborhood in $G_1[U]$ satisfies* $\left|N_{G_1[U]}(X)\right| \geq \frac{\log n}{1000}|X|$.

Proof. Suppose to the contrary that there exists a set $X \subset U$ of size at most $|X| \leq n/\log n$ such that its neighborhood in G_1 satisfies $|N_{G_1}(X)| < \frac{\log n}{1000}|X|$. By the definition of U, the number of edges incident to X in G_1 is at least

$$e_{G_1}(X, N_{G_1}(X)) \geq |X| \log n/200.$$

But Lemma 1.3.1 tells us that

$$\mathbf{Pr}\left(\exists A, B \subset [n],\, |A| \le \frac{n}{\log n}, |B| < \frac{\log n}{1000}|A| :\, e_{G_1}(A,B) \ge |A|\log n/200\right)$$
$$< \sum_{A,B\subset[n],\,|A|\le\frac{n}{\log n},\,|B|=\frac{\log n}{1000}|A|} \left(e\cdot\frac{\mathbb{E}\left(|e_{G_1}(A,B)|\right)}{|A|\log n/200}\right)^{|A|\log n/200}$$
$$< \sum_{A,B\subset[n],\,|A|\le\frac{n}{\log n},\,|B|=\frac{\log n}{1000}|A|} \left(\frac{200e|A||B|p_1}{|A|\log n}\right)^{|A|\log n/200}$$
$$< \sum_{a\le\frac{n}{\log n}} \binom{n}{a}\binom{n}{\frac{a\log n}{1000}}\left(\frac{3ap_1}{5}\right)^{a\log n/200}$$
$$< \sum_{a\le\frac{n}{\log n}} \left(\frac{3ap_1}{5}\right)^{a\log n/400} = o(1),$$

providing the claim.

□

We are now ready to prove Property ***(P5)***.

Proof of **(P5)**. Provided that the graphs G_1 and G_2 are such as Claims 5, 4, and 3 prove them to be a.a.s., we obtain for every set $X \subset U$ of size $|X| \le n/\log n$

$$\left|N_{G[U]}(X)\setminus(W\cup P)\right| \ge \left|N_{G[U]}(X)\right| - |X|\max_{x\in X}\{e(x, W\cup P)\}$$
$$\overset{\text{Claim 5, Claim 4}}{\ge} |X|\left(\frac{\log n}{1000} - \frac{3\log n}{\log\log n}\right) > \frac{\log n}{5\log\log n}|X|.$$

Furthermore, every set $X \subset U$ of size $|X| \le n\log\log\log n/\log n$ contains a subset $X' \subset X$ of size $|X'| = n/\log n$. Therefore, again applying Claim 5, we obtain

$$\left|N_{G[U]}(X)\setminus(W\cup P)\right| \ge \left|N_{G[U]}(X')\setminus(W\cup P)\right| \ge n/1000 > \frac{\log n}{5\log\log n}|X|.$$

Finally, for a set $X \subset U$ of size $n\log\log\log n/\log n \le |X| \le 2m$, we obtain using Claim 3

$$\left|N_{G[U]}(X)\setminus(W\cup P)\right| \ge |N_G(X)| - |D\cup W\cup P| \ge |N_{G_1}(X)| - |D\cup W\cup P| = n - o(n) > \frac{\log n}{5\log\log n}|X|.$$

□

The expansion of sets of potential leaves into W is a key property to prove the existence of bounded degree spanning trees with many leaves. In fact, Property ***(P6)*** is the reason we introduce W.

Proof of **(P6)**. The property is increasing, hence it is again sufficient to prove it to hold in G_1.

By the construction of W, G_1 is a.a.s. such that every vertex $x \in U\setminus(W\cup P)$ has more than $8\log\log n$ neighbors in $W_1 \subseteq W$. Furthermore, x has at most one neighbor in $P\cup D$ by Property ***(P2)***. Hence, x has at least $\left\lfloor\frac{\log\log n}{8}\right\rfloor$ neighbors in W; in the following, we omit the flooring signs. Now, for the event complementary to the statement of Property ***(P6)*** we obtain using Chernoff's inequality (Lemma 1.3.1)

$$\begin{aligned}
&\mathbf{Pr}\left(\exists X \subset U \setminus (W \cup P),\, |X| \leq \frac{2n \log\log n}{\log n} \,:\, |N_{G_1}(X) \cap W| \leq |X|\right) \\
&\leq \mathbf{Pr}\left(\exists X, Y \subset [n],\, X \cap Y = \emptyset,\, |X| = |Y| \leq \frac{2n \log\log n}{\log n} \,:\, e_{G_1}(X,Y) \geq 8 \log\log n |X|\right) \\
&< \sum_{k \leq \frac{2n \log\log n}{\log n}} \binom{n}{k}^2 \left(\frac{e \cdot k^2 p_1}{8k \log\log n}\right)^{8k \log\log n} \\
&< \sum_{k \leq \frac{2n \log\log n}{\log n}} \left(\frac{ne}{k}\right)^{2k} \left(\frac{ek \log n}{8n \log\log n}\right)^{8k \log\log n} \\
&= \sum_{k \leq \frac{2n \log\log n}{\log n}} (ne)^{2k} \left(\frac{e \log n}{8n \log\log n}\right)^{8k \log\log n} k^{8k \log\log n - 2k} \\
&< \sum_{k \leq \frac{2n \log\log n}{\log n}} (ne)^{2k} \left(\frac{e \log n}{8n \log\log n}\right)^{8k \log\log n} \left(\frac{2n \log\log n}{\log n}\right)^{2k - 8k \log\log n} \\
&= \sum_{k \leq \frac{2n \log\log n}{\log n}} \left(\frac{e \log n}{2 \log\log n}\right)^{2k} (e/4)^{8k \log\log n} = o(1),
\end{aligned}$$

completing the proof of Property ***(P6)***. □

Proof of **(P7)**. It suffices to show ***(P7)*** for G_1, since ***(P7)*** is a monotone increasing graph property. Furthermore, we only need to show it for sets of size m, since all larger sets contain subsets of size m.

$$\begin{aligned}
&\mathbf{Pr}\left(\exists X, Y \subset [n],\, X \cap Y = \emptyset,\, |X| = |Y| = m :\, e_{G_1}(X,Y) = 0\right) \leq \binom{n}{m}^2 (1 - p_1)^{m^2} \\
&< \left(\frac{ne}{m}\right)^{2m} e^{-p_1 m^2} = \left(\frac{e \log n}{2 \log\log n}\right)^{2m} e^{-2m \log\log n (1 - \log\log n / \log n)} \\
&= \left(\frac{e}{2 \log\log n} e^{(\log\log n)^2 / \log n}\right)^{2m} = o(1).
\end{aligned}$$

□

□

5.2 Embedding Paths and Trees in Expanders

The first breakthrough result on embedding trees of bounded degree into $G(n,p)$ was achieved by Friedman and Pippenger [35]. A later refinement of their result by Haxell [49] implies a possibility to embed almost spanning bounded degree trees in sparse random graphs. The following result is a corollary of Theorem 3 from [7], basically setting $L = 0$ there. We state it because our technique is a modification of the approach in [49] and [7].

Theorem 5.2.1 (Corollary of Theorem 3 in [7]). *Let $d, m \in \mathbb{N}$ and let T be a tree with maximum degree at most d. Suppose that G is a graph satisfying the following conditions:*

- *$N_G(X) \geq d|X| + 1$ for every set $X \subset V(G)$ with $1 \leq |X| \leq m$ and*

- $N_G(X) \geq d|X| + |V(T)|$ *for every set* $X \subset V(G)$ *with* $m+1 \leq |X| \leq 2m$.

Then G *contains a copy of* T *as a subgraph.*

Another important tool of our proof that will be essential for the embedding of a spanning tree with a long bare path is the result of Hefetz, Krivelevich, and Szabó [52] on Hamilton connectivity of expanders. The following theorem is an adaptation of Theorem 1.2 in [52] which takes into account that in the result of Hefetz et al. $N_G(X)$ is defined as the *external* neighborhood of X, while in our case X can be part of its own neighborhood. In fact, we just plug in $(d+2)$ for every d in [52] and estimate $|N_G(X) \setminus X| \geq |N_G(X)| - |X|$ for a graph G and a vertex set $X \subseteq V(G)$ as well as $\frac{2\log(d+2)}{d+2} \geq \frac{\log d}{d}$ and $\frac{\log(d+2)}{5000} \leq \frac{\log d}{4130}$ for $d \geq 14$.

Theorem 5.2.2 (Corollary of Theorem 1.2 in [52])**.** *Let* $n \in \mathbb{N}$ *and let* $d = d(n)$ *satisfy* $14 \leq d \leq e^{\sqrt[3]{\log n}}$. *Suppose that* G *is a graph on* n *vertices satisfying the following conditions:*

(H1) $|N_G(X)| \geq (d-1)|X|$ *for every set* $X \subset V(G)$ *with* $|X| \leq \frac{2n \log\log n \log d}{d \log n \log\log\log n}$ *and*

(H2) $e_G(X,Y) > 0$ *for every two disjoint sets* $X, Y \subset V(G)$ *with* $|X| \geq \frac{n \log\log n \log d}{5000 \log n \log\log\log n}$ *and* $|Y| \geq \frac{n \log\log n \log d}{5000 \log n \log\log\log n}$.

Then G *is Hamilton connected, provided that* n *is sufficiently large.*

The following definition formalizes the crucial expansion property used for obtaining almost spanning trees.

Definition 5.2.3. Let $d, m \in \mathbb{N}$ satisfy $m \geq 1$ and $d \geq 3$, let G be a graph, and let $S \subseteq G$ be a subgraph of G. We say that S is (d,m)-*extendable* if S has maximum degree at most d and

$$D(X;S) := |N_G(X) \setminus V(S)| - (d-1)|X| + \sum_{x \in X \cap V(S)} \big(\deg_S(x) - 1\big) \geq 0$$

holds for all sets $X \subseteq V(G)$ with $|X| \leq 2m$.

Furthermore, we call all sets $X \subseteq V(G)$ of size at most $2m$ with $D(X;S) = 0$ *critical with respect to* S (or shortly S-*critical*).

The following is the key lemma in working with crucial sets. It states that crucial sets are closed under union and intersection.

Lemma 5.2.4 (Criticality Lemma)**.** *Let* $d, m \in \mathbb{N}$ *satisfy* $m \geq 1$ *and* $d \geq 3$, *let* G *be a graph, and let* S *be a* (d,m)-*extendable subgraph of* G. *Suppose that* G *satisfies*

$$|N_G(X)| \geq |V(S)| + 2dm + 1$$

for all $X \subseteq V(G)$ *with* $m \leq |X| \leq 2m$. *Then*

(i) all S*-critical sets have size at most* m, *and*

(ii) the union and the intersection of two critical sets are critical.

Consequently, we may define the unique inclusion-minimal S*-critical set* $X(S)$ *that contains all* S*-critical sets as subsets and has size at most* m.

Proof. To show (i), consider an S-critical set $X \subseteq V(G)$. By definition, we have $|X| \leq 2m$ and

$$|N_G(X) \setminus V(S)| = (d-1)|X| - \sum_{x \in X \cap V(S)} \big(\deg_S(x) - 1\big) \leq d|X|.$$

Since $|N_G(X) \setminus V(S)| \geq |N_G(X)| - |V(S)|$, this implies

$$|N_G(X)| \leq |V(S)| + d|X|.$$

Thus, as $|N_G(X)| \geq |V(S)| + 2dm + 1 \geq |V(S)| + d|X| + 1$ holds in the case that $m \leq |X| \leq 2m$, we have that X is at most of size m.

To show (ii), let us first observe that the function $D(\cdot, S)$ is submodular, that is, that

$$D(X \cup Y; S) + D(X \cap Y; S) \leq D(X; S) + D(Y; S)$$

holds for all $X, Y \subseteq V(G)$. The reason for this is that we have

$$|N_G(X \cup Y) \setminus V(S)| + |N_G(X \cap Y) \setminus V(S)| \leq |N_G(X) \setminus V(S)| + |N_G(Y) \setminus V(S)|,$$

$$\sum_{x \in (X \cup Y) \cap V(S)} \big(\deg_S(x) - 1\big) + \sum_{x \in (X \cap Y) \cap V(S)} \big(\deg_S(x) - 1\big) = \sum_{x \in X \cap V(S)} \big(\deg_S(x) - 1\big) + \sum_{x \in Y \cap V(S)} \big(\deg_S(x) - 1\big),$$

and

$$(d-1)|X \cup Y| + (d-1)|X \cap Y| = (d-1)|X| + (d-1)|Y|$$

for all $X, Y \subseteq V(G)$. Now, suppose that $X, Y \subseteq V$ are critical. Then, by (i), $|X| \leq m$ and $|Y| \leq m$. Thus, it holds that $|X \cup Y| \leq 2m$ and $|X \cap Y| \leq 2m$ and therefore also that $D(X \cup Y; S) \geq 0$ and $D(X \cap Y; S) \geq 0$. Hence, since $D(\cdot; S)$ is submodular, we obtain

$$0 \leq D(X \cup Y; S) + D(X \cap Y; S) \leq D(X; S) + D(Y; S) = 0,$$

that is, both $X \cup Y$ and $X \cap Y$ are critical, too. □

The following proposition will be crucial for the upcoming lemmas in this section. Here we calculate the change of the value $D(X; S)$ when S gets an additional vertex attached to the previous vertex set by an edge.

Proposition 5.2.5. *Let $d \in \mathbb{N}$, let G be a graph, and let S be a subgraph of G. For arbitrary $s \in V(S)$ and $y \in N_G(s) \setminus V(S)$, let S_y be the subgraph of G obtained from S by adding y to the vertex set of S_y and adding $\{y, s\}$ to the edge set of S_y. Then, for every $X \subseteq V(G)$, we obtain*

$$D(X; S_y) = D(X; S) - I\,[y \in N_G(X)] + I[s \in X].$$

Proof. For every $X \subseteq V(G)$ we have

$$|N_G(X) \setminus V(S_y)| = |N_G(X) \setminus V(S)| - I\,[y \in N_G(X)]$$

and

$$\sum_{x \in X \cap V(S_y)} \big(\deg_{S_y}(x) - 1\big) = \sum_{x \in X \cap V(S)} \big(\deg_{S_y}(x) - 1\big) + (1-1) = \sum_{x \in X \cap V(S)} \big(\deg_S(x) - 1\big) + I[s \in X].$$

□

The following lemma is crucial for our embedding. It ensures that as long as we did not embed too large proportion of the vertices, we can embed one more vertex without destroying the (d, m)-extendability of the image subgraph.

