Graph bootstrap percolation and additive combinatorial constructions

Dissertation

zur Erlangung des Grades eines Doktors der Naturwissenschaften
(Dr. rer. nat.)

am Fachbereich Mathematik und Informatik

der Freien Universität Berlin

vorgelegt von

David Fabian

Berlin, 2023

Erstgutachter: Prof. Tibor Szabó, PhD

Zweitgutachter: Prof. Dr. Yury Person

Tag der Disputation: 31.10.2023

FREIE UNIVERSITÄT BERLIN

Abstract

Graph bootstrap percolation and additive combinatorial constructions

by David Fabian

Given a (usually small) graph H and an n-vertex graph G the *H-bootstrap process* on G is defined to be the sequence of graphs G_t, $t \geq 0$ which starts with $G_0 := G$ and in which G_{t+1} is obtained from G_t by adding every edge that completes a copy of H. This process eventually stabilises. How many steps it takes before the process stabilises depends on H and G. We investigate the *maximum running time $M_H(n)$*, which is the largest number of steps an H-bootstrap process on an n-vertex graph can take before it has stabilised, for several choices of H and initiate the study of which graph parameters determine the asymptotic growth of $M_H(n)$ as a function of n. The sublinear range is characterised by the question "Does $M_H(n) = o(n)$ hold?". We will see that this range encompasses graphs such as trees and cycles, and we will provide sufficient conditions for answering the question above in the negative. On the other hand the superlinear range is given by the question whether $M_H(n) = \omega(n)$. In this range we will encounter graphs of high connectivity or high density, but also sparse graphs such as when H is distributed as the Erdős-Rényi random graph for certain edge probabilities. Within the superlinear range we put particular emphasis on graphs H with $M_H(n) = \Theta(n^2)$. To provide such quadratic bounds we will generalise a construction introduced by Balogh, Kronenberg, Pokrovskiy, and Szabó to study the maximum running time of complete graphs.

Extremal constructions from additive combinatorics such as Sidon sets or 3-AP-free sets turned out to provide some of the best-known lower bounds on running times in graph bootstrap percolation. In the second part of the thesis we focus on an extremal additive problem. We study α-strong B_h-sets in the integers, a generalisation of Sidon sets, where $h \geq 2$ is an integer and $0 < \alpha < 1$ is a real parameter. In an α-strong B_h-set sums of the form $x_1 + \ldots + x_h$, where $x_1, \ldots, x_h \in A$ and $x_1 \geq \ldots \geq x_h$, have pairwise distances of at least x_1^{α}. By elaborating on a construction of Cilleruelo we give an infinite α-strong B_h-set S with counting function $S(n)$ that provides the first improvement over the greedy construction for $\alpha > 5.76 \cdot 10^{-5}$. Building on work of Kohayakawa, Lee, Moreira, and Rödl, we then use that construction to prove the existence of B_h-sets of a certain density in sparse random sets of integers.

Finally, we consider the problem of splitting matchings, that is, given a k-regular n-vertex graph G whose edge set is the union of perfect matchings $M_1, \ldots, M_k$, we want to determine the tuples $(a_1, \ldots, a_k)$ of positive integers for which there exists a matching M in G satisfying $|M \cap M_i| = a_i$ for all $i \in [k]$. We are particularly interested in those tuples $(a_1, \ldots, a_k)$ for which a suitable M exists no matter what the initial matchings $M_1, \ldots, M_k$ are. This question was introduced by Arman, Rödl, and Sales. Two special cases are *fair splits* and *perfect splits*. In the former case one has $a_1 = \ldots = a_k$, while in the latter $a_1 + \ldots + a_k = n$. We will give necessary conditions on the existence of perfect splits as well as fair splits, and show that in the case $k = 3$ we can realise every triple (a_1, a_2, a_3) with $a_1 + a_2 + a_3 \leq n - 2$ for every choice of M_1, M_2, and M_3.

FREIE UNIVERSITÄT BERLIN

Zusammenfassung

Graph bootstrap percolation and additive combinatorial constructions

von David Fabian

Zu einem (üblicherweise kleinen) Graphen H und einem Graphen G mit n Knoten definieren wir den *H-Bootstrap-Prozess* auf G als die Folge $(G_t)_{t\geq 0}$, die mit $G_0 := G$ beginnt und in der G_{t+1} aus G_t hervorgeht, indem wir jede Kante, die eine Kopie von H vervollständigt, hinzufügen. Dieser Prozess stabilisiert sich, d.h., nach einer gewissen Anzahl an Schritten werden keine weiteren Kanten mehr hinzugefügt. Wir untersuchen für verschiedene H die *maximale Laufzeit* $M_H(n)$, welche die größtmögliche Anzahl an Schritten ist, die ein H-Bootstrap-Prozess auf einem Graphen mit n Knoten benötigt, bevor er stabil wird, und begründen das Studium der Graphenparameter, welche das asymptotische Wachstum von $M_H(n)$ als Funktion von n bestimmen. Der sublineare Bereich ist durch die Frage "Gilt $M_H(n) = o(n)$?" charakterisiert und wir zeigen, dass dieser Graphen wie etwa Bäume und Kreise umfasst. Ferner präsentieren wir hinreichende Bedingungen, die eine negative Antwort auf obige Frage garantieren. Weiterhin betrachten wir den superlinearen Bereich, welcher durch die Frage, ob $M_H(n) = \omega(n)$ vorliegt, bestimmt ist. Hier begegnen wir Graphen mit hoher Zusammenhangszahl, aber auch dünnen Graphen wie etwa den Erdős-Rényi-Zufallsgraphen für kleine Kantenwahrscheinlichkeiten. Eine für uns besondere Rolle im superlinearen Bereich nehmen Graphen H mit $M_H(n) = \Theta(n^2)$ ein. Um solche quadratischen Schranken zu erhalten, verallgemeinern wir zwei Konstruktionen, die von Balogh, Kronenberg, Pokrovskiy und Szabó zur Untersuchung von $M_{K_r}(n)$ für $r \geq 5$ eingeführt wurden.

Konstruktionen aus der additiven Kombinatorik wie maximale Sidonmengen oder Mengen ohne arithmetische Folgen der Länge drei liefern mehrere der derzeit besten unteren Schranken für die maximale Laufzeit bestimmter Bootstrap-Prozesse. Im zweiten Teil dieser Arbeit untersuchen wir *α-starke B_h-Mengen* in $\mathbb{Z}$, wobei $h \geq 2$ eine ganze Zahl und $0 \leq \alpha < 1$ ein reeller Parameter ist. In solchen Mengen haben Summen der Form $x_1 + \ldots + x_h$, wobei $x_1 \geq \ldots \geq x_h$, paarweise Abstände von mindestens x_1^α. Unter Nutzung einer Konstruktion von Cilleruelo beschreiben wir eine unendliche, α-starke B_h-Menge mit einer Zählfunktion, deren Wachstum für $\alpha > 5.76 \cdot 10^{-5}$ die erste bekannte Verbesserung gegenüber der gierigen Konstruktion ist. Anschließend nutzen wir aufbauend auf Arbeiten von Kohayakawa, Lee, Moreira und Rödl diese Konstruktion, um die Existenz von B_h-Mengen gewisser Dichten in dünnen zufälligen Teilmengen von $\mathbb{Z}$ nachzuweisen.

Schließlich betrachten wir das Problem des Teilens von Matchings. Zu einem Graphen G mit $2n$ Knoten, dessen Kantenmenge die Vereinigung k perfekter Matchings $M_1, \ldots, M_k$ ist, suchen wir jene Tupel $(a_1, \ldots, a_k) \in \mathbb{N}^k$, für die ein Matching M in G mit $|M \cap M_i| = a_i$ für alle $i \in [k]$ existiert. Wir sind insbesondere an solchen $(a_1, \ldots, a_k)$ interessiert, für die es ein geeignetes M unabhängig von den gegebenen $M_1, \ldots, M_k$ gibt. Diese Fragestellung geht zurück auf Arman, Rödl und Sales. Zwei Spezialfälle sind *faire Aufteilungen* und *perfekte Aufteilungen*. Bei Ersteren gilt $a_1 = \ldots = a_k$ während Letztere $a_1 + \ldots + a_k = n$ erfüllen. Wir präsentieren eine notwendige Bedingung für die Existenz solcher Aufteilungen, und zeigen, dass im Fall $k = 3$ jedes Tripel (a_1, a_2, a_3) mit $a_1 + a_2 + a_3 \leq n - 2$ für alle M_1, M_2 und M_3 realisierbar ist.

Acknowledgements

My first thanks go to my advisor Tibor Szabó for his support of my PhD project and particularly for organising our annual group workshops in Wilhelmsaue. These workshops have always been a great source of interesting mathematics.

I am particularly grateful to my coauthor Patrick Morris for the productive discussions and long lists of remarks during the writing of our articles on graph bootstrap percolation.

Further, I want to thank my other coauthors Christoph Spiegel, Juanjo Rué, Michael Anastos, and Alp Müyesser for the fruitful collaborations.

Finally I want to express my gratitude for my other (former) colleagues from my working group in Berlin, that is, Ralph-Hardo Schulz, Shagnik Das, Tamás Mészáros, Anurag Bishnoi, Leticia Mattos, Olaf Parczyk, Simona Boyadzhiyska, and Ander Lamaison for providing an inspiring academic environment.

Contents

List of Figures

Notation

In this section we introduce most of the standard concepts and notations used throughout the text. More specialised notation that is unique to an individual section will be introduced at the beginning of the respective section.

Natural numbers and integers. In this text a natural number is the same as a positive integer, that is we do not consider zero a natural number. We denote the set of natural numbers by $\mathbb{N}$ and write $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ for the set of non-negative integers.

Intervals in the integers. All intervals we are going to encounter in this thesis are discrete. Therefore there should not be any confusion with the usual notation for intervals in the real numbers. Given $a, b \in \mathbb{R}$, we write

$$[a,b] := \{x \in \mathbb{Z} : a \leq x \leq b\} \qquad (a,b] := \{x \in \mathbb{Z} : a < x \leq b\}$$
$$[a,b) := \{x \in \mathbb{Z} : a \leq x < b\} \qquad (a,b) := \{x \in \mathbb{Z} : a < x < b\}$$

Note that we do not require a and b to be integers. We are sometimes only interested in the odd integers from a given interval. In that case we write

$$[a,b]_1 := \{x \in [a,b] : x \equiv 1 \mod 2\}.$$

We also use the shorthand notation $[n] := [1,n]$ and $[n]_1 := [1,n]_1$ for $n \in \mathbb{N}$.

Floors and ceilings. We make frequent use of floors and ceilings. Here we recall a few rules that facilitate calculations involving floors and ceilings. Given real numbers x and y, and an integer n,

$$\lfloor x+n \rfloor = \lfloor x \rfloor + n \qquad \text{and} \qquad \lceil x+n \rceil = \lceil x \rceil + n.$$

If $x+y$ is an integer then

$$x+y = \lfloor x \rfloor + \lceil y \rceil.$$

In particular, if $k \in \mathbb{Z}$ then

$$k - \lfloor x \rfloor = \lceil k-x \rceil \quad \text{and} \quad k - \lceil x \rceil = \lfloor k-x \rfloor.$$

Note that $\lfloor x \rfloor + 1$ is the smallest integer strictly greater than x, and $\lceil x \rceil - 1$ is the largest integer strictly smaller than x.

Cyclic groups. Given a positive integer n we write $\mathbb{Z}_n$ for the cyclic group $\mathbb{Z}/n\mathbb{Z}$ with n elements. If n is clear from context we set

$$\mathbf{0} := 0 + n\mathbb{Z} \quad , \quad \mathbf{1} := 1 + n\mathbb{Z} \quad , \quad \mathbf{2} := 2 + n\mathbb{Z}.$$

If $x \in \mathbb{Z}_n$ and $k \in \mathbb{Z}$ then by $x+k$ we simply mean the element $(x+k)+n\mathbb{Z}$. For example we have that $\mathbf{1}+1 = \mathbf{2}$ and $x+2 = x+\mathbf{2}$ for all $x \in \mathbb{Z}_n$.

Sum sets and difference sets. Given two subsets $A, B \subseteq G$ of an abelian group G with group operation $+$ we define the *sum set* (also referred to as the *Minkowski sum*) $A+B$ as

$$A+B := \{a+b : a \in A\,,\, b \in B\}$$

Similarly we define the *difference set* $A-B$ as

$$A-B := \{a-b : a \in A\,,\, b \in B\}$$

If h is an integer we define the *dilate* $h \cdot A$ via

$$h \cdot A := \{h \cdot a : a \in A\}$$

and, if h is non-negative, the *h-fold sum set* hA by

$$0A := \{0\} \quad , \quad hA := (h-1)A + A,$$

or equivalently

$$hA = \{a_1 + \ldots + a_h : a_1, \ldots, a_h \in A\}.$$

Observe that while $h \cdot A \subseteq hA$ the converse does not hold in general.

Frobenius numbers. The *Frobenius number* $F(x,y)$ of two positive, coprime integers x, y is the largest natural number that cannot be expressed as an integral linear combination of x and y with non-negative coefficients, i.e.

$$F(x,y) := \max\left(\mathbb{Z} \setminus \{\alpha x + \beta y : \alpha, \beta \in \mathbb{N}_0\}\right).$$

A thorough treatise of Frobenius numbers and their generalisations by Ramírez Alfonsín can be found in [5]. The precise formula $F(x,y) = xy - x - y$ is well-known. We are interested in $F(k-2,k)$ for odd integers $k \geq 3$, in which case the formula above gives

$$F(k-2,k) = k^2 - 4k + 2. \tag{1}$$

If k is even we set $F'(k-2,k)$ to be the largest multiple of $\gcd(k-2,k)=2$ that cannot be written as an integral linear combination of $k-2$ and k with non-negative coefficients, i.e.

$$F'(k-2,k) := 2 \cdot F\left(\frac{k-2}{2}, \frac{k}{2}\right) = \frac{k^2}{2} - 3k + 2. \tag{2}$$

Landau notation. We use the common asymptotic notation. For any two real-valued, non-negative functions $f,g : \mathbb{N} \to \mathbb{R}$ we write

- $f = O(g)$ if there exists a positive constant $C > 0$ such that $f(n) \le C \cdot g(n)$ for all $n \in \mathbb{N}$, or equivalently, if $\limsup_{n\to\infty} \frac{f(n)}{g(n)} < \infty$
- $f = o(g)$ if for every $c > 0$ there exists $n_0 \in \mathbb{N}$ such that $f(n) < c \cdot g(n)$ for all $n \ge n_0$, or equivalently, if $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 0$.
- $f = \Omega(g)$ if there exists a positive constant $c > 0$ such that $f(n) \ge c \cdot g(n)$ for all $n \in \mathbb{N}$, or equivalently, if $g = O(f)$.
- $f = \omega(g)$ if for every $C > 0$ there exists $n_0 \in \mathbb{N}$ such that $f(n) > C \cdot g(n)$ for all $n \ge n_0$, or equivalently, if $g = o(f)$.
- $f = \Theta(g)$ if both $f = O(g)$ and $g = O(f)$.

We usually describe the function g by a term with a single variable. For example by $f = O(n^2)$ we mean that $f = O(g)$ where $g : \mathbb{N} \to \mathbb{R}, g(n) := n^2$. Unless specified otherwise we use the letter n to denote the asymptotic variable. Sometimes asymptotic expressions occur in isolation within an equality or inequality. If the expression can be found on the right hand side it means an unspecified function of the indicated growth. For example, an equality of the form $f(n) = n^{3/2-o(1)}$ is to be read as "there exists a non-negative function $g : \mathbb{N} \to \mathbb{R}$ such that $g = o(1)$ and $f(n) = n^{3/2-g(n)}$ for all $n \in \mathbb{N}$". If the expression occurs on the left hand side it stands for all function with the indicated asymptotics. For example, the inequality $n^{1+o(1)} + O(\log n) \le n^{1+o(1)}$ means that for every non-negative $f,g : \mathbb{N} \to \mathbb{R}$ with $f = o(1)$ and $g = O(\log n)$ there exists a non-negative $h : \mathbb{N} \to \mathbb{R}$ with $h = o(1)$ such that $n^{1+f(n)} + g(n) \le n^{1+h(n)}$.

Graphs. A *graph* G is a pair (V,E) consisting of a finite set V and a set $E \subseteq \binom{V}{2}$, where $\binom{V}{2}$ denotes the set of two-element subsets of V. The elements of V are called *vertices* of G, the elements of E are the *edges* of G. We denote the vertex set of a graph G by $V(G)$, its edge set by $E(G)$, and define $v(G) := |V(G)|$ and $e(G) := |E(G)|$. If $X,Y \subset V(G)$ we denote the set of edges between X and Y by $E_G(X,Y)$ or just $E(X,Y)$ if the underlying graph is clear from context, i.e.

$$E_G(X,Y) := \{xy : x \in X, y \in Y, xy \in E(G)\}$$

Given two graphs $G = (V,E)$ and $G = (V',E')$ we define their *union* $G \cup G'$ and their *intersection* $G \cap G'$ by

$$G \cup G' := \left(V \cup V, E \cup E'\right) \qquad , \qquad G := \left(V \cap V', E \cap E'\right).$$

We denote their (external) *disjoint union* by $G \sqcup G'$, that is,

$$G \sqcup G' := \big((V \times \{1\}) \cup (V' \times \{2\})\ ,\ \{(x,1)(y,1) : xy \in E(G)\} \cup \{(x,2)(y,2) : xy \in E(G')\}\big)\,.$$

For any vertex x of G we denote the graph obtained by removing x and its incident edges from G by $G - x$, i.e.

$$G - x := (V \setminus \{x\}\,, E \setminus \{e \in E : x \in e\})$$

Similarly, given an edge $e \in E$, we let $G - e$ denote the graph $(V, E \setminus \{e\})$. The vertex set and the edge set of a graph will always be disjoint so there will be no clash of notation. If $e \in \binom{V}{2} \setminus E$ we write $G \cup \{e\}$ for the graph $(V, E \cup \{e\})$. More generally, for $U \subseteq \binom{V}{2} \setminus E$ we let $G \cup U := (V, E \cup U)$ and $G \setminus U := (V, E \setminus U)$. If H is an edge-transitive graph, that is, for any two edges $e, e' \in E(H)$ there exists an automorphism φ of H such that $\varphi(e) = e'$, we denote the (unlabelled) graph obtained by removing an arbitrary edge from H by H^-.

Neighbourhoods and degrees. We write $N_G(v)$ for the set of neighbours of v in G. We sometimes call the elements of $N_G(v)$ G-neighbours of v when we are dealing with multiple graphs simultaneously and want to emphasize that we talk about adjacency in G. Similarly we write $d_G(v)$ for the degree of v is G. We denote the minimum degree of G by $\delta(G)$ and the maximum degree by $\Delta(G)$.

Graph homomorphisms. Given two graphs G and H, a *homomorphism* φ from G to H, written $\varphi : G \to H$, is a map $\varphi : V(G) \to V(H)$ such that for all $x, y \in V(G)$ with $xy \in E(G)$ one has $\varphi(x)\varphi(y) \in E(H)$. We denote the set of graph homomorphisms from H to G by $\mathrm{Hom}(H,G)$. An *embedding* of H into G is an injective graph homomorphism from H to G. If $H' \subseteq H$ and $\varphi \in \mathrm{Hom}(H,G)$ we write

$$\varphi(H') := (\varphi(V(H)), \{\varphi(x)\varphi(y) : xy \in E(H)\})\,.$$

We call a homomorphism $\varphi : G \to H$ an *isomorphism* if it is bijective and for all $x, y \in V(G)$ with $xy \notin E(G)$ one has $\varphi(x)\varphi(y) \notin E(H)$. An *automorphism* of H is an isomorphism from H to itself. We denote the automorphism set of H by $\mathrm{Aut}(H)$.

Paths and distances. The *path* P_n on n vertices is the graph defined by

$$V(P_n) = \{0, \dots, n-1\} \qquad , \qquad E(P_n) = \{\{i, i+1\} : 0 \le i < n-1\}.$$

Given a graph G and $x, y \in V(G)$ a *path* of length ℓ from x to y (or xy-path for short) in G refers to either a sequence $v_0 \dots v_\ell$ of vertices of G with $v_0 = x$, $v_\ell = y$ and $v_i v_{i+1} \in E(G)$ for $i \in [\ell]$ or a subgraph of G that is isomorphic to $P_{\ell+1}$ via an isomorphism φ with $\varphi(0) = x$ and $\varphi(\ell) = y$. We use these notions of a path interchangeably since every sequence $v_0 \dots v_\ell$ with the properties above yields a suitable isomorphism φ via $\varphi(i) := v_i$ for $0 \le i \le \ell$ and vice versa.

We denote the distance, that is, the length of a shortest path between x and y in G by $\mathrm{dist}_G(x,y)$, and write $\mathrm{diam}(G)$ for the diameter of G.

Rooted trees and rooted forests. A *rooted tree* is a tree T together with a designated vertex $z \in V(T)$ called the *root*. In a rooted tree the neighbours of a vertex $x \in V(T)$ whose distance to the root is larger than the distance from the root to x are called the *children* of x. If $x \in V(T) \setminus \{z\}$, the unique neighbour of x that is closer to z is called the *parent vertex* of x. The *height* of a vertex is the length of a longest downward path from the vertex to a leaf of T where downward path of length ℓ means any sequence $v_0 \dots v_\ell$ of vertices such that v_i is a child of v_{i-1} for $1 \leq i \leq \ell$. The *height* of the tree T is defined as the height of the root. We denote the height of a vertex $x \in V(T)$ by $\mathrm{ht}_z(x)$ and the height of T by $\mathrm{ht}_z(T)$. We do not specify T in the expression $\mathrm{ht}_z(x)$ as in our application the underlying tree will be clear from context.

A *rooted forest* is a forest F together with designated vertices $z_1,\dots,z_s$, one from each of the s components of F. In other words it is a disjoint union of rooted trees. The height of a vertex x in F, denoted by $\mathrm{ht}_{\mathbf{z}}(x)$, where $\mathbf{z} := (z_1,\dots,z_s)$, is simply its height in the component containing it. The height $\mathrm{ht}_{\mathbf{z}}(F)$ of F is defined as the maximum height among its components.

Wheel graphs. We denote the wheel graph on $k+1$ vertices by W_k, that is,

$$V(W_k) = \{w_1,\dots,w_k,v\} \quad , \quad E(W_k) = \{w_1w_2,\dots,w_{k-1}w_k,w_kw_1\} \cup \{vw_i : i \in [k]\}$$

If $k \geq 4$ we call the unique universal vertex v the *hub* and refer to the cycle formed by $w_1,\dots,w_k$ as the *outer cycle*.

Edge colourings. Given a graph H, an *edge-colouring* is a map $\chi : E(H) \to A$, where A is a finite set to which we refer as the set of *colours*. An edge-colouring χ is said to be *monochromatic* (abbreviated as m.c.) if $\chi(e) = \chi(f)$ for all $e,f \in E(H)$. We call a subgraph $H' \subseteq H$ *monochromatic* (under χ) if the restriction of χ to H' is monochromatic. For example we say that χ contains a m.c. cycle if there exists a subgraph $C \subseteq H$ that is a cycle and satisfies $\chi(e) = \chi(f)$ for all $e,f \in E(C)$. We say that an edge-colouring is *proper* if every two incident edges receive different colours.

Extremal numbers. Given a graph H, another graph G is called *H-free* if there are no injective graph homomorphisms from H to G, i.e. G does not contain any copies of H. The *extremal number* $\mathrm{ex}(n,H)$, also referred to as the *Turán number* in the literature, is defined to be the largest number of edges an H-free graph on n vertices can have.

Probability and random graphs. When we work with randomness we denote the probability measure of the underlying probability space by $\mathbb{P}$.

We denote the binomial random graph (often referred to as the Erdős-Renyi random graph in the literature) on n vertices with edge probability $p = p(n)$ by $G(n,p)$. In that model every edge occurs independently with probability p. For our purposes we let $[n]$ be the vertex set of $G(n,p)$. For each graph on G' on $[n]$ we then have $\mathbb{P}(G(n,p) = G') = p^{e(G')} \cdot (1-p)^{\binom{n}{2} - e(G')}$.

We say that a property of $G(n,p)$ holds *with high probability* (abbreviated as *w.h.p*) if the probability of $G(n,p)$ having that property tends to 1 as $n \to \infty$.

Typographic remarks

We shall mark the end of the proof of a major result (i.e. those stated in Chapter 1) by a filled box ■ and the end of a nested proof (e.g. that of a distinguished claim within the proof of another theorem) by an empty box □.

Chapter 1

Introduction

Extremal graph theory and additive combinatorics, the theory of the structure of set addition, are two areas of combinatorics which over the course of the last decades have frequently enriched each other. The most famous example of such a connection is Szemerédi's development of the Regularity Lemma, one of the most powerful tools in modern graph theory, in his theorem on the existence of k-term arithmetic progression in subsets of $\mathbb{N}$ with positive upper density.

In this text we investigate three types of problem, two of them from graph theory, one from additive combinatorics and point out bridges between the involved additive combinatorial problems and the extremal problems from graph theory.