Lemma 5.2.6 (Vertex Extension Lemma)**.** *Let $d, m \in \mathbb{N}$ satisfy $m \geq 1$ and $d \geq 3$, let G be a graph, and let S be a (d, m)-extendable subgraph of G. Assume that G satisfies*

$$|N_G(X)| \geq |V(S)| + 2dm + 1$$

for all $X \subseteq V(G)$ with $m \leq |X| \leq 2m$. Then, for every vertex $s \in V(S)$ with $\deg_S(s) \leq d - 1$, there exists a vertex $y \in N_G(s) \setminus V(S)$ such that the graph $S_y := (V(S) \cup \{y\}, E(S) \cup \{\{y, s\}\})$ is (d, m)-extendable, too.

Proof. Let $Y = N_G(s) \setminus V(S)$. Since S is (d, m)-extendable, we have

$$0 \leq D(\{s\}; S) = |N_G(s) \setminus V(S)| - (d-1) + \big(\deg_S(s) - 1\big) \leq |N_G(s) \setminus V(S)| - 1$$

and therefore the set Y is non-empty.

For the sake of contradiction, suppose that for every $y \in Y$, the subgraph S_y defined in the statement of the lemma is not (d, m)-extendable. Consider an arbitrary $y \in Y$. Since we supposed that S_y is not (d, m)-extendable, there exists a set X_y of size $|X_y| \leq 2m$ such that

$$D(X_y; S_y) < 0. \tag{5.2}$$

We have seen in Proposition 5.2.5 that

$$D(X; S_y) = D(X; S) - I[y \in N_G(X)] + I[s \in X]$$

holds for all $X \subseteq V(G)$. Thus, in order to satisfy inequality (5.2), the three conditions

(i) $D(X_y; S) = 0$ (that is, X_y has to be S-critical),

(ii) $y \in N_G(X_y)$, and

(iii) $s \notin X_y$

have to be satisfied. Now, Lemma 5.2.4 implies that $X^* = \bigcup_{y \in Y} X_y$ also satisfies these three conditions (i)–(iii) in place of X_y, too. Because of (ii), we know that $N_G(s) \setminus V(S) = Y \subseteq N_G(X^*)$ and therefore

$$|N_G(X^* \cup \{s\}) \setminus V(S)| = |N_G(X^*) \setminus V(S)|. \tag{5.3}$$

Because of (iii), we know that

$$(d-1)|X^* \cup \{s\}| = (d-1)|X^*| + d - 1. \tag{5.4}$$

Finally, we have that

$$\sum_{x \in (X^* \cup \{s\}) \cap V(S)} \big(\deg_S(x) - 1\big) \leq \sum_{x \in X^* \cap V(S)} \big(\deg_S(x) - 1\big) + d - 2. \tag{5.5}$$

Thus, together (5.3), (5.4), and (5.5) imply that

$$D(X^* \cup \{s\}; S) \leq D(X^*; S) - 1.$$

However X^* is critical by condition (i) and therefore

$$D(X^* \cup \{s\}; S) < 0.$$

But this is a contradiction to the fact that S is (d, m)-extendable and that by Lemma 5.2.4 $|X^* \cup \{s\}| \leq m + 1 \leq 2m$. Hence, there exists $y \in Y$ such that S_y is (d, m)-extendable, too. □

Lemma 5.2.7 (Removal Lemma)**.** *Let $d, m \in \mathbb{N}$ satisfy $m \geq 1$ and $d \geq 3$, let G be a graph, and let S be a subgraph of G. Furthermore, assume that there exist vertices $s \in V(S)$ and $y \in N_G(S) \setminus V(S)$ such that the graph $S_y := (V(S) \cup \{y\}, E(S) \cup \{\{y, s\}\})$ is (d, m)-extendable. Then S is (d, m)-extendable, too.*

Proof. By Proposition 5.2.5, we have

$$D(X; S) = D(X; S_y) + I[y \in N_G(X)] - I[s \in X].$$

for all $X \subseteq V(G)$. Since y and s are neighbors, it holds that $I[s \in X] \leq I[y \in N_G(X)]$. Therefore,

$$D(X; S) \geq D(X; S_y) \geq 0$$

for every $X \subseteq V(G)$ with $|X| \leq 2m$. □

Lemma 5.2.8 (Edge Insertion Lemma)**.** *Let $d, m \in \mathbb{N}$ satisfy $m \geq 1$ and $d \geq 3$, let G be a graph, and let S be a (d,m)-extendable subgraph of G. If $s, t \in V(S)$ such that $\deg_S(s), \deg_S(t) \leq d-1$ and $\{s,t\} \in E(G)$, then $S + \{s,t\}$ is a (d,m)-extendable subgraph of G, too.*

Proof. Lemma 5.2.8 follows directly from the definition of $D(X; \cdot)$, since

$$\begin{aligned} D\left(X; S + \{s,t\}\right) &= |N_G(X) \setminus V(S)| - (d-1)|X| + \sum_{x \in X \cap V(S)} \left(\deg_{S+\{s,t\}}(x) - 1\right) \\ &= D(X;S) + I[s \in X] + I[t \in X] \geq 0 \end{aligned}$$

holds for all $X \subseteq V(G)$. □

Lemma 5.2.9 (Connection Lemma)**.** *Let $n, d, m \in \mathbb{N}$ satisfy $m \geq 1$ and $d \geq 3$, let G be a graph on n vertices satisfying*

$$e_G(X,Y) > 1$$

for all disjoint $X, Y \subseteq V(G)$ with $|X| \geq m$ and $|Y| \geq m$, and let S be a (d,m)-extendable subgraph of G on at most $n - 10dm$ vertices. Let $k = \lceil \log(2m)/\log(d-1) \rceil$ and let $j \in \{1, \ldots, k\}$. Assume that there are two disjoint vertex sets A and B in S such that both, A and B, are of size at least $2m/(d-1)^j$ and all vertices in A and B are of degree at most $d/2$ in S. Then there exists a path P of length $2j+1$ in G with endpoints in A and B such that all vertices of P except the two endpoints do not belong to S, and $S + P$ is (d,m)-extendable.

Note that for $j = k$, the two sets A and B both consist of a single vertex, that is, we may connect any two vertices of S by a path of length $2k+1$.

Proof. We first choose two sets $A' \subseteq A$ and $B' \subseteq B$, each of size $\lceil m/(d-1)^j \rceil$. Next, for each of the vertices in A' and B', we attach by an edge $\lceil \frac{d-1}{2} \rceil$ complete rooted $(d-1)$-ary trees of depth $j-1$ by repeatedly applying the Vertex Extension Lemma (Lemma 5.2.6). In order to apply this lemma, we have to make sure that S does not grow by too much (the degree constraint is satisfied since every vertex in A' and B' is of degree at most $d/2$ in S). This is shown by the following calculation. We attach at most $(d-1)$ trees to each of the at most $2\lceil m/(d-1)^j \rceil$ vertices of each set and at most $2(d-1)^{j-1}$ vertices for each tree, that is, at most $8(d-1)m$ vertices in total. Therefore, since $e_G(X,Y) > 1$ holds for all disjoint $X, Y \subseteq V(G)$ with $|X| \geq m$ and $|Y| \geq m$, we have

$$|N_G(X)| \geq n - 2m + 1 \geq |V(S)| + 10dm - 2m + 1 > \left(|V(S)| + 8(d-1)m\right) + 2dm + 1$$

for every set $X \subseteq V(G)$ with $m \leq |X| \leq 2m$. Thus, we can indeed apply the Vertex Extension Lemma to attach the $(d-1)$-ary trees.

In total the subtrees attached to each of the sets have at least m leaves each, and every leaf is at distance j from the vertex in A' or B' the respective $(d-1)$-ary tree is attached to. Thus, there is an edge between them closing a path of length $2j+1$ in G between A and B. We add this edge to S by applying the Edge Insertion Lemma (Lemma 5.2.8). Finally, we repeatedly apply the Removal Lemma (Lemma 5.2.7) to remove from the attached trees all vertices that do not lie on the path closed by the additional edge, effectively adding only this single path of length $2j+1$ between A' and B' to S. □

5.3 Hitting Times of Spanning Trees with Long Bare Paths

In this section we prove Theorem 1.2.9 and Observation 1.2.10.

As we already mentioned in Chapter 1, in our tree we need a set of leaves that are sufficiently far from each other. We formalize it in the following definition.

Definition 5.3.1. Let $\ell \in \mathbb{N}$. For a forest F and a set $A \subseteq V(F)$, we call the *skeleton* of A in F the subforest of F induced by all vertices in $V(F)$ that lie on a path between two vertices of A (which includes all vertices of A). Furthermore, we say that A is *ℓ-scattered* in F if there exists an ordering $u_1, \dots, u_{|A|}$ of the vertices of A such that, for all $j \in \{1, \dots, |A|-1\}$, the vertex u_{j+1} does not lie in the skeleton of $A_j := \{u_1, \dots, u_j\}$ and the unique path (if any exists) between u_{j+1} and the skeleton of A_j has length at least ℓ.

First let us observe a connection between an ℓ-scattered set A and the pairwise distances between vertices from A.

Observation 5.3.2. Let T be a tree, ℓ a positive integer, and $A \subseteq V(T)$ a set of vertices with pairwise distances at least 2ℓ. Then A is ℓ-scattered.

Proof. We show the observation by induction on $|A|$. In case $|A| \leq 2$, the statement is trivial. If $|A| \geq 3$, we denote by T_A the skeleton of A and consider it to be rooted at an arbitrary vertex. Let u be a vertex of largest distance from the root among all vertices of degree at least 3 in T_A. Then T_A consists of a tree T' with $|A| - \deg_{T_A}(u) + 2$ leaves, with u being a leaf of T', as well as $\deg_{T_A}(u) - 1 =: i$ edge-disjoint paths between u and the vertices $\{a_1, \dots, a_i\} \subset A$. Since vertices from A have pairwise distance at least 2ℓ, we can assume without loss of generality that the path between u and a_1 has length at least ℓ.

By induction, the set $A \setminus \{a_1\}$ is ℓ-scattered. Furthermore, the path between a_1 and u is exactly the connection between u and the skeleton of $A \setminus \{a_1\}$. And since this path is of length at least ℓ, the set A is ℓ-scattered as well. $\square$

Observation 5.3.2 implies that Theorem 1.2.9 is an immediate corollary of the following more technical statement, where the requirement on the pairwise distances is replaced by the corresponding scatterness.

Theorem 5.3.3. *In the random graph process on n vertices, a.a.s. the following holds. For every $t < n$, if $G^{(t)}$ is connected, then it contains every n-vertex tree of maximum degree at most $\frac{\log n}{2 \log\log n \log\log\log n}$ containing a $(2.1 \log n / \log\log n)$-scattered t-set of leaves and a bare path of length at least $\frac{23n}{\log\log\log n}$.*

Proof. Let $t \in \mathbb{N}$. For $n \in \mathbb{N}$, let $\Delta = \frac{\log n}{2 \log\log n \log\log\log n}$, $m = \frac{2n \log\log n}{\log n}$, $d = \left\lceil \frac{\log n}{\log\log n \log\log\log n} \right\rceil$, and $\ell = 2 \left\lceil \frac{\log(2m)}{\log(d-1)} \right\rceil + 3$. Then, for sufficiently large n, we have $\ell < \frac{2.1 \log n}{\log\log n}$ and $11dm \leq \frac{23n}{\log\log\log n}$.

By Lemma 5.1.1 and the connectivity threshold from [?], the random graph $G^{(t)}$ is a.a.s. either not connected or satisfies the properties **(P1)**–**(P7)**. Thus, let $n \in \mathbb{N}$ be sufficiently large and let G be a graph with sets U, W, D, and P as in Lemma 5.1.1 that satisfies the properties **(P1)**–**(P7)**. (Notice that we actually do not use the set W in this section, so we could leave out W, **(P4)**, and **(P6)** here.) Let L be the set of vertices of degree one in G and let L' be the set of neighbors of L. Assume that the set L is of size exactly t. Furthermore, let F be an n-vertex tree of maximum degree at most Δ containing an ℓ-scattered set of leaves L_F of size t and a bare path B in F of length at least $11dm$. Let L'_F be the set of neighbors of L_F in F and let F_1 and F_2 be the two subtrees of F intersecting B in one vertex each such that $F_1 + B + F_2 = F$. In order to prove Theorem 5.3.3, we show that there exists an embedding of F into G.

Notice that $L \subseteq D$ by Property **(P5)**, since for every vertex $x \in U$, applying Property **(P5)** with $X = \{x\}$, we see that $d_G(x) \geq \frac{\log n}{5 \log\log n} > 1$. Thus, $L' \subseteq P$ by Property **(P3)**.

To be able to apply the results from Section 5.2, we create an auxiliary graph G' from G by deleting every vertex from D as well as all edges incident to D, and for every $x \in D$ of degree at least two, adding an auxiliary edge $\{p_1, p_2\}$ between the two neighbors p_1 and p_2 of x in P. Note that by Property **(P2)** no two vertices in D share a common neighbor in P. Correspondingly, we delete L_F from F and replace B by a bare path B' of length $|E(B)|-|D|+|L| \overset{\textbf{(P1)}}{\geq} (11-o(1))dm$ to obtain a tree F' with the corresponding subtrees $F_1' + B' + F_2' = F'$.

If there exists an embedding of F' into G' such that all vertices from L_F' are embedded into L' and all auxiliary edges are edges of the image of B', we can find in G a subgraph isomorphic to F as follows:

- we take the embedding of F' into G', and for every auxiliary edge p_1p_2, replace it by the two edges p_1x and p_2x, where x is the (unique) neighbor of p_1 and p_2 in D, and
- for every vertex $y \in L'$, we add the edge xy between y and its (unique) neighbor $x \in L$.

As a first step, we start the embedding of F into G' by setting S to be the subgraph of G' consisting of the auxiliary edges and the vertices from L'. Note that due to our construction P is the vertex set of S.

We will modify S using the statements in Section 5.2. However, during the embedding process, while we will be using these statements, at least $10.5dm$ vertices from B will be not embedded into S. Thus, we have to check that as long as $|V(S)| \leq n - 10.5dm$, the following two conditions that are necessary to apply the statements in Section 5.2 are satisfied.

First, we check that the initial graph S is (d,m)-extendable. Since S consists of isolated edges and vertices, the maximum degree condition is clearly satisfied. Furthermore, since $V(S) = P$, notice that Properties **(P3)** and **(P2)** imply that (in both G and G') every vertex has at most two neighbors in $V(S)$. Hence, for every $X \subseteq V(G')$, we have $|N_{G'}(X) \cap V(S)| \leq 2|X|$. Armed with this note and keeping in mind that every vertex has at most one neighbor in $P = V(G) \setminus V(G')$, we see that in G', for every set $X \subseteq V(G')$ of size at most $2m$, Property **(P5)** implies

$$D(X;S) \geq |N_{G'}(X) \setminus V(S)| - (d-1)|X| \geq |N_G(X) \setminus V(S)| - |X| - (d-1)|X| \overset{\textbf{(P5)}}{\geq} \left(\frac{\log n}{5 \log\log n} - 2 - d\right)|X| \geq 0,$$

and therefore S is indeed (d,m)-extendable.

For the second condition needed in order to apply lemmas from Section 5.2, we see that as long as $|V(S)| \leq n - 10.5dm$, we obtain for every $X \subset V(G')$ of size $m \leq |X| \leq 2m$

$$|N_{G'}(X)| \geq |N_G(X) \setminus D| \overset{\textbf{(P7)}}{\geq} n - |D| - 2m + 1 \overset{\textbf{(P1)}}{>} |V(S)| + 2dm + 1.$$

This implies that as long as at least $10.5dm$ vertices of B' are not embedded into G', we can apply the Vertex Extension Lemma (Lemma 5.2.6) and the Connection Lemma (Lemma 5.2.9, here only used with $j = k$) to extend the embedded subgraph.

Recall that we start with a graph S consisting of isolated vertices and (auxiliary) edges. First, we connect (using only the Connection Lemma) all auxiliary edges to a path $\hat{B}$ that does not contain any other vertex from S. We do so by connecting the auxiliary edges one by one to the current path, each time extending the path by $\ell - 1$ additional edges. Note that $\hat{B}$ is of length at most $(\ell-1)|P| = o(dm)$.

Then we consider the set of leaves $A = V(F_1') \cap L_F'$. We embed A to arbitrary $|A|$ vertices in $L' \subseteq V(S)$. Since L_F was ℓ-scattered in F, we see that $A \subset L_F'$ is $(\ell-2)$-scattered in F'. Thus, in G' we can connect (using the Vertex Extension Lemma and the Connection Lemma) the images of vertices from A by paths of length at least $\ell - 2$ in the ordering given by Definition 5.3.1 to obtain an embedding of the skeleton of A. We then extend this skeleton to a copy of F_1' (using the Vertex Extension Lemma). Similarly, we find a copy of F_2' in G' that maps $V(F_2') \cap L_F'$ to the remaining vertices of L' and is vertex-disjoint from the previously embedded part. Finally, we again apply the Connection Lemma to connect the path $\hat{B}$ by a path of length $\ell - 2$ with the image of the end-vertex of B' in F_1'. Thus, $\hat{B}$ becomes part of the embedding of B.

We have now a (d,m)-extendable subgraph S of G' consisting of two trees such that connecting two designated vertices x and y from S' by a path from G' using exactly the vertices that are not yet in $V(S)$ yields a copy of F' in G'. In other words, we are looking for a Hamilton path between x and y in $G'\left[V\left(G'\right)\setminus\left(V\left(S'\right)\setminus\{x,y\}\right)\right] =: \hat{G}$. Notice that $\hat{G}$ is a subgraph of G. We assume without loss of generality that x and y have degree exactly one in F'; otherwise, we just embed the preimages of their neighbors in B using the Vertex Extension Lemma. Notice that, as predicted before the beginning of the embedding, at least $10.5dm$ vertices from B are not embedded yet.

To finish up the proof, we use Theorem 5.2.2. To that end, notice that $\hat{n} := \left|V(\hat{G})\right| \geq (11-o(1))dm$, hence we obtain $\frac{2\hat{n}\log\log\hat{n}\log d}{d\log\hat{n}\log\log\log\hat{n}} = o(m)$ and $\frac{\hat{n}\log\log\hat{n}\log d}{5000\log\hat{n}\log\log\log\hat{n}} = \omega(m)$. Since S is (d,m)-extendable and every $v \in V(S)\cap V\left(\hat{G}\right)$ has degree exactly one in S, we obtain for every set $X \subset V\left(\hat{G}\right)$ of size $|X| \leq 2m$

$$\left|N_{\hat{G}}(X)\right| \geq |N_{G'}(X)\setminus V(S)| \geq (d-1)|X| - \sum_{x\in V(S)\cap X}(\deg_S(x)-1) = (d-1)|X|,$$

thus Property (H1) of Theorem 5.2.2 holds. Furthermore, Property (H2) of Theorem 5.2.2 holds for $\hat{G}$ by Property **(P7)**. Hence, $\hat{G}$ is Hamilton connected by Theorem 5.2.2, and especially there exists a necessary Hamilton path between x and y, completing the copy of F' in G', and therefore proving the theorem. □

Finally, we prove Observation 1.2.10, giving some intuition on the tightness of Theorem 1.2.9.

Proof of Observation 1.2.10 . Fix an integer $j \geq k$, and let us denote $G = G^{(j)}$. Suppose for the sake of contradiction that there exists an embedding of T into G; fix an arbitrary one. Clearly, a vertex of degree 1 in G can be only used to embed a leaf of T. Thus, there exists a set J of size j of leaves of T that are embedded into the set of vertices of degree one in G. Since G is connected and contains vertices of degree 1, a.a.s. the properties from Lemma 5.1.1 hold in G. Now, the vertices in J a.a.s. have pairwise distance at least $\frac{2\log n}{3\log\log n}$ by **(P2)**, contradicting the assumption about the structure of T. □

5.4 Hitting Times of Spanning Trees with Many Leaves

The main purpose of this section is to present a proof of Theorem 1.2.11. To this end, we show the following result.

Theorem 5.4.1. *Let $\Delta \in \mathbb{N}$ and $\varepsilon > 0$. Then there exists an absolute constant $n_0 = n_0(\Delta,\varepsilon)$ such that for every $n \geq n_0$ the following holds. Let $d = \left\lfloor\frac{\varepsilon\log n}{30\Delta\log\log n}\right\rfloor$ and let $m = \frac{2n\log\log n}{\log n}$. Furthermore, let H be a host graph on at least $n-m$ vertices containing a set of* critical *vertices C of size at most $8m$ and let T be a tree of maximum degree at most Δ which has at most $(1-\varepsilon/\Delta)n$ vertices and contains a set Q of* special *vertices of size at least $\varepsilon n/\Delta$.*

Assume that H satisfies the following two conditions:

(E1) *$|N_H(X)\setminus C| \geq d|X|$ holds for all $X \subseteq V(H)$ with $1 \leq |X| \leq 2m$,*

(E2) *$e_H(X,Y) > 0$ holds for all disjoint $X, Y \subseteq V(H)$ with $|X| \geq m$ and $|Y| \geq m$.*

Then there exists an embedding of T into H with image S such that every critical vertex in H is the image of a special vertex of T and $|N_H(X)\setminus V(S)| \geq d|X|$ holds for all $X \subseteq V(H)$ with $1 \leq |X| \leq 2m$.

Next, before giving the proof of Theorem 5.4.1, we argue how it allows us to prove Theorem 1.2.11.

5.4.1 Proof of Theorem 1.2.11

According to Lemma 5.1.1, the random graph G_{M^*} a.a.s. satisfies the properties **(P1)**–**(P7)**. Thus, we need to show that if G is an n-vertex graph satisfying **(P1)**–**(P7)** and if T' is a n-vertex tree of maximum degree at most Δ and has at least εn many leaves, then we can find a copy of T' in G. For this, let D, U, W, and P be the vertex sets of G provided by Lemma 5.1.1.

Let H be the induced subgraph $G[U]$ and let C be the union of the sets W and P' , where P' is a subset of P that contains exactly one neighbors for each vertex in D. Note that Property **(P3)** together with Property **(P2)** ensure the existence of P' and that each vertex in D has distinct neighbors in P'. Furthermore, let Q be the set of vertices in T' that are parents of leaves and let T be the subtree of T' obtained by deleting from T' one leave attached to each of the vertices in Q. Note that the set Q has size at least $\varepsilon n/\Delta$.

Then T is a tree of maximum degree at most Δ and of size at most $(1-\varepsilon/\Delta)n$ which contains the set of special vertices Q. If we consider the graph H with the set of critical vertices C, then the properties **(P1)**–**(P7)** guarantee that we can apply Theorem 5.4.1. Thus, we can find an embedding of T into H (and thus into G) with image S such that every vertex in $W \cup P'$ (the critical vertices in H) is embedded to a vertex in Q (the special vertices in T).

To embed the remainder of T', we first embed the missing leaves attached to the vertices in Q that have been mapped to P' to the respective (unique) neighbor in D. Afterwards, it remains to show that we can embed the remaining leaves in T' without T to the vertices of G that are not covered by $S + D$. In other words, we need to find a perfect matching between the set A of vertices in S that lie in the image of Q but not in P' and the set B of vertices in H that are not covered by $S + D$. Observe crucially that A contains W.

By Hall's theorem (see, e.g., [21]), it is sufficient to show that the condition

$$|N_H(X) \cap B| \geq |X| \tag{5.6}$$

holds for all $X \subseteq A$. To see this, we distinguish three cases based on the size of X. Inequality (5.6) holds for $1 \leq |X| \leq m$ by Theorem 5.4.1, for $m \leq |X| \leq |A| - m$ because $e_H(X,Y) > 0$ holds for all $Y \subseteq B$ with $|Y| \geq m$, and for $|A| - m \leq |X| \leq |A|$ because of property **(P6)** in Lemma 5.1.1.

Thus, we find a matching between A and B in H which allows us to extend our current embedding to the missing leaves of T'.

The remainder of this section is devoted to the proof of Theorem 5.4.1.

5.4.2 Setup for the Proof of Theorem 5.4.1

Let us fix $\Delta \in \mathbb{N}$ and $0 < \varepsilon < 1$ and let us assume that n is sufficiently large with respect to Δ and $1/\varepsilon$. Furthermore, let $d = \frac{\varepsilon \log n}{30\Delta \log\log n}$, $m = \frac{2n \log\log n}{\log n}$, $k = \left\lceil \frac{\log(2m)}{\log(d-1)} \right\rceil$, and $\ell = 2k + 1$.

Let H be a host graph on n vertices and let C be a set of critical vertices in H. Finally, let T be a tree of maximum degree Δ which has at most $\varepsilon n/\Delta$ that is, at most $n - 11dm$ vertices and let Q be a set of *special* vertices in T of size at least $\varepsilon n/\Delta$.

Assume that H satisfies the conditions **(E1)** and **(E2)** in Theorem 5.4.1. Our aim is to give an embedding of T into H such that every critical vertex in H is embedded to a special vertex of T. To this end, we embed T successively into H. We say an embedding of a subforest F of T into H is *feasible* if every critical vertex of H in the image S of F is the image of a special vertex and, furthermore, the subgraph $S + C'$ of H is (d,m)-extendable, where C' is the set of critical vertices in H that are not in $V(S)$.

At each point of the following embedding process, we denote by F the current subforest of T for which we have already found a feasible embedding into H, and by S the corresponding image, and let $C' = C \setminus V(S)$. Whenever we extend the embedding of F, we ensure that the new embedding is still feasible. We guarantee this by only applying the operations from Section 5.2

to extend $S + C'$. In this, the challenge is to ensure at any point of the embedding process that we are still able to cover the non-covered critical vertices C' by the remaining special vertices in $T - F$.

Our two main tools are the Vertex Extension Lemma (Lemma 5.2.6) and the Connection Lemma (Lemma 5.2.6). Recall that T and therefore at any point of the proof, also $S + C'$ has at most $n - 11dm$ vertices. Thus, $|N_H(X)| \geq |V(S)| + 2dm + 1$ holds for all $X \subseteq V(G)$ with $m + 1 \leq |X| \leq 2m$, that is, throughout the remainder of this section, whenever we apply the Vertex Extension Lemma or the Connection Lemma, the corresponding conditions are satisfied.

At any time in the construction of the embedding of T, we can (repeatedly) apply the Vertex Extension Lemma to embed into H a subtree of $T - F$ that has a neighbor in F. This operation extends $S + C'$ by the image of the embedded subtree while maintaining the property of being (d, m)-extendable. In our proof this operation never touches the non-covered critical vertices in C', thus the resulting embedding is always feasible. To cover the so far uncovered vertices C', we apply the Connection Lemma in a rather involved construction which forms the core in our proof of Theorem 5.4.1 and is laid out in the following subsections.

5.4.3 Meta-Trees

The first ingredient to our construction is the following decomposition of T into vertex-disjoint subtrees, such that all but one of them contains a special vertex at a minimum depth but only few special vertices in total. For this, we consider T to be rooted at an arbitrary special vertex. This root of T never changes throughout the remainder of this section and implies a root vertex for every subtree of T which is the vertex in the subtree that is closest to the root of T.

For the moment, let $j \in \{1, \dots, k\}$ be fixed and consider the following procedure to find a subtree of T. Starting with the root of T as the current vertex, we successively branch into the root of a subtree of T pending from the current vertex, provided this subtree contains a special vertex at depth at least $2j$. Whenever this procedure stops, the current vertex is the root of a subtree of T that contains a special vertex at depth $2j$ and none at a higher depth. Thus, this subtree contains at most $2\Delta^{2j}$ special vertices.

To decompose T into subtrees, we repeatedly apply the above procedure to find a subtree of T with the above properties, remove it from T (together with the edge connecting it to T), add add it to the set of subtrees that forms the decomposition. This process stops when none of the subtrees pending from the root of T contains a special vertex at depth at least $2j$. The remainder of T is the last subtree of the decomposition, it is the only subtree that might not contain any special vertex.

Since all subtrees of T generated by the above process are connected, these subtrees together with the edges of T that lie between them again form a tree structure. We call this structure the (rooted) *meta-tree* [1] $\mathcal{T}_j$ of T *with parameter* j. That is, every (meta-)vertex of $\mathcal{T}_j$ is one of the subtrees of T given by the decomposition (with the subtree containing the root of T becoming the meta-root of the meta-tree) and every meta-edge in $\mathcal{T}_j$ corresponds to an edge in T connecting two meta-vertices. With slight abuse of notation, we identify a subforest $\mathcal{F}$ of $\mathcal{T}_j$ with the corresponding subtree F of T that is spanned by all vertices in T contained in the meta-vertices of $\mathcal{F}$.