Our first main object of study are *graph bootstrap percolation processes*, which are a type of cellular automaton, a concept introduced by von Neumann [79] in his lectures on self-reproducing automata, following a suggestion of Ulam [98]. A cellular automaton is a discrete dynamical process on a collection of cells that change their states at each step according to a fixed local rule. We will focus on the question of determining the amount of time needed for a cellular automaton with finitely many cells and a monotone update rule to reach a state that does not change under said update rule anymore.

The second main object of study in this thesis are strong B_h-sets, a generalisation of B_h-sets in the integers, which in turn generalise the notion of a Sidon set in the integers. The latter are named after Simon Sidon [93], who investigated them in his studies of Fourier series, and are defined as subsets of $\mathbb{Z}$ in which the pairwise sums of the elements are distinct (up to the obvious permutation of the summands). In B_h-sets one demands that the sums of any h elements are pairwise distinct. In strong B_h-sets the sums are not only required to be pairwise distinct but also sufficiently far apart in absolute value. We will be interested in the extremal questions of finding dense infinite B_h-sets both in the integers and in sparse subsets thereof.

Our final object of study are splits of matchings. Those splits are themselves matchings that intersect a given familiy of disjoint perfect matchings of an underlying graph in a prescribed way. Another way of imagining splits is to think of an edge-coloured regular graph such that no edges of the same colour share an endpoint and then ask for a matching with a given distribution of colours among its edges. We will study conditions on the colour distributions that guarantee the existence of splits no matter what the underlying regular graph and its edge-colouring look like.

1.1 Bootstrap percolation

The term *percolation process* has its origin in the work of Broadbent and Hammersley [26] where it was used to desribe a model of the random flow of a liquid through a medium. The idea of percolation was that it is not the fluid that behaves randomly (which leads to the concept of a *diffusion process*) but the medium.

The notion of *bootstrap percolation* was introduced by Chalupa, Leath, and Reich [28] in 1979 in their study of ferromagnetic phenomena. In graph-theoretic terminology their problem reads as follows: Given the lattice with vertex set $\mathbb{Z}^d$ and any two vertices being adjacent if their difference is a standard unit vector, one chooses a random subset A_0 of $\mathbb{Z}^d$ by picking every element independently with probability p. One then fixes a parameter $r \in \mathbb{N}$ and defines a process $(A_t)_{t\in\mathbb{N}_0}$ by $A_t := A_{t-1} \cup \{x \in \mathbb{Z}^d \setminus A_{t-1} : |N(x) \cap A_{t-1}| \geq r\}$. This type of process is known as *r-neighbour bootstrap percolation* (or just neighbourhood bootstrap percolation) and, together with its variants for other graphs than lattices, constitutes one of the most studied type of bootstrap process. The perhaps most common type question in neighboorhood percolation asks for the *critical probability* of a certain property of the process, that is, the infimum over all $0 \leq p \leq 1$ for which the process above has the desired property with probability at least one half. For example, if the underlying graph is the grid $[n]^d$ instead of the lattice $\mathbb{Z}^d$ it was a major open problem to determine the critical probability $p_c([n]^d, r)$ of the property that eventually $A_t = [n]^d$, until in 2011, Balogh, Bollobás, Duminil-Copin, and Morris [14] found sharp bounds on $p_c([n]^d, r)$.

The applications of bootstrap processes in other areas of science are numerous. For thorough treatises we recommend the survey [89] by Saberi and the books [90, 95] by Sahimi and Stauffer-Aharony. See also the article [1] by Adler and Lev.

There are many other models of bootstrap percolation. For a more detailed overview on combinatorial models of bootstrap percolation we refer to the survey article of Morris [77] from 2017. In the following we concentrate on the model called graph bootstrap percolation.

1.1.1 Graph bootstrap percolation

In 1968 Bollobás [23] introduced the concept of *weakly saturated graphs*. Starting from a graph G add all non-adjacent pairs e of vertices to the edge set for which e is contained in a k-clique in $G \cup \{e\}$. Given a positive integer k an n-vertex graph is called *weakly k-saturated* if by repeating the procedure above one eventually reaches the complete graph K_n. In the literature this property is often stated in the following equivalent form: An n-vertex graph G is called weakly k-saturated if there exists an ordering $e_1, \ldots, e_{\binom{n}{2}-e(G)}$ of all non-adjacent pairs of vertices such that adding pairs as edges to the graph one by one in the specified order produces a new k-clique in each step, that is, for every $1 \leq t \leq \binom{n}{2} - e(G)$, there exists a k-clique in $G \cup \{e_1, \ldots, e_t\}$ containing e_t.

Bollobás defined the quantity $\mathrm{wsat}_{K_r}(n)$ as the smallest number of edges in an n-vertex weakly k-saturated graph. Nowadays the additional assumption that G does not contain a copy of H is often included in the definition of weakly saturated graphs. This distinction does not affect the numbers $\mathrm{wsat}_{K_r}(n)$ since a graph attaining the minimum cannot contain a copy of K_r. Bollobás conjectured that $\mathrm{wsat}_{K_r}(n) = (r-2)n - \binom{r-1}{2}$. This conjecture was proved independently by Alon [7], Frankl [49] and Kalai [59]. We remark that the extremal construction with $\mathrm{wsat}_{K_r}(n)$ edges is obtained by removing the edges of a clique of order $n-r+2$ from K_n.

A more general case of weak saturation is the following. For a fixed graph H a spanning subgraph G of K_n is called *weakly H-saturated* if the non-adjacent pairs of vertices of G can be ordered as $e_1, \dots, e_{\binom{n}{2}-v(G)}$ such that adding them one by one in order creates a copy of H at each step. The smallest number of edges in a weakly H-saturated n-vertex graph is denoted by $\mathrm{wsat}_H(n)$. We refer to Section 10 of [47] for a short and concise survey of weak sauturation and the problem of determining $\mathrm{wsat}_H(n)$ for various H.

Balogh, Bollobás, and Morris [15] phrased the general case in terms of a graph process by taking the definition of weakly k-saturated graphs used in [23] and replacing k-cliques by arbitrary H. They named the general process *H-bootstrap process*

Definition 1.1.1 (Graph bootstrap processes)**.** Let H be a fixed graph, and let G be another graph. Denote the number of copies of H in G by $n_H(G)$. The *H-bootstrap process*, or *H-process* for short, on G is the sequence $(G_t)_{t\geq 0}$ of graphs defined by $G_0 := G$ and

$$V(G_t) = V(G),$$
$$E(G_t) = E(G_{t-1}) \cup \left\{ e \in \binom{V(G)}{2} : n_H(G_{t-1} \cup \{e\}) > n_H(G_{t-1}) \right\}$$

for $t \geq 1$. We also refer to G as the *starting graph* of the process.

When the graph H is clear from context we often omit it and write *bootstrap process* (or sometimes even just *process*) instead of H-bootstrap process. Informally, the H-bootstrap process is the process that starts with G and in every step adds every edge that completes a copy of H. As $\binom{V(G)}{2}$ is a finite set it is clear that any bootstrap process stabilises after a finite number of steps, i.e. there exists $t^* \in \mathbb{N}$ such that $G_t = G_{t^*}$ for all $t \geq t^*$.

As an example, Figure 1.1 depicts the K_3-process on the path of length three. Note that, while in this example the process reaches a complete graph, in general this is not necessarily the case.

Definition 1.1.2 (Stable graphs and final graphs)**.** A graph G is called *H-stable* if for every $e \in E(H)$ and every embedding $\varphi : H - e \to G$, one has $\varphi(e) \in E(G)$. The graph $\bigcup_{t\geq 0} G_t$ is called the *final graph* of the H-process on G and denoted by $\langle G \rangle_H$. We say that G *percolates* (with respect to H) if $\langle G \rangle_H$ is a complete graph.

In these terms a weakly k-saturated graph as defined in [23] is a graph that percolates with respect to the complete graph K_k, and similarly, a weakly H-saturated graph is one that percolates in the H-process. An equivalent way of saying that a graph is H-stable is that the only copies of H minus an edge in G are those obtained by removing an edge from a copy of H in G. The final

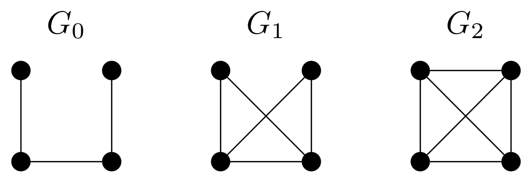

FIGURE 1.1: The K_3-bootstrap process on $G = P_3$. At time 1 the two diagonal edges are added. This allows the top edge to be added after one more step. The process has stabilised at time 2 as it has reached a complete graph.

graph of an H-process must be H-stable by definition of the H-process. Furthermore, $\langle G\rangle_H$ is the smallest H-stable graph containing G as a subgraph, since by a simple inductive argument every H-stable graph containing G must also contain each graph of the H-process.

We mention that graph bootstrap percolation is an instance of the following more general type of bootstrap process introduced by Balogh, Bollobás and Morris in [16]. Given a hypergraph $\mathscr{G}$ one starts with a set $A_0 \subset V(\mathscr{H})$, and at each step adds all vertices that are the last missing vertex in a hyperedge. More precisely, for $t \geq 1$, one sets $A_t := A_{t-1} \cup \{u \in V(\mathscr{G}) : \{u\} = e \setminus A_{t-1}$ for some $e \in E(\mathscr{G})\}$. The H-bootstrap process on G for a (simple) graph H, and an n-vertex graph G can be expressed in this general model by choosing $\mathscr{G}$ to be the $e(H)$-uniform hypergraph whose vertices are the edges of K_n and whose hyperedges are the edge sets of copies of H in K_n, and by selecting the edges of G (viewed as subgraph of K_n) as A_0.

The most studied type of extremal question in graph bootstrap percolation asks for conditions that either force the starting graph to percolate or guarantee that it does not percolate. The question for the minimum size of a percolating graph is the above mentioned problem of determining $\mathrm{wsat}_H(n)$. Motivated by the random setting considered in [28] Balogh, Bollobás, and Morris [15] studied the *critical probability* $p_c(n,H)$ which is defined as

$$p_c(n,H) = \inf\left\{p : \mathbb{P}(\langle G(n,p)\rangle_H = K_n) \geq \frac{1}{2}\right\}. \tag{1.1}$$

They determined the critical probability $p_c(n,K_r)$ up to a factor of $(\log n)^2$ for $r \geq 4$, and gave the exact value of $p_c(n,K_4)$ up to a constant factor. Later, Kolesnik [68] refined their result for the case $r = 4$ by giving the asymptotically sharp value of $p_c(n,K_4)$. Further research on the critical probability was done in [17, 20] for complete bipartite graphs.

1.1.2 The running time of graph bootstrap percolation

Most of the literature on graph bootstrap percolation revolves around the question under which conditions the H-process on a given n-vertex starting graph eventually reaches K_n. Another, substantially less investigated question asks how long it takes before the process stabilises. In particular, one is interested in the maximum number of required steps in the H-process over all graphs of a given order. This problem, posed by Bollobás, was first investigated in [25] and [76]. Przykucki [83] and later Benevides-Przykucki [19] studied the analogous question for neighbourhood percolation. Koch, Gunderson, and Przykucki [54] considered the critical probability for percolation by a time $t \geq 0$. Given H and a graph G that is distributed as $G(n,p)$ they were interested in the probability

$$p_{\mathrm{c}}(n,r,t) := \inf\left\{p : \mathbb{P}(G_t = K_n) \geq \frac{1}{2}\right\},$$

where $(G_t)_{t\geq 0}$ is the H-process on G, and showed that for $r \geq 4$, $t \in \left[\log\log(n)/(3\log(\binom{r}{2}-1))\right]$, and n suffciently large one has

$$\frac{n^{-\lambda(r,t)}}{\omega(n)} \leq p_{\mathrm{c}}(n,r,t) \leq n^{-\lambda(r,t)} \cdot \log n$$

for some positive constant $\lambda(r,t)$. This problem is quite different from determining the maximum number of steps because graphs maximising that number might come from a set of graphs that occur only with tiny probability. In fact the maximising graphs might not even percolate. We note that the time required to reach K_n has also been studied in the context of neighbourhood percolation. See for example [12] for neighbourhood percolation on the grid $[n]^2$ or [58] for random starting graphs.

In this thesis the focus of our discussion lies on the problem of maximising the number of steps in the H-process over all n-vertex graphs.

Definition 1.1.3 (Running time of a bootstrap process)**.** The *running time* of the H-bootstrap process $(G_t)_{t\geq 0}$ on a graph G is

$$\tau_H(G) := \min\{t \in \mathbb{N} : G_t = G_{t+1}\}.$$

For $n \in \mathbb{N}$, we define $M_H(n)$ to be the maximum running time of the H-bootstrap process over all starting graphs on n vertices. That is,

$$M_H(n) := \max_{|V(G)|=n} \tau_H(G).$$

The maximum running time always obeys the trivial bound $M_H(n) \leq \binom{n}{2}$. Let us consider a simple example to familiarise ourselves with the definition above.

Example 1.1.4 (The running time of the K_3-process)**.** The K_3-bootstrap process on a graph G can be described as follows: Start with G and in every step add a new edge between two vertices if and only if their distance is two. Then in the K_3-process $(G_t)_{t\geq 0}$ on G one has $\mathrm{diam}(G_t) = \lceil \mathrm{diam}(G_{t-1})/2 \rceil$ for all $t \geq 1$ because for any path $v_0v_1 \ldots v_{2\ell}$ of even length in G_{t-1}, $v_0v_2 \ldots v_{2\ell-2}v_{2\ell}$ is a path with the same endpoints in G_t. This implies

$$M_{K_3}(n) = \lceil \log_2(n-1) \rceil$$

for every positive integer n. The maximum running time is realised by the path on n vertices and, more generally, by any n-vertex graph G with $\lceil \log_2 \mathrm{diam}(G) \rceil = \lceil \log_2(n-1) \rceil$.

A natural extension of the case $H = K_3$ is the investigation of complete graphs K_r for $r \geq 4$. The precise value of $M_{K_4}(n)$ has been determined by Matzke [76] and, independently, by Bollobás-Przykucki-Riordan-Sahasrabudhe [25].

Theorem 1.1.5 (Matzke 2015, Bollobás-Przykucki-Riordan-Sahasrabudhe 2017)**.** *The maximum running time of the K_4-process is $M_{K_4}(n) = n-3$ for all $n \geq 3$.*

The latter set of authors also showed, using a random construction, that $M_{K_r}(n) \geq n^{2-\frac{r-2}{\binom{r}{2}}-o(1)}$ as $n \to \infty$ for all $r \geq 5$ and conjectured that $M_{K_r}(n)$ is subquadratic for all $r \geq 5$.

Conjecture 1.1.6 ([25] Conjecture 1)**.** For all $r \geq 5$ we have $M_{K_r}(n) = o(n^2)$.

Two years later Balogh, Kronenberg, Pokrovskiy, and Szabó [13] disproved the case $r \geq 6$ of Conjecture 1.1.6 by constructing graphs G with $\tau_{K_r}(G) = \Omega(n^2)$.

Theorem 1.1.7 (Balogh, Kronenberg, Pokrovskiy, Szabó 2019)**.** *For every $r \geq 6$ and large enough n, we have $M_{K_r}(n) \geq \frac{n^2}{2500}$.*

In the same paper they gave a construction that attains $M_{K_5}(n) \geq n^{2-O(1/\sqrt{\log n})}$ by relating the problem of finding lower bounds on $M_{K_5}(n)$ to the additive combinatorial function $r_3(n)$, which denotes the size of a largest subset of $[n]$ that is free of three-term arithmetic progression (that is, non-trivial solutions to the equation $x+z=2y$).

Theorem 1.1.8 (Balogh, Kronenberg, Pokrovskiy, Szabó 2019)**.** *The maximum running time of the K_5-process is at least $M_{K_5}(n) \geq \frac{nr_3(n)}{1200}$. In particular,*

$$M_{K_5}(n) \geq n^{2-O(1/\sqrt{\log n})}. \tag{1.2}$$

Determining the asymptotics of $r_3(n)$ is one of the most fruitful and well-studied problems in additive combinatorics. As to lower bound the most well-known example is the Behrend construction [18], which transfers the fact that in the Euclidean plane no line can intersect a sphere in more than two points to the integers via digit representations. That construction offered the best known lower bound $r_3(n) \geq n \cdot e^{-O(\sqrt{\log n})}$ for about six decades until the improvement by a factor of $\Theta(\sqrt{\log n})$ by Elkin [36] (see [53] also by Green and Wolf for a shorter proof of the same result). Note that Elkin's bound is still of the form $n \cdot e^{-O(\sqrt{\log n})}$. The upper bound $r_3(n) = o(n)$ is the famous theorem of Roth [84]. There have been several gradual improvements of Roth's bound over the years. A brief historic overview is given in the introduction of [92]. The currently best upper bound $n \cdot e^{-\Omega((\log n)^{1/11})}$ comes from a recent article by Kelley and Meka [61]. An exposition of their proof from a more additive combinatorial perspective is given by Bloom and Sisask in [22]. In the context of graph bootstrap percolation the upper bound on $r_3(n)$ prevents Theorem 1.1.8 from giving a quadratic lower bound.

The constructions employed to prove Theorems 1.1.7 and 1.1.8 use building blocks similar to those in the random construction in [25] but do so in a deterministic way, and constitute the initial ground for our discussions and results in Chapter 3. The question whether $M_{K_5}(n) = o(n^2)$

remains an open problem, but we will later provide more evidence for a positive answer by considering the analogous question for wheel graphs.

Recently the study of the running time *hypergraph bootstrap processes* has gained increased attention. Those processes and their running times are defined analogously to the common graph bootstrap processes: Given hypergraphs $\mathscr{G}$ and $\mathscr{H}$ the $\mathscr{H}$-process on $\mathscr{G}$ is the sequence $(\mathscr{G}_t)_{t\geq 0}$ that starts with $\mathscr{G}_0 := \mathscr{G}$ and in which at each step every hyperedge that is the only missing hyperedge in a copy of $\mathscr{H}$ is added. Then $M_{\mathscr{H}}(n)$ is the largest number of steps an $\mathscr{H}$-process on an n-vertex hypergraph can take before it stabilises. Note that this type of process is not the general model in [16] that describes graph bootstrap processes in terms of hypergraphs. There one adds vertices completing hyperedges, whereas here hyperedges completing copies of a hypergraph $\mathscr{H}$ are introduced at each step. However, hypergraph bootstrap processes can be described by the model in [16] by regarding copies of $\mathscr{H}$ as hyperedges of another hypergraph. Noel and Ranganathan [80] showed that for $r \geq 3$ and $k \geq r+2$ the complete r-uniform hypergraph K_k^r on k vertices satisfies

$$M_{K_k^r}(n) = \Theta(n^r) \tag{1.3}$$

whereas for $k = r+1$ they obtained the lower bound $M_{K_{r+1}^r}(n) = \Omega(n^{r-1})$. In the concluding remarks of their article they conjecture that $M_{K_4^3}(n) = O(n^2)$, and ask whether (1.3) extends to the case $k = r+1$ when r is sufficiently large. In the same year Hartarsky and Lichev [57] and independently Espuny Díaz, Janzer, Kronenberg, and Lada [41] answered that question positively for any $r \geq 3$, thereby disproving the conjecture from [80]. The second set of authors further introduced a variant of the hypergraph bootstrap process in which they complete copies of $\mathscr{H}$ with more than one missing edge. More precisely, a copy of $\mathscr{H}$ is called m-completable in $\mathscr{G}$ if all but at most m hyperedges of $\mathscr{H}$ lie in $\mathscr{G}$. They then define the $(\mathscr{H}, m)$-bootstrap process on $\mathscr{G}$ as the process that starts with $\mathscr{G}$ and in each step adds all missing edges of every m-completable copy of $\mathscr{H}$. Note that since complete hypergraphs are edge-transitive the $(K_k^r, 2)$-process on any $\mathscr{G}$ is the same as the $K_k^r - e$-process on $\mathscr{G}$ for any edge e of K_k^r. Using that type of process, the authors of [41] obtain $M_{K_5^3 - e}(n) = \Theta(n)$ and the exact result $M_{K_4^3}(n) = 2n - \lfloor \log_2(n-2) \rfloor - 6$ where e is an arbitrary edge of the respective complete hypergraph, and thus show that bootstrap processes for r-uniform hypergraphs allow for more types of running times than just $\Theta(n^r)$. In the following we will concentrate on simple graphs and not go deeper into questions about hypergraphs.

In the setting of simple graphs two ranges of asymptotic running times stand out. The first is the sublinear range, which consists of those graphs H for which the question whether $M_H(n) = o(n)$ is answered in the positive. The second is the superlinear range and is defined by the question whether $M_H(n) = \omega(n)$. There are also graphs for which $M_H(n) = \Theta(n)$ and so both of the questions above are answered in the negative. Understanding the properties of H which control the asymptotic behaviour of $M_H(n)$ is the main goal of our work on graph bootstrap percolation presented in the next two sections and Chapters 2 and 3.

1.2 Sublinear running times

In this section we state the contributions of this thesis that belong to the sublinear range. All results are based on joint work with Patrick Morris and Tibor Szabó [42, 43]. An extended abstract of that work has appeared in [45].

1.2.1 Constant and logarithmic running times

By looking at the trivial scenario $H = K_2$ one can see that there exist graphs H for which $M_H(n)$ only depends on H but not on n. Playing around with small examples such as paths and stars one finds that there are also non-trivial choices of H with constant maximum running time. In fact, this behaviour is common to all trees and, more generally, to all forests.

Theorem 1.2.1 (The maximum running time for forests)**.** *Every forest F on k vertices satisfies,*

$$M_F(n) \leq \frac{1}{8} \cdot (k^2 + 6k + 76) \tag{1.4}$$

In the literature on maximum running times of simple graphs so far the focus was on complete graphs. Besides them, another family of graphs that generalise K_3 are the cycles C_k for $k \geq 3$. Determining $M_{C_k}(n)$ for any $k \geq 3$ is thus a natural generalisation of the problem of finding $M_{K_3}(n)$. Recall that in the K_3-process (cf. Example 1.1.4) on an n-vertex graph the diameter decreased by roughly a factor of two in each step since a triangle minus an edge is just a path of length two. This resulted in a running time of at most $\lceil \log_2(n-1) \rceil$. For any $k \geq 3$, a k-cycle minus an edge is a path of length $k-1$. An intuitive guess would be that apart from a constant number of steps that might be necessary to deal with small examples the diameter in any C_k-process decreases by a factor of about $k-1$ in every step until the process stabilises. While that guess proves true from an asymptotic point of view, it turns out that once one is interested in the precise values of $M_{C_k}(n)$ the situation differs slightly depending on the parity of k.

Theorem 1.2.2. *Let $k \geq 3$. For sufficiently large $n \in \mathbb{N}$ we have*

$$M_{C_k}(n) = \begin{cases} \lceil \log_{k-1}(n + k^2 - 4k + 2) \rceil & \text{if } k \text{ is odd;} \\ \lceil \log_{k-1}\left(2n + k^2 - 5k\right) \rceil & \text{if } k \text{ is even.} \end{cases} \tag{1.5}$$

Moreover, for any $n \geq k \geq 5$,

$$M_{C_k}(n) \geq \left\lfloor \frac{k}{2} \right\rfloor + 1.$$

Remark 1.2.3. In the theorem above *sufficiently large* means that the terms on the right hand side of (1.5) exceed the lower bound $\lfloor k/2 \rfloor + 1$ given by the second part of the theorem, which happens when n is larger than roughly $k^{k/2}$. For smaller n the behaviour is different as a single k-cycle with a well-placed chord achieves a longer running time than the extremal constructions that determine the running time for larger n.

For odd k the value $\lceil \log_{k-1}(n+k^2-4k+2) \rceil$ comes from the path P_n. With the exception of those n for which $M_{C_k}(n) > M_{C_k}(n-1)$ the extremal construction is not unique as, for example, one could take a slightly shorter path.

As to even k the behaviour of $M_{C_k}(n)$ is similar. Again $M_{C_k}(n)$ eventually becomes (essentially) a shifted and rounded logarithm. However the points where the function jumps up by one do not lie close to powers of $k-1$ but halfway between subsequent powers. This difference is attributed to a combination of two facts. First, completing an even cycle in a bipartite graph does not destroy the bipartiteness. Second, the C_k-process (and, in larger generality, the H-process of any connected H) is a local process in the sense, that whether a new edge between two vertices will be added at some time t only depends on the $(k-1)^{\text{th}}$ neighbourhoods of the two vertices at time $t-1$. Now imagine a graph with large diameter that percolates in the C_k-process and is not bipartite but can be made bipartite by removing a single edge. In such a graph the non-bipartiteness would have to spread throughout the graph over the course of the process. We will see that the graphs which maximise the running time of the C_k-process are those non-bipartite graphs in which the diameter is large and the non-bipartiteness spreads as slowly as possible.