The meta-tree $\mathcal{T}_j$ has two key properties. First, every meta-vertex other than the meta-root contains a special vertex of T at depth exactly $2j$. Second, every meta-vertex (including the meta-root) contains less than $2\Delta^{2j}$ special vertices. Therefore, the set of meta-vertices $\mathcal{V}_j$ of $\mathcal{T}_j$

[1] Strictly speaking, for a fixed j there may be several possible decompositions of T (choice of the root, choice of the subtrees while branching). However, we do not care for the particular decomposition and implicitly assume it is chosen in some canonical way. In other words, we suppose that there exists a fixed meta-tree $\mathcal{T}_j$ of T with parameter j.

has size at least

$$|\mathcal{V}_j| > \frac{|Q|}{2\Delta^{2j}} \geq \frac{\varepsilon n}{2\Delta^{2j+1}}. \tag{5.7}$$

Proposition 5.4.2. *Let $j \in \{1, \ldots, k-1\}$. Let F be a subforest of T and assume that there exists a feasible embedding of F with image S and that the set C' of critical vertices outside S is of size at least $\frac{m}{(d-1)^j}$. Assume further that there exists a set $P = \{v_1, \ldots, v_r\}$ of vertices in F with $\frac{m}{(d-1)^j} \leq r \leq |C'|$ and distinct (non-root) meta-vertices $M_1, \ldots, M_r$ in $\mathcal{T}_j$, such that F and the M_i's are vertex-disjoint and such that, for every $i \in \{1, \ldots, r\}$, the root of M_i is connected by an edge e_i to the vertex v_i. Then we can extend the feasible embedding of F to a feasible embedding of $F + \sum_{i=1}^{r}(e_i + M_i)$ containing at least $r - \left\lfloor \frac{m}{(d-1)^j} \right\rfloor$ vertices from C'.*

Proof. We first show that there exists an $i \in \{1, \ldots, r\}$ such that we can extend the embedding of F to a feasible embedding of $F + e_i + M_i$.

For this, let A be the image of P in S. By the prerequisites of Proposition 5.4.2, both A and C' are of size at least $\frac{m}{(d-1)^j}$. Thus, we may apply the Connection Lemma (Lemma 5.2.6) to connect the vertex sets A and C' by a path of length $2j+1$ outside of $S + C'$. Let $c \in C'$ be the one end-point of this path and let $i \in \{1, \ldots, r\}$ be the index for which the other end-point a_i is the image of the vertex v_i in the embedding of F.

Now, since the subtree M_i is a meta-vertex in the meta-tree $\mathcal{T}_j$, it contains a special vertex w at depth $2j$. Thus, the vertices v_i and w are connected by a path of length $2j+1$ in T. We embed this path between v_i and w in T to the path of the same length between a_i and b in H.

Next, we apply the Vertex Extension Lemma (Lemma 5.2.6) to embed the remainder of M_i. Note that the preimage of the critical vertex b is the special vertex w. Together with the Connection Lemma and the Vertex Extension Lemma, this guarantees that our embedding of $F + e_i + M_i$ is feasible.

So far, we have seen how to find one index $i \in \{1, \ldots, r\}$ such that we can extend the embedding of F to a feasible embedding of $F + e_i + M_i$. However, as long as the sets P and C' are of size at least $\frac{m}{(d-1)^j}$ each, we can repeatedly do so, each time removing v_i from P and b from C' afterwards. Thus, we can embed the pending subtrees of $r - \left\lfloor \frac{m}{(d-1)^j} \right\rfloor$ vertices in P, such that each time a vertex in C' becomes the image of a special vertex in T. Hence, the resulting embedding is feasible.

Finally, we embed the remaining $\left\lfloor \frac{m}{(d-1)^j} \right\rfloor$ pending M_i's by applying only the Vertex Extension Lemma, not covering any further of the remaining vertices in C'. □

Observe that a bare (meta)-path in the meta-tree that does not contain the meta-root has a canonical ordering of its meta-vertices where each (except for the first) meta-vertex in the meta-path is a rooted subtree of T pending by an edge from a vertex in the preceding meta-vertex.

Proposition 5.4.3. *Let $j \in \{1, \ldots, k-1\}$. Let F be a subforest of T and assume that there exists a feasible embedding of F with image S and remaining critical vertices $C' := C \setminus V(S)$. Furthermore, suppose that there exist two vertices v and w in F with images a and b in S, respectively, and a bare meta-path (not containing the meta-root) of length at least 2ℓ in $\mathcal{T}_j$ such that in T the root of the first meta-vertex is connected by an edge to v and the last meta-vertex in the path is connected by an edge to w. Then we can extend the feasible embedding of F to a feasible embedding of F plus the meta-path plus the two edges connecting the meta-path to v and w. Moreover, if C' is non-empty, we can embed a special vertex of T that is contained in the bare meta-path to one of the vertices in C'.*

Proof. Consider the $(\ell{+}1)$-st meta-vertex in the bare meta-path. Since it is not the meta-root, it contains a special vertex u of T. Let us denote by u' the (unique) vertex lying on all three paths between v and w, between u and v, and between u and w in T. Since the $(\ell+1)$-st meta-vertex has distance (at least) ℓ in the bare meta-path from both (meta-)end-vertices, the two paths in T, one from v to u' and one from u' to w, have length at least ℓ.

We now distinguish two cases, depending on whether C' is empty or not. If C' is empty, we apply the Connection Lemma (Lemma 5.2.9) with parameter $j = k$ to find (and add to S) a path of length $\mathrm{dist}_T(v, w)$ between a and b. Strictly speaking, since the Connection Lemma only allows us to embed paths of length exactly ℓ, we first embed all but the last ℓ vertices of this path using the Vertex Extension Lemma (Lemma 5.2.6) and only then apply the Connection Lemma. If C' is non-empty and $c \in C'$, we use the same argument to find (and add to S) two paths in H, one path of length $\mathrm{dist}_T(v, u)$ between a and c and one path of length $\mathrm{dist}_T(u', w)$ connecting the previous path and b. Like in Proposition 5.4.2, we afterwards apply the Vertex Extension Lemma to embed the rest of the bare meta-path. □

5.4.4 Many Leaves Versus Long Bare Paths

The second ingredient to our construction is a standard tool to find an embedding a given tree in pseudo-random graphs (for example, used in [59, 66]). It guarantees that every tree has (i) many leaves or (ii) a large number of vertex-disjoint bare paths of a certain length. The following version is a corollary of Lemma 2.1 in [66].

Lemma 5.4.4 (Corollary of Lemma 2.1 in [66])**.** *Every tree F on a sufficiently large vertex set $[n]$ that does not contain a collection of at least $\frac{|V(F)|}{10\ell}$ vertex-disjoint bare paths of length 3ℓ has at least $\frac{|V(F)|}{10\ell}$ leaves.*

A typical application of Lemma 5.4.4 is a case distinction between the case that the embedded tree has many leaves and the case that it has many bare paths of a certain length (note that these cases do not need to be exclusive). In our proof, we also make a case distinction. However, instead of applying Lemma 5.4.4 directly to the tree T, we apply it iteratively to the meta-trees $\mathcal{T}_1, \ldots, \mathcal{T}_k$.

In case that the meta-tree with parameter j has many bare meta-paths of length roughly 3ℓ, we apply Proposition 5.4.3 to cover all (remaining) critical vertices with these bare meta-paths. In case that the meta-tree with parameter j has many meta-leaves, however, we apply Proposition 5.4.2 to cover all but roughly m/d^j remaining critical vertices with these meta-leaves. Afterwards, we iterate the same case distinction on the meta-tree with parameter $j + 1$.

Case 1: *$\mathcal{T}_1$ has a (meta-)subtree on at least $|\mathcal{V}_1|/2$ meta-vertices that contains a collection of at least $\frac{|\mathcal{V}_1|}{20\ell}$ vertex-disjoint bare meta-paths of length 3ℓ each.*

Proposition 5.4.5. *If T satisfies the condition of Case 1, then there exists a feasible embedding of T into H.*

We postpone the proof of this proposition to the next subsection.

Next, assume that we are not in Case 1. Then Lemma 5.4.4 implies that $\mathcal{T}_1$ contains a collection of at least $\frac{|\mathcal{V}_1|}{20\ell}$ meta-leaves. Now, we perform the following iterative procedure. Starting with $\mathcal{T}_1$, we repeatedly remove $\frac{|\mathcal{V}_1|}{20\ell}$ meta-leaves from the current meta-tree, until we remove at least $10\Delta m$ meta-leaves in total. We call the meta-leaves that were removed in the i-th iteration of this process the i-th level of meta-leaves. Since we remove less than $10\Delta m + \frac{|\mathcal{V}_1|}{20\ell}$ meta-leaves in total and since

$$10\Delta m + \frac{|\mathcal{V}_1|}{20\ell} < 11\Delta m = \frac{22\Delta n \log\log n}{\log n} \ll \frac{\varepsilon n}{6\Delta^3} \overset{(5.7)}{<} \frac{|\mathcal{V}_1|}{3},$$

the current meta-tree has more than $|\mathcal{V}_1|/2$ meta-vertices throughout the process. Thus, we can indeed remove $\frac{|\mathcal{V}_1|}{20\ell}$ meta-leaves in each iteration until we removed at least $10\Delta m$ (and at most

$10\Delta m + \frac{|\mathcal{V}_1|}{20\ell} < 11\Delta m$) meta-vertices in total. Notice that there are in total at most

$$\frac{11\Delta m}{\frac{|\mathcal{V}_1|}{20\ell}} < \frac{22n\log\log n \cdot 2\Delta^3 \cdot 40\Delta \log n}{\log n \cdot \varepsilon n \cdot \log\log n} < \frac{2000\Delta^4}{\varepsilon} \tag{5.8}$$

layers.

Finally, we make a further case distinction, depending on whether one of the meta-trees $\mathcal{T}_2, \ldots, \mathcal{T}_k$ contains a collection of at least $\frac{|\mathcal{V}_1|}{10\ell}$ vertex-disjoint bare meta-paths of length 3ℓ each.

Case 2: *$\mathcal{T}_1$ contains a (meta-)subtree that can be extended to $\mathcal{T}_1$ by successively adding layers of at least $\frac{|\mathcal{V}_1|}{20\ell}$ meta-leaves, such that between $10\Delta m$ and $11\Delta m$ meta-vertices are added in total. Furthermore, there exists an index $h \in \{2, \ldots, k\}$ such that $\mathcal{T}_h$ contains a collection of at least $\frac{|\mathcal{V}_h|}{10\ell}$ vertex-disjoint bare meta-paths of length 3ℓ each.*

Proposition 5.4.6. *If T satisfies the conditions of Case 2, then there exists a feasible embedding of T into H.*

We also postpone the proof of this proposition to the next subsection.

Case 3: *$\mathcal{T}_1$ contains a (meta-)subtree that can be extended to $\mathcal{T}_1$ by successively adding layers of at least $\frac{|\mathcal{V}_1|}{20\ell}$ meta-leaves, such that between $10\Delta m$ and $11\Delta m$ meta-vertices are added in total. Furthermore, for every index $j \in \{2, \ldots, k\}$, $\mathcal{T}_j$ does not contain a collection of at least $\frac{|\mathcal{V}_j|}{10\ell}$ vertex-disjoint bare meta-paths of length 3ℓ each.*

Proposition 5.4.7. *If T satisfies the conditions of Case 3, then there exists a suitable embedding of T into H.*

We again postpone the proof of this proposition to the next subsection.

5.4.5 The Embedding

Finally, we construct the actual embeddings for the three cases in the previous subsection, that is, we give the proofs of Propositions 5.4.5, 5.4.6, and 5.4.7.

Proof of Proposition 5.4.5. Let $\mathcal{T}_1^*$ be a (meta-)subtree of $\mathcal{T}_1$ containing a collection of $\frac{|\mathcal{V}_1|}{20\ell} - 1$ vertex-disjoint bare meta-paths of length 3ℓ (none of which contains the root of $\mathcal{T}_1$) and let $\mathcal{F}$ be the (meta-)subforest of $\mathcal{T}_1^*$ obtained by removing from $\mathcal{T}_1^*$ all meta-vertices in these $\frac{|\mathcal{V}_1|}{20\ell} - 1$ bare meta-paths except for the end-meta-vertices. Then, $\mathcal{F}$ contains exactly $\frac{|\mathcal{V}_1|}{20\ell}$ components that are connected in $\mathcal{T}_1^*$ by these bare meta-paths.

Let us denote by F the subforest of T corresponding to $\mathcal{F}$. We first embed F into H. Since each component in $\mathcal{F}$ contains at least one non-root meta-vertex (indeed, every component of $\mathcal{F}$ contains at least two meta-vertices by construction), every component of F contains a special vertex. We embed each of these $\frac{|\mathcal{V}_1|}{20\ell}$ special vertices of T to a distinct (arbitrary) critical vertex of H. (Note that the set C of critical vertices has size $8m$ and there are less components in F). Let S' be the image of this embedding and let C' be the set of critical vertices that are not in S'. Then, by Property **(E1)** of Theorem 5.4.1, the subgraph $S' + C'$ of H is (d, m)-extendable. Thus, we can apply the Vertex Extension Lemma (Lemma 5.2.6) to extend $S' + C'$ to a feasible embedding of $F + C'$.

It remains to embed the bare meta-paths such that the preimage of C' contains only special vertices. We orient the chosen bare meta-paths arbitrarily and number the meta-vertices in each of them from 1 to $3\ell - 1$. For $i \in [3\ell - 1]$, we denote by the *i-th layer of subtrees* the set of all i-th meta-vertices in the bare meta-paths.

In the first stage, we repeatedly apply Proposition 5.4.2 with the currently embedded subforest being F, $r = \frac{|\mathcal{V}_1|}{20\ell}$, and the next ($i$th) layer of subtrees being $M_1, \ldots, M_r$, to embed these subtrees of T layer by layer, each time (except in the last iteration) covering $\frac{|\mathcal{V}_1|}{20\ell} - 1 - \frac{m}{d-1}$ additional critical vertices in H. We proceed until at most $\frac{m}{d-1}$ critical vertices are not in the image of the embedding.

In the second stage, we repeatedly apply Proposition 5.4.3 to embed one by one the remainders of the bare meta-paths. Since we have more that $\frac{m}{d-1}$ bare meta-path, and their remainders are of length at least 2ℓ each, Proposition 5.4.3 allows us to cover the remaining non-covered critical vertices of H in the process. □

The following two claims are fundamental for our proof. They roughly state that, although the meta-trees $\mathcal{T}_1, \ldots, \mathcal{T}_k$ do not seem to necessarily have much of a common structure, we can somewhat control their interaction where we need to do so. The statements of the claims are not tight but they suffice for our purposes.