Theorem 1.2.2 together with the observation that paths have constant maximum running time (we will encounter a proof of this fact in Chapter 2) determine, up to a small additive constant, the maximum running time of any connected graph with maximum degree at most 2 as well as any disconnected graph with a path component. The only missing part to determining the maximum running times for all graphs of maximum degree at most 2 are disjoint unions of cycles. We give the following more general result that includes all H with a cycle component:

Theorem 1.2.4. *Let $k \geq 3$, and let H be a graph, one of whose components is a k-cycle, that is, $H \cong \tilde{H} \sqcup C_k$ for some other graph H'. Then there exists a constant $\kappa = \kappa(H)$*

$$M_H(n) \leq \log_{k-1}(n) + \kappa.$$

Remark 1.2.5. The magnitude of κ obtained in the proof of Theorem 1.2.4 presented in Section 2.6 has a tower-type dependency on $v(H)$. This dependency is a consequence of our method and should be far from optimal. For example, in [43] it is shown that if H is a union of s disjoint cycles of lengths $k_1, \ldots, k_s$ one can bound κ by a polynomial in $k_1 + \ldots + k_s$.

1.2.2 Necessary conditions for sublinear running time

The results on trees and cycles stated above provide two examples of sublinear running time. In both cases the asymptotics of the maximum running time are determined by how fast the diameter of the graph on which the process is run decreases at each step. The proof of Theorem 1.2.1 relies heavily on the fact that every tree has a vertex of degree one. This made the diameter decrease to at most $v(H)$ after just one step. We will see that, in accordance with our intuition, in the C_k-process on any graph the diameter decreases by a factor of $k-1$ at each step unless it is already below k. The extremal examples that realise the maximum running time are graphs that maximise the maximum distance or the related notion of maximum length of a shortest odd

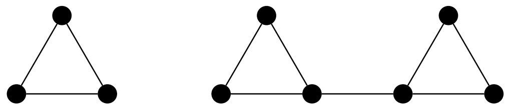

FIGURE 1.2: A graph H with maximum running time $O(1)$ whose two components have logarithmic or linear running time, respectively. Given a copy H_1 of H at time 1 of the H-process on some graph G, the neighbourhood of every vertex in $V(G) \setminus V(H_1)$ will be a clique at time 2.

walk between two vertices. While investigating the local behaviour was necessary to derive the precise value of $M_{C_k}(n)$, the asymptotics only depended on the observations that the diameter is divided by roughly $k-1$ in every step of the C_k-process, and the final graph will be K_n. If the graph H does not have a degree-one vertex and is not a cycle we can build starting graphs in whose H-process the diameter decreases merely by an additive constant in each step. The next result tells us that such graphs have at least linear running time.

Theorem 1.2.6. *Let H be a graph such that each component of H has minimum degree at least two and maximum degree at least three. Then*

$$M_H(n) = \Omega(n).$$

If H is bipartite the bound can be achieved by a bipartite starting graph.

Remark 1.2.7. In Theorem 1.2.6 it is necessary that the degree assumptions hold for every component of H. If it was dropped for just one component the graph in Figure 1.2 would be a counterexample.

As a consequence of Theorems 1.2.2 and 1.2.6 we can see that for connected H the existence of a pendent vertex in H is necessary to make $M_H(n)$ asymptotically constant. However, a minimum degree of one turns out not to be a sufficient condition, even if the maximum degree is small. In fact it does not even imply sublinear running time.

Proposition 1.2.8. *There exists a connected graph H with minimum degree one and maximum degree three satisfying $M_H(n) = \Omega(n)$.*

If we focus on maximising $M_H(n)$ subject to the condition $\delta(H) = 1$ and drop the assumption $\Delta(H) = 3$ we can improve the lower bound from $\Omega(n)$ to $\Theta(n^2)$.

The goal of our investigations of sublinear running times was to pinpoint the properties of H that determine whether $M_H(n) = o(n)$. Theorems 1.2.6 and 1.2.4 provide the partial answer that for $\delta(H) \geq 2$, one has $M_H(n) = o(1)$ if H has a cycle component, and $M_H(n) = \Omega(n)$ otherwise. Furthermore Proposition 1.2.8 made clear that this criterion cannot be directly extended to include the case $\delta(H) = 1$.

All results presented in this section will be proven in Chapter 2. At the end of that chapter we continue our discussion of sublinear running times with consideration of the ideas introduced in the proofs.

1.3 Linear and superlinear running times

In this section we introduce the results on graphs with linear or superlinear running times. All of them were obtained as joint work with Patrick Morris and Tibor Szabó [44].

The common theme of our lower bounds is how far we can generalise the chain-based construction of [13] to yield superlinear, and in the best case quadratic, lower bounds on the maximum running time. As to superlinear upper bounds, we will encounter two ways of obtaining non-trivial, i.e. subquadratic, results. One is based on Turán numbers of bipartite graphs while the other relies on the Triangle Removal Lemma or equivalent applications of Szemerédi's Regularity Lemma.

The results are split into two parts. First, we investigate bipartite H such as complete bipartite graphs and cubes. Second, we move to non-bipartite H. These will include graphs of high density, random graphs, certain 3-connected graphs, as well as wheel graphs.

1.3.1 Bipartite graphs

A crucial aspect of constructing graphs with high running times is to avoid undesired copies of H minus an edge. It is intuitive that the Turán number $\mathrm{ex}(n,H)$ or $\mathrm{ex}(n,H-e)$ for $e \in E(H)$ should in some way restrict the achievable lower bounds because we cannot hope to avoid undesired copies of $H-e$ if the number of steps, and thereby the number of edges added during the process, greatly exceeds $\mathrm{ex}(n,H)$ and thus lots of copies of H start to appear. Indeed we have the following relation between the quantities $M_H(n)$ and $\mathrm{ex}(n,H)$.

Theorem 1.3.1. *Let H be a graph with at least 2 edges. Then we have that*

$$M_H(n) \leq 2 \cdot \mathrm{ex}(n,H).$$

To put Theorem 1.3.1 into context we recall the following two fundamental results of extremal graph theory. The first is the famous Erdős-Stone-Simonovits Theorem [39, 37] which determined the Turán number of any non-bipartite graph up to lower order terms.

Theorem 1.3.2 (Erdős-Stone 1946, Erdős-Simonovits 1966)**.** *Let $\chi(H)$ denote the chromatic number of a graph H. For every H we have that*

$$\mathrm{ex}(n,H) = \left(1 - \frac{1}{\chi(H)-1}\right) \cdot \binom{n}{2} + o(n^2).$$

This tells us that for non-bipartite H Theorem 1.3.1 merely yields an estimate of $M_H(n) \leq \frac{1}{2}\binom{n}{2} + o(n^2)$ at best. This is not significantly better than the trivial bound $\binom{n}{2}$. Therefore Theorem 1.3.1 provides meaningful bounds only if H is bipartite. The second fundamental result we need for our purposes is the Kővari-Sós-Turán Theorem [62] which gives an asymptotic upper bound on the Turán number of any complete bipartite graph.

Theorem 1.3.3 (Kővari-Sós-Turán, 1954)**.** *Let $r,s \in \mathbb{N}$ with $2 \leq r < s$. Then*

$$\mathrm{ex}(n, K_{r,s}) \leq \frac{1}{2}(s-1)^{\frac{1}{r}} \cdot n^{2-\frac{1}{r}} + \frac{1}{2}(r-1) \cdot n.$$

When combined with the Kővari-Sós-Turán Theorem and the fact that Turán numbers are monotone with respect to taking subgraphs, Theorem 1.3.1 results in the following bound on $M_H(n)$ for all bipartite H.

Corollary 1.3.4. *Let H be a bipartite graph such that the two partite sets of H have size r and s, respectively, where $1 \leq r \leq s$. Then*

$$M_H(n) = O(n^{2-\frac{1}{r}}).$$

While Corollary 1.3.4 gives a general upper bound for bipartite graphs it does not tell us for which graphs, if for any, that bound is asymptotically best possible. In the case $H = K_{2,s}$, $s \geq 3$ we can find the following improvement on the extremal number bound.

Theorem 1.3.5. *For every $s \geq 3$,*

$$M_{K_{2,s}}(n) = \Theta(n).$$

For complete bipartite graphs $K_{r,s}$ with $3 \leq r \leq s$ the subquadratic upper bound of Corollary 1.3.4 can be partially complemented by a superlinear lower bound obtained from a construction in the spirit of Theorem 1.1.8. The idea of associating undesired copies of H minus an edge with non-trivial solutions to linear equations also has a bipartite instance. Here the role of the Behrend constructions is played by K-fold Sidon sets, a generalisation of Sidon sets that builds on the concept of k-fold Sidon sets, which in turn were introduced by Lazebnik and Verstraëte in [73] to study hypergraphs of girth five. We will introduce those sets in more detail in Chapter 3.

Theorem 1.3.6. *Let $3 \leq r \leq s$. The maximum running time $M_{K_{r,s}}(n)$ is bounded from below via*

$$M_{K_{r,s}}(n) \geq n^{3/2-o(1)}.$$

Another example of a bipartite graph whose running time lies strictly between linear and quadratic is the three-dimensional cube.

Theorem 1.3.7. *The running time of the cube Q_3 is bounded from below by $M_{Q_3}(n) = \Omega(n^{3/2})$ and from above by $M_{Q_3}(n) = O(n^{8/5})$.*

In the next section we will encounter more graphs H satisfying both $M_H(n) = \omega(n)$ and $M_H(n) = o(n^2)$, and we will see a general criterion one of whose consequences is that most bipartite graphs fall into this asymptotic range.

1.3.2 Non-bipartite graphs

The crucial property of K_r, $r \geq 6$, in [13] that lead to $M_{K_r}(n) = \Theta(n^2)$ was not that any two vertices are adjacent but the fact that the minimum degree is large. What is the smallest minimum degree a graph H can have that still allows us to use the construction introduced in [13] for complete graphs?

Theorem 1.3.8. *Let H be a graph such that $v(H) \geq 6$ and $\delta(H) > \frac{3}{4}v(H)$. Then*

$$M_H(n) \geq (1 - o(1))\frac{n^2}{4v(H)^2}. \tag{1.6}$$

For $v(H) \in \{6,7,8\}$ the only H with $\delta(H) > \frac{3}{4}v(H)$ are the complete graphs K_6, K_7, and K_8. Therefore the first case where Theorem 1.3.8 gives something new is when $v(H) = 9$.

So far all graphs playing the role of H either were specifically chosen small graphs or belonged to commonly encountered families of graphs such as cycles and complete graphs. Another direction is to ask about the asymptotic behaviour of the running time when H is chosen at random according to the random graph model $G(k, p)$ where $p = p(k)$ is a function of k. While it is customary to use the letter n to denote the number of vertices, this letter is already reserved for the order of the starting graph in bootstrap processes. For this reason we will denote the order of the random graph by k. Thorough treatments of random graphs can be found in [50, 24]. Note that $M_H(n)$ is not monotone in H. For example K_4 plus a pendent vertex has constant maximum running time whereas $M_{K_4}(n)$ is linear. Therefore the property of having at least a certain asymptotic running time is not a monotone graph property, that is, in general it is not preserved under the addition of edges. In this section and the previous one we have seen various types of running times such as constant, logarithmic, linear or quadratic. Surprisingly, it turns out that random graphs have with high probability either constant or quadratic running time.

Theorem 1.3.9. *Let H be distributed as $G(k, p)$. Then with high probability as $k \to \infty$,*

- *$M_H(n) \leq 3$ if $p = o\left(\frac{\log k}{k}\right)$ and n is sufficiently large in terms of k,*
- *$M_H(n) = \Theta(n^2)$ if $p = \omega\left(\frac{\log k}{k}\right)$.*

Note that the first part of this theorem simply comes from the fact that for $p = o(\log(k)/k)$, $G(k, p)$ contains w.h.p. an isolated edge (for a proof of this fact see [24]). It is a short exercise to show that $M_H(n) \leq 3$ if H has an isolated edge and $n \geq 2v(H)$. Theorem 1.3.9 has two interesting consequences. First, by setting $p = 1/2$ we obtain that as $v(H)$ tends to infinity almost all graphs H have $M_H(n) = \Theta(n^2)$. Second, choosing, say, $p = \log(k)^2/k$ shows that a high edge density is not a necessary condition for quadratic running time.

Given the consequences of Theorem 1.3.9 we are particularly interested in finding non-bipartite graphs with both superlinear and subquadratic running time, and determining the graph properties responsible for such asymptotic behaviour.

While we do not resolve Conjecture 1.1.6 the following theorem gives further evidence towards the conjectured subquadratic upper bound in the sense that there exist graphs H for which $M_H(n)$ is subquadratic and has at least the order of magnitude on the right hand side of (1.2).

Theorem 1.3.10. *Let $k \geq 7$ be odd. The wheel graph W_k satisfies $M_{W_k}(n) = o(n^2)$ and $M_{W_k}(n) \geq n^{2-O(1/\sqrt{\log n})}$.*

The approach to Theorem 1.3.10 is inspired by the discussion on K_5 at the end of [25], which suggests to use the Triangle Removal Lemma (see for example [48]) to show that it is not possible to add a quadratic number of edges during the K_5-process. Currently the wheel graphs above are the only family of graphs H known to satisfy $M_H(n) = o(n^2)$ and $M_H(n) \geq n^{2-o(1)}$.

We conclude this section with a criterion for superlinear running time whose role is similar to that of Theorem 1.2.6 for sublinear running times.

Definition 1.3.11. A graph H is called *inseparable* if $H - e$ is 3-connected for each $e \in E(H)$, that is, H cannot be disconnected by removing two vertices and an edge. We call H *bipartite-inseparable* if it is bipartite and for each $e \in E(H)$, one cannot disconnect $H - e$ by removing at most one vertex from each partite set.

We remark that being bipartite-inseparable is weaker than being both bipartite and inseparable. For example, $K_{3,3}$ is bipartite-inseparable but not inseparable since removing two vertices from the same side and an arbitrary edge from the remaining star results in a disconnected graph.

Theorem 1.3.12. *If H is an inseparable or bipartite-inseparable graph, we have that*

$$M_H(n) = \Omega\left(n^{1+1/\left\lceil \frac{6v(H)-7}{4} \right\rceil}\right).$$

On the other hand, if there exists $e \in E(H)$ such that $H - e$ is not 3-connected, and $\langle H \rangle_H$ is a complete graph, then

$$M_H(n) = O(n).$$

Note that the first part of the theorem does not require H to be non-bipartite. The connectivity condition in Theorem 1.3.12 is sharp in the sense that for infinitely many k there exists a k-vertex graph H and an edge $e \in E(H)$ such that $M_H(n) = \Theta(n)$ and for $e' \in E(H) \setminus \{e\}$, $H - e'$ is 3-connected. One such graph is depicted in Figure 1.3. Unfortunately, that figure together with our previous examples of graphs with superlinear running time shows that we cannot hope for a degree-based criterion like the one for sublinear running times.

In the random setting the connectivity criterion of Theorem 1.3.12 provides lower bounds for some random graph models other than $G(k, p)$. For example, when $d \geq 4$, a uniformly chosen random d-regular graph on n vertices, with nd even for divisibility reasons, is with high probability d-connected (cf. Section 7.6 of [24]) and thus has superlinear running time. Furthermore for $p(k) = \omega(\log(k)/k)$ the random bipartite graph $G(k, k, p)$ is w.h.p. 4-connected. Hence w.h.p. $M_{G(k,k,p)}(n) = \omega(n)$.

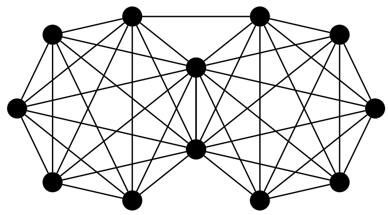

FIGURE 1.3: A graph H with linear running time (cf. Theorem 1.3.12) such that $H - e$ is 3-connected for all but one $e \in E(H)$. The cliques of size seven can be replaced by larger cliques to obtain a graph of higher order with the same above properties. Note that the H-process on H results in $K_{v(H)}$ after a single step.

We shall prove the results introduced in this section and discuss further directions of research in Chapter 3.

1.4 Strong B_h-sets

Extremal constructions from additive combinatorics play an important role in finding lower bounds on the maximum running time of several families of graphs. Indeed, the lower bound in Theorem 1.1.8 and 1.3.10 come from the largest size of a 3-AP-free set in $[n]$ while Theorem 1.3.7 and 1.3.6 rely on large Sidon sets or k-fold Sidon sets. In this section and Chapter 4 we focus our attention on another generalisation of Sidon sets, namely *α-strong B_h-sets*, where $0 \le \alpha < 1$ is a real parameter. These sets combine the concept of a B_h-set with the concept of an α-strong Sidon set. Let us explain those terms. Recall that a Sidon set in the integers is a set $A \subset \mathbb{Z}$ in which any four elements $x_1, x_2, x_3, x_4 \in A$ satisfy

$$|x_1 + x_2 - (x_3 + x_4)| \ge 1 \tag{1.7}$$

unless $\{x_1, x_2\} = \{x_3, x_4\}$. A B_h-set in $\mathbb{Z}$ is a subset $S \subset \mathbb{Z}$ such that for any $x_1, \ldots, x_h \in S$ and $y_1, \ldots, y_h \in S$ with $\{x_1, \ldots, x_h\} \ne \{y_1, \ldots, y_h\}$ one has

$$|x_1 + \ldots + x_h - (y_1 + \ldots + y_h)| \ge 1. \tag{1.8}$$

One of the most studied extremal problems regarding Sidon sets and B_h-sets is to determine their maximum size (in the finite setting of subsets of $[n]$ or finite Abelian groups) or their asymptotic growth (in the infinite setting). In the latter case one is interested in the *counting function* of a set $A \subset \mathbb{Z}$ given by

$$A(n) := |A \cap [n]|.$$

A thorough exposition of the literature on Sidon sets up to the year 2004 can be found in [81]. A more recent article by Eberhard and Manners [35] deals with the structure of dense Sidon sets and offers a unifying perspective on many of the known constructions. We will not go into details on Sidon sets in this text. However, to give a meaningful context for the results on α-strong B_h-sets we collect some of the most important results in that area of study. In the finite case Erdős and Turán [40] showed that for each $\varepsilon > 0$ and sufficiently large n the size of a largest Sidon set in $[n]$ lies between $(\frac{1}{\sqrt{2}}-\varepsilon)\sqrt{n}$ and $\sqrt{n}+O(n^{1/4})$. As to the growth of an infinite Sidon set S one easily obtains an upper bound of $O(\sqrt{n})$ by considering the intersection $S \cap [n]$. Erdős, in a letter to Stöhr [97], proved the stronger bound

$$\liminf_{n\to\infty} \frac{S(n)}{\sqrt{n}} = 0, \tag{1.9}$$

which provided a clear distinction between the finite and the infinite setting. On the lower bound side the best known constructions come from Ruzsa [86] and Cilleruelo [29]. Both achieved the same lower bound $A(n) \geq n^{\sqrt{2}-1+o(1)}$. Ruzsa used a random construction based on digit representation of $\log(p)$ for primes p, whereas Cilleruelo gave a deterministic contruction. By introducing a random argument Cilleruelo, in the same article, also extended his construction to B_h-sets.

Theorem 1.4.1 (Cilleruelo 2014)**.** *Let $h \geq 2$. There exists a B_h-set $S \subset \mathbb{N}$ satisfying*

$$S(n) \geq n^{\sqrt{(h-1)^2+1}-(h-1)+o(1)}.$$

In [65] Kohayakawa, Lee, Moreira, and Rödl introduced the concept of an strong Sidon set as a tool for studying Sidon sets in certain sparse random subsets of $\mathbb{Z}$. Those are sets given by

$$|x_1+x_2-(x_3+x_4)| \geq \max\{x_1^\alpha, x_2^\alpha, x_3^\alpha, x_4^\alpha\} \tag{1.10}$$

for $\max\{x_1,x_2\} \neq \max\{x_3,x_4\}$.

Note that it is important that the right hand side of (1.10) depends on the elements appearing on the left hand side. Indeed, increasing the right hand side of (1.7) by merely a constant factor $\lambda > 0$ does not introduce a relevant new concept since for any Sidon set $S \subset \mathbb{Z}$ the dilate λS satisfies

$$|x_1+x_2-(x_3+x_4)| \geq \lambda$$

unless $\max\{x_1,x_2\} \neq \max\{x_3,x_4\}$, and vice versa any set $S \subset \mathbb{Z}$ satisfying (1.4) is also a Sidon set.

Intuitively an α-strong Sidon set is a Sidon set in which one is allowed to displace each element a little without losing the Sidon property. The larger the element the more it may be moved around. The parameter α determines how quickly the allowed displacement grows with the size of the displaced element. The effect on finding large Sidon sets in random infinite subsets of the integers is roughly as follows: Pick a suitable α-strong Sidon set S and show that with high probability a random set R contains many elements that are close to elements of S. Using the displacement property one can then show that the elements of R close to S form a Sidon set.

As shown in [66] the greedy construction $S = \{s_1, s_2, \ldots\}$, which starts with $s_1 := 1$ and defines s_k to be the smallest positive integer such that $\{s_1, \ldots, s_k\}$ is an α-strong Sidon set, yields

$$S(n) \geq \frac{1}{2} n^{(1-\alpha)/3}. \tag{1.11}$$

On the other hand the authors of [66] extended the upper bound of $O(\sqrt{n})$ for Sidon set in n that is obtained by comparing the number of formal sums with the number of possible values of sums extends to the α-strong setting.

Theorem 1.4.2 (Kohayakawa, Lee, Moreira, Rödl, 2021)**.** *Let $0 \le \alpha < 1$. There exists a constant $c = c(\alpha)$ such that every α-strong Sidon set $S \subset \mathbb{N}$ satisfies*

$$S(n) \le n^{(1-\alpha)/2} \tag{1.12}$$

for sufficiently large n.

Recall that Kohayakawa et al. defined α-strong Sidon sets for their study of random subsets of the integers. They were interested in the following quantities: Given a real parameter $0 < \delta \le 1$ and a random subset R_δ of $\mathbb{N}$ given by choosing each element m independently with probability $p_m := \frac{1}{m^{1-\delta}}$, what is the largest real number $f(\delta)$ such that with probability tending to 1 there exists a Sidon set S in R_δ satisfying $S(n) \ge n^{f(\delta)-o(1)}$. Furthermore what is the smallest constant $g(\delta)$ such that with probability tending to 1 every Sidon set satisfies $S(n) \le n^{g(\delta)+o(1)}$. The statement of their main result is quite elaborate and long, which is why we will not print it here in full but refer to the original source [66]. The crucial ingredient in providing bounds on $f(\delta)$ is the construction of a strong Sidon set whose counting function grows a quickly as possible.

Theorem 1.4.3 (Kohayakawa, Lee, Moreira, Rödl, 2021)**.** *For every $0 \le \alpha \le 10^{-4}$, there exists an α-strong Sidon set $S \subset \mathbb{N}$ such that*

$$S(n) \ge n^{(\sqrt{2}-1+o(1))/(1+32+\sqrt{\alpha})} \tag{1.13}$$

Compare the general case of this theorem and the greedy bound (1.11) to the original setting of Sidon sets, that is, the case $\alpha = 0$. Due to the dependence on α, the bound (1.13) does not necessarily beat the greedy construction anymore. Indeed, the two constructions yield the same asymptotic bound when α is about 5.75×10^{-5}, and (1.13) beats (1.11) for smaller α whereas for larger α it is the other way around. The construction used by Kohayakawa et al. is a black box approach, that is, they provide a general construction that, given a Sidon set A, produces an α-strong Sidon set S with counting function (1.13). The advantage of that approach is that any improvement on the growth of an infinite Sidon set automatically gives a corresponding improvement in the α-strong setting. The disadvantage is that one cannot use properties individual to the underlying Sidon set that might improve the obtained lower bound.

The purpose of our work is to generalise the notion of strong Sidon sets to B_h-set for arbitrary $h \ge 2$, and for the rest of Section 1.4 as well as in Chapter 4 we concentrate on those sets.

Definition 1.4.4 (Strong B_h-sets)**.** Let $0 \le \alpha < 1$. A set $A \subset \mathbb{Z}$ is called an *α-strong B_h-set* if

$$|(x_1 + \ldots + x_h) - (y_1 + \ldots + y_h)| \ge \max\{x_1^\alpha, y_1^\alpha, \ldots, x_h^\alpha, y_h^\alpha\} \tag{1.14}$$

for any $x_1, y_1, \ldots, x_h, y_h \in A$ satisfying $\max\{x_1, \ldots, x_h\} \neq \max\{y_1, \ldots, y_h\}$.

Remark 1.4.5. If $\alpha = 0$ one recovers the original notion of a B_h-set. The additional condition $\max\{x_1,\dots,x_h\} \neq \max\{y_1,\dots,y_h\}$ is necessary to avoid trivial counterexamples where the maximums cancel and hence the right hand side of (1.14) can be made arbitrarily big (when A is infinite) by choosing $x_h = y_h$ large while the right hand side just consists of small fixed $x_1,\dots,x_{h-1}$ and $y_1,\dots,y_{h-1}$. Note that (1.14) also applies to differences of sums with less than h summands. Indeed, given $x_1,y_1,\dots,x_s,y_s \in A$ for some $s \in [h]$ one can simply set $x_i := \min A$ for $s < i \leq h$ and use (1.14). When Kohayakawa et al. first introduced α-strong Sidon sets, they imposed the condition $x_2 < x_3 \leq x_4 < x_1$ instead of the more symmetric $\max\{x_1,x_2\} \neq \max\{x_3,x_4\}$. This results in a slightly weaker definition compared to the above for $h = 2$. We believe that our symmetric definition is the more natural one when it comes to the case $h \geq 3$ as there is no obvious ordering of $x_1,y_1,\dots,x_h,y_h$.