Claim 6. *For any two indices $i, j \in [k]$, $i < j$, every bare 3ℓ-meta-path in $\mathcal{T}_j$ is vertex-disjoint (in T) from all but at most $7\ell\Delta^{2j}$ meta-vertices of $\mathcal{T}_i$.*

Proof. Let us fix an arbitrary bare ℓ-meta-path in $\mathcal{T}_j$. The main observation of the proof is that almost every meta-vertex of $\mathcal{T}_i$ is either contained in this bare meta-path, or it is vertex-disjoint (in T) from the meta-path. There are at most three exceptional meta-vertices in $\mathcal{T}_i$ that intersect the bare meta-path, but are not contained in it: these are exactly the meta-vertices containing the (at most) two edges connecting the bare meta-path with the rest of T, and the meta-root of $\mathcal{T}_i$. The proof is now a simple double-counting argument. Every meta-vertex of $\mathcal{T}_j$ contains at most $2\Delta^{2j}$ special vertices. On the other hand, every meta-vertex of $\mathcal{T}_i$ except of the meta-root of $\mathcal{T}_i$ contains at least one special vertex. Hence, at most $6\ell\Delta^{2j}$ meta-vertices of $\mathcal{T}_i$ (not counting the meta-root of $\mathcal{T}_i$) are completely contained in the chosen bare ℓ-meta-path in $\mathcal{T}_j$. Adding the at most three exceptional meta-vertices of $\mathcal{T}_i$, we obtain the statement of the claim. □

Claim 7. *For any two indices $i, j \in [k]$, $i < j$, every meta-leaf in $\mathcal{T}_j$ is vertex-disjoint (in T) from all but at most $3\Delta^{2j}$ meta-vertices of $\mathcal{T}_i$.*

Proof. The proof proceeds similarly to the proof of Claim 6. Let us fix an arbitrary meta-leaf in $\mathcal{T}_j$. Almost every meta-vertex of $\mathcal{T}_i$ is either contained in this meta-leaf, or it is vertex-disjoint (in T) from this meta-leaf. The only possible exception is the meta-vertex of $\mathcal{T}_i$ containing the edge from T that connects the chosen meta-leaf of $\mathcal{T}_j$ to the rest of the tree. The double-counting argument from the proof of Claim 6 (again taking care of the meta-root of $\mathcal{T}_i$) provides the statement of the claim. □

We are now ready to prove the final two propositions.

Proof of Proposition 5.4.6. In Case 2, there exists an index $h \in \{2, \ldots, k\}$ such that $\mathcal{T}_h$ contains a collection of at least $\frac{|\mathcal{V}_h|}{10\ell}$ vertex-disjoint bare meta-paths of length 3ℓ. Without lost of generality, h is the smallest index from $\{2, \ldots, k\}$ with this property. Then, for $j \in \{2, \ldots, h-1\}$, Lemma 5.4.4 implies that the meta-tree $\mathcal{T}_j$ has at least $\frac{|\mathcal{V}_j|}{10\ell}$ meta-leaves. Moreover, by the conditions given in Case 2, the-meta tree $\mathcal{T}_1$ contains at least $10\Delta m$ and at most $11\Delta m$ meta-vertices that can be successively stripped from $\mathcal{T}_1$ in layers of meta-leaves of size at least $\frac{|\mathcal{V}_1|}{20\ell}$.

The core of this proof is a decomposition of T into a forest F, bare meta-paths between the components of F, and meta-leaves attached (in layers) to F.

First, we choose a collection of $\left\lceil \frac{m}{(d-1)^h} \right\rceil$ vertex-disjoint bare meta-paths of length 3ℓ each in $\mathcal{T}_h$ not containing the meta-root. Second, for j ranging from $h-1$ down to 2, we iteratively reserve $\frac{\Delta m}{(d-1)^{j-1}}$ (non-root) meta-leaves of $\mathcal{T}_j$, such that these meta-leaves neither intersect (in T)

the chosen bare meta-paths of $\mathcal{T}_h$ nor the reserved meta-leaves of the meta-trees $\mathcal{T}_{j+1}, \dots, \mathcal{T}_{h-1}$. Finally, we reserve layers of meta-leaves of $\mathcal{T}_1$ such that these meta-leaves neither intersect the chosen bare meta-paths of $\mathcal{T}_h$ nor the reserved meta-leaves of the meta-trees $\mathcal{T}_{j+1}, \dots, \mathcal{T}_{h-1}$. Furthermore, we require every layer to contain at least $\frac{|\mathcal{V}_1|}{21\ell}$ meta-vertices of $\mathcal{T}_1$, and in total, at least $9.5\Delta m$ meta-vertices of $\mathcal{T}_1$ should be reserved.

To see that we can indeed do so, let us fix an arbitrary index $j \in \{1, \dots, h-1\}$ and assume by induction that we already reserved $\frac{\Delta m}{(d-1)^{i-1}}$ meta-leaves in the meta-trees $\mathcal{T}_i$ for $i \in \{j+1, \dots, h-1\}$. We call a meta-vertex in $\mathcal{T}_j$ *forbidden* if it intersects (in T) the chosen meta-paths of $\mathcal{T}_h$ or the reserved meta-leaves of $\mathcal{T}_{j+1}, \dots, \mathcal{T}_{h-1}$. We denote by f_j the number of forbidden vertices in $\mathcal{T}_j$ and note that by Claims 6 and 7, we can estimate

$$f_j \leq 7\ell\Delta^{2h} \left\lceil \frac{m}{(d-1)^h} \right\rceil + \sum_{i=j+1}^{h-1} \frac{3\Delta^{2j+1} m}{(d-1)^{i-1}} \ll \frac{|\mathcal{V}_j|}{\ell}.$$

Now, for j ranging from $h-1$ down to two, from the reserved $\frac{\Delta m}{(d-1)^{j-1}}$ non-root meta-leaves of $\mathcal{T}_j$, we choose $\frac{m}{(d-1)^{j-1}}$ meta-leaves such that their roots have distinct neighbors in T. Furthermore, from the layers of reserved meta-leaves of $\mathcal{T}_1$, we choose sublayers of at least $\frac{|\mathcal{V}_1|}{21\Delta\ell}$ (not unwanted) meta-leaves, such that the roots of all meta-vertices from one sublayer have distinct neighbors in T, and the total number of chosen meta-vertices of $\mathcal{T}_1$ is at least $9.5m$.

Furthermore, since we are aiming at embedding the meta-vertices starting with $\mathcal{T}_1$ and going up to $\mathcal{T}_h$, chosen meta-vertices of $\mathcal{T}_1$ lying above a chosen meta-vertex of $\mathcal{T}_j$ in T for any $j \geq 2$ might cause us additional technical difficulties. We call such meta-vertices of $\mathcal{T}_1$ *unwanted*. Luckily, for every $j \geq 2$, every meta-vertex M of $\mathcal{T}_j$, and every layer of chosen meta-vertices of $\mathcal{T}_1$, there exists at most one meta-vertex of $\mathcal{T}_1$ from this layer such that M lies below it in T. Furthermore, for every $j \geq 2$, every bare meta-path of $\mathcal{T}_j$, and every layer of chosen meta-vertices of $\mathcal{T}_1$, there exists at most one meta-vertex of $\mathcal{T}_1$ from this layer such that some vertex of $\mathcal{T}_j$ from the bare meta-path lies below it in T. Hence, using (5.8) we see that the number of unwanted meta-vertices from one layer of leaves of $\mathcal{T}_1$ is at most

$$\sum_{j=2}^{h} \frac{2000\Delta^4}{\varepsilon} \left\lceil \frac{\Delta m}{(d-1)^{j-1}} \right\rceil \ll \frac{|\mathcal{V}_1|}{\ell}.$$

Thus, we can decompose T as desired into

- a subforest F (containing the root of T) which contains all vertices of T except for the following,
- the (non-endpoint) meta-vertices of the $\left\lceil \frac{m}{(d-1)^h} \right\rceil \left(\ll \frac{|\mathcal{V}_h|}{\ell} \right)$ bare meta-paths of $\mathcal{T}_h$ of length 3ℓ each connecting the components of F,
- $\frac{m}{(d-1)^{j-1}} \left(\ll \frac{|\mathcal{V}_j|}{\ell} \right)$ meta-leaves of $\mathcal{T}_j$ pending from distinct vertices of F for every $j \in \{2, \dots, h-1\}$, and
- meta-vertices of $\mathcal{T}_1$ pending from F (and from meta-vertices of $\mathcal{T}_1$ from previous layers) in at most $2000\Delta^4/\varepsilon$ layers, such that the total number of these meta-vertices of $\mathcal{T}_1$ is at least $9m$. Furthermore, every layer contains a sublayer of size at least $\frac{|\mathcal{V}_1|}{22\Delta\ell}$ such that roots of meta-vertices from one sublayer have distinct neighbors in T.

With this decomposition at hand, we construct a feasible embedding of T. Similarly to the proof of Proposition 5.4.5, we proceed in several stages.

In the first stage, we proceed exactly as in Proposition 5.4.5 and apply the Vertex Extension Lemma (Lemma 5.2.6) to find a feasible embedding of F into H.

In the second stage, we embed the layers of pending subtrees corresponding to the layers of meta-leaves of $\mathcal{T}_1$ such that all but m of the critical vertices are covered. We do this layer by layer, starting with the one that was stripped of last. For every layer, we first apply Proposition 5.4.2 to the chosen sublayer, i.e., the subtrees M_i are the meta-vertices of the sublayer, P is the set of neighbors of the roots of these meta-vertices in T, and $j = 1$. After that, we complete the embedding of the remainder of the layer by applying the Vertex Extension Lemma. With each layer, we cover the sublayers size minus $\frac{m}{d-1}$ many critical vertices by the embedding. Thus, since the number of layers is bounded by $2000\Delta^4/\varepsilon$ by (5.8) and the cumulative size of all sublayers is at least $9m \gg \frac{m}{d-1}$, we can cover all but $\frac{m}{d-1}$ of the at most $8m$ critical vertices outside F during this stage.

In the third stage, we use the remaining subtrees pending from F to reduce the number of non-covered critical vertices in H to $\frac{m}{(d-1)^{h-1}}$. This is done iteratively. For j ranging from 2 to $h-1$, we use the $\frac{m}{(d-1)^{j-1}}$ subtrees corresponding to the chosen meta-leaves of $\mathcal{T}_j$ to reduce the number of non-covered critical vertices in H from $\frac{m}{(d-1)^{j-1}}$ to $\frac{m}{(d-1)^{j}}$. In fact, we can do so by a single application of Proposition 5.4.2, expectedly setting M_i's to be the chosen meta-leaves.

In the fourth stage, we are only left with the meta-vertices of the chosen bare meta-paths of $\mathcal{T}_h$ to cover the remaining $\frac{m}{(d-1)^{h-1}}$ yet non-covered critical vertices in H. Similarly to the proof of Proposition 5.4.5, in the first sub-stage we apply Proposition 5.4.2 to embed the bare meta-paths layer by layer until only $\left\lceil \frac{m}{(d-1)^h} \right\rceil$ critical vertices remain non-covered, after which we complete the paths one by one (using Proposition 5.4.2) in the second sub-stage. We omit the details now, as they are completely identical to those exposed in the proof of Proposition 5.4.5. $\square$

Proof of Proposition 5.4.7. This proof closely follows the proof of Proposition 5.4.6. The only difference is that there is no index $h \in \{2, \dots, k\}$ such that $\mathcal{T}_h$ contains $\left\lceil \frac{|\mathcal{V}_h|}{10\ell} \right\rceil$ vertex-disjoint bare meta-paths of length 3ℓ each. However, instead of choosing $\left\lceil \frac{m}{(d-1)^h} \right\rceil$ vertex-disjoint bare meta-paths in $\mathcal{T}_h$, we only choose one (non-root) meta-leaf M of $\mathcal{T}_k$. Observe crucially that $\left\lceil \frac{m}{(d-1)^k} \right\rceil = 1$.

The complete embedding process then proceeds exactly as in the proof of Proposition 5.4.6 with $h = k$, until we come to the forth stage and are left with exactly one remaining non-covered critical vertices in H and the chosen meta-leaf M of $\mathcal{T}_k$ to embed it. To do so, we apply Proposition 5.4.2 with $r = 1$ and $M_1 = M$. This provides the statement of the proposition. $\square$

5.5 Concluding remark

The only aim of this section is to present the following conjecture. We hope that it will lead the research in this area to new insights in the random graph process.

Conjecture 5.5.1. For every $\Delta \in \mathbb{N}$, in the random graph process, when the graph contains a Hamilton path, it a.a.s. contains all spanning trees with maximum degree at most Δ.

6 On extremal hypergraphs for Hamilton cycles

6.1 Proofs

6.1.1 Outline of the Proofs

In the following we give a brief overview over the structure of the proofs of Theorems 1.2.12 and 1.2.13. For this section we define for every $k \in \mathbb{N}$

$$\varepsilon = \frac{1}{22(1280k^3)^{k-1}}$$

and

$$\rho = (22\varepsilon)^{\frac{1}{k-1}}.$$

Suppose $H = (V, E)$ is a k-graph on n vertices with $\delta_1 \geq (1-\varepsilon)\binom{n-1}{k-1}$ and n sufficiently large. By an *end* of a path $P_t^{(k,l)}$ we mean the tuple consisting of its first $k-1$ vertices, $(x_1, \dots, x_{k-1})$, or the tuple consisting of its last $k-1$ vertices in reverse order, $(x_t, \dots, x_{t-k+2})$, considering the ordered vertices. For an i-tuple $(x_1, \dots, x_i)$ in H we write $\boldsymbol{x_i}$, $1 \leq i \leq n$. We call $\boldsymbol{x_{k-1}}$ *good* if all x_is are pairwise distinct and for all $i \in \{1, \dots, k-1\}$ it holds that

$$\deg(x_1, \dots, x_i) \geq \left(1 - \rho^{k-i}\right)\binom{n-i}{k-i}. \tag{6.1}$$

A path is called *good* if both of its ends are good.

Outline of the proofs and some definitions:

1. At first, we prove the existence of one l-tight good path or several vertex-disjoint good tight paths containing the vertices of small degree, see Claim 8. (Note that we do not need this step in the proof of Theorem 1.2.13.)

2. We say that a tuple $\boldsymbol{x_{2k-2}}$ *absorbs* a vertex $v \in V$ if both $\boldsymbol{x_{2k-2}}$ and $(x_1, \dots, x_{k-1}, v, x_k, \dots, x_{2k-2})$ induce good paths in H, meaning that the corresponding ordering of the paths is $\boldsymbol{x_{2k-2}}$ or $(x_1, \dots, x_{k-1}, v, x_k, \dots, x_{2k-2})$, respectively, and the ends are good. Lemma 6.1.2 ensures a set $\mathcal{A}$, such that any remaining vertex can be absorbed by many tuples of $\mathcal{A}$. We call an element of $\mathcal{A}$ an *absorber*.