1.4.1 An infinite strong B_h-set of integers

The results presented in this section deal with infinite B_h-sets and are joint work with Juanjo Rué and Christoph Spiegel [46]. As pointed out above, Theorem 1.4.3 uses any given infinite Sidon set as a black box. Another approach, which we follow here, relies on the known construction of Cilleruelo's established in [29]. Indeed, choosing that construction as a base point allows us to build infinite α-strong B_h-sets with non-trivial growth functions for any $h \geq 2$.

Theorem 1.4.6. *For every real $0 \leq \alpha < 1$ and every integer $h \geq 2$ there exists an α-strong B_h-set $S \subset \mathbb{N}$ satisfying*

$$S(n) \geq n^{\sqrt{(h-1+\frac{\alpha}{2})^2+1-\alpha}-(h-1+\frac{\alpha}{2})+o(1)}. \tag{1.15}$$

We remark that a discussion of Theorem 1.4.6 for the case of α-strong Sidon sets (i.e. $h = 2$) can be found in Spiegel's thesis [94]. The arguments for that case are completely deterministic whereas the general case requires the use of randomness.

Compare (1.15) to the greedy bound $S(n) = \Omega(n^{\frac{1-\alpha}{2h-1}})$. One has

$$\sqrt{\left(h-1+\frac{\alpha}{2}\right)^2+1-\alpha}-\left(h-1+\frac{\alpha}{2}\right) > \frac{1-\alpha}{2h-1} \tag{1.16}$$

for all $h \geq 2$ and $0 \leq \alpha < 1$. To see this, one can, for fixed h, turn (1.16) into a quadratic equation looking for the value of α (in all reals) for which the right hand side and the left hand side coincide. In the resulting equation, $\alpha = 1$ is a root with multiplicity two. Therefore Theorem 1.4.6 always beats the greedy construction, though the difference of the exponents diminishes as α tends to 1.

Just like the simple double-counting upper bound $O(n^{1/2})$ for infinite Sidon sets can be generalised to $O(n^{(1-\alpha)/2})$ for infinite α-strong Sidon sets (as done in [66]), it is possible to extend the upper bound $O(n^{1/h})$ for infinite B_h-sets to the α-strong setting.

Theorem 1.4.7. *For every $0 \leq \alpha < 1$ and each $h \geq 2$ there exists a constant $c = c(\alpha, h)$ such that for every α-strong B_h-set $S \subset \mathbb{N}$,*

$$S(n) \leq c \cdot n^{\frac{1-\alpha}{h}} \tag{1.17}$$

Finally we transfer the result of Theorem 1.4.6 to the problem of finding α-strong B_h-sets in sparse random subsets of the integers along the lines of the results of Kohayakawa et al. [65].

Theorem 1.4.8. *For any $h \geq 2$ and $0 < \delta \leq 1$ there exists, with probability 1, a B_h-set A in the infinite random set R_δ satisfying*

$$A(n) \geq n^{\sqrt{(h-1+\frac{1-\delta}{2})^2+\delta} \, - \, (h-1+\frac{1-\delta}{2})+o(1)}. \tag{1.18}$$

The proofs of Theorems 1.4.6, 1.4.7, and 1.4.8 along with a few remarks on further problems are presented in Chapter 4.

1.5 Fairly split matchings

Theorems 1.3.6 and 1.3.10 rely on constructions from additive combinatorics to provide lower bounds on the maximum running time $M_H(n)$ when H is a wheel or a complete bipartite graph. In Chapter 3 we will see that the former uses certain types of Sidon sets whereas the latter involves sets free of three-term arithmetic progressions. One has to translate these arithmetic properties to suitable graph properties to make them applicable. The resulting graph properties are that in every edge-colouring of the graph H minus an edge one can find certain non-monochromatic cycles. Those properties are instances of the following more general type of problem: Given a graph G, a family $\mathscr{G}$ of subgraphs of G, and a proper edge-colouring χ of G, under which conditions can we find a rainbow member of $\mathscr{G}$, or more generally, a member of $\mathscr{G}$ with a given colour distribution?

In this section we consider the question above when G is a regular graph that can be decomposed into perfect matchings, and $\mathscr{G}$ is the family of all matchings in G. This particular question was asked by Arman, Rödl, and Sales [11] in 2021.

Question 1.5.1 ([11], Question 1.1)**.** Let G be a graph on $2n$ vertices whose edge set is the union of k pairwise disjoint perfect matchings $M_1, \ldots, M_k$. For which tuples $(a_1, \ldots, a_k)$ of non-negative integers with $a_1 + \ldots + a_k \leq n$ can we always (that is, no matter what the k initial matchings are) find a new matching M in G such that $|M \cap M_i| \geq a_i$ for all $i \in [k]$?

There are two famous instances of Question 1.5.1 when G is bipartite, $k \in \{n-1, n\}$ and $a_1 = \ldots = a_k = 1$. These are Ryser's Conjecture [88] if n is odd and $a_1 + \ldots + a_n = n$, and the Brualdi-Stein Conjecture [27, 96] if $a_1 + \ldots + a_n = n - 1$. See [60] for a short and concise overview of the latter, including the several approximate versions that have been established over the past decades. The same article also contains the currently best approximate result, which we state below.

Theorem 1.5.2 (Keevash-Pokrovskiy-Sudakov-Yepremyan, 2022)**.** *For any decomposition of the edges of $K_{n,n}$ into perfect matchings $M_1,\dots,M_n$ one can find a set $I \subseteq [n]$ of size $O(\frac{\log n}{\log\log n})$ and a matching M such that $|M \cap M_i| = 0$ for $i \in I$ and $|M \cap M_i| = 1$ for $i \in [n] \setminus I$.*

Note that the statement of this theorem cannot be phrased as a special case of Question 1.5.1 because in the latter the a_i are ordered whereas the former does not tell us for which $i \in [n]$ the intersection $M \cap M_i$ is non-empty. Ryser's conjecture does not encompass the case when n is even because there exists an elegant and simple construction coming from addition tables of cyclic groups (cf. [55] and [99] Theorem 2) that does not admit any rainbow perfect matchings.

The Ryser-Brualdi-Stein Conjecture is an example of Question 1.5.1 when $k = n-1$. As to their original question Arman et al. obtained the following relaxed answer for the case when k is at least a constant factor away from n.

Theorem 1.5.3 (Arman-Rödl-Sales, 2021)**.** *For any real $0 < \varepsilon < 1$ there exists $n_0 \in \mathbb{N}$ such that for $n \geq n_0$ the following holds. Let $a_1,\dots,a_k$ be positive integers with $a_1 + \ldots + a_k < (1-\varepsilon)n$ and either $\max_{i\in[k]} \leq n^{1-\varepsilon}$ or $\min_{i\in[k]} a_i \geq n^{\varepsilon}$. Then there exists an integer $\ell \in [k,(1+\varepsilon)k]$ such that for any ℓ pairwise disjoint matchings $M_1,\dots,M_\ell$ of K_{2n} one can find $I \subseteq [\ell]$ of size k and a matching M satisfying $|M \cap M_i| \geq a_i$ for all $i \in [k]$.*

From this theorem they deduce the following partial answer to their question.

Corollary 1.5.4 ([11], Corollary 5.1)**.** *For any real $\alpha > 0$ and $\varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$ the following holds. Let $a_1,\dots,a_k \in \mathbb{N}_0$ such that $a_1 + \ldots + a_k \leq (1-\varepsilon)n$ and $\min_{i\in[k]} a_i \geq \alpha n$. Then for any k pairwise disjoint perfect matchings $M_1,\dots,M_k$ of K_{2n} there exists a matching M satisfying $|M \cap M_i| \geq a_i$ for $i \in [k]$.*

A similar type of question has been investigated by Aharoni, Alon, and Berger [2] where they studied independence complexes of graphs. Recall that a simplicial complex $\mathscr{C}$ is a hypergraph with the property that $e \in E(\mathscr{C})$ implies $f \in E(\mathscr{C})$ for all $f \subseteq e$ The collection of independent sets of a given graph G form a simplicial complex, the *independence complex* of G. A *matching complex* is the collection of matchings in a given graph or equivalently the independence complex of its line graph.

Theorem 1.5.5 (Aharoni-Alon-Berger, 2016)**.** *If G is the line graph of a graph and $V_1,\dots,V_k$ is a partition of $V(G)$ then there exists an independent set S such that $|S \cap V_i| \geq \left\lfloor \frac{|V_i|}{\Delta(G)+2} \right\rfloor$ for every $i \in [k]$.*

In the language of splitting matchings and for the case of graphs that are the union of disjoint perfect matchings, Theorem 1.5.5 can be reformulated as follows: If $a_i \leq \lfloor \frac{n}{2k} \rfloor$ for all $i \in [k]$ then a matching M with $|M \cap M_i| \geq a_i$ for $i \in [k]$ exists.

In the following we are particularly interested in the case of Question 1.5.1 when k is small and the sum $a_1 + \ldots + a_k$ is at most an additive constant away from n. The latter restriction is important because for a fixed k, Theorem 1.5.3 with $\varepsilon < 1/k$ always gives $\ell = k$ and thus the desired matching M exists whenever $a_i \geq n^{\varepsilon}$ for all i or $a_i \leq n^{1-\varepsilon}$ for all i.

1.5.1 Non-realisable splits and almost arbitrary splits of three matchings

The results presented below constitute joint work with Michael Anastos, Alp Müyesser, and Tibor Szabó [10].

In the context of Question 1.5.1 we refer to M as a *split* with *multiplicities* $(a_1,\dots,a_k)$. We call the split *fair* if $a_1 = \dots = a_k$, and *perfect* if $a_1 + \dots + a_k = n$.

A construction similar to the one that ruled out perfect, fair splits for even n can be used to answer Question 1.5.1 when $a_1 + \dots + a_k = n$ and one is neither in the situation of Ryser's conjecture nor in the situation that a_i is n for precisely one $i \in [k]$ and zero for all others (in which case one can just take the initial matching corresponding to the nonzero-coordinate as M).

Proposition 1.5.6. *Let $k,n \in \mathbb{N}$ and $a_1,\dots,a_k \in [0,n-1]$ such that $a_1 + \dots + a_k = n$. If $k < n$, or n is even, or $\min_{i\in[k]} a_i = 0$, then we can find pairwise disjoint perfect matchings $M_1,\dots,M_k$ on a common vertex set of size $2n$ such that there exists no matching $M \subseteq M_1 \cup \dots \cup M_n$ satisfying $|M \cap M_i| = a_i$ for $i \in [n]$.*

Since according to Proposition 1.5.6 arbitrary perfect splits are not always possible, it is natural to ask how close to a perfect split one can get. The following theorem tells us that if $k = 3$ one can always realise a split with multiplicities (a_1,a_2,a_3) where $a_1 + a_2 + a_3 \le n-2$.

Theorem 1.5.7. *Let $n \in \mathbb{N}$, and let $a_1,a_2,a_3 \in [0,n]$ such that $a_1 + a_2 + a_3 \le n-2$. For any three edge-disjoint perfect matchings M_1, M_2, and M_3 on a common vertex set of size $2n$ there exists a matching $M \subseteq M_1 \cup M_2 \cup M_3$ such that*

$$|M \cap M_1| = a_1 \qquad , \qquad |M \cap M_2| = a_2 \qquad , \qquad |M \cap M_3| = a_3. \tag{1.19}$$

We prove Proposition 1.5.6 and Theorem 1.5.7 in Chapter 5. In that chapter we will also discuss several generalisations of those two results, potential extensions to the case $a_1 + a_2 + a_3 \le n-1$, as well as related directions of research.

Chapter 2

Sublinear running times in graph bootstrap percolation

This chapter collects the proofs of our results related to sublinear running times and introduced in Section 1.2, and is organised as follows. In Section 2.1 we develop several simple but important auxiliary results that we will frequently use troughout the remaining sections of this chapter. In Section 2.2 we prove the claimed constant upper bound on the running time for forests (Theorem 1.2.1). The proof of the precisely values of $M_{C_k}(n)$ is the most elaborate in this thesis and split over three sections. We set up the strategy of the proof in Section 2.3. Rather than splitting the proof according to the parity of k we have a separate theorem that consists of the upper bounds for both even and odd k as well as a theorem that collects the extremal constructions needed to prove the lower bounds both for odd and even k. The former is presented in Section 2.5, the latter in Section 2.4. We then move to the proof Theorem 1.2.4 in Section 2.6. In Section 2.7 we prove Theorem 1.2.6. Section 2.8 contains the proof of Proposition 1.2.8. We conclude the chapter with a discussion of open questions concerning sublinear running times in general and some aspects of our proofs in particular.

In the following, given a graph H and the H-boostrap process $(G_t)_{t\geq 0}$ on a graph G we say that a copy H' of H is *completed* by an edge *e at time t* if $H' \subseteq G_t$, $e \in E(H') \setminus E(G_{t-1})$, and $H' - e \subseteq G_{t-1}$. If the time t is clear from context or not important we simply say that H' is completed by e. Note that this notion only covers those copies that are directly obtained from the definition of the bootstrap process. There might be copies of H that are contained in G_t for some $t \geq 1$ such that G_{t-1} misses at least two of their edges.

More generally, whenever the bootstrap process is clear from context we say that a property holds *at time t* if G_t has that property.

2.1 Auxiliary results

Let us begin with three trivial results which we are going to use throughout the text. First, any H-process on any graph G on less than $v(H)$ vertices is just a constant sequence, hence $M_H(n) = 0$ for $n < v(H)$. Therefore all starting graphs we consider will have at least $v(H)$ vertices. Second, $M_H(n) \geq 1$ for every H and $n \geq v(H)$ since the process on $H - e$ for any $e \in E(H)$ together with

$n - v(H)$ isolated vertices needs at least one step before it stabilises. For this reason whenever we have to show an upper bound that holds for all G we will not deal with H-stable starting graphs and assume that $\tau_H(G) \geq 1$. Third, if H has an isolated vertex, that is, there exists a graph $\tilde{H}$ such that $H \cong \tilde{H} \sqcup K_1$, then $M_H(n) = M_{\tilde{H}}(n)$ for all $n \geq v(H)$. This follows from the simple fact that for any $e \in E(\tilde{H})$, any copy of $\tilde{H} - e$ in a graph of order at least $v(H)$ can be extended to a copy of $H - e$ by an arbitrary additional vertex. Vice versa if a copy of H is completed by an edge e we can remove an isolated vertex from that copy to see that the same edge completes a copy of $\tilde{H}$. The last of the three claims can be applied once for each isolated vertex of H and thereby allows us to ignore those isolated vertices when determining $M_H(n)$ for $n \geq v(H)$.

Observe that if a graph G contains a copy of some H minus an edge as a subgraph then so does every supergraph of G. Therefore graph bootstrap processes behave well with respect to taking subgraphs. In fact the following more general observation holds.

Observation 2.1.1. Let $\varphi : G' \to G$ be an injective graph homomorphism, and let $(G_t)_{t\geq 0}$, $(G'_t)_{t\geq 0}$ be the respective H-processes on G and G'. Then $\varphi \in \mathrm{Hom}(G'_t, G_t)$ for every $t \geq 0$.

Proof. The claim holds for $t = 0$ because $G_0 = G$, $G'_0 = G'$. Let $t \geq 0$ and suppose that $\varphi \in \mathrm{Hom}(G'_{t-1}, G_{t-1})$. Let $e \in E(G'_t) \setminus E(G'_{t-1})$. There exists $H_t \subseteq G'_t$ such that $H_t \cong H$ and $H_t - e \subseteq G'_{t-1}$. We have $\varphi(H_t) - \varphi(e) = \varphi(H_t - e)$ because φ is injective, and $\varphi(H_t - e) \subseteq G_{t-1}$ since $\varphi \in \mathrm{Hom}(G'_{t-1}, G_{t-1})$. Thus, $\varphi(e) \in E(G_t)$ by definition of the H-process on G. $\square$

If $G' \subseteq G$ are nested graphs, the restriction of the identity map $\mathrm{id} : V(G) \to V(G)$ to $V(G')$ is an embedding of G' into G, and hence Observation 2.1.1 indeed implies $G'_t \subseteq G_t$ for all $t \geq 0$. Another consquence is that for any bootstrap process $(G_t)_{t\geq 0}$ one has $\mathrm{Aut}(G_t) \subseteq \mathrm{Aut}(G_{t+1})$ for all $t \geq 0$.

The observation above is useful because it turns out that in order to show that a graph G percolates in the H-process it is often sufficient to look at the H-process on a suitable subgraph G' and show that G' percolates.

Our last result in this section is that bipartiteness is preserved throughout an H-process provided that H is itself bipartie and 2-edge-connected.

Lemma 2.1.2. *Let H be a 2-edge-connected bipartite graph. If G is a bipartite graph with partite sets $X, Y \subset V(G)$, so is $\langle G \rangle_H$.*

Proof. Let $(G_t)_{t\geq 0}$ be the H-process on G, and suppose for a contradiction that the final graph was not bipartite. Pick the smallest t for which G_t contains an edge e whose endpoints lie in the same part, and let H' be a copy of H completed by e. As H is 2-edge-connected there exists a path P of length $k-1$ between the endpoints of e in $H' - e$. By minimality of t, G_{t-1} is bipartite with partite sets X,Y, and thus P must alternate between vertices in X and vertices in Y. Since $e \subseteq X$ or $e \subseteq Y$, P must be even. But then adding e to H' would close an odd cycle which contradicts the assumption that H is bipartite. $\square$

2.2 Trees and forests

Before starting with the actual proof of Theorem 1.2.1 let us consider two motivating examples.

Example 2.2.1 (Graphs with a pendent path of length two)**.** Suppose that H has a vertex w of degree one that is adjacent to a vertex v of degree two. Let $(G_t)_{t\geq 0}$ be the H-process on a graph G that is not H-stable. There exists a copy H_1 of H in G_1 and a vertex $w_1 \in V(H_1)$ such that $d_{H_1}(w_1) = 1$ and the unique H_1-neighbour v_1 of w_1 has degree two in H_1. Let u_1 be the other H_1-neighbour of v_1. Given an arbitrary $w' \in V(G) \setminus V(H_1)$, $H_1 - w_1$ and w' form a copy of $H - vw$ where v_1 and w' play the roles of v and w, hence $v_1 w' \in E(G_2)$. At time 2, $u_1 w'$ is the only missing edge of the copy

$$\big(V(H_1) \setminus \{w_1\} \cup \{w'\}\,,\, E(H_1) \setminus \{v_1 w_1, u_1 v_1\} \cup \{v_1 w', u_1 w'\}\big)\,.$$

In the latter copy the role of v played by w' while v_1 plays the role of the leaf w. This means that for each $w' \in V(G) \setminus V(H_1)$ there exists a copy of H in G_3 in which w' is the unique neighbour of a leaf. Therefore, at time 4 each such w' is adjacent to all but at most $v(H) - 2$ vertices of G. If n is sufficiently large (in our case a suitable polynomial in $v(H)$ is enough) we can thus find a clique K of size at least $v(H) - 1$ in G_4. Each vertex in K is universal at time 5 since we can arbitrarily embed H minus a leaf in a clique so for each $v' \in K$ and $w' \notin K$, $v'w'$ completes a copy of H in which w' plays the role of a leaf. In G_5 every pair of non-adjacent vertices can be extended to a copy of $H - vw$ using $v(H) - 2$ universal vertices, so $G_6 = K_n$. As G was an arbitrary graph with $\tau_H(G) > 0$ we obtain $M_H(n) \leq 6$.

In the introduction we have mentioned that paths have constant maximum running time. This is an immediate corollary of the above. In Example 2.2.1 it was a small induced subgraph that determined the behaviour of the running time. The resulting constant bound on $M_H(n)$ did not depend on H. The next example shows that $M_H(n)$ may depend linearly on the order of H.

Example 2.2.2 (The bootstrap processes of stars)**.** Let $s \in \mathbb{N}$. In the $K_{1,s}$-process $(G_t)_{t\geq 0}$ on a graph G every vertex that has degree $s-1$ at time t will be universal at time $t+1$. For this reason G_{s-1} is either $K_{1,s}$-stable or has at least $s-1$ universal vertices. In the latter case every vertex of G_{s-1} has degree at least $s-1$ and hence will be universal in G_s. Therefore,

$$M_{K_{1,s}}(n) \leq s \tag{2.1}$$

If n is at least $2s$, we have equality in (2.1): Let G be a bipartite graph with $V(G) = X \cup Y$ where $X = \{x_1, \ldots, x_s\}$ and $Y = \{y_1, \ldots, y_s\}$ are two djisjoint s-element sets and

$$E(G) = \big\{x_i y_j : i \in [s-2], j \in [s-1], i \leq j\big\} \cup \{x_{s-1} y_s\}.$$

Note that x_i will be universal at time i for $i \in [s]$ since $d_G(x_i) = s - i$. Therefore, G_{s-1} has $s-1$ universal vertices, which implies $G_s = K_{2s}$. Let U_t be the set of universal vertices in G_t for $t \in [s]$. Then $N_{G_t}(v) = N_G(v) \cup U_t$ for $v \in V(G) \setminus U_t$. Since Y is an independent set in G we have $Y \cap U_t \neq \varnothing$ only if $|N_{G_{t-1}}(y) \cap X| \geq s-1$ for some $y \in Y$. This implies $U_t = \{x_1, \ldots, x_t\}$

for $t \in [s-2]$, and $U_{s-1} = \{x_1, \dots, x_{s-1}, y_s\}$ so the bootstrap process has not stabilised at time $s-1$. We have seen that $M_{K_{1,s}}(2s) = \tau_{K_{1,s}}(G) = s$. To extend this to arbitrary n we just observe that adding isolated vertices to G does not decrease the running time of the $K_{1,s}$-process.

In the following we will frequently build copies of a forest by taking an already existing copy and replacing one of the vertices by a new one. For that purpose we shall use the following notation: Given a copy $F_0 \subset G$ of a forest F in a graph G and vertices $x \in V(F_0)$, $y \in V(G) \setminus V(F_0)$ we define the graph $F_0^{(x\to y)}$ via

$$\begin{aligned} V(F_0^{(x\to y)}) &= V(F_0) \setminus \{x\} \cup \{y\} \\ E(F_0^{(x\to y)}) &= E(F_0) \setminus \{e \in E(F_0) : x \in e\} \cup \{yz : z \in N_{F_0}(x)\}. \end{aligned}$$

We can think of $F_0^{(x\to y)}$ as the graph obtained by replacing x with y in F_0. Note that $F_0^{(x\to y)}$ is a subgraph of G if and only if $N_{F_0}(x) \subseteq N_G(y)$. If $\varphi : V(F) \to V(G)$ is an embedding of F into G with $\varphi(F) = F_0$ we write $\varphi^{(x\to y)}$ for the map

$$V(F) \to V(F_0) \setminus \{x\} \cup \{y\}\ ,\ v \mapsto \begin{cases} \phi(v) & , \phi(v) \neq x; \\ y & , \phi(v) = x. \end{cases}$$

We are now ready to prove Theorem 1.2.1. Although the theorem is stated for forests it is instructive for a first read-through to think of F as a tree rather than a forest with multiple components. There is no significant difference between the two cases.

As we have already dealts with stars in Example 2.2.2, in the following we assume F is not a star. This assumption will be important only at the end of the proof. Further, following the discussion at the beginning of Section 2.1, we assume that F does not contain any isolated vertices. This does not cause any problems because the right hand side of (1.4) is monotone in k. Moreover let G be a graph on n vertices which maximises $\tau_F(G)$ among all n-vertex graphs, that is, $\tau_F(G) = M_F(n)$. Let s be the number of components of F, and fix internal (that is, no leaves) vertices $z_1, \dots, z_s \in F$ such that they come from s different components. Consider F as a rooted tree with roots $\mathbf{z} := (z_1, \dots, z_s)$. In the following we make use of a certain vertex cover which is specified in the claim below:

Claim 2.2.3. *There exists a subset $U \subset V(F)$ such that*

(1) U is a smallest vertex cover of F.

(2) No vertex in U is a leaf of F.

(3) A vertex $v \in V(F)$ lies in U if and only if it has a child that is not contained in U.

Proof. For $u \in V(F)$ let $\mathrm{dist}_F(u, \mathbf{z}) := \mathrm{dist}_F(u, z_j)$ where $j \in [s]$ is the unique index such that z_j is the root of the component containg u. Pick a smallest vertex cover U that minimises $\sum_{u\in U} \mathrm{dist}_F(u, \mathbf{z})$ among all smallest vertex covers of F. Property (1) is satisfied by the choice of U. Property (2) is a direct consequence of (3) since leaves do not have children. The if direction of Property (3) holds because U is a vertex cover. The only if direction of (3) will be achieved by contradiction: Suppose that $v \in U$ and all its children lie in U. Note that this includes the case when v is a leaf. If v is a root we can remove it from U to arrive at a smaller vertex cover and hence at a contradiction. Suppose that $v \notin \{z_1, \dots, z_s\}$, and let w be the parent vertex of v. The set $U \setminus \{v\} \cup \{w\}$ is another smallest vertex cover since every edge involving v is covered by either w or a child of v. However,

$$\begin{aligned}\sum_{u \in U \setminus \{v\} \cup \{w\}} \operatorname{dist}_F(u, \mathbf{z}) &= \sum_{u \in U} \operatorname{dist}_F(u, \mathbf{z}) \;-\; \operatorname{dist}_F(v, \mathbf{z}) + \operatorname{dist}_F(w, \mathbf{z}) \\ &= \sum_{u \in U} \operatorname{dist}_F(u, \mathbf{z}) \;-\; 1,\end{aligned}$$

which contradicts the minimality of U. Therefore Property (3) must hold. □

Choose a set $U \subset V(F)$ as given by Claim 2.2.3. Let

$$\mu := M_{F[U]}(|U|). \tag{2.2}$$

Note that since U is a smallest vertex cover of F, removing all but one of the vertices in U from F results in the disjoint union of a star and some isolated vertices. Let

$$\delta := \min_{u \in U} |N_F(u) \setminus U|$$

be the smallest number of neighbours outside U a vertex in U can have. In other words, δ is the size of the smallest (non-empty) star that can remain if all vertices of U but one are removed.

Let $(G_t)_{t \geq 0}$ be the F-process on G. If there are $|U| + \delta - 2$ universal vertices in G_t for some t and x,y are two distinct, non-adjacent vertices in G_t then xy is the only missing edge in a copy of $K_{1,\delta}$ with centre x whose vertices are x,y and $\delta - 1$ universal vertices. The remaining $|U| - 1$ universal vertices together with any $t - |U| - \delta$ vertices from $V(G)$ other than x, y and the $|U| + \delta - 2$ universal vertices can be used to extend the copy of $K_{1,\delta}$ to a copy of F. As x and y were arbitrary, G_{t+1} is a complete graph. For this reason we will show that unless the F-process stabilises within the first

$$t^* := 2 + 3\lceil \mathrm{ht}_{\mathbf{z}}(F)/2 \rceil + \mu + \max\{\delta, 2\} \tag{2.3}$$

steps, G_{t^*} contains $|U| + \delta - 2$ universal vertices. Suppose $E(G_{t+1}) \setminus E(G_t) \neq \varnothing$ for $0 \leq t \leq t^*$. Fix a copy $F_1 \subseteq G_1$ of F, an isomorphism $\phi_1 : F \to F_1$, and set $U_1 := \phi_1(U)$. Let us rewrite t^* in the following slightly more cumbersome form:

$$t^* = \left(2 + 2\left\lceil \frac{\mathrm{ht}_{\mathbf{z}}(F)}{2} \right\rceil\right) + (\mu + 1) + \left(1 + \left\lceil \frac{\mathrm{ht}_{\mathbf{z}}(F)}{2} \right\rceil\right) + \max\{\delta - 2, 0\} \tag{2.4}$$

The summands on the right hand side correspond to different stages of the process:

(i) At time $2 + 2\left\lceil \frac{\mathrm{ht}_{\mathbf{z}}(F)}{2} \right\rceil$ there will be a complete bipartite graph between U_1 and $V(G) \setminus U_1$.

(ii) After at most $\mu + 1$ more steps we can find a copy F_2 of F and and isomorphism $\phi_2 : F \to F_2$ such that $\phi_2(U) \neq U_1$.

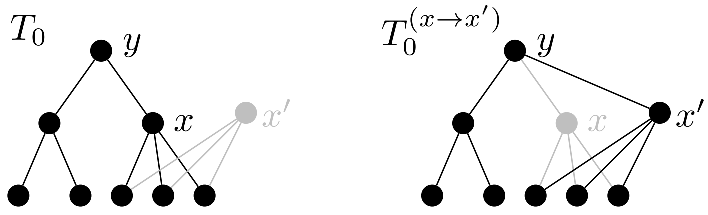

FIGURE 2.1: The situation of Observation 2.2.4 for a concrete choice of F_0. For simplicity we suppose that F_0 is a tree T_0. On the left we have the copy T_0 (black vertices and edges) at some time t. The vertex x' is adjacent to all T_0-neighbours of x but y. On the right we see the copy $T_0^{(x\to x')}$ that is completed by $x'y$ at time $t+1$.

(iii) After at most $1+\left\lceil \frac{\mathrm{ht}_{\mathbf{z}}(F)}{2} \right\rceil$ further steps U_1 will be a set of universal vertices.

If $\delta \leq 2$ then $|U| \geq |U| + \delta - 2$ so we are done after the first three stages. For $\delta > 2$ we need to consider one more stage.

(iv) Once there is a set of $|U|$ universal vertices, $\delta - 2$ additional universal vertices will occur within at most $\delta - 2$ steps.

A useful approach to showing that two given vertices $u, w \in V(G)$ become adjacent in the F-process is to start with a copy of F containing u but not w and replace a single vertex with w to obtain a copy of F minus an edge which is complemented by uw. We formalise this approach in the observation below.

Observation 2.2.4. Let $F_0 \subset G_t$ be a copy of F for some $t \in \mathbb{N}$, and let $x \in V(F_0)$ and $x' \in V(G) \setminus V(F_0)$. If $y \in N_{F_0}(x)$ such that all F_0-neighbours of x but y are adjacent to x' at time t, then $x'y \in E(G_{t+1})$ because it completes $F_0^{(x\to x')}$.

A pictorial description of Observation 2.2.4 is given in Figure 2.1. The following claim is a generalisation of the observation that the neighbour of a leaf in a copy of F becomes almost universal after one more step. It tells us that in a copy of F all vertices which correspond to elements of U become adjacent to every vertex outside the copy of F after a number of steps that only depends on F:

Claim 2.2.5. *Let $t_0 \in \mathbb{N}$ and let $F_0 \subset G_{t_0}$ be a copy of F with an isomorphism $\phi : F \to F_0$. At time $t_0 + \lceil \mathrm{ht}_{\mathbf{z}}(F)/2 \rceil$, every vertex in $\phi(U)$ is adjacent to every vertex in $V(G) \setminus V(F_0)$.*

Proof. Let $w \in V(G) \setminus V(F_0)$. We show by induction on t that $\phi(u)w \in E(G_{t_0+t})$ for every $0 \leq t \leq \lceil \mathrm{ht}_{\mathbf{z}}(F)/2 \rceil$ and every $u \in U$ with $\mathrm{ht}_{\mathbf{z}}(u) \in \{2t-1, 2t\}$. This claim holds vacuously for $t = 0$ because U does not contain any leaves (cf. property (2) of Claim 2.2.3).

Given $u \in U$ with $\mathrm{ht}_{\mathbf{z}}(u) \in \{2t-1, 2t\}$ where $t \geq 1$, there exists a child v of u that is not contained in U by Claim 2.2.3 (3). As U is a cover, all children of v lie in U, so their images under ϕ are adjacent to w at time $t_0 + t - 1$ due to the induction hypothesis. Observation 2.2.4 with $x = \phi(v)$, $x' = w$, $y = \phi(u)$ implies $\phi(u)w \in E(G_{t_0+t})$. $\square$

We are going to apply Claim 2.2.5 to different copies of F that occur during the process to obtain a set of $|U|$ universal vertices.

Stage (i): Because of Claim 2.2.5, every vertex in U_1 is adjacent to every vertex in $V(G) \setminus V(F_1)$ at time $1+\lceil \mathrm{ht}_{\mathbf{z}}(F)/2\rceil$. Recall that U_1 is a vertex cover of F_1 and thus every F_1-neighbour of any $w \in V(F_1) \setminus U_1$ is an element of U_1. Therefore, for any $w \in V(F_1) \setminus U_1$ and $w' \in V(G) \setminus V(F_1)$, F_1 satisfies the conditions of Observation 2.2.4 with $x = w$ and $x' = w'$ so the copy $F_1^{(w \to w')}$ is contained in $G_{2+\lceil \mathrm{ht}_{\mathbf{z}}(F)/2\rceil}$. Claim 2.2.5 applied to $F_1^{(w \to w')}$ and $w \in V(F_1) \setminus U_1$ and a fixed w' guarantees that in $G_{2+2\lceil \mathrm{ht}_{\mathbf{z}}(F)/2\rceil}$, w is adjacent to all vertices in U_1 and so the sets U_1 and $V(G) \setminus U_1$ are the partite sets of a complete bipartite graph.

Stage (ii): For $t > 2+2\lceil \mathrm{ht}_{\mathbf{z}}(F)/2\rceil$ all edges in $E(G_t) \setminus E(G_{t-1})$ have either both their endpoints in U_1 or both their endpoints in $V(G) \setminus U_1$. We want to choose a copy F_2 of F and an isomorphism $\phi_2 : F \to F_2$ such that F_2 is completed by an edge e_2 at time t_2 for some $t_2 \in \mathbb{N}$ with $3+2\lceil \mathrm{ht}_{\mathbf{z}}(F)/2\rceil \le t_2 \le 3+2\lceil \mathrm{ht}_{\mathbf{z}}(F)/2\rceil + \mu$ and $\phi_2(U) \ne U_1$. It is not obvious that such a copy exists: Suppose that for every t_2 in the range above and every choice of F_2, ϕ_2, e_2 one had $\phi_2(U) = U_1$. In that case e_2 has an endpoint in U_1 and hence must be fully contained in U_1. Then $F_2[U_1]$ is a copy of $F[U]$ in G_{t_2} with $F_2[U_1] - e_2 \subset G_{t_2-1}$. Moreover, any copy of $F[U]$ minus an edge in $G_{t_2-1}[U_1]$ can be extended to a copy of F minus an edge using an arbitrary set of $t - |U|$ vertices from $V(G) \setminus U_1$. Therefore, the graphs $G_{3+2\lceil \mathrm{ht}_{\mathbf{z}}(F)/2\rceil+t}[U_1]$, $0 \le t \le \mu$, are the first $\mu+1$ elements of the $F[U]$-process on $G_{3+2\lceil \mathrm{ht}_{\mathbf{z}}(F)/2\rceil}[U_1]$. This however contradicts (2.2) because $3+2\lceil \mathrm{ht}_{\mathbf{z}}(F)/2\rceil + \mu < t^*$ and thus the $\mu+1$ graphs above are distinct. Now that we have seen that the desired choices of t_2, F_2, ϕ_2 exist, fix one such choice.

Stage (iii): Pick $u \in U$ such that $\phi_2(u) \notin U_1$ and $\mathrm{ht}_{\mathbf{z}}(u)$ is minimised among all vertices in $U \setminus \phi_2^{-1}(U_1)$. Let v be a child of u in F that does not lie in U, which exists by Claim 2.2.3 (3). By minimality of $\mathrm{ht}_{\mathbf{z}}(u)$, the children of $\phi_2(v)$, if there are any, lie in U_1 and thus are adjacent to all vertices in $V(G) \setminus U_1$. Consequently, Observation 2.2.4 with $x = \phi_2(v)$, $x' = w$ for $w \in V(G) \setminus (U_1 \cup V(F_2))$ and $y = \phi_2(u)$ implies $F_2^{(\phi_2(v) \to w)} \subset G_{t_2+1}$ for every $w \in V(G) \setminus (U_1 \cup V(F_2))$ so $\phi_2(u)$ is adjacent to every vertex in $V(G) \setminus (U_1 \cup V(F_2))$ at time t_2+1. Choose an arbitrary set $W \subset V(G) \setminus (U_1 \cup V(F_2))$ of size $t - |U_1|$ and a map $\phi_3 : V(F) \to U_1 \cup W$ such that $\phi_3(u) = \phi_1(u)$ for all $u \in U$. Since $V(F) \setminus U$ is an independent set and all edges between U_1 and $V(G) \setminus U_1$ are present at time t_2, in particular those between $\phi_2(u)$ and U_1, $F_3 := \phi_3(F)$ is a copy of F in G_{t_2+1} that lies in the G_{t_2+1}-neighbourhood of $\phi_2(u)$. Then for every $u' \in U_1$, we can replace u' by $\phi_2(u)$ in F_3 to obtain another copy of F, that is $F_3^{(u' \to \phi_2(u))} \subset G_{t_2+1}$. Now Claim 2.2.5 implies that at time $t_2+1+\lceil \mathrm{ht}_{\mathbf{z}}(F)/2\rceil$ every $u' \in U_1$ is adjacent to every vertex in $\phi_3^{(u' \to \phi_2(u))}(U) = U_1 \setminus \{u'\} \cup \{\phi_2(u)\}$ and hence is a universal vertex. Recall that $t_2+1+\lceil \mathrm{ht}_{\mathbf{z}}(F)/2\rceil \le 4+3\lceil \mathrm{ht}_{\mathbf{z}}(F)/2\rceil + \mu$. We have shown that U_1 is a set of universal vertices at time $4+3\lceil \mathrm{ht}_{\mathbf{z}}(F)/2\rceil + \mu$.

Stage (iv): If $\delta > 2$ the remaining $\delta - 2$ universal vertices will be obtained by applying the next claim $(\delta - 2)$ times.

Claim 2.2.6. *Let $t \in \mathbb{N}$, $0 \le \delta^* < \delta - 2$ and assume that G_t has precisely $|U| + \delta^*$ universal vertices. Then there exists $v \in V(G)$ which is universal in G_{t+1} but not in G_t.*

Proof. Denote the set of universal vertices in G_t by U^* and observe that $|U^*| \geq \delta^* + 1$. If $x \in V(G) \setminus U^*$ has at least $\delta - 2 - \delta^*$ G_t-neighbours outside U^* then x together with any $\delta - 2 - \delta^*$ of its neighbours outside U^* and any $\delta^* + 1$ vertices from U^* forms a copy of $K_{1,\delta-1}$ with centre x. By the definition of δ we can, for any $y \in V(G) \setminus N_{G_t}(x)$, use the remaining vertices of U^* and y to extend this star to a copy of F minus an edge such adding xy completes a copy of F. Thus x will be universal at time $t+1$. For this reason, we are going to prove that such a vertex x exists. Suppose it did not, i.e. $|N_{G_t}(x) \setminus U^*| < \delta - 2 - \delta^*$ for each $x \in V(G) \setminus U^*$. Choose a copy F^* of F that is completed at time $t+1$ and an isomorphism $\phi : F \to F^*$. The set $\phi(U) \setminus U^*$ cannot be empty because U is a vertex cover and a universal vertex of G_t cannot be part of an edge in $E(G_{t+1}) \setminus E(G_t)$. Each $x \in \phi(U) \setminus U^*$ satisfies

$$\begin{aligned}
|N_{F^*}(x) \cap U^* \setminus \phi(U)| &= |N_{F^*}(x) \setminus \phi(U)| - |(N_{F^*}(x) \setminus \phi(U)) \setminus U^*| \\
&\geq |N_{F^*}(x) \setminus \phi(U)| - |N_{G_t}(x) \setminus U^*| \\
&\geq \delta - (\delta - 2 - \delta^*) \\
&= \delta^* + 2.
\end{aligned}$$

Thus,

$$\begin{aligned}
|E_{F^*}(\phi(U) \setminus U^*, U^* \setminus \phi(U))| &= \sum_{x \in \phi(U) \setminus U^*} |N_{F^*}(x) \cap U^* \setminus \phi(U)| \\
&\geq (2 + \delta^*) \cdot |\phi(U) \setminus U^*|
\end{aligned}$$

and, since in any forest the number of edges is strictly smaller than the number of vertices,

$$|E_{F^*}(\phi(U) \setminus U^*, U^* \setminus \phi(U))| \leq |\phi(U) \setminus U^*| + |U^* \setminus \phi(U)| - 1$$

Combining the last two estimates results in

$$(1 + \delta^*) \cdot |\phi(U) \setminus U^*| \leq |U^* \setminus \phi(U)| - 1 \tag{2.5}$$

However, we also have

$$|U^* \setminus \phi(U)| + |\phi(U) \cap U^*| = |U^*| = |U| + \delta^*$$

and hence

$$|U^* \setminus \phi(U)| = |U| - |\phi(U) \cap U^*| + \delta^* = |\phi(U) \setminus U^*| + \delta^*. \tag{2.6}$$

The inequalities (2.5) and (2.6) together imply

$$(1 + \delta^*) \cdot |\phi(U) \setminus U^*| < |\phi(U) \setminus U^*| + \delta^*,$$

which is a contradiction as $\phi(U) \setminus U^*$ is non-empty and δ^* is non-negative. $\square$

Since there are at least $|U|$ universal vertices at time $4 + 3\lceil \mathrm{ht}_{\mathbf{z}}(F)/2 \rceil + \mu$, Claim 2.2.6 guarantees the existence of at least $|U| + \delta - 2$ universal vertices at time $2 + 3\lceil \mathrm{ht}_{\mathbf{z}}(F)/2 \rceil + \mu + \max\{\delta, 2\}$

and so, as pointed out in the discussion preceding (2.3), the process stabilises after at most one more step. As G was chosen to maximise $\tau_F(G)$ this yields the estimate

$$M_F(n) = \tau_F(G) \leq 4 + 3\left\lceil \frac{\mathrm{ht}_{\mathbf{z}}(F)}{2} \right\rceil + \mu + \delta \tag{2.7}$$

The next step is to bound the right hand side of (2.7) from above. Recall that the roots $z_1, \ldots, z_s$ were fixed but arbitrary and that both μ and δ depend on U, which depends on $\mathbf{z}$.

Claim 2.2.7. *There exists a component T of F and a choice of roots $\mathbf{z}$ such that* $\mathrm{ht}_{\mathbf{z}}(F) = \lceil \mathrm{diam}(T)/2 \rceil$ *and*

$$\delta \leq \frac{v(T) - 1}{\lceil \mathrm{diam}(T)/2 \rceil} + 2.$$

Proof. Let T be a component of F that maximises $\mathrm{diam}(T)$ among the components of F. Fix a path P of length $\mathrm{diam}(T)$ in T and denote its endpoints by x and y, respectively. Let $z \in V(P)$ such that its distance to x is $\lfloor \mathrm{diam}(T)/2 \rfloor$ and its distance to y is $\lceil \mathrm{diam}(T)/2 \rceil$ Let P_x be the subpath of P from x to z and let P_y be the subpath from z to y. Any path Q from z to a leaf of T intersects either P_x or P_y in exactly the root z. Therefore $Q \cup P_x$ or $Q \cup P_y$ is a path in T. Since any path in T has length at most $\mathrm{diam}(T)$, Q has length at most $\mathrm{diam}(T) - \lfloor \mathrm{diam}(T)/2 \rfloor$. This implies

$$\mathrm{ht}_z(T) = \mathrm{diam}(T) - \lfloor \mathrm{diam}(T)/2 \rfloor = \lceil \mathrm{diam}(T)/2 \rceil.$$

Any vertex cover of T, in particular the set U from Claim 2.2.3 intersected with $V(T)$, must contain at least $\lceil \mathrm{diam}(T)/2 \rceil$ vertices of P. As T does not contain any cycles, P must be an induced path. This allows us to bound δ as follows:

$$\begin{aligned}
\delta &= \min_{u \in U} |N_T(u) \setminus U| \\
&\leq \min_{u \in U \cap V(P)} d_T(u) \\
&\leq \frac{|\{(u,v) : u \in U \cap V(P), v \in V(T), uv \in E(T)\}|}{|U \cap V(P)|} \\
&\leq \frac{2(\mathrm{diam}(T) - 1) + v(T) - (\mathrm{diam}(T) - 1)}{\lceil \mathrm{diam}(T)/2 \rceil} \\
&\leq \frac{v(T) - 1}{\lceil \mathrm{diam}(T)/2 \rceil} + 2
\end{aligned}$$

Denote the components of F by $T_1, \ldots, T_s$ and pick $\mathbf{z} = (z_1, \ldots, z_s)$ with $\mathrm{ht}_{z_j}(T_j) = \lceil \mathrm{diam}(T_j)/2 \rceil$ for all $j \in [s]$. With this choice $\mathrm{ht}_{\mathbf{z}}(F) = \mathrm{ht}_z(T) = \lceil \mathrm{diam}(T)/2 \rceil$. $\square$

Combining Claim 2.2.7 with (2.7) yields

$$M_F(n) \leq 4 + 3\left\lceil \frac{\mathrm{diam}(T)}{4} \right\rceil + \frac{v(T) - 1}{\lceil \mathrm{diam}(T)/2 \rceil} + 2 + \mu \tag{2.8}$$

where T is the component given by Claim 2.2.7. If T is a star, then F must have at least two components and, by the definition of δ, $v(T)$ is at most $k/2$. As a consequence (2.8) yields

$$M_F(n) \leq 7 + \frac{k}{4} + \mu.$$

In the case that T is not a star, its diameter must lie between 3 and $k-1$. Then the term $3\lceil \mathrm{diam}(T)/4\rceil + (k-1)/\lceil \mathrm{diam}(T)/2\rceil$ is bounded from above by $(3k+14)/4$, so

$$M_F(n) \leq 9 + 3\frac{k+2}{4} + \mu.$$

We can bound μ by $\binom{|U|}{2}$, which in turn can be bounded by $k^2/8$ since the vertex cover number of any forest of order k without isolated vertices is at most $k/2$ (this bound is sharp for the path on k vertices). Therefore in both cases,

$$M_T(n) \leq \frac{1}{8} \cdot (k^2 + 6k + 76).$$

This completes the proof of Theorem 1.2.1.

■

2.3 Cycles: Strategy

In this section we present the strategy we follow to prove Theorem 1.2.2. and give a short proof of the second part of the same theorem on starting graphs of small order. As mentioned in the outline of this chapter, we split the proof the first part of Theorem 1.2.2 into an upper bound part (Theorem 2.3.1), where we show that for any graph G the C_k-process on G stabilises after at most the number of steps given in (1.5), and a lower bound part (Theorem 2.3.2) where we give starting graphs whose running times attain the values in (1.5). The proof of the lower bound part is given in Section 2.4 whilst the upper bound part is shown in Section 2.5. In this section we also provide a few auxiliary statements that are necessary to prove Theorems 2.3.1 and 2.3.2 and to combine them into a proof of Theorem 1.2.2.

Theorem 2.3.1 (Upper bound part)**.** *Let $k \geq 3$, and let G be a connected graph on at least $k+1$ and at most n vertices with C_k-process $(G_t)_{t\geq 0}$ such that $\langle G\rangle_{C_k} \neq G$. Define*

$$r = r(n,k) := \begin{cases} \lceil \log_{k-1}(n + k^2 - 4k + 2)\rceil & , k \text{ odd};\\ \lceil \log_{k-1}\left(2n + k^2 - 5k\right)\rceil & , k \text{ even}. \end{cases} \tag{2.9}$$

If n is sufficiently large the following hold:

(i) If k is odd, then $xy \in E(G_r)$ for every distinct $x, y \in V(G)$.

(ii) If k is even and G is bipartite with parts $X, Y \subset V(G)$ then $xy \in E(G_r)$ for each $x \in X$, $y \in Y$.

(iii) For even k and non-bipartite G, we have $xy \in E(G_r)$ for any distinct $x, y \in V(G)$.

By definition of r, (1) and (2), r is the unique natural number satisfying

$$(k-1)^{r-1} - F(k-2,k) \;\leq\; n-1 \;<\; (k-1)^r - F(k-2,k), \tag{2.10}$$

when k is odd. Likewise

$$\frac{(k-1)^{r-1}-(k-1)}{2} - F'(k-2,k) + 2 \leq n < \frac{(k-1)^{r}-(k-1)}{2} - F'(k-2,k) + 2, \tag{2.11}$$

when k is even. To obtain a lower bound of the form $M_{C_k}(n) \geq r$ we need to specify a starting graph G and an edge $e \in \binom{V(G)}{2}$ such that e is present at time r but not at time $r-1$. In view of Theorem 2.3.1 it suffices to give a pair of vertices (from different partite sets if G is bipartite and k is even) that is not adjacent at time $r-1$.

Theorem 2.3.2 (Lower bound part)**.** *Let $k \geq 3$, and let G be a graph with C_k-process $(G_t)_{t\geq 0}$. Define r as in (2.9), and set*

$$\ell = \ell(n,k) := \frac{(k-1)^{r-1}-(k-1)}{2} - F'(k-2,k) - 1 \tag{2.12}$$

when k is even. Then the following hold:

1. *If k is odd and $G = P_n$, then $\{0, (k-1)^{r-1} - F(k-2,k)\} \notin E(G_t)$ for $t < r$.*
2. *If k is even and $G = P^{\Delta}$ (see Figure 2.2) on $\ell + 3 \leq n$ vertices then for the vertices $v_\ell, w_\ell \in V(P^{\Delta})$ we have that $\{v_\ell, w_\ell\} \notin E(G_t)$ for $t < r$.*

The upper bound part requires the starting graph G to be connected. However, in general G might be disconnected. The following observation reduces the running times on disconnected G to running times on connected starting graphs.

Observation 2.3.3. Let G be a graph with connected components $G^{(1)}, \ldots, G^{(s)}$, and let $(G_t)_{t\geq 0}$ be its C_k-process. Then $G_t = G_i^{(1)} \cup \ldots \cup G_i^{(s)}$, and hence

$$\langle G \rangle_{C_k} = \langle G^{(1)} \rangle_{C_k} \cup \ldots \cup \langle G^{(s)} \rangle_{C_k}$$

and

$$\tau_{C_k}(G) = \max\left\{\tau_{C_k}(G^{(1)}), \ldots, \tau_{C_k}(G^{(s)})\right\}.$$

Proof. Suppose that at some step in the process the number of components decreases. Take the smallest i for which there exists an edge $e \in E(G_t)$ whose endpoints lie in distinct components of G. At time $i-1$ there must be path of length $k-1$ between the endpoints of e. This is a contradiction since by the minimality of i the vertex sets of the components of G_{t-1} are precisely those of G. $\square$

Any component with less than k vertices is C_k-stable and thus does not affect the process. Therefore,

$$M_{C_k}(n) = \max\{\tau_{C_k}(G) : G \text{ connected}, k \leq v(G) \leq n\}. \tag{2.13}$$

For even k another graph property that is preserved throughout the process is bipartiteness. This follows from Lemma 2.1.2 since C_k is 2-edge connected.