3. For $\boldsymbol{x_i}, \boldsymbol{y_j} \in V^{k-1}$ we define

$$\boldsymbol{x_i} \Diamond \boldsymbol{y_j} := (x_1, \dots, x_i, y_1, \dots, y_j).$$

   Let $\boldsymbol{x_{k-1}}$ and $\boldsymbol{y_{k-1}}$ be good. We say that a tuple $\boldsymbol{z_{k-1}}$ *connects* $\boldsymbol{x_{k-1}}$ with $\boldsymbol{y_{k-1}}$ if $(x_{k-1}, \dots, x_1) \Diamond \boldsymbol{z_{k-1}} \Diamond \boldsymbol{y_{k-1}}$ induces a path in H with respect to the order. Notice that the connecting-operation is not symmetric. Lemma 6.1.3 guarantees a set $\mathcal{C}$ such that any pair of $(k-1)$-tuples in H can be connected by many elements of $\mathcal{C}$. We call the elements of $\mathcal{C}$ *connectors*.

4. We modify $\mathcal{A}$ and $\mathcal{C}$ such that $\mathcal{A}$, $\mathcal{C}$ and the element(s) of Step 1 are pairwise vertex-disjoint.

5. In Lemma 6.1.4 we create a good tight path that contains all elements of the modified $\mathcal{A}$, respecting their ordering.

6. Using Lemma 6.1.5, we extend the path from Step 5 until it covers almost all of the remaining vertices that neither participate in (l-tight or tight) good paths of Step 1 nor in the modified $\mathcal{C}$.

7. Using connectors, we create a cycle containing the (l-tight or tight) good paths from Step 1 and the good path from Step 6.

8. In the final step all remaining vertices are absorbed by the absorbers in the cycle.

6.1.2 Auxiliary Lemmas

In this part we derive the main tools used to prove Theorems 1.2.12 and 1.2.13. For this subsection let $H = (V, E)$ be a k-graph on n vertices with

$$\delta_1 \geq (1-\varepsilon)\binom{n-1}{k-1}. \tag{6.2}$$

Recall $\varepsilon = \frac{1}{22(1280k^3)^{k-1}}$ and n sufficiently large.

The following lemma provides us with an essential tool which we use to prove other statements in this subsection.

Lemma 6.1.1. *Let $\boldsymbol{x_{2k-2}}$ be chosen u.a.r. from V^{2k-2}. The probability that all x_is are pairwise distinct and both $(x_1, \dots, x_{k-1})$ and $(x_{2k-2}, \dots, x_k)$ are good is at least $\frac{8}{11}$.*

Proof. Let a be the number of $(k-1)$-tuples that are not good and have $k-1$ distinct entries, i.e. that are taken from V without repetition. Further, let b_j be the number of j-tuples $\boldsymbol{y_j}$ with $\deg(y_1, \dots, y_j) < (1-\rho^{k-j})\binom{n-j}{k-j}$, $j \in \{1, \dots, k-1\}$, and all y_js are again pairwise distinct. Thus, by the definition of a good tuple, for each tuple $\boldsymbol{y_{k-1}}$ that is not good and has pairwise distinct entries, there exists a j such that $\boldsymbol{y_j}$ is one of the b_j tuples with small degree. Furthermore, for every $\boldsymbol{y_j}$ there are at most $\frac{(n-j)!}{(n-k+1)!}$ different $(k-1)$-tuples $(y_1, \dots, y_j, z_1, \dots, z_{k-1-j})$ with pairwise distinct $z_j \in V \backslash \{y_1, \dots, y_j\}$. Hence,

$$a \leq \sum_{j=1}^{k-1} \frac{(n-j)!}{(n-k+1)!} b_j.$$

The second time we apply double counting, we recall that H has at most $\varepsilon\binom{n}{k}$ non-edges. Each of the b_j j-tuples is by definition in at least $\rho^{k-j}\binom{n-j}{k-j}$ non-edges, and from every non-edge one obtains $\binom{k}{j}j!$ different j-tuples. Thus,

$$\rho^{k-j}\binom{n-j}{k-j} b_j \leq \binom{k}{j} j! \varepsilon \binom{n}{k}.$$

Putting the two bounds together, we obtain for a $vectw_{k-1}$ chosen u.a.r. from V^{k-1}

$$
\begin{aligned}
\mathbf{Pr}\left[\boldsymbol{w_{k-1}} \text{ is not good and has pairwise distinct entries}\right] &= \frac{a}{n^{k-1}} \\
&\leq \sum_{j=1}^{k-1} \frac{(n-j)!}{(n-k+1)!} b_j \frac{1}{n^{k-1}} \\
&\leq \sum_{j=1}^{k-1} \frac{(n-j)!}{(n-k+1)!} \frac{\binom{k}{j} j! \varepsilon \binom{n}{k}}{\rho^{k-j}\binom{n-j}{k-j}} \frac{1}{n^{k-1}} \\
&\leq \varepsilon \sum_{i=1}^{k-1} \frac{1}{\rho^{k-i}} < \frac{2\varepsilon}{\rho^{k-1}} = \frac{1}{11}.
\end{aligned}
$$

Then for a $vectx_{2k-2}$ chosen u.a.r. from V^{2k-2}

$$
\begin{aligned}
&\mathbf{Pr}[(x_1,\ldots,x_{k-1}) \text{ and } (x_{2k-2},\ldots,x_k) \text{ are good}] \\
&\quad \geq \frac{n(n-1)\ldots(n-2k+3)}{n^{2k-2}} - \frac{2}{11} \geq 1 - \frac{3}{11} = \frac{8}{11}.
\end{aligned}
$$

□

For a given set $\mathcal{X}$ of tuples or graphs, we write X when considering the corresponding vertex set.

Lemma 6.1.2. *For all γ, $0 < \gamma \leq \frac{1}{64k^2}$, there exists a set $\mathcal{A}$ of size at most $2\gamma n$ consisting of disjoint $(2k-2)$-tuples, each inducing a good path with respect to its order, such that for each vertex $v \in V$ at least $\frac{\gamma n}{4}$ tuples in $\mathcal{A}$ absorb v.*

Proof. By Lemma 6.1.1, we know that there are at least $\frac{8}{11}n^{2k-2}$ tuples $\boldsymbol{x_{2k-2}} \in V^{2k-2}$, such that the x_is are pairwise distinct and both $(x_1,\ldots,x_{k-1})$ and $(x_{2k-2},\ldots,x_k)$ are good. We denote the set of such tuples by $\mathcal{A}'$.

Let v be a vertex from V and denote by $\mathcal{A}_v$ the set of tuples $\boldsymbol{x_{2k-2}}$ from $\mathcal{A}'$ such that, in addition,

- $\{x_j,\ldots,x_{j+k-1}\} \in E(H)$, $1 \leq j \leq k-1$, and
- $\{v, x_j,\ldots,x_{j+k-2}\} \in E(H)$, $1 \leq j \leq k$.

Therefore, the set $\mathcal{A}_v$ consists of those tuples that can absorb the vertex v. From the minimum degree condition on H, see (6.2), it follows that

$$|\mathcal{A}_v| \geq \frac{8}{11}n^{2k-2} - 2k\varepsilon n^{2k-2} \geq \frac{7}{11}n^{2k-2}. \tag{6.3}$$

Fix γ with $0 < \gamma \leq \frac{1}{64k^2}$. Let $\mathcal{A}$ be the set obtained by choosing each $(2k-2)$-tuple $\boldsymbol{x_{2k-2}} \in V^{2k-2}$ from $\mathcal{A}'$ independently with probability $\frac{\gamma}{n^{2k-3}}$.

The expected size of $|\mathcal{A}|$ is at most γn and we apply Chernoff's inequality:

$$\mathbf{Pr}\left[|\mathcal{A}| - \gamma n > \gamma n\right] < e^{-\gamma n}. \tag{6.4}$$

This way, with high probability we obtain at most $2\gamma n$ many $(2k-2)$-tuples.

Let Y be the random variable taking the value 1 whenever a pair of tuples in $\mathcal{A}$ is not vertex-disjoint and the value 0 else. Thus,

$$\mathbb{E}[Y] \leq (2k-2)^2 n^{4k-5} \frac{\gamma^2}{n^{4k-6}} \leq 4k^2\gamma^2 n. \tag{6.5}$$

Applying Markov's inequality, we obtain:

$$\mathbf{Pr}\left[Y > 8k^2\gamma^2 n\right] < \frac{1}{2}. \tag{6.6}$$

From (6.3) we infer by Chernoff's inequality that

$$\mathbf{Pr}\left[|\mathcal{A}_v \cap \mathcal{A}| < \frac{\gamma n}{2}\right] < e^{-\frac{1}{100}\gamma n} \tag{6.7}$$

since $\mathbb{E}[|\mathcal{A}'_v \cap \mathcal{A}|] \geq \frac{7}{11}\gamma n$.

With (6.3), (6.6) and (6.7), we see that if we delete from $\mathcal{A}$ those pairs of tuples that have vertices in common, we obtain with probability at least

$$1/2 - e^{-\gamma n} - ne^{-\frac{1}{100}\gamma n} > 1/4$$

a new set satisfying the conditions of the lemma (note that $\frac{\gamma n}{2} - 8k^2\gamma^2 n \geq \frac{\gamma n}{4}$). □

The following lemma provides us with the essential tool to close the cycle.

Lemma 6.1.3. *For all β, $0 < \beta \leq \frac{1}{64k^2}$, there exists a set $\mathcal{C}$ of size at most $2\beta n$ consisting of pairwise disjoint $(k-1)$-tuples, such that for each pair of good $(k-1)$-tuples there exist at least $\frac{\beta n}{4}$ elements in $\mathcal{C}$ that connect this pair.*

Proof. For two good vertex-disjoint tuples $\boldsymbol{x_{k-1}}, \boldsymbol{y_{k-1}}$ in H, let $\mathcal{C}_{\boldsymbol{x}_{k-1},\boldsymbol{y}_{k-1}}$ be the set of all connectors that connect $\boldsymbol{x}_{k-1}$ with $\boldsymbol{y}_{k-1}$ and are vertex-disjoint from $\boldsymbol{x}_{k-1}, \boldsymbol{y}_{k-1}$. Recall that the following conditions hold for $z \in \mathcal{C}_{\boldsymbol{x}_{k-1},\boldsymbol{y}_{k-1}}$:

- $\{x_{k-i}, \ldots, x_1, z_1, \ldots, z_i\} \in E(H)$ for $i \in \{1, \ldots, k-1\}$, and
- $\{z_{i-k+1}, \ldots, z_{k-1}, y_1, \ldots, y_{i-k+1}\} \in E(H)$ for $i \in \{k, \ldots, 2k-2\}$.

From the condition (6.1), the definition of good tuples, and from (6.2), the minimum degree of H, we infer

$$|\mathcal{C}_{\boldsymbol{x}_{k-1},\boldsymbol{y}_{k-1}}| \geq \left(1 - 2\sum_{i=1}^{k-1}\rho^i\right) n^{k-1} - o(n^{k-1}) \geq (1 - 4\rho k)n^{k-1}.$$

Now, take β as asserted by the lemma and let $\mathcal{C}' := \bigcup \mathcal{C}_{\boldsymbol{x}_{k-1},\boldsymbol{y}_{k-1}}$, where the union is over all vertex-disjoint good $(k-1)$-tuples $\boldsymbol{x_{k-1}}$ and $\boldsymbol{y_{k-1}}$. Define $\mathcal{C}$ to be the set obtained by choosing each $\boldsymbol{z_{k-1}} \in \mathcal{C}'$ independently with probability $\frac{\beta}{n^{k-2}}$.

Similarly to (6.4), by Chernoff's inequality:

$$\mathbf{Pr}\left[|\mathcal{C}| - \beta n > \beta n\right] < e^{-\beta n}.$$

With probability at least $\frac{1}{2}$ at most $4k^2\beta^2 n \leq \beta n/4$ of the $(k-1)$-tuples have to be removed from $\mathcal{C}$ to obtain a set of vertex-disjoint tuples, analogously to (6.6).

Analogously to (6.7), for two good vertex-disjoint tuples $\boldsymbol{x_{k-1}}, \boldsymbol{y_{k-1}}$ in H,

$$\mathbf{Pr}\left[|\mathcal{C}_{\boldsymbol{x_{k-1}},\boldsymbol{y_{k-1}}} \cap \mathcal{C}| < \frac{\beta n}{2}\right] < e^{-\frac{1}{16}\gamma n}.$$

Therefore, we deduce with positive probability that after removing from $\mathcal{C}$ all tuples that are not vertex-disjoint, we are left with a set that satisfies the conditions in the lemma. □

The next lemma helps us to connect a linear amount of small paths into a single path avoiding a small forbidden vertex subset.

Lemma 6.1.4. *For any set $\mathcal{X}$ of vertex-disjoint $(2k-2)$-tuples that each induce a good path in H, $|\mathcal{X}| \leq \frac{1}{4k^2}n$, and any forbidden set $F \subset V$ of size at most $\frac{1}{8k}n$, there exists a path P containing all tuples of $\mathcal{X}$, respecting their individual ordering, such that $(V(P) \backslash X) \cap F = \emptyset$.*

Proof. For arbitrary $\boldsymbol{x_{2k-2}}, \boldsymbol{y_{2k-2}} \in \mathcal{X}$, we choose a $\boldsymbol{z_{k-1}} \in V^{k-1}$ uniformly at random and define the events

$$E^1 = \{\boldsymbol{x_{2k-2}} \Diamond \boldsymbol{z_{k-1}} \text{ induces a path, respecting the ordering}\}$$

and

$$E^2 = \{\boldsymbol{z_{k-1}} \Diamond \boldsymbol{y_{2k-2}} \text{ induces a path, respecting the ordering}\}.$$

With E_i^2 being the event that $\{z_i, \ldots, z_{k-1}, y_1, \ldots, y_i\} \in E$, $i \in \{1, \ldots, k-1\}$, we obtain that

$$\mathbf{Pr}\left[E_i^2\right] \geq 1 - \rho^{k-i} - o(1),$$

since $\boldsymbol{y_{2k-2}}$ induces a good path, and the probability that at least two of the k vertices coincide is $o(1)$. Therefore,

$$\mathbf{Pr}\left[E^2\right] \quad \geq \quad 1 \; - \; \sum_{i=1}^{k-1} (1 - \mathbf{Pr}\left[E_i^2\right]) \quad \geq \quad 1 \; - \; \sum_{i=1}^{k-1} \left(\rho^{k-i} + o(1)\right) \quad \geq \quad 1 \; - \; 2\rho.$$

The same holds for E^1. Hence, by the union bound

$$\mathbf{Pr}\left[E^1 \cap E^2\right] \geq 1 - 4\rho.$$

We choose an arbitrary ordering of $\mathcal{X}$. Iteratively, we consider two consecutive elements $\boldsymbol{x_{2k-2}}, \boldsymbol{y_{2k-2}}$ of $\mathcal{X}$. The probability that a u.a.r. chosen $\boldsymbol{z_{k-1}} \in V^{k-1}$ connects $\boldsymbol{x_{2k-2}}$ with $\boldsymbol{y_{2k-2}}$ (meaning that both E^1 and E^2 hold) is at least $1 - 4\rho$ and the probability that it is not vertex-disjoint to an already chosen element (connecting previous pairs of elements of $\mathcal{X}$), to X or to F is at most

$$\left(\frac{k}{4k^2} + \frac{2k}{4k^2} + \frac{1}{8k}\right) k = \frac{7}{8} < 1 - 4\rho$$

by the union bound.