Given Theorems 2.3.1 and 2.3.2, and Observation 2.3.3 we can deduce the first part of the main theorem.

Proof of Theorem 1.2.2. The upper bound of (2.13) tells us that we can restrict ourselves to connected starting graphs on at most n and at least k vertices. We assume that n is sufficiently large so that Theorem 2.3.1 holds. When k is odd or the starting graph is non-bipartite, then the desired upper bound follows from parts (i) and (iii) of Theorem 2.3.1, which state that by round r our process reaches the complete graph, which is C_k-stable. If k is even and the starting graph is bipartite with parts X and Y, then part (ii) tells us that at time r there is a complete bipartite graph between X and Y, which by Lemma 2.1.2 must be the final graph of the process. To obtain the lower bounds observe that $\ell \leq n-3$ by definition of ℓ and r and that the edges specified in parts (1) and (2) of Theorem 2.3.2 are not present at time $r-1$, but will be added eventually by Theorem 2.3.1 (in fact in the next step). So the process is not finished after $r-1$ steps. □

A crucial ingredient of both the lower and the upper bound part is the aforementioned decrease of the diameter by a factor of $k-1$ in each step.

Lemma 2.3.4. *Let $(G_t)_{t\geq 0}$ be the C_k-process on a graph G, and let $x,y \in V(G)$. For each $i \geq 1$, the distance $\mathrm{dist}_{G_t}(x,y)$ satisfies*

$$\mathrm{dist}_{G_0}(x,y) \leq (k-1)^t \mathrm{dist}_{G_t}(x,y),$$

and

$$\mathrm{dist}_{G_t}(x,y) \leq \left\lfloor \frac{\mathrm{dist}_{G_0}(x,y)}{(k-1)^t} \right\rfloor + k - 2.$$

When $\mathrm{dist}_{G_0}(x,y)$ is a multiple of $(k-1)^t$ the above can be improved to

$$\mathrm{dist}_{G_t}(x,y) \leq \frac{\mathrm{dist}_{G_0}(x,y)}{(k-1)^t}. \tag{2.14}$$

Proof. Observe that for any edge $e \in E(G_t) \setminus E(G_{t-1})$ one can find a path of length $k-1$ between its endpoints in G_{t-1}. Given a shortest xy-path in G_t, replacing every edge on the path which is not present at time $t-1$ by a suitable path of length $k-1$ yields an xy-walk of length at most $(k-1) \cdot \mathrm{dist}_{G_t}(x,y)$ in G_{t-1}. From this we deduce

$$\mathrm{dist}_{G_{t-1}}(x,y) \leq (k-1)\mathrm{dist}_{G_t}(x,y)$$

and thus

$$\mathrm{dist}_{G_0}(x,y) \leq (k-1)^t \mathrm{dist}_{G_t}(x,y).$$

To obtain the upper bound on $\mathrm{dist}_{G_t}(x,y)$ write $\mathrm{dist}_{G_{t-1}}(x,y) = q \cdot (k-1) + r$ for suitable $q, r \in \mathbb{N}_0$, $0 \leq r \leq k-2$, and choose a path $u_0 \dots u_{q(k-1)+r}$ from x to y in G_{t-1}. In G_t, we have that

$u_0 u_{k-1} \dots u_{q(k-1)} \dots u_{q(k-1)+r}$ is a path of length $q+r$ from x to y. Since $r \leq k-2$, we obtain

$$\operatorname{dist}_{G_t}(x,y) \leq q+r \leq \frac{q\cdot(k-1)+r-(k-2)}{k-1}+k-2 = \frac{\operatorname{dist}_{G_{t-1}}(x,y)-(k-2)}{k-1}+k-2. \quad (2.15)$$

We can bound the left hand side by just q whenever $\operatorname{dist}_{G_t}(x,y)$ is divisible by $k-1$. An inductive application of (2.15) yields the claim. □

2.3.1 Small starting graphs

Suppose that $n \geq k \geq 5$. Let G be given by $V(G) = \{v_i : i \in \mathbb{Z}_k\}$ and $E(G) = \{v_i v_{i+1} : i \in \mathbb{Z}_k \setminus \{\ell\}\} \cup \{v_{-\mathbf{1}} v_{\mathbf{1}}\}$ where $\ell := \lfloor \frac{k}{2} \rfloor \cdot \mathbf{1}$, that is, G is a path of length $k-1$ with an additional between two vertices that are not endpoints of the path and whose distance on the path is two.

In the following we refer to any edge that is not of the form $\{v_i v_{i+1}\}$ for some $i \in \mathbb{Z}_k$ as a *chord*. Our goal is to show that the edge $v_{\ell-1}v_{\ell+1}$ lies in $\langle G \rangle_{C_k}$ but is not present at time $\lfloor k/2 \rfloor$. A copy of C_k^- in a k-vertex graph is a Hamilton path. Observe that since $k \geq 5$ we have that $\{\ell, \ell+1\} \cap \{-\mathbf{1}, \mathbf{1}\} = \varnothing$. For this reason, $d_G(v_\ell) = d_G(v_{\ell+1}) = 1$ and so in the first step of the process only the edge $v_\ell v_{\ell+1}$ is added, that is,

$$E(G_1) = \{v_i v_{i+1} : i \in \mathbb{Z}_k\} \cup \{v_{-1} v_1\}. \quad (2.16)$$

Any Hamilton path with endpoints $v_{\ell-1}$, $v_{\ell+1}$ in G_{t-1} for some $t \in \mathbb{N}$ must use a chord that is incident to v_ℓ. Therefore $v_{\ell-1}v_{\ell+1} \notin E(G_t)$ for $t \leq t_\ell$ where

$$t_\ell := \min\{t \in \mathbb{N} : d_{G_t}(v_\ell) > 2\}.$$

We may assume that t_ℓ is well-defined for otherwise we could immediately conclude that $v_{\ell-1}v_{\ell+1}$ does not lie in $E(G_t)$ for any $t \geq 0$. [1] It remains to check that $t_\ell \geq \lfloor k/2 \rfloor$. The next claim tells us that a vertex that is not incident to a chord receives a new neighbour in the process only if one of its neighbours is already incident to a chord.

Claim 2.3.5. *Let $t \geq 1$. If for some $i \in \mathbb{Z}_k$ none of the vertices v_{i-1}, v_i, v_{i+1} is incident to a chord at time t, then v_i is not incident to a chord at time $t+1$.*

Proof. Suppose for a contradiction that v_i is adjacent to v_j for some $j \in V(G) \setminus \{i-1, i+1\}$ at time $t+1$. By assumption the edge $v_i v_j$ is not present at time t. Therefore there exists a Hamilton path P with endpoints v_i and v_j in G_t. Both v_{i-1} and v_{i+1} are internal vertices of P. However since neither v_{i-1} nor v_{i+1} is incident to a chord at time t both of them must be adjacent to v_i in P. But this is a contradiction because v_i is an endpoint of P and as such has only one P-neighbour. □

For $i \in \mathbb{Z}_k \setminus \{\ell\}$ define

$$t_i := \min\{t \in \mathbb{N} : d_{G_t}(v_i) > 2\},$$

[1] This assumption is only necessary because we have not shown the upper bound part, yet. In that part, we will see that a k-cycle with a chord that is not bipartite percolates in the C_k-process, which tells us that t_ℓ is always well-defined.

and observe that by (2.16) one has $t_{\mathbf{1}} = 1$ and $t_i > 1$ for $i \notin \{-\mathbf{1}, \mathbf{1}\}$. We will see in a moment that these numbers are well-defined. They satisfy $t_i = t_{-i}$ for all $i \in \mathbb{Z}_k$ because the map $V(G) \to V(G)$, $v_i \mapsto v_{-i}$ is an automorphism of G_1. As a consequence of Claim 2.3.5 one has

$$t_i > \min\{t_{i+1}, t_{i-1}\} \tag{2.17}$$

for all $i \in \{\mathbf{1}, \ldots, \ell\}$. If k is odd, the relations $\ell + 1 = -\ell$ tells us that $t_\ell > \min\{t_{-\ell}, t_{\ell-1}\} = \min\{t_\ell, t_{\ell-1}\}$ and thus $t_\ell > t_{\ell-1}$. In the even case the relation $\ell - 1 = -(\ell+1)$ yields $t_{\ell+1} = t_{\ell-1}$, so $t_\ell > \min\{t_{\ell+1}, t_{\ell-1}\} = t_{\ell-1}$. Iterating (2.17) gives us

$$t_\ell > t_{\ell-1} > \ldots > t_1 \geq 1$$

We can see that the t_i are well-defined and the desired inequality $t_\ell \geq \lfloor k/2 \rfloor$ follows. Therefore $v_{\ell-1}v_\ell + 1 \notin E(G_{\lfloor k/2 \rfloor})$. Too see that this edge is eventually present we observe that if $v_i v_j \in G_t$ for some $i, j \in \mathbb{Z}_k$ with $j \notin \{i-1, i, i+1\}$ then $v_{i+1}, \ldots, v_j v_i v_{i-1} \ldots v_{j+1}$ is a Hamilton path, and so $v_{i+1}v_{j+1} \in E(G_{t+1})$. By iterating this observation (starting from the edge $v_{-\mathbf{1}}v_{\mathbf{1}}$) and noting that $\lfloor k/2 \rfloor + 1$ and $\lfloor k/2 \rfloor - 1$ have the same parity we obtain $v_{\ell-1}v_{\ell+1} \in E(\langle G \rangle_{C_k})$. This completes the proof of the second part of Theorem 1.2.2.

2.4 Lower bound part for cycles

In this section we prove Theorem 2.3.2. Part (i) of Theorem 2.3.2 is shown in Section 2.4.1, and the proof of Part (ii) can be found in Section 2.4.2.

In both of the latter two parts the following set will be convenient to get a handle on when a pair of vertices is an edge at time t of the C_k-process.

$$A_t := \left\{(k-1)^t - \alpha \cdot (k-2) - \beta \cdot k : \alpha, \beta \in \mathbb{N}_0\right\} \tag{2.18}$$

Note that when k is even, A_t consists of odd numbers while for odd k there is no restriction on the parity. The A_t form a increasing sequence because for any $\alpha, \beta \in \mathbb{N}_0$,

$$(k-1)^t - \alpha(k-2) - \beta k = (k-1)^{t+1} - (\alpha + (k-1)^t) \cdot (k-2) - \beta k \in A_{t+1}.$$

2.4.1 The odd case

Since during any C_k-process the diameter decreases by a factor of $k-1$ in each step, starting graphs with large diameter are natural candidates for obtaining high running times. This is why we look at paths. Let $(P^t)_{t \geq 0}$ be the C_k-process on P_n.

Lemma 2.4.1. *If $xy \in E(P^t)$ for some $x, y \in V(P_n), t \geq 0$ then $y - x \in A_t$.*

Proof. We prove the claim by induction on $t \geq 0$.

$t = 0$: All edges in P^0 are of the form $\{x, x+1\}$ and we can write

$$1 = (k-1)^1 - (k-2) - 0 \cdot k, \qquad -1 = (k-1)^1 - 0 \cdot (k-2) - k.$$

$t \geq 1$: Let $xy \in E(P^t)$. If xy was already present at time $t-1$, the induction hypothesis and the inclusion $A_{t-1} \subset A_t$ give $y - x \in A_t$. Suppose $xy \notin E(P^{t-1})$. Let $v_0, \ldots, v_{k-1}$ be a path from $v_0 := x$ to $v_{k-1} := y$ in P^{t-1}. By the induction hypothesis there exist $\alpha_1, \ldots, \alpha_{k-1}, \beta_1, \ldots, \beta_{k-1}$ such that $v_j - v_{j-1} = (k-1)^{t-1} - \alpha_j \cdot (k-2) - \beta_j \cdot k$ for $j \in [k-1]$. Then

$$y - x = \sum_{j=1}^{k-1} v_j - v_{j-1} = \sum_{j=1}^{k-1} \left((k-1)^{t-1} - \alpha_j \cdot (k-2) - \beta_j \cdot k\right) = (k-1)^t - \sum_{j=1}^{k-1} \alpha_j \cdot (k-2) - \sum_{j=1}^{k-1} \beta_j \cdot k.$$

□

Lemma 2.4.1 assures that whenever $d \in \mathbb{N}$ is an integer that cannot be expressed as $d = \alpha(k-2) + \beta k$ for suitable $\alpha, \beta \in \mathbb{N}_0$ then $(k-1)^t - d$ does not lie in A_t and hence any edge xy with $y - x = (k-1)^t - d$ cannot be present at time t. Therefore the edge $\{0, (k-1)^{r-1} - F(k-2,k)\}$ cannot be present in P^{r-1}. This shows part (1).

2.4.2 The even case

To show part (ii) we have to introduce the graph P^Δ: We assume that n, and thus ℓ, is sufficiently large so we do not run into degenerate cases when defining P^Δ. Let P be a path with vertex and edge sets

$$\begin{aligned} V(P) &= \{v_0, \ldots, v_\ell, w_0, \ldots, w_\ell\}, \\ E(P) &= \{w_\ell w_{\ell-1}, \ldots, w_1 w_0, w_0 v_0, v_0 v_1, \ldots, v_{\ell-1} v_\ell\}, \end{aligned}$$

and consider the map

$$\varphi : V(P) \to V(P^\Delta) \quad , \quad \varphi(v) := \begin{cases} v_j & , \text{ if } v = w_j \text{ for some } j \in [\ell-1]; \\ v & , \text{ otherwise}, \end{cases},$$

which identifies v_j and w_j for $j \in [\ell - 1]$. Define $P^\Delta := \varphi(P)$, i.e.

$$\begin{aligned} V(P^\Delta) &= \{v_0, \ldots, v_\ell, w_0, w_\ell\}, \\ E(P^\Delta) &= \{v_j v_{j+1} : 0 \leq j \leq \ell - 1\} \cup \{v_0 w_0, w_0 v_1, v_{\ell-1} w_\ell\}. \end{aligned}$$

We can think of P^Δ as a graph which maximises the length of a shortest odd walk between two vertices for fixed n.

Let $(P^{\Delta,t})_{t \geq 0}$ be the C_k-process on P^Δ. Recall that our goal is to show $v_\ell w_\ell \notin E(P^{\Delta,r-1})$. To do so we want to set up an analogue of Lemma 2.4.1 for P^Δ. Call an edge $v_j v_{j'}$, $v_j w_{j'}$ or $w_j w_{j'}$ *even* is $j - j'$ is even, and *odd* if $j - j'$ is odd. Lemma 2.4.3 below is the analogue of Lemma

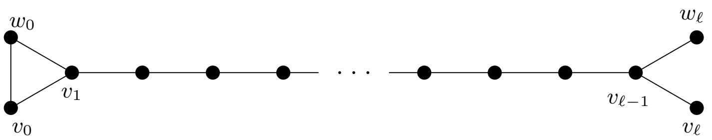

FIGURE 2.2: A visualisation of P^Δ.

2.4.1 dealing with odd edges, while Lemma 2.4.4 deals with the even edges. Both rely on the following auxiliary statement:

Lemma 2.4.2. *For every $t \geq 0$, the largest $j \in [\ell]$ such that v_j or w_j is an endpoint of an even edge in $P^{\Delta,t}$ is at most $(k-1)^t - 1$.*

Proof. The only even edge in $P^{\Delta,0}$ is v_0w_0 so the claim holds for $t = 0$. Let $t \geq 1$ and suppose the claim holds for $t-1$. Since $(k-1)^t - 1 > (k-1)^{t-1} - 1$ it suffices to show that whenever v_j or w_j is the endpoint of an edge in $E(P^{\Delta,t}) \setminus E(P^{\Delta,t-1})$ one has $j \leq (k-1)^t - 1$. Let $u_{j_0}u_{j_{k-1}}$ be an even edge in $E(P^{\Delta,t}) \setminus E(P^{\Delta,t-1})$ and let $u_{j_0} \dots u_{j_{k-1}}$ be a path in $P^{\Delta,t-1}$ such that $u_{j_i} \in \{v_{j_i}, w_{j_i}\}$ for $0 \leq i \leq k-1$. For parity reasons there exists at least one even edge on that path. Let $s \in [k-1]$ such that $j_s - j_{s-1} \equiv 0 \mod 2$. The first part of Lemma 2.3.4 gives

$$\mathrm{dist}_{P^\Delta}(u_{j_i}, u_{j_{i-1}}) \leq (k-1)^{t-1}\mathrm{dist}_{P^{\Delta,t-1}}(u_{j_i}, u_{j_{i-1}}) = (k-1)^{t-1}$$

for $s+1 \leq i \leq k-1$. In P^Δ we have $\mathrm{dist}_{P^\Delta}(u_{j_i}, u_{j_{i-1}}) = |j_i - j_{i-1}|$ whenever $j_i \neq j_{i-1}$. Now the inductive hypothesis implies

$$j_{k-1} = j_s + \sum_{i=s+1}^{k-1} j_i - j_{i-1} \leq (k-1)^{t-1} - 1 + (k-1-s) \cdot (k-1)^{t-1} \leq (k-1)^t - 1.$$

□

Recall the definition of A_t in (2.18).

Lemma 2.4.3. *Let $t \geq 1$ and $j, j' \in [\ell]$ with $j \not\equiv j' \mod 2$. If $u_j \in \{v_j, w_j\}$, $u_{j'} \in \{v_{j'}, w_{j'}\}$ and $u_ju_{j'} \in E(P^{\Delta,t})$, then $j - j' \in A_t$.*

Proof. We induct on t with $t \in \{1, 2\}$ being our base cases.

$t = 1$: Any path of length $k-1$ in $P^{\Delta,0}$ whose endpoints form an edge at time 2 must not use v_0w_0 because of parity, and thus misses at least one of v_0,w_0. This implies $j - j' \in \{-(k-1), -1, 1, (k-1)\} \subset A_1$ (cf. base case of Lemma 2.4.1).

$t = 2$: If $u_ju_{j'}$ is present at time 1 we are done because $A_1 \subset A_2$ and $j - j' \in A_1$ by the induction hypothesis. Suppose that the edge does not lie in $E(P^{\Delta,1})$. Let $Q = u_{j_0} \dots u_{j_{k-1}}$ be a path in $P^{\Delta,1}$ with $j_0 = j$, $j_{k-1} = j'$ and $u_{j_i} \in \{v_{j_i}, w_{j_i}\}$ for $1 \leq i \leq k-2$. There has to be an even number of even edges in Q because $j - j'$ is odd and Q has odd length, and there cannot be more than two because the only even edges of $P^{\Delta,1}$ are v_0w_0, v_0v_{k-2}, w_0v_{k-2} and they form a triangle. If all edges of Q are odd we proceed as in the inductive step of Lemma 2.4.1. Otherwise there are precisely two even edges on Q. These two edges must share a common endpoint considering that the even edges in $P^{\Delta,1}$ form a triangle. Let $s \in [k-1]$ such that $u_{j_{s-1}}u_{j_s}$ and $u_{j_s}u_{j_{s+1}}$ are the even edges. We have either $j_{s-1} = j_{s+1} = 0$ or $\{j_{s-1}, j_{s+1}\} = \{0, k-2\}$. For $i \in [k-1] \setminus \{s, s+1\}$, choose $\alpha_i, \beta_i \in \mathbb{N}_0$ such that $j_i - j_{i-1} = (k-1) - \alpha_i(k-2) - \beta_i k$. This allows us to express $j_{k-1} - j_0$ as follows:

$$\begin{aligned}
j_{k-1} - j_0 &= \sum_{i \in [k-1] \setminus \{s,s+1\}} j_i - j_{i-1} \;+\; j_s - j_{s-1} + j_{s+1} - j_s \\
&= \sum_{i \in [k-1] \setminus \{s,s+1\}} ((k-1) - \alpha_i(k-2) - \beta_i k) \;+\; j_{s+1} - j_{s-1} \\
&= (k-3) \cdot (k-1)^1 - \sum_{i \in [k-1] \setminus \{s,s+1\}} \alpha_i(k-2) - \sum_{i \in [k-1] \setminus \{s,s+1\}} \beta_i k \;+\; j_{s+1} - j_{s-1} \\
&= \begin{cases} (k-1)^2 - \sum_i \alpha_i(k-2) - \sum_i \beta_i k - 2(k-2) - k & \text{, if } j_{s+1} - j_{s-1} = -(k-2); \\ (k-1)^2 - \sum_i \alpha_i(k-2) - \sum_i \beta_i k - (k-2) - k & \text{, if } j_{s+1} - j_{s-1} = 0; \\ (k-1)^2 - \sum_i \alpha_i(k-2) - \sum_i \beta_i k - k & \text{, if } j_{s+1} - j_{s-1} = k-2. \end{cases}
\end{aligned}$$

Therefore $j_j - j_{i-1} \in A_t$, as required.

$t \geq 3$: We handle the case $u_j u_{j'} \in E(P^{\Delta,t-1})$ as before and so assume that $u_j u_{j'}$ is an odd edge not in $P^{\Delta,t-1}$. Let $u_{j_0} \dots u_{j_{k-1}}$ be a $u_j u_{j'}$-path in $P^{\Delta,t-1}$ where $u_{j_i} \in \{v_{j_i}, w_{j_i}\}$ for $1 \leq i \leq k-2$, and let $J := \{i \in [0, k-2] : j_i \equiv j_{i+1} \mod 2\}$. Since $j - j'$ and $k-1$ are odd, $|J|$ must be even. If J is empty, that is, if Q consists of odd edges we can again proceed as in Lemma 2.4.1. Suppose that $|J| \geq 2$ and let $s := \min J$. Lemma 2.4.2 yields $j_s \leq (k-1)^{t-1} - 1$ while the induction hypothesis guarantees $j_i - j_{i+1} \leq \max A_{t-1} = (k-1)^{t-1}$ for $i \notin J$. Therefore,

$$\begin{aligned}
j - j' \leq j_0 = j_s + \sum_{i=0}^{s-1} j_i - j_{i+1} & \\
&\leq (k-1)^{t-1} - 1 + s \cdot (k-1)^{t-1} \\
&\leq (k-1)^{t-1} - 1 + (k-3) \cdot (k-1)^{t-1} \\
&\leq (k-1)^t - (k-1)^{t-1} - 1 \\
&< (k-1)^t - F'(k-2, k).
\end{aligned}$$

The last inequality uses (2) and $t \geq 3$. We now have $j - j' \in A_t$ by (2) and because $(k-1)^t$ and $j - j'$ are odd. □

Lemma 2.4.4. *Let $1 \leq t < r$, and let $j, j' \in \{0, \dots, \ell\}$ such that $j \equiv j' \mod 2$ and $j + j' \geq (k-1)^t - (k-1) - 2 \cdot F'(k-2, k) - 2$. If $u_j \in \{v_j, w_j\}$, $u_{j'} \in \{v_{j'}, w_{j'}\}$ and $u_j u_{j'} \in E(P^{\Delta,t}) \setminus E(P^{\Delta,t-1})$, there exist $\alpha, \gamma \in \mathbb{Z}_{\geq -1}$, $\beta, \delta, \lambda, \mu \in \mathbb{N}_0$ with $\lambda + \mu = (k-1)^{t-1} - 1$ such that*

$$j = \lambda(k-1) - \alpha(k-2) - \beta k \qquad , \qquad j' = \mu(k-1) - \gamma(k-2) - \delta k.$$

Proof. We induct on t.

Base case $t=1$: The only even edges in $E(P^{\Delta,1})\setminus E(P^{\Delta,0})$ are v_0v_{k-2} and w_0v_{k-2}. Both of them satisfy the hypothesis $j+j'\geq (k-1)^1-(k-1)-2\cdot F'(k-2,k)-2$. The claim now holds with either $\alpha=-1$ and $\beta,\gamma,\delta,\lambda,\mu$ equal to zero or $\gamma=-1$ and $\alpha,\beta,\delta,\lambda,\mu$ equal to zero.

Inductive step: Let $Q=u_{j_0}\dots u_{j_{k-1}}$ be a path in $P^{\Delta,t-1}$ such that $j_0=j$, $j_{k-1}=j'$, and $u_{j_i}\in\{v_{j_i},w_{j_i}\}$ for $1\leq i\leq k-2$. We first show that Q has exactly one even edge. The number of even edges in Q is odd for otherwise we have $j\not\equiv j' \mod 2$. If $t=2$ the only three even edges in $P^{\Delta,t-1}$ are v_0w_{k-2}, $v_{k-2}w_0$ and v_0w_0. A path cannot contain all three of them so Q has precisely one even edge. If $t\geq 3$ we proceed as follows: Suppose there are at least three even edges in Q and let $s,s'\in[k-1]$ such that $u_{j_s}u_{j_{s+1}}$ is the first and $u_{j_{s'-1}}u_{j_{s'}}$ is the last even edge in Q. Then $s+(k-1-s')\leq k-4$. By Lemma 2.4.2

$$j_s\leq (k-1)^{t-1}-1 \qquad , \qquad j_{s'}\leq (k-1)^{t-1}-1.$$

Combining this with Lemma 2.4.3 and $\max A_{t-1}=(k-1)^{t-1}$ gives us

$$\begin{aligned}
j+j'=j_0+j_{k-1}&=\sum_{i=0}^{s-1}(j_i-j_{i+1})+j_s+j_{s'}+\sum_{i=s'+1}^{k-1}(j_i-j_{i-1})\\
&\leq 2(k-1)^{t-1}-2+(s+k-1-s')\cdot(k-1)^{t-1}\\
&\leq (k-2)\cdot(k-1)^{t-1}-2\\
&=(k-1)^t-(k-1)^{t-1}-2\\
&<(k-1)^t-(k-1)-2\cdot F(k-2,k)-2,
\end{aligned}$$

which contradicts the assumption $j+j'\geq (k-1)^t-(k-1)-2\cdot F'(k-2,k)-2$. Here we used that $t\geq 3$ and so $(k-1)^{t-1}>2\cdot F'(k-2,k)+(k-1)$ by (2). We have thus shown that Q has precisely one even edge.

Take the unique s^* for which $u_{j_{s^*-1}}u_{j_{s^*}}$ is an even edge. We have

$$\begin{aligned}
j_{s^*-1}+j_{s^*}&=j+j'-\sum_{i=1}^{s^*-1}(j_{i-1}-j_i)-\sum_{i=s^*+1}^{k-1}(j_i-j_{i-1})\\
&\geq (k-1)^t-(k-1)-F'(k-2,k)-2-(k-2)\cdot(k-1)^{t-1}\\
&=(k-1)^{t-1}-(k-1)-F'(k-2,k)-2.
\end{aligned}$$

In particular, the assumption $t\geq 3$ gives $j_{s^*-1}+j_{s^*}>0$. If the edge $u_{j_{s^*-1}}u_{j_{s^*}}$ already appeared at time $t-2$, then by Lemma 2.4.2

$$j_{s^*-1}\leq (k-1)^{t-1}-1\leq (k-1)^{r-2}-1<\ell \quad \text{and} \quad j_{s^*}\leq (k-1)^{t-1}-1\leq (k-1)^{r-2}-1<\ell$$

provided that r is sufficiently large. Therefore $j_{s^*-1}\neq j_{s^*}$ and hence $j_{s^*-1}+j_{s^*}<2(k-1)^{t-2}-2$. This, however, is a contradiction because

$$2(k-1)^{t-2}-2\leq (k-1)^{t-1}-(k-1)-F'(k-2,k)-2.$$

For this reason $u_{j_{s^*-1}}u_{j_{s^*}} \in E(P^{\Delta,t-1}) \setminus E(P^{\Delta,t-2})$. By the induction hypothesis there exist $\alpha^*, \gamma^* \in \mathbb{Z}_{\geq -1}$, $\beta^*, \delta^*, \lambda^*, \mu^* \in \mathbb{N}_0$ such that $\lambda^* + \mu^* = (k-1)^{t-2} - 1$ and

$$j_{s^*-1} = \lambda^*(k-1) - \alpha^*(k-2) - \beta^* k \qquad , \qquad j_{s^*} = \mu^*(k-1) - \gamma^*(k-2) - \delta^* k.$$

By Lemma 2.4.3 we have that whenever $u_{j_i}u_{j_{i-1}}$ is an odd edge we can find $\alpha_i, \beta_i \in \mathbb{N}_0$ such that

$$j_i - j_{i-1} = (k-1)^{t-1} - \alpha_i(k-2) - \beta_i k$$

for $s^* < i \leq k-1$ and

$$j_{i-1} - j_i = (k-1)^{t-1} - \alpha_i(k-2) - \beta_i k$$

for $1 \leq i < s^*$. Therefore,

$$j = \sum_{i=1}^{s^*-1} (j_{i-1} - j_i) + j_{s^*-1} = \lambda(k-1) - \alpha(k-2) - \beta k,$$
$$j' = \sum_{i=s^*+1}^{k-1} (j_i - j_{i-1}) + j_{s^*} = \mu(k-1) - \gamma(k-2) - \delta k,$$

where

$$\begin{aligned}
\lambda &:= (s^*-1)\cdot(k-1)^{t-2} + \lambda^*, & \mu &:= (k-1-s^*)\cdot(k-1)^{t-2} + \mu^*,\\
\alpha &:= \alpha_0 + \ldots + \alpha_{s^*-2} + \alpha^*, & \beta &:= \beta_0 + \ldots + \beta_{s^*-2} + \beta^*,\\
\gamma &:= \alpha_{s^*+1} + \ldots + \alpha_{k-1} + \gamma^*, & \delta &:= \beta_{s^*+1} + \ldots + \beta_{k-1} + \delta^*.
\end{aligned}$$

Moreover,

$$\lambda + \mu = (k-2)(k-1)^{t-2} + \lambda^* + \mu^* = (k-1)^{t-1} - 1,$$

which completes the induction. □

Take the smallest $t_0 \in \mathbb{N}$ for which the even edge $v_\ell w_\ell$ lies in $E(P^{\Delta,t_0})$ and suppose that $t_0 \leq r-1$. Lemma 2.4.2 and (2.12) yield

$$2(k-1)^{t_0} - 2 \geq \ell + \ell = (k-1)^{r-1} - (k-1) - 2 \cdot F'(k-2,k) - 2,$$

and so $t_0 \geq r-1$ when n and thus r is sufficiently large. It remains to rule out the case $t_0 = r-1$. Suppose that $t_0 = r-1$. By Lemma 2.4.4 there exist $\alpha, \gamma \in \mathbb{Z}_{\geq -1}$, $\beta, \delta, \lambda, \mu \in \mathbb{N}_0$ with $\lambda + \mu = (k-1)^{r-2} - 1$ such that

$$\ell = \lambda(k-1) - \alpha(k-2) - \beta k = \mu(k-1) - \gamma(k-2) - \delta k. \tag{2.19}$$

By symmetry we can assume that $\lambda \leq \mu$. From (2.19) and the definition of ℓ we obtain

$$F'(k-2,k) = \left(\frac{(k-1)^{r-2}-1}{2} - \lambda - 1\right) \cdot (k-1) + (\alpha+1) \cdot (k-2) + \beta k, \tag{2.20}$$

$$F'(k-2,k) = \left(\frac{(k-1)^{r-2}-1}{2} - \mu - 1\right) \cdot (k-1) + (\gamma+1) \cdot (k-2) + \delta k. \tag{2.21}$$

If we take (2.20) modulo 2 we can see that

$$\frac{(k-1)^{r-2}-1}{2} - \lambda \equiv 1 \mod 2. \tag{2.22}$$

The condition $\lambda + \mu = (k-1)^{r-2} - 1$ implies

$$\lambda \leq \frac{(k-1)^{r-2}-1}{2} \leq \mu. \tag{2.23}$$

We cannot have equality in (2.23) because of (2.22). Therefore

$$\frac{(k-1)^{r-2}-1}{2} - \lambda \geq 1.$$

Since $2(k-1)$ can be written as $(k-2)+k$, by (2.20) we have

$$F'(k-2,k) = (\alpha + 1 + \tilde{\alpha}) \cdot (k-2) + \left(\beta + \tilde{\beta}\right) \cdot k,$$

where

$$\tilde{\alpha} := \frac{1}{2}\left(\frac{(k-1)^{r-2}-1}{2} - \lambda - 1\right) \quad \text{and} \quad \tilde{\beta} := \frac{1}{2}\left(\frac{(k-1)^{r-2}-1}{2} - \lambda - 1\right).$$

However, this contradicts the definition of $F'(k-2,k)$. Consequently, $v_\ell w_\ell \notin E(P^{\Delta,r-1})$.

2.5 Upper bound part for cycles

We now give the proof of Theorem 2.3.1. Throughout this section we assume that G is always a connected graph on at least $k+1$ and at most n vertices that is not C_k-stable. Let $(G_t)_{t\geq 0}$ be the C_k-process on G.

We start with a couple of general results on C_k-processes in Section 2.5.1, followed by another investigation of the C_k-process on paths in Section 2.5.2. We will prove parts (i) and (ii) of Theorem 2.3.1 in Section 2.5.3. Part (iii) of Theorem 2.3.1 will be shown in Section 2.5.4

2.5.1 General results

Lemma 2.5.1. *Let $\tilde{G}$ be a connected graph with $\tau_{C_k}(G) \geq 1$. Then in the C_k-process on $\tilde{G}$ every vertex contained in a k-cycle at time 2.*

Proof. Suppose that $\tau_{C_k}(\tilde{G}) \geq 2$, and let $(\tilde{G})_{t\geq 0}$ be the C_k-process on $\tilde{G}$. Since $\tau_{C_k}(\tilde{G}) \neq 0$, there exists a k-cycle C in $\tilde{G}_1$. Let $x \in V(\tilde{G}) \setminus V(C)$, and let Q be a shortest path from x to $V(C)$ in $\tilde{G}_1$. If Q has length at least $k-1$ the first k vertices of Q starting from x form a path of length $k-1$ with endpoint x, hence x lies in a cycle at time 2. If the length of Q is smaller than $k-1$ we can extend Q to a path of length $k-1$ using vertices of C. Note that the only vertex of $Q \cap C$ is the other endpoint of Q for otherwise Q would not be a shortest path, so none of the vertices of C we use to extend Q are already in Q. The vertices of the extended path, one of which is x, form a k-cycle in $\tilde{G}_2$. □

Lemma 2.5.2. *Let $k \geq 3$, and let $z, z' \in V(K_{\lfloor k/2 \rfloor, \lceil k/2 \rceil})$ be vertices from the same partite set of $K_{\lfloor k/2 \rfloor, \lceil k/2 \rceil}$. Then $\tau_{C_k}(K_{\lfloor k/2 \rfloor, \lceil k/2 \rceil} \cup \{zz'\}) \leq 2$ and $\langle K_{\lfloor k/2 \rfloor, \lceil k/2 \rceil} \cup \{zz'\}\rangle_{C_k} = K_k$.*

Proof. Let $\tilde{G} := K_{\lfloor k/2 \rfloor, \lceil k/2 \rceil} \cup \{zz'\}$ and denote the partite sets of $K_{\lfloor k/2 \rfloor, \lceil k/2 \rceil}$ by X and Y such that $|X| = \lceil k/2 \rceil$ and $|Y| = \lfloor k/2 \rfloor$. If k is odd, then for any two distinct $x, x' \in X$ we can find a Hamilton path, which has length $k-1$, from x to x' in $K_{\lfloor k/2 \rfloor, \lceil k/2 \rceil}$. Thus X is a clique after one step in the C_k-process on $\tilde{G}$. At time 1, $X \setminus \{x\}$ and $Y \cup \{x\}$ are partite sets of a complete bipartite graph of size $\lfloor k/2 \rfloor$ and $\lceil k/2 \rceil$, respectively. Therefore $Y \cup \{x\}$ is a clique at time 2. This shows the claim for odd k. Now assume that k is even, in particular, $k \geq 4$ so both $|X| \geq 2$ and $|Y| \geq 2$. Since $|X| = |Y|$ we may further assume that $z, z' \in X$. For any distinct $y, y' \in Y$ we can pick a Hamilton path from y to z in the complete bipartite graph $\tilde{G} - y' - z'$ and extend that path to a yy'-path of length $k-1$ in $\tilde{G}$ by zz' and $z'y'$. Then Y must be a clique at time 1. Analogous arguments show that X is a clique after one more step and hence the claim follows. □

Lemma 2.5.3. *Let $\tilde{G}$ be a connected graph of order at least $k+1$ which contains a cycle. The final graph $\langle \tilde{G} \rangle_{C_k}$ is a clique if k is odd or $\tilde{G}$ is non-bipartite, and a complete bipartite graph if k is even and $\tilde{G}$ is bipartite.*

Proof. In $\langle \tilde{G} \rangle_{C_k}$ the endpoints of any path of length $k-1$ are adjacent. Therefore the shortest path between any two vertices has length less than $k-1$. Choose vertices v_j, $j \in [0, k-1]$, in $\tilde{G}$ that form a k-cycle C with edges $v_j v_{j+1}$. Here and for the rest of this proof addition and subtraction in the subscript are always performed modulo $k-1$. Every $x \in V(\tilde{G}) \setminus \{v_0, \ldots, v_{k-1}\}$ has a $\langle \tilde{G} \rangle_{C_k}$-neighbour on C because a shortest path from x to C in $\tilde{G}$ can always be extended to a path of length $k-1$ by vertices of C. If $xv_j \in E(\langle \tilde{G} \rangle_{C_k})$ then $xv_j v_{j-1} \ldots v_{j+2}$ is a path of length $k-1$ so $xv_{j+2} \in E(\langle \tilde{G} \rangle_{C_k})$. In case that k is odd the above implies that every vertex of C is adjacent to every other vertex of $\tilde{G}$ in $\langle \tilde{G} \rangle_{C_k}$. Thus for any two distinct vertices x, y we can find a k-cycle containing x but not y. Repeating the argument above for such a cycle gives $xy \in E(\langle \tilde{G} \rangle_{C_k})$.

Now assume that k is even. Let

$$X := \{v_j : j \equiv 0 \mod 2\} \qquad , \qquad Y := \{v_j : j \equiv 1 \mod 2\}.$$

Then every vertex outside C is adjacent in $\langle \tilde{G} \rangle_{C_k}$ to all vertices in X or all vertices in Y. Define

$$X' := \left\{ z \in V(\tilde{G}) \setminus V(C) : Y \subseteq N_{\langle \tilde{G} \rangle_{C_k}}(z) \right\} \qquad , \qquad Y' := \left\{ z \in V(\tilde{G}) \setminus V(C) : X \subseteq N_{\langle \tilde{G} \rangle_{C_k}}(z) \right\}.$$

One of these two sets, say X', must be non-empty. For any $x \in X'$, $y \in Y'$, $yv_0v_1 \dots v_{k-3}x$ is an xy-path of length $k-1$ in $\langle \tilde{G} \rangle_{C_k}$. Furthermore for any $j, j' \in [0, k-1]$, with $v_j \in X$, $v_{j'} \in Y \setminus \{v_{j-1}, v_{j+1}\}$ and any $x \in X'$

$$v_{j'}v_{j'+1} \dots v_{j-1}xv_{j'-2} \dots v_{j+1}v_j$$

is an $v_jv_{j'}$-path of length $k-1$. Therefore $\langle \tilde{G} \rangle_{C_k}$ contains a complete bipartite graph whose partite sets are $X \cup X'$ and $Y \cup Y'$. If $\tilde{G}$ is bipartite we are done. Otherwise the claim follows from Lemma 2.5.2. □

We remark that Lemmas and 2.5.3 and 2.3.4 already suffice to establish an upper bound of the form $\log_{k-1}(n) + c_k$ for some constant $c_k > 0$.

2.5.2 Results on paths

Let $n' \in \mathbb{N}$, and $(P^i_{n'})_{t \geq 0}$ be the C_k-process on $P_{n'}$. We write P^t instead of $P^t_{n'}$ when n' is clear from context. The sets

$$D_t = D_t(n') := \{\ell \in [n'-1] : xy \in E(P^t) \text{ whenever } y - x = \ell\}$$

plays a central role in proving upper bounds on $\tau_{C_k}(P_{n'})$. Clearly $D_t \subseteq D_{t+1}$. If $D_t = [n'-1]$, then the percolation process is over by the t^{th} step. When k is even then the process is already over when D_t contains just the odd integers up to $n'-1$, since then P^t has stabilised at $K_{\lfloor n/2 \rfloor, \lceil n/2 \rceil}$. Flipping the vertices of $P_{n'}$, i.e. the map $\sigma : V(P_{n'}) \to V(P_{n'}), x \mapsto n' - 1 - x$, is an automorphism of $P_{n'}$ and hence of P^t for all $t \geq 0$ by Observation 2.1.1. For any $x, y \in V(P_{n'})$ one has $x + y \leq n' - 1$ or $\sigma(x) + \sigma(y) \leq n' - 1$. This allows us to write D_t as follows:

$$D_t = \left\{\ell \in [n'-1] : xy \in E(P^t) \text{ whenever } y - x = \ell \text{ and } x + y \leq n' - 1\right\} \tag{2.24}$$

The next couple of lemmas state further simple properties about how D_t develops during the C_k-process.

Lemma 2.5.4. *For every $t \geq 0$, $(k-1)D_t \cap [n'-1] \subseteq D_{t+1}$, where $(k-1)D_t$ is the $(k-1)$-fold sumset of D_t.*

Proof. Let $\ell \in (k-1)D_t \cap [n'-1]$. Choose $d_1, \dots, d_{k-1} \in D_t$ such that $\ell = d_1 + \dots + d_{k-1}$. For any $x \in V(P_{n'-1})$ with $x + \ell \in V(P_{n'-1})$,

$$x, x + d_1, \dots, x + (d_1 + \dots + d_{k-1})$$

is a path of length $k-1$ from x to $x + \ell$ in G_t since $d_1, \dots, d_{k-1} \in D_t$. Thus x and $x + \ell$ are adjacent in G_{t+1} and the claim follows by the definition of D_{t+1}. □

Lemma 2.5.5. *If $n' \geq 3(k-1)$, then $[k]_1 \subset D_2$.*

Proof. We have $D_0 = \{1\}$ and $D_1 = \{1, k-1\}$. For every odd $3 \leq \ell \leq k$ and $x \in V(P_{n'})$ with $x + (x+\ell) \leq n' - 1$,

$$x, \dots, x + \frac{\ell-1}{2}, x + (k-1) + \frac{\ell-1}{2}, x + (k-1) + \frac{\ell-1}{2} - 1, \dots, x + \ell$$

is a path of length $k-1$ from x to $x+\ell$ in P^1 due to the hypothesis $3(k-1) - 1 \leq n' - 1$. Therefore (2.24) gives $[k]_1 \subset D_2$. □

Recall that Lemma 2.4.1 in the proof of Theorem 2.3.2 gave us some idea how differences that occur as edges at time t look like and thereby helped us to obtain lower bounds on the running time. For upper bounds we need a converse statement telling us for which parameters α, β the differences $(k-1)^t - \alpha \cdot (k-2) - \beta \cdot k$ do appear. To this end we define a subset

$$A'_t := \left\{(k-1)^t - \alpha \cdot (k-2) - \beta \cdot k : \alpha, \beta \in \mathbb{N}_0, \alpha + \beta \leq (k-1)^{t-2} \cdot (k-2)\right\}$$

of $A_t = \{(k-1)^t - \alpha \cdot (k-2) - \beta \cdot k : \alpha, \beta \in \mathbb{N}_0\}$. This set lies in the intersection of A_t and the interval $[(k-1)^{t-2}, (k-1)^t]$. The upper end of the interval is attained when $\alpha = \beta = 0$ whereas the lower end is achieved by $\alpha = 0$, $\beta = (k-1)^{t-2} \cdot (k-2)$. The next lemma states that a slightly smaller interval piece of A_t however is already contained in A_t.

Lemma 2.5.6. *For every $t \geq 3$,*

$$[(k-1)^{t-2} + 2(k-1), (k-1)^t] \cap A_t \subseteq A'_t.$$

Proof. Let $t \geq 3$ and $\ell \in [(k-1)^{t-2} + 2(k-1), (k-1)^t] \cap A_t$. Then there exist $\alpha, \beta \in \mathbb{N}_0$ satisfying $\ell = (k-1)^t - \alpha(k-2) - \beta k$. We may assume that $\alpha \leq k-1$ because $(\alpha - k) \cdot (k-2) + (\beta + k - 2)k = \alpha(k-2) + \beta k$. From

$$(\alpha+\beta)k - 2\alpha = (k-1)^t - \ell \leq (k-1)^t - (k-1)^{t-2} - 2(k-1) = (k-1)^{t-2} \cdot (k-2)k - 2(k-1)$$

we infer

$$\alpha + \beta \leq (k-1)^{t-2} \cdot (k-2) + \frac{2\alpha}{k} - \frac{2(k-1)}{k} \leq (k-1)^{t-2} \cdot (k-2),$$

hence $\ell \in A'_t$. □

The next lemma ensures that the relevant piece of A'_t is contained in D_t. This fact will play a crucial role in us showing that a bootstrap process has ended. In the proof the advantage of the somewhat technical choice of the upper bound on $\alpha + \beta$ in the inductive proof becomes visible.

Lemma 2.5.7. *Given $n' \geq 3(k-1)$ we have that for every $t \geq 0$,*

$$A'_t \cap [n'-1] \subseteq D_t.$$

Proof. We induct on t. We have that $A'_0 = \{1\} = D_0$ and $A'_1 = \{k-1\} \subset D_1$. Let $t = 2$, and let $x, y \in V(P_{n'})$ with $y - x = (k-1)^2 - \alpha(k-2) - \beta k$ for some $\alpha + \beta \leq (k-2)$. Write $s :=$

$k-1-\alpha-\beta$ and note that $s \geq 1$. If $x \geq \beta$ then

$$x, x-1, \dots, x-\beta, x-\beta+(k-1), \dots, x-\beta+s(k-1), x-\beta+s(k-1)+1, \dots, y$$

is a path of length $k-1$ from x to $y = x-\beta+s(k-1)+\alpha$ in P^1. If $x \leq \beta$ we define $x_{ij} := x+\alpha+i(k-1)-j$ for $i \in [s], j \in [0,\beta]$ and consider the xy-path

$$x, \dots, x+\alpha, x_{10}, x_{11}, \dots, x_{1\beta}, x_{2\beta}, \dots, x_{s\beta}.$$

This path is well-defined because $x+\alpha+(k-1) \leq \beta+\alpha+(k-1) \leq 2k-3 \leq n'-1$ and $s \geq 1$. Thus, $xy \in E(P^2)$. Since x, y were arbitrary, $A_2' \cap [n'-1] \subset D_2$. For every $t \geq 3$, the induction hypothesis and Lemma 2.5.4 imply

$$A_t' \cap [n'-1] \subseteq (k-1)A_{t-1}' \cap [n'-1] \subseteq (k-1)D_{t-1} \cap [n'-1] \subseteq D_t,$$

where the inclusion $A_t' \subseteq (k-1)A_{t-1}'$ follows from the fact that for any $\alpha, \beta \in \mathbb{N}_0$ with $\alpha+\beta \leq (k-1)^{t-2} \cdot (k-2)$ we can find

$$\alpha_1, \dots, \alpha_{k-1} \in \left\{ \left\lfloor \frac{\alpha}{k-1} \right\rfloor, \left\lceil \frac{\alpha}{k-1} \right\rceil \right\} \quad , \quad \beta_1, \dots, \beta_{k-1} \in \left\{ \left\lfloor \frac{\beta}{k-1} \right\rfloor, \left\lceil \frac{\beta}{k-1} \right\rceil \right\}$$

such that $\alpha = \alpha_1 + \ldots + \alpha_{k-1}$, $\beta = \beta_1 + \ldots + \beta_{k-1}$ and $\alpha_s + \beta_s \leq (k-1)^{t-3} \cdot (k-2)$ for all $s \in [k-1]$. □

Proposition 2.5.8. *If k is odd and $3(k-1) \leq n' \leq (k-1)^{\rho} - F(k-2,k)$ for some integer $\rho \geq 4$ then $P_{n'}^{\rho}$ is the complete graph on n' vertices.*

Proof. Our goal is to show $D_\rho = [n'-1]$. To do so we write $[n'-1]$ as the union

$$[n'-1] = \big([n'-1] \cap [3(k-1)]\big) \cup [3(k-1), n'-1]$$

and show $[n'-1] \cap [3(k-1)] \subseteq D_4$ and $[3(k-1), n'-1] \subseteq D_\rho$. Lemma 2.5.5 yields $k-2 \in D_2$ because $k-2$ is odd. Then for each even $2 \leq \ell \leq k-1$ and each vertex x with $x+(x+\ell) \leq n'-1$,

$$x, \dots, x+\frac{\ell}{2}, x+(k-2)+\frac{\ell}{2}, x+(k-2)+\frac{\ell}{2}-1, \dots, x+\ell$$

is a path of length $k-1$ from x to $x+\ell$ in $P_{n'}^2$. Here it is important that $+(k-2)+\ell/2 \leq n'-1$ so all vertices of the path indeed belong to $P_{n'}$. Now (2.24) implies $[k] \subset D_3$. Applying Lemma 2.5.4 gives

$$D_4 \supseteq [(k-1) \cdot k] \cap [n'-1] \supseteq [3(k-1)] \cap [n'-1].$$

The inclusion $[3(k-1), n'-1] \subset D_\rho$ follows from Lemmas 2.5.6 and 2.5.7: As k is odd we have $A_t \supseteq (-\infty, (k-1)^t - F(k-2,k) - 1]$ for all $t \geq 0$. This allows us to write

$$\begin{aligned}
[3(k-1), n'-1] &= [n'-1] \cap [3(k-1), (k-1)^\rho - F(k-2,k) - 1] \\
&= [n'-1] \cap \bigcup_{t=3}^{\rho} [(k-1)^{t-2} + 2(k-1), (k-1)^t - F(k-2,k) - 1] \\
&\subseteq [n'-1] \cap \bigcup_{t=3}^{\rho} [(k-1)^{t-2} + 2(k-1), (k-1)^t] \cap A_t \\
&\subseteq [n'-1] \cap \bigcup_{t=3}^{\rho} A'_t \\
&\subseteq \bigcup_{t=3}^{\rho} D_t \\
&= D_\rho.
\end{aligned}$$

The second equality holds since $(k-1)^{3-2} + 2(k-1) = 3(k-1)$ and $(k-1)^t - F(k-2,k) \geq (k-1)^{t+1-2} + 2(k-1)$ for $t \geq 3$ by (1). We used Lemma 2.5.6 in the fourth line and Lemma 2.5.7 in the fifth. We have shown that $D_\rho = [n'-1]$, so $P^\rho_{n'}$ is a complete graph. $\square$

The bipartite version of Proposition 2.5.8 reads as follows.