Thus, we choose a $\boldsymbol{z_{k-1}} \in V^{k-1}$ satisfying the conditions in the lemma, and iterate. $\square$

The next lemma helps us find an almost spanning path in the hypergraph H.

Lemma 6.1.5. *For every good path P and every set $F \subset V$ of size at most $k\rho n$, there exists a good path P' that contains P and covers all vertices except those from F and at most $k\rho n$ further vertices.*

Proof. Consider the longest good path P' that contains P and suppose that $|V(P') \cup F| < n - k\rho n$. Then choose one end $\boldsymbol{x}_{k-1}$ of P'. Note that from (6.1), i.e. from the condition $\deg(x_1, \ldots, x_i) \geq \left(1 - \rho^{k-i}\right)\binom{n-i}{k-i}$ for every i, it follows that, for every i, the number of vertices $v \in V(H)$ such that

$$\deg(v, x_1, \ldots, x_i) \geq (1 - \rho^{k-i-1})\binom{n-i-1}{k-i-1} \tag{6.8}$$

is at least $n - \rho n$, implying that $|V(P') \cup F| \geq n - k\rho n$.

Indeed, suppose for contradiction that there exists an i, such that the number of vs satisfying (6.8) is less than $n - \rho n$. Then,

$$\begin{aligned}
\deg(x_1, \ldots, x_i) &= \sum_{v \in V \backslash \{x_1, \ldots, x_i\}} \frac{1}{k-i} \deg(v, x_1, \ldots, x_i) \\
&< \frac{(n - \rho n)}{k-i}\binom{n-i-1}{k-i-1} + \frac{\rho n - i}{k-i}(1 - \rho^{k-i-1})\binom{n-i-1}{k-i-1} \\
&\leq \left(1 - \rho^{k-i}\right)\binom{n-i}{k-i},
\end{aligned}$$

contradicting (6.1). □

6.2 Proofs of Theorem 1.2.12 and Theorem 1.2.13

Proof of Theorem 1.2.12. Suppose $H = (V, E)$ is a k-graph on n vertices, n sufficiently large, with at least

$$\binom{n-1}{k} + \mathrm{ex}\left(n-1, P(k,l)\right)$$

edges and no vertex with a $P(k,l)$-free link. Then the vertex set can be partitioned into two sets $V = V' \cup V$" with $|V'| = n'$ and $V'' = \{v_1, \ldots, v_t\}$ such that

$$\delta_1\left(H'\right) \geq (1-\varepsilon)\binom{n'}{k-1} \tag{6.9}$$

with $H' = H[V']$ and $\varepsilon = \frac{1}{22(1280k^3)^{k-1}}$. To obtain V', we iteratively delete vertices $v_1, \ldots, v_t$ of minimum degree from H till the δ_1-condition (6.9) holds. Counting the non-edges one observes that $t \leq \frac{2}{\varepsilon}$.

The following claim provides an embedding of the vertices of V''.

Claim 8. *There exists a set $\mathcal{S}$ of t paths of type $P_{2k-2}^{(k,k-1)}$, if $t \geq 2$, or of type $P(k,l)$, if $t = 1$, such that v_i is in each edge of the i^{th} element of $\mathcal{S}$, $1 \leq i \leq t$, for some ordering of $\mathcal{S}$.*

We apply Lemma 6.1.2 for $\gamma = \frac{1}{64k^2}$ and Lemma 6.1.3 for $\beta = \frac{1}{1280k^3}$ to H'. As we want disjoint sets A, C, and S, we delete all elements from $\mathcal{A}$ that are not vertex-disjoint to an element from $\mathcal{C} \cup \mathcal{S}$. Thus, we delete at most $2k\beta n' + \frac{4k}{\varepsilon} \leq \frac{\gamma n'}{20}$ absorbers overall, and for every vertex $v \in V'$ there are at least $\frac{\gamma n'}{5}$ elements in the new set $\mathcal{A}$ absorbing v. Similarly, we make C disjoint from S still keeping at least $\frac{\beta n'}{5}$ connectors in $\mathcal{C}$ for each pair of good $(k-1)$-tuples.

Applying Lemma 6.1.4 on the new set $\mathcal{A}$, we obtain a good tight path in H' containing all elements of $\mathcal{A}$ and no vertex from $C \cup S$. We extend this path to one good path with Lemma 6.1.5 such that it covers all but $k\rho n'$ vertices from $V' \backslash (C \cup S)$ and does not contain any vertex from $C \cup S$.

As a next step, we use connectors from $\mathcal{C}$ to connect the elements of $\mathcal{S}$ and the extended path to one cycle. This cycle absorbs the remaining vertices including the unused connectors, since $2\beta n' + k\rho n' < \frac{\gamma n'}{5}$. If $\mathcal{S}$ contains only one element, we obtain a Hamilton cycle that is tight except for the l-tight path from $\mathcal{S}$. Otherwise, we obtain a tight Hamilton cycle. Hence, there exists an l-tight Hamilton cycle. □

Note that we actually prove the bound for Hamilton cycles that are tight except in the link of at most one vertex.

Now deliver the missing proof of the above claim.

Proof of the Claim 8. We consider two cases.

Case 1 ($\deg(v_1) < \frac{\varepsilon}{2}\binom{n-1}{k-1}$)**.** In this case, the number of missing edges yields a sufficient minimum degree in $H - v_1$, hence $t = 1$.

Let a be the number of hyperedges $\{x_1, \ldots, x_{k-1}\}$ in $H(v_1)$ that contain a subset $\{x_1, \ldots, x_j\}$, $j \in \{1, \ldots, k-1\}$, satisfying $\deg_{H'}(x_1, \ldots, x_j) < (1 - \rho^{k-j})\binom{n-1-j}{k-j}$, i.e. if there is a tuple $(x_1, \ldots, x_{k-1})$ obtained by an ordering of the edge that is not good in H'. We denote the number of such j-sets by b_j and observe that each of them lies in at most $\binom{n-1-j}{k-1-j}$ edges in $H(v_1)$. Hence, we obtain

$$a \leq \sum_{j=1}^{k-1} \binom{n-1-j}{k-1-j} b_j.$$

We further call those a edges *bad*.

The second time we apply double counting, we set c to be the number of non-edges in H'. By definition, each of the b_j j-sets lies in at least $\rho^{k-j}\binom{n-1-j}{k-j}$ non-edges of H'. Note that each non-edge of H' has exactly 2^k subsets. Henceforth,

$$\sum_{j=1}^{k-1} \rho^{k-j}\binom{n-1-j}{k-j} b_j \leq 2^k c.$$

Combining the two bounds, there are at most

$$\begin{aligned} a &\leq \sum_{j=1}^{k-1}\binom{n-1-j}{k-1-j} b_j = \sum_{j=1}^{k-1} \rho^{k-j}\binom{n-1-j}{k-j} b_j \frac{k-j}{n-k}\rho^{j-k} \\ &\leq \frac{k-1}{n-k}\rho^{1-k}\sum_{j=1}^{k-1}\rho^{k-j}\binom{n-1-j}{k-j} b_j \\ &\leq \frac{k-1}{n-k}\rho^{1-k}2^k c \leq c \end{aligned}$$

edges in $H(v_1)$, which have an ordering producing a tuple that is not good in H'. Observe that equality can only be obtained with $c = 0$.

For $c = 0$, there exists a $P(k,l)$ in the link of v_1 by assumption on H, and this path is good, hence, we are done. For $c > 0$, we obtain $\deg(v_1) > c + \mathrm{ex}(n-1, P(k,l))$. We disregard bad hyperedges in $H(v_1)$ and using $a < c$, we still find a $P(k,l)$ in the link of v_1. The obtained path is good, proving the claim.

Case 2 ($\deg(v_1) \geq \frac{\varepsilon}{2}\binom{n-1}{k-1}$)**.** In this case, we have for all $1 \leq i \leq t$, $\deg(v_i) \geq \frac{1}{3}\binom{n-1}{k-1}$ holds because the v_is are chosen greedily with ascending degree. In this case, we actually show that each of the v_is can be matched to a good tight path such that the assigned paths are pairwise vertex-disjoint. Since the proportion of k-sets that are edges in H' is $1 - o(1)$, we know that there are $o(1)\binom{n'}{k-1}$ tuples that are not good in H'. By the result from [46] mentioned in the introduction it holds that

$$\mathrm{ex}\left(n', P_{2k-2}^{(k-1,k-2)}\right) \leq (k-1)\binom{n'}{k-2} = o(1)\binom{n'}{k-1}.$$

There are at most $O(1)\binom{n-2}{k-2}$ edges including at least two vertices from V''. We assign iteratively vertex-disjoint good $(2k-1)$-paths to each of the v_i, $1 \leq i \leq t$, such that v_i is in each of its edges. This is possible, since we disregard at most $o(1)\binom{n'}{k-1}$ many edges in the link of each v_i that contain a tuple that is not good or a vertex contained in a previously assigned path or another vertex from V''.

□

Proof of Theorem 1.2.13. The proof of Theorem 1.2.13 follows the same pattern as the proof of Theorem 1.2.12 without making use of Claim 8. Therefore, we only give a brief sketch of it.

Suppose H is a k-graph on n vertices, n sufficiently large, with $\delta_1 \geq (1-\varepsilon)\binom{n-1}{k-1}$ and $\varepsilon = \frac{1}{22(1280k^3)^{k-1}}$. Similarly to the proof of Theorem 1.2.12, we apply Lemmas 6.1.2 and 6.1.3 and obtain via deletion of elements of $\mathcal{A}$ two vertex-disjoint sets A and C such that $\mathcal{A}$ and $\mathcal{C}$ have the desired properties. Using Lemma 6.1.4 we find a good path containing all elements of $\mathcal{A}$ such that $\mathcal{C}$ is vertex-disjoint from it. We extend this path with Lemma 6.1.5 such that it contains all but at most $\rho k n$ vertices from $V \backslash C$ and no vertex from C. Using a connector, we connect the ends of this path, obtaining a cycle. As $2\beta n + k\rho n < \frac{\gamma n}{5}$ holds, we absorb the remaining vertices and obtain a tight Hamilton cycle. □

6.3 Concluding Remarks

The edge-density of extremal non-hamiltonian hypergraphs is $1 - o(1)$ (unlike the density of F-extremal graphs for fixed k-graphs F), since a Hamilton cycle is a spanning substructure. In [40], we conjectured that an extremal graph of any bounded spanning structure consists of an $(n-1)$-clique and a further extremal graph.

Conjecture 6.3.1. For any $k \in \mathbb{N}$ there exists an n_0 such that for every k-graph H on $n \geq n_0$ vertices without a spanning subgraph isomorphic to a forbidden hypergraph F of bounded maximum vertex degree,

$$|e(H)| \leq \binom{n-1}{k} + \operatorname{ex}\left(n-1, \{F(v) : v \in V\}\right)$$

holds, and the bound is tight.

However, recently Alon and Yuster [4] showed that the statement is wrong in this generality, but true if we restrict it to 2-graphs.

A 2-graph is called *pancyclic*, if for any c with $3 \leq c \leq n$ it contains a c-cycle. Similarly to the spanning structure of hamiltonian l-tight cycles, Katona and Kierstead [60] defined l-tight cycles of any length. This allows us to generalize the concept of pancyclicity by calling a k-graph l-*pancyclic*, if for any c with $3 \leq c \leq n/(k-l)$ it contains an l-tight cycle on c edges. In his famous metaconjecture [17], Bondy claimed for 2-graphs that almost any non-trivial condition on a graph which implies that the graph is hamiltonian also implies that the graph is pancyclic. (There may be a simple family of exceptional graphs.)

It is not hard to see that both the condition in Theorem 1.2.12 and the condition in Theorem 1.2.13 imply not only Hamiltonicity but also pancyclicity.

Bibliography

[1] M. Ajtai, J. Komlós, and E. Szemerédi, *The longest path in a random graph*, Combinatorica **1** (1981), 1–12. (Cited on pages 8 and 14.)

[2] N. Alon, M. Krivelevich, and B. Sudakov, *Embedding nearly-spanning bounded degree trees*, Combinatorica **27** (2007), 629–644. (Cited on page 13.)

[3] N. Alon and J. H. Spencer, *The probabilistic method*, 3rd ed., Interscience Series in Discrete Mathematics and Optimization, Wiley, 2008. (Cited on page 18.)

[4] N. Alon and R. Yuster, *The Turán number of sparse spanning graphs*, preprint. (Cited on page 92.)

[5] B. Alspach, *The wonderful walecki construction*, Bulletin of the Institute of Combinatorics and its Applications **52** (2008), 7–20. (Cited on page 11.)

[6] B. Alspach, J.-C. Bermond, and D. Sotteau, *Decompositions into cycles I: Hamilton decompositions*, Proceedings of 1987 Cycles and Rays Colloquium, Kluwer Academic Publishers, 1990, pp. 9–18. (Cited on page 11.)

[7] J. Balogh, B. Csaba, M. Pei, and W. Samotij, *Large bounded degree trees in expanding graphs*, Electronic Journal of Combinatorics **17** (2010), R6. (Cited on pages 13 and 68.)

[8] J. Beck, *Remarks on positional games. I*, Acta Mathematica Hungarica **40** (1982), 65–71. (Cited on pages 47 and 48.)

[9] S. Ben-Shimon, A. Ferber, D. Hefetz, and M. Krivelevich, *Hitting time results for Maker-Breaker games*, Random Structures and Algorithms **41** (2012), 23–46. (Cited on pages 9 and 12.)

[10] ———, *Hitting time results for maker-breaker games*, Random Structures and Algorithms **41** (2012), 23–46. (Cited on pages 13 and 49.)

[11] S. Ben-Shimon, M. Krivelevich, and B. Sudakov, *On the resilience of Hamiltonicity and optimal packing of Hamilton cycles in random graphs*, SIAM Journal on Discrete Mathematics **25** (2011), 1176–1193. (Cited on pages 8, 9, 11, 26, 30, and 64.)

[12] J.C. Bermond, A. Germa, M.C. Heydemann, and D. Sotteau, *Hypergraphes hamiltoniens*, Problèmes combinatoires et théorie des graphes, Orsay 1976, Colloq. int. CNRS **260** (1978), 39–43. (Cited on page 15.)