Proposition 2.5.9. *If k is even and $3(k-1) \leq n' \leq (k-1)^\rho$ for some $\rho \in \mathbb{N}$ then any $x, y \in V(P_{n'})$ with $|x-y| \in A_\rho$ are adjacent in $P^\rho_{n'}$. This implies that if $n' \leq (k-1)^\rho - F'(k-2,k)$, $P^\rho_{n'}$ is a copy of $K_{\lfloor n'/2 \rfloor, \lceil n'/2 \rceil}$.*

Proof. Let $x, y \in V(P_{n'})$ such that $|x-y| \in A_\rho$. We want to show that $xy \in E(P^\rho_{n'})$. Since n' is at least $3(k-1)$ we may invoke Lemma 2.5.5 to obtain

$$[k]_1 \subseteq D_2.$$

Lemma 2.5.4 then gives

$$[3(k-1)]_1 \cap [n'-1] \ \subseteq \ [k(k-1)]_1 \cap [n'-1] \subseteq D_3 \subseteq D_\rho.$$

This takes care of the case $|x-y| \leq 3(k-1)$. Suppose that $|x-y| > 3(k-1)$. For $t \geq 3$, write

$$I_t := [(k-1)^{t-2} + 2(k-1), (k-1)^t - F'(k-2,k) - 1].$$

The intervals I_t and I_{t+1} intersect whenever $t \geq 3$. Therefore by (2)

$$[3(k-1), (k-1)^\rho] = \bigcup_{t=3}^{\rho} I_t \ \cup \ [(k-1)^\rho - F'(k-2,k), (k-1)^\rho].$$

If $(k-1)^\rho - F'(k-2,k) \leq |x-y|$ we can use Lemma 2.5.6 for $t = \rho$ because $|x-y| \in A_\rho$ and thereby obtain $|x-y| \in A'_\rho$. Now Lemma 2.5.7 tells us that $|x-y| \in D_\rho$. In the case when

$(k-1)^\rho - F'(k-2,k) > |x-y|$, there exists $3 \le t \le \rho$ such that $|x-y| \in I_t$ so we have $|x-y| \in A_t$ by definition of $F'(k-2,k)$ and can apply Lemmas 2.5.6 and 2.5.7 to conclude $|x-y| \in D_t \subseteq D_\rho$. The claim now follows from the definitions of D_ρ and A_ρ. □

This completes our preliminary investigation of the C_k-process on paths.

2.5.3 Proof of parts (i) and (ii) of Theorem 2.3.1

Suppose that k is odd, and let $x, y \in V(G)$. If there exists an xy-path Q of length at least $3(k-1)-1$ it satisfies the hypotheses of Proposition 2.5.8 with $s = r$ and n' being the length of Q. In that case the vertices of Q must form a clique in G_r. In particular x and y are adjacent. If every path from x to y in G has length less than $3(k-1)-1$, invoke Lemma 2.5.1 to fix a cycle $C \subset G_2$ containing x, a shortest path Q from y to $V(C)$ in G and an arbitrary vertex $z \in V(G) \setminus V(C)$ with $N_G(z) \cap V(C) \neq \varnothing$. The latter vertex exists because G was assumed to be connected, and is needed because we cannot rule out that $Q \subset C$, and Lemma 2.5.3 requires a $(k+1)$-vertex graph. Apply Lemma 2.5.3 to $G_2[V(C) \cup V(Q) \cup \{z\}]$. As $\tau_{C_k}(G_2[V(C) \cup V(Q) \cup \{z\}])$ is trivially bounded by $(|V(C)| + |V(Q)| + 1)^2/2$ we obtain that $xy \in E(G_r)$ when n and hence r is sufficiently large.

We will see that part (ii) is analogous to part (i) with the roles of cliques being played by complete bipartite graphs.

Now suppose that k is even and G is bipartite with partite sets X, Y, and let $x \in X$, $y \in Y$. If there is a path of length at least $3(k-1)-1$ from x to y, then Proposition 2.5.9 with $\rho = r$ and n' the length of that path implies $xy \in E(G_r)$. Should there be no such path, Lemma 2.5.1 again allows us to choose a cycle $C \subset G_2$ containing x, a shortest path Q from y to $V(C)$ in G, and $z \in V(G) \setminus V(C)$ such that $N_G(z) \cap V(C) \neq \varnothing$. Lemma 2.5.3 applied to $G_2[V(C) \cup V(Q) \cup \{z\}]$ tells us that at time r each vertex of $X \cap (V(C \cup Q) \cup \{z\})$ neighbours each vertex of $Y \cap (V(C \cup Q) \cup \{z\})$.

2.5.4 Proof of part (iii) of Theorem 2.3.1

Assume that in the following k is even and G is not bipartite. When we dealt with the upper bound for odd cycles and wanted to show that an edge xy from the final graph occurs at a certain time in the process it was sufficient to restrict ourselves to an xy-path in the starting graph. In the case of even k and non-bipartite G one has to modify the approach since all xy-paths in G could have even length while the final graph of the C_k-process on a path is not a clique but a complete bipartite graph (cf. Proposition 2.5.9), and thus the restricted C_k-process does not yield the desired edge. To deal with this issue we consider carefully chosen odd walks instead of odd paths, where by carefully we mean that the odd walk contains sufficiently long subwalks without repeated vertices. More precisely we restrict our attention to odd walks which can be expressed as the union of two paths as specified in the next claim:

Claim 2.5.10. *Let $e \in E(\langle G \rangle_{C_k})$. Then there exist $\ell, \ell' \in \mathbb{N}_0$ with $\ell' \le \ell \le n-2$, $\ell' \le n-3$, and vertices $v_0, \ldots, v_\ell, w_0, \ldots, w_{\ell'} \in V(G)$ such that $w_0 v_0 \ldots v_\ell$ and $v_0 w_0 \ldots w_{\ell'}$ are paths, $v_\ell \ldots v_0 w_0 \ldots w_{\ell'}$ is a shortest odd walk between the endpoints of e in G, and $v_j \neq w_{j'}$ for $j \neq j'$.*

Proof. Take a shortest odd walk $u_0 \dots u_m$ between the endpoints of e in G. The claim is clearly satisfied with $\ell' = 0$ and $\ell = m-1$ when the shortest odd walk is already a path. We therefore assume that the walk has at least one repeated vertex. Observe that for any $0 \le j < j' \le m$ with $u_j = u_{j'}$, $j - j'$ must be odd for otherwise $u_0 \dots u_j u_{j'+1} \dots u_m$ would be a shorter odd walk. This implies that no vertex occurs more than twice on $u_0 \dots u_m$. Set

$$j_0 := \max\{j \in [m] \mid \exists j' > j : u_j = u_{j'}\},$$
$$j_1 := \min\{j \in [m] \mid \exists j' < j : u_j = u_{j'}\},$$

and let $j_0', j_1' \in [m]$ be the unique integers satisfying $j_0' > j_0, u_{j_0} = u_{j_0'}$ and $j_1' < j_1, u_{j_1} = u_{j_1'}$. Then $j_0 < j_1$ since otherwise $u_0 \dots u_{j_1'-1} u_{j_1} \dots u_{j_0} u_{j_0'+1} \dots u_m$ would be an odd walk of length less than m. By the extremality of j_0 and j_1 we have $j_1' \le j_0$ and $j_1 \le j_0'$. Equality is attained in both of the last two inequalities. Indeed, if one of them was strict the walk $u_0 \dots u_{j_1'-1} u_{j_1} \dots u_{j_0} u_{j_0'+1} \dots u_m$ would have length

$$j_1' + (j_1 - j_0) + m - j_0' < j_1' + (j_0' - j_1') + m - j_0' = m$$

so

$$0 \equiv j_1' + (j_1 - j_0) + m - j_0' \equiv (j_1 - j_1') + (j_0' - j_0) + m \mod 2$$

by the minimality of m. But this would contradict the fact that $j_1 - j_1'$, $j_0' - j_0$ and m are all odd. Therefore $j_1' = j_0$ and $j_0' = j_1$, which implies $u_{j_0} = u_{j_1}$. We conclude $j_1 - j_0 \ge 3$ because $j_1 - j_0$ must be odd. By definition of j_0 and j_1, both $u_0 \dots u_{j_1-1}$ and $u_{j_0+1} \dots u_m$ are paths and each of the vertices $u_{j_0+1}, \dots, u_{j_1-1}$ occurs precisely once on $u_0 \dots u_m$. Define $v_j, 0 \le j \le \ell := \lfloor (j_0 + j_1)/2 \rfloor$, and $w_{j'}$, $0 \le j' \le \ell' := m - \lceil (j_0 + j_1)/2 \rceil$ by

$$v_j := u_{\left\lfloor \frac{j_0+j_1}{2} \right\rfloor - j} \quad , \quad w_{j'} := u_{\left\lceil \frac{j_0+j_1}{2} \right\rceil + j'}.$$

Then both $v_\ell \dots v_0 w_0$ and $v_0 w_0 \dots w_{\ell'}$ are paths. Therefore $\ell, \ell' \le n-2$. Now we check that $\ell' \le \ell$ and $v_j \ne w_{j'}$ whenever $j \ne j'$. Suppose there were $j \ne j'$ with $v_j = w_{j'}$. Due to the definition of v_j, w_j' and the minimality of m we have

$$\left(\left\lfloor \frac{j_0 + j_1}{2} \right\rfloor - j \right) - \left(\left\lceil \frac{j_0 + j_1}{2} \right\rceil + j' \right) \equiv 1 \mod 2,$$

and thus, as $j_0 - j_1 \equiv 1 \mod 2$,

$$\left\lfloor \frac{j_0 + j_1}{2} \right\rfloor - j - j_0 \equiv \left\lceil \frac{j_0 + j_1}{2} \right\rceil + j' - j_1 \mod 2.$$

But then replacing the longer of the two walks $u_{j_0} \dots u_{\left\lfloor \frac{j_0+j_1}{2} \right\rfloor - j}$ and $u_{j_1} \dots u_{\left\lceil \frac{j_0+j_1}{2} \right\rceil + j'}$ by the shorter one creates an odd walk between the endpoints of e whose length is less than m. If $\ell' \le \ell$ we are done. Otherwise we simply relabel the path by interchanging the roles of ℓ and ℓ' and turning v_i into w_i and vice versa.

We cannot have $\ell' = \ell = n-2$ as in that case $w_{\ell'} \notin \{v_0, \dots, v_\ell\}$ gives $|\{v_0, \dots, v_\ell, w_0, w_{\ell'}\}| = n+1$. $\square$

Remark 2.5.11. The property $v_j \neq w_{j'}$ for $j \neq j'$ in Claim 2.5.10 guarantees that whenever $J \subset [\ell]$ and $J' \subset [\ell']$ are disjoint, one has $\{v_j : j \in J\} \cap \{w_{j'} : j' \in J'\} = \varnothing$.

Let $x, y \in V(G)$ be distinct vertices. If the length of a shortest odd walk between them is smaller than $3(k-1)^2$ we can fix a subgraph of G_2 on at least $k+1$ vertices that contains x, y and a k-cycle (the existence of such a subgraph is guaranteed by Lemma 2.5.1) and apply Lemma 2.5.3. From now on let the length of a shortest odd walk from x to y in G be at least $3(k-1)^2$. We are done once we have shown $xy \in E(G_r)$. Let $v_\ell \dots v_0 w_0 \dots w_{\ell'}$ be a shortest odd walk from x to y or y to x as given by Claim 2.5.10. Note that $\ell \equiv \ell' \mod 2$. The remaining proof is divided into the Claims 2.5.12, 2.5.13 and 2.5.14.

Claim 2.5.12. *If $\ell + \ell' \leq (k-1)^r - (k-1) \cdot F'(k-2,k) - 3(k-1)^2$, we have $xy \in E(G_r)$.*

Proof. Write $\ell' = q'(k-1) + s'$ and $\ell = q(k-1) + s$ and where $q, q', s, s' \in \mathbb{N}_0$, $0 \leq s \leq k-2$, $0 \leq s' \leq k-2$. Recall that $\ell \geq \ell'$ and $\ell + \ell' \geq 3(k-1)^2 - 1$, hence $q \geq q'$ and $q > 1$. At time 1,

$$P := w_0 \dots w_{s'} w_{s'+(k-1)} \dots w_{s'+q'(k-1)}$$

is a path of length $q' + s'$ from w_0 to $w_{\ell'}$. It does not contain v_ℓ as $\ell' \leq \ell$. If $s' = 0$,

$$Q_0 := v_0 v_1 v_{1+(k-1)} \dots v_{1+q(k-1)} v_{1+q(k-1)+1} \dots v_{1+q(k-1)+s-1}$$

is a path of length $q+s$ from v_0 to v_ℓ which is vertex-disjoint from P as the indices of the vertices on P are multiples of $k-1$ whereas the indices of vertices on $Q_0 - v_0 - v_\ell$ are not. The union of P, Q_0 and $v_0 w_0$ is an xy-path of length $q' + (q+s) + 1$. Note that

$$q' + (q+s) + 1 \equiv q'(k-1) + q(k-1) + s + 1 \equiv (\ell - \ell' + 1) \equiv 1 \mod 2$$

and

$$q' + q + s + 1 \leq \frac{\ell' + \ell}{k-1} + k - 1 \leq (k-1)^{r-1} - F'(k-2,k) - 2(k-1) < (k-1)^{r-1} - F'(k-2,k).$$

If $s' > 0$, recall that $q > 1$ and consider the path

$$Q_{s'} := \begin{cases} v_0 v_{k-1} \dots v_{q(k-1)} v_{q(k-1)-1} \dots v_{(q-1)(k-1)+s} v_\ell & \text{, if } s > s'; \\ v_0 v_{k-1} \dots v_{q(k-1)} v_{q(k-1)+1} \dots v_\ell & \text{, if } s \leq s'. \end{cases}$$

This path has length $q + (k-1) - s + 1$ or $q + s$. It is vertex-disjoint from P since $w_j \equiv s' \mod k-1$ for all $j \geq s'$ with $w_j \in V(P)$, whereas all $j \in [0, \ell']$ with $v_j \in V(Q_{s'}) \setminus \{v_0, v_\ell\}$ satisfy $j > s'$ and $j \not\equiv s' \mod k-1$. For the case $\ell = \ell'$ it is important that v_ℓ never lies on P because $\ell \geq \ell'$ and $v_\ell \neq w_{\ell'}$ by assumption.

As $v_0w_0 \in E(G_1)$, the union of the paths P,$Q_{s'}$ and the edge v_0w_0 is an xy-path of length $q'+s'+q+k-s$ or $q'+s'+q+s$, where

$$q'+s'+q+k-s+1 \equiv q'+s'+q+s+1 \equiv \ell+\ell'+1 \mod 2.$$

The length of $P \cup Q_{s'} \cup \{v_0w_0\}$ is bounded from above by

$$\frac{\ell+\ell'}{k-1}+2(k-2)+1<(k-1)^{r-1}-F'(k-2,k).$$

In both cases we have an odd xy-path of length at least $q+q'$ and less than $(k-1)^{r-1}-F'(k-2,k)$. Using

$$q+q'=\left\lfloor\frac{\ell}{k-1}\right\rfloor+\left\lfloor\frac{\ell}{k-1}\right\rfloor\geq\left\lfloor\frac{\ell+\ell'}{k-1}\right\rfloor\geq 3(k-1)$$

we can apply Proposition 2.5.9 with $\rho=r-1$ to either $P\cup Q_0\cup\{v_0w_0\}$ or $P\cup Q_{s'}\cup\{v_0w_0\}$ to deduce $xy \in E(G_{1+r-1})=E(G_r)$. □

Our next claim will serve as an auxiliary statement in the proof of Claim 2.5.14.

Claim 2.5.13. *Set $w_{-1}:=v_0$. Then*

$$v_{\frac{(k-1)^t-(k-1)}{2}}w_{\frac{(k-1)^t-(k-1)}{2}+k-2}\in E(G_t),$$

whenever $t\geq 1$ with $\frac{(k-1)^t-(k-1)}{2}\leq\ell$ and $\frac{(k-1)^t-(k-1)}{2}+k-2\leq\ell'$, and

$$v_{\frac{(k-1)^t-(k-1)}{2}+k-1}w_{\frac{(k-1)^t-(k-1)}{2}-1}\in E(G_t)$$

whenever $\frac{(k-1)^t-k-1}{2}+(k-1)\leq\ell$ and $\frac{(k-1)^t-(k-1)}{2}-1\leq\ell'$.

Proof. The size constraints are only necessary to guarantee that the vertices occurring in the statement actually exist. We induct on t. When $t=1$ the claim reads $v_0w_{k-2}, v_{k-1}w_{-1}\in E(G_1)$, which holds since v_0 and w_{k-2}, and similarly v_{k-1} and v_0, are clearly endpoints of paths of length $k-1$ in G. Suppose that $t\geq 2$ and the above size constraints are satisfied. Set

$$j_s:=\frac{(k-1)^{t-1}-(k-1)}{2}+s\cdot(k-1)^{t-1} \quad , \quad 0\leq s\leq (k-2)/2.$$

The induction hypothesis implies

$$v_{j_0}w_{j_0+k-2}, v_{j_0+k-1}w_{j_0-1}\in E(G_{t-1}),$$

and Lemma 2.3.4 assures that any two vertices of distance $(k-1)^{t-1}$ in G are adjacent at time $t-1$. We have

$$j_s+k-2\equiv k-2\not\equiv 0\equiv j_{s'} \mod k-1$$

for any s, s' hence $w_{j_s+k-2} \neq v_{j_{s'}}$. Similarly, $w_{j_s-1} \neq v_{j_{s'}+k-1}$. Therefore

$$v_{j_{(k-2)/2}} \dots v_{j_1} v_{j_0} w_{j_0+k-2} w_{j_1+k-2} \dots w_{j_{(k-2)/2}+k-2}$$

and

$$v_{j_{(k-2)/2}+k-1} \dots v_{j_1+k-1} v_{j_0+k-1} w_{j_0-1} w_{j_1-1} \dots w_{j_{(k-2)/2}-1}$$

are paths of length $k-1$ in G_{t-1}. The claim now follows from the observation that

$$j_{(k-2)/2} = \frac{(k-1)^t - (k-1)}{2}.$$

□

Claim 2.5.14. *Suppose that* $\ell + \ell' > (k-1)^r - (k-1) \cdot F'(k-2,k) - 3(k-1)^2$. *Then* $xy \in E(G_r)$.

Proof. Recall that $\ell \equiv \ell' \mod 2$ since $v_\ell \dots v_0 w_0 \dots w_{\ell'}$ is an odd walk. Our plan is to find an xy-path of length $k-1$ in G_{r-1}. By (2) and the upper bound in (2.11) we have

$$\begin{aligned}
\ell \geq \ell' &\geq \ell' + \ell - (n-2) \\
&> (k-1)^r - (k-1) \cdot F'(k-2,k) - 3(k-1)^2 - \frac{(k-1)^r - (k-1)}{2} + F'(k-2,k) \\
&= \frac{(k-1)^r}{2} - (k-2) \cdot F'(k-2,k) - 3(k-1)^2 + \frac{k-1}{2} \\
&\geq \frac{(k-1)^{r-1} - (k-1)}{2} + k - 1. \qquad (2.25)
\end{aligned}$$

This and the assumption that r is sufficiently large allows us to apply Claim 2.5.13 with $t = r-1$. Choose

$$j_0 \in \left\{ \frac{(k-1)^{r-1} - (k-1)}{2}, \frac{(k-1)^{r-1} - (k-1)}{2} + k - 1 \right\}$$

and

$$j_0' \in \left\{ \frac{(k-1)^{r-1} - (k-1)}{2} - 1, \frac{(k-1)^{r-1} - (k-1)}{2} + k - 2 \right\}$$

such that $\ell - j_0 \equiv \ell' - j_0' \equiv (k-2)/2 \mod 2$. The congruences $\ell - j_0 \equiv \ell' - j_0' \mod 2$ and $\ell \equiv \ell' \mod 2$ together imply $j_0 \equiv j_0' \mod 2$. Thus $v_{j_0} w_{j_0'}$ is one of the edges whose presence at time $r-1$ is guaranteed by Claim 2.5.13 with $t = r-1$.

We are now going to construct a $w_{\ell'} w_{j_0'}$-path $P \subset G_{r-1}$ of length $(k-2)/2$ avoiding v_{j_0} and v_ℓ, as well as a $v_{j_0} v_\ell$-path $Q \subset G_{r-1}$ of length $(k-2)/2$ that is disjoint from P. Once we have found those paths we will be done because the union of P, $w_{j_0'} v_{j_0}$ and Q will be a $w_{\ell'} v_\ell$-path of length $k-1$. Recall the upper bound given in (2.11). Since $\ell + \ell' > (k-1)^r - (k-1) \cdot F'(k-2,k) -$

$3(k-1)^2$ and $\ell \le n-2$, $\ell' \le n-3$ we get the bounds

$$\begin{aligned}\ell - j_0 &\ge \ell + \ell' - (n-3) - j_0 \\ &> \frac{(k-1)^r}{2} - (k-2)\cdot F'(k-2,k) - 3(k-1)^2 + \frac{k-1}{2} - j_0 \\ &\ge \frac{k-2}{2}\cdot (k-1)^{r-1} - (k-2)\cdot F'(k-2,k) - 3(k-1)^2 \end{aligned} \tag{2.26}$$

and

$$\begin{aligned}\ell - j_0 &< \frac{(k-1)^r - (k-1)}{2} - F'(k-2,k) + 2 - 2 - \frac{(k-1)^{r-1} - (k-1)}{2} \\ &= \frac{k-2}{2}\cdot (k-1)^{r-1} - F'(k-2,k). \end{aligned} \tag{2.27}$$

Similarly, $\ell' - j_0'$ satisfies

$$\begin{aligned}\ell' - j_0' &\ge \ell' + \ell - (n-2) - j_0' \\ &\ge \frac{k-2}{2}(k-1)^{r-1} - (k-2)\cdot F'(k-2,k) - 3(k-1)^2 + 1 \end{aligned}$$

and

$$\begin{aligned}\ell' - j_0' &\le (n-3) - \frac{(k-1)^{r-1} - (k-1)}{2} + 1 \\ &< \frac{(k-1)^r - (k-1)}{2} - F'(k-2,k) + 2 - 3 - \frac{(k-1)^{r-1} - (k-1)}{2} + 1 \\ &= \frac{k-2}{2}(k-1)^{r-1} - F'(k-2,k). \end{aligned} \tag{2.28}$$

Therefore, by definition of $F'(k-2,k)$ and using that $\ell - j_0 \equiv \ell' - j_0' \equiv (k-2)/2 \mod 2$, there exist $\alpha, \beta, \gamma, \delta \in \mathbb{N}_0$ such that

$$\ell - j_0 = \frac{k-2}{2}\cdot (k-1)^{r-1} - \alpha(k-2) - \beta k \tag{2.29}$$

$$\ell' - j_0' = \frac{k-2}{2}\cdot (k-1)^{r-1} - \gamma(k-2) - \delta k.$$

By combining (2.26) and (2.29) we infer

$$\alpha + \beta \le F'(k-2,k) + \frac{3(k-1)^2}{k-2} < \frac{k^2}{2} + 4. \tag{2.30}$$

Define

$$j_s' := j_0' + s\cdot (k-1)^{r-1}$$

for $1 \le s < \frac{k-2}{2} - 1$. Then $w_{j_{s-1}'} w_{j_s'} \in E(G_{r-1})$ for $1 \le s < \frac{k-2}{2} - 1$ by Lemma 2.3.4. Using (2.25), (2.28) and the estimate

$$\frac{(k-1)^{r-1} - (k-1)}{2} + k - 1 > 3(k-1)$$

we can find a subpath P' of $w_0 \dots w_{\ell'}$ of length more than $3(k-1)$ and less than $(k-1)^{r-1} - F(k-2,k)$ that contains both $w_{j'_{(k-2)/2-1}}$ and $w_{\ell'}$. Applying Proposition 2.5.9 with $\rho = r-1$ and n' being the length of P' yields $w_{j'_{(k-2)/2-1}} w_{\ell'} \in E(G_{r-1})$, so

$$P := w_{j'_0} w_{j'_1} \dots w_{j'_{(k-2)/2-1}} w_{\ell'}$$

is a path of length $(k-2)/2$ that avoids v_{j_0} and v_ℓ because $j'_0 \neq j_0 < j'_1$ and $w_{\ell