[13] B. Bollobás, *The evolution of sparse graphs*, Graph Theory and Combinatorics, Academic Press, 1984, pp. 35–57. (Cited on pages 8, 11, and 14.)

[14] ———, *Random graphs*, 2nd ed., Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2001. (Cited on pages 7, 14, 24, 29, 30, and 64.)

[15] B. Bollobás, T. I. Fenner, and A. M. Frieze, *An algorithm for finding hamilton paths and cycles in random graphs*, Combinatorica **7** (1987), 327–341. (Cited on pages 27 and 37.)

[16] B. Bollobás and A. M. Frieze, *On matchings and Hamiltonian cycles in random graphs*, Annals of Discrete Mathematics **28** (1985), 23–46. (Cited on page 11.)

[17] J.A. Bondy, *Pancyclic graphs: recent results*, Infinite and finite sets, Bolyai Society Mathematical Studies, vol. 10, Springer, 1975, pp. 181–188. (Cited on page 92.)

[18] V. Chvátal and P. Erdős, *Biased positional games*, Annals of Discrete Mathematics **99** (1978), 221–228. (Cited on pages 12, 48, and 57.)

[19] D. Clemens, A. Ferber, M. Krivelevich, and A. Liebenau, *Fast strategies in Maker-Breaker games played on random boards*, preprint. (Not cited.)

[20] C. Cooper and A. Frieze, *Almost all regular graphs are Hamiltonian*, Journal of Graph Theory **13** (1989), 719–735. (Cited on pages 9, 10, and 30.)

[21] R. Diestel, *Graph theory*, 4th ed., Graduate Texts in Mathematics, Springer, 2010. (Cited on pages 63 and 76.)

[22] G.A. Dirac, *Some theorems on abstract graphs*, Proceedings of the London Mathematical Society, Series 3 **2** (1952), 69–81. (Cited on pages 9 and 17.)

[23] A. Dudek and A. Frieze, *Tight Hamilton cycles in random uniform hypergraphs*, preprint. (Not cited.)

[24] ______, *Loose Hamilton cycles in random k-uniform hypergraphs*, Electronic Journal of Combinatorics **18** (2011), P48. (Cited on page 9.)

[25] A. Dudek, A. Frieze, P. Loh, and S. Speiss, *Optimal divisibility conditions for loose Hamilton cycles in random hypergraphs*, preprint. (Cited on page 9.)

[26] G. Egorychev, *The solution of the Van der Waerden problem for permanents*, Doklady Akademii Nauk SSSR **258** (1981), 1041–1044. (Cited on page 22.)

[27] P. Erdős and H. Hanani, *On a limit theorem in combinatorial analysis*, Publicationes Mathematicae Debrece **10** (1963), 10–13. (Cited on page 10.)

[28] P. Erdős and A. Rényi, *On random graphs. I*, Publicationes Mathematicae Debrecen **6** (1959), 290–297. (Cited on page 7.)

[29] ______, *On the evolution of random graphs*, Publication of the Mathematical Institute of the Hungarian Academy of Sciences **5** (1960), 17–61. (Cited on page 7.)

[30] D. Falikman, *A proof of the Van der Waerden problem for permanents of a doubly stochastic matrix*, Matematicheskie Zametki **29** (1981), 931–938. (Cited on page 22.)

[31] A. Ferber, R. Glebov, M. Krivelevich, and A. Naor, *Biased games on random boards*, in preparation. (Cited on pages 8, 9, and 47.)

[32] A. Ferber and D. Hefetz, *Weak and strong k-connectivity game*, preprint. (Cited on page 12.)

[33] W. Fernandez de la Vega, *Long paths in random graphs*, Studia Scientiarum Mathematicarum Hungarica **14** (1979), 335–340. (Cited on page 13.)

[34] ______, *Trees in sparse random graphs*, Journal of Combinatorial Theory, Series B **45** (1988), 77–85. (Cited on page 13.)

[35] J. Friedman and N. Pippenger, *Expanding graphs contain all small trees*, Combinatorica **7** (1987), 71–76. (Cited on pages 13 and 68.)

[36] A. Frieze, *Loose Hamilton cycles in random 3-uniform hypergraphs*, Electronic Journal of Combinatorics **17** (2010), N28. (Cited on page 9.)

[37] A. Frieze and M. Krivelevich, *Hamilton cycles in random subgraphs of pseudo-random graphs*, Discrete Mathematics **256** (2002), 137–150. (Cited on pages 20, 27, and 37.)

[38] ———, *On packing Hamilton cycles in ε-regular graphs*, Journal of Combinatorial Theory, Series B **94** (2005), 159–172. (Cited on pages 9, 11, and 46.)

[39] ———, *On two Hamilton cycle problems in random graphs*, Israel Journal of Mathematics **166** (2008), 221–234. (Cited on pages 8, 9, 11, and 48.)

[40] ———, *On extremal hypergraphs for Hamiltonian cycles*, European Journal of Combinatorics **33** (2012), 544–555. (Cited on pages 9 and 92.)

[41] Z. Füredi, *Matchings and covers in hypergraphs*, Graphs and Combinatorics **4** (1988), 115–206. (Cited on page 10.)

[42] H. Gebauer and T. Szabó, *Asymptotic random graph intuition for the biased connectivity game*, Random Structures and Algorithms **35** (2009), 431–443. (Cited on pages 12 and 55.)

[43] R. Glebov, D. Johannsen, and M. Krivelevich, *Hitting time appearence of certain spanning trees in the random graph process*, in preparation. (Cited on pages 8 and 63.)

[44] R. Glebov and M. Krivelevich, *On the number of Hamilton cycles in sparse random graphs*, to appear. (Cited on pages 8, 19, and 66.)

[45] R. Glebov, M. Krivelevich, and T. Szabó, *On covering expander graphs by Hamilton cycles*, to appear. (Cited on pages 8, 9, 27, and 33.)

[46] E. Györi, G. Y. Katona, and N. Lemons, *Hypergraph extensions of the Erdős-Gallai theorem*, Discrete Mathematics **36** (2010), 655–662. (Cited on pages 17 and 91.)

[47] H. Hàn, Y. Person, and M. Schacht, *On perfect matchings in uniform hypergraphs with large minimum vertex degree*, SIAM Journal on Discrete Mathematics **23** (2009), 732–748. (Cited on page 18.)

[48] H. Hàn and M. Schacht, *Dirac-type results for loose Hamilton cycles in uniform hypergraphs*, Journal of Combinatorial Theory, Series B **100** (2010), 332–346. (Cited on page 17.)

[49] P. E. Haxell, *Tree embeddings*, Journal of Graph Theory **36** (2001), 121–130. (Cited on pages 13 and 68.)

[50] D. Hefetz, M. Krivelevich, M. Stojaković, and T. Szabó, *Fast winning strategies in Maker-Breaker games*, Journal of Combinatorial Theory, Series B **99** (2009), 39–47. (Cited on page 12.)

[51] ———, *A sharp threshold for the Hamilton cycle Maker-Breaker game*, Random Structures and Algorithms **34** (2009), 112–122. (Cited on page 9.)

[52] D. Hefetz, M. Krivelevich, and T. Szabó, *Hamilton cycles in highly connected and expanding graphs*, Combinatorica **29** (2009), 547–568. (Cited on pages 27, 30, 37, 39, 42, and 69.)

[53] ———, *Sharp threshold for the appearance of certain spanning trees in random graphs*, RSA **41** (2012), 391–412. (Cited on pages 14 and 15.)

[54] D. Hefetz, D. Kühn, J. Lapinskas, and D. Osthus, *Optimal covers with Hamilton cycles in random graphs*, preprint. (Cited on pages 9 and 46.)

[55] D. Hefetz and S. Stich, *On two problems regarding the Hamilton cycle game*, Electronic Journal of Combinatorics **16** (2009), R28. (Cited on page 12.)

[56] S. Janson, *Hamilton cycles and perfect matchings in a random graph*, Combinatorics, Probability and Computing **3** (1994), 97–126. (Cited on pages 9, 10, and 30.)

[57] ———, *Random regular graphs: asymptotic distributions and contiguity*, Combinatorics, Probability and Computing **4** (1995), 369–405. (Cited on page 10.)

[58] S. Janson, T. Łuczak, and A. Ruciński, *Random graphs*, 1st ed., Interscience Series in Discrete Mathematics and Optimization, Wiley, 2000. (Cited on page 12.)

[59] D. Johannsen, M. Krivelevich, and W. Samotij, *Expanders are universal for the class of all spanning trees*, to appear. (Cited on pages 13, 14, and 79.)

[60] G.Y. Katona and H.A. Kierstead, *Hamiltonian chains in hypergraphs*, Journal of Graph Theory **30** (1999), 205–212. (Cited on pages 8, 15, 16, 17, and 92.)

[61] F. Knox, D. Kühn, and D. Osthus, *Edge-disjoint Hamilton cycles in random graphs*, preprint. (Cited on pages 9, 11, 20, and 46.)

[62] ———, *Approximate Hamilton decompositions of random graphs*, Random Structures and Algorithms **40** (2012), 133–149. (Cited on pages 9, 11, 20, 25, 34, 45, and 46.)

[63] J. Komlós and E. Szemerédi, *Hamilton cycles in random graphs*, Infinite and finite sets, Bolyai Society Mathematical Studies, vol. 10, Springer, 1975, pp. 1003–1011. (Cited on page 7.)

[64] ———, *Limit distribution for the existence of hamiltonian cycles in a random graph*, Discrete Mathematics **43** (1983), 55–63. (Cited on pages 7, 8, 11, 14, 27, and 37.)

[65] A. Korshunov, *A solution of a problem of Erdős and Rényi on Hamilton cycles in non-oriented graphs*, Soviet Mathematics. Doklady **17** (1976), 760–764. (Cited on pages 7, 8, and 14.)

[66] M. Krivelevich, *Embedding spanning trees in random graphs*, SIAM Journal on Discrete Mathematics **24** (2010), 1495–1500. (Cited on pages 13, 14, and 79.)

[67] ———, *The critical bias for the Hamiltonicity game is* $(1+o(1))n/\ln n$, Journal of Journal of the American Mathematical Society **24** (2011), 125–131. (Cited on pages 12 and 55.)

[68] ———, *On the number of hamilton cycles in pseudo-random graphs*, Electronic Journal of Combinatorics **19** (2012), P25. (Cited on pages 10, 28, 30, and 46.)

[69] M. Krivelevich and W. Samotij, *Optimal packings of Hamilton cycles in sparse random graphs*, SIAM Journal on Discrete Mathematics **26** (2012), 964–982. (Cited on pages 9, 11, 20, and 46.)

[70] M. Krivelevich and B. Sudakov, *Pseudo-random graphs*, More sets, graphs and numbers, Bolyai Society Mathematical Studies, vol. 15, Springer, 2006, pp. 199–262. (Cited on page 46.)

[71] D. Kühn, R. Mycroft, and D. Osthus, *Hamilton ℓ-cycles in uniform hypergraphs*, Journal of Combinatorial Theory, Series A **117** (2010), 910–927. (Cited on page 17.)

[72] D. Kühn, D. Osthus, and A. Treglown, *Matchings in 3-uniform hypergraphs*, preprint. (Cited on page 18.)

[73] C. Lee and B. Sudakov, *Dirac's theorem for random graphs*, Random Structures and Algorithms **41** (2012), 293–305. (Cited on pages 8 and 9.)

[74] A. Lehman, *A solution to the shannon switching game*, Journal of the Society for Industrial and Applied Mathematics **12** (1964), 687–725. (Cited on page 12.)

[75] K. Markstrom and A. Ruciński, *Perfect matchings and Hamilton cycles in hypergraphs with large degrees*, European Journal of Combinatorics **32** (2011), 677–687. (Cited on page 17.)

[76] Ø. Ore, *Arc coverings of graphs*, Annali di Matematica Pura ed Applicata, IV. Serie **55** (1961), 315–321. (Cited on pages 9 and 15.)

[77] ———, *Theory of graphs*, American Mathematics Society College Publishers, 1962. (Cited on page 21.)

[78] L. Pósa, *Hamiltonian circuits in random graphs*, Discrete Mathematics **14** (1976), 359–364. (Cited on pages 7, 27, 33, and 37.)

[79] R. Robinson and N. Wormald, *Almost all cubic graphs are Hamiltonian*, Random Structures and Algorithms **3** (1992), 117–125. (Cited on page 10.)

[80] ———, *Almost all regular graphs are Hamiltonian*, Random Structures and Algorithms **5** (1994), 363–374. (Cited on page 10.)

[81] V. Rödl, *On a packing and covering problem*, European Journal of Combinatorics **6** (1985), 69–78. (Cited on page 10.)

[82] V. Rödl and A. Ruciński, *Dirac-type questions for hypergraphs - a survey (or more problems for Endre to solve)*, An Irregular Mind, Bolyai Society Mathematical Studies, vol. 21, Springer, 2010, pp. 561–590. (Cited on pages 17 and 18.)

[83] V. Rödl, A. Ruciński, and E. Szemerédi, *A Dirac-type theorem for 3-uniform hypergraphs*, Combinatorics, Probability and Computing **15** (2006), 229–251. (Cited on page 17.)

[84] ———, *Perfect matchings in uniform hypergraphs with large minimum degree*, European Journal of Combinatorics **27** (2006), 1333–1349. (Cited on page 18.)

[85] ———, *An approximate Dirac-type theorem for k-uniform hypergraphs*, Combinatorica **28** (2008), 229–260. (Cited on page 17.)

[86] V. Rödl, A. Ruciński, and E. Szemerédi, *Dirac-type conditions for Hamiltonian paths and cycles in 3-uniform hypergraphs*, Advances in Mathematics **227** (2011), 1225–1299. (Cited on page 17.)

[87] M. Stojaković and T. Szabó, *Positional games on random graphs*, Random Structures and Algorithms **26** (2005), 204–223. (Cited on pages 9 and 12.)

[88] B. Sudakov and V.H. Vu, *The local resilience in random graphs*, Random Structures and Algorithms **33** (2008), 409–433. (Cited on pages 8 and 9.)

[89] M. Stojaković T. Müller, *A threshold for Maker-Breaker Clique game*, preprint. (Cited on page 12.)

[90] Z. Tuza, *Steiner system and large non-Hamiltonian hypergraphs*, Matematiche **61** (2006), 179–183. (Cited on page 16.)

[91] D. B. West, *Introduction to graph theory*, APrentice Hall, 2001. (Cited on page 47.)

[92] D. Woitas, *Permutationsgruppen als Codes - das Bailey-Verfahren*, Diploma Thesis, Freie Universität Berlin, 2010. (Cited on pages 15 and 16.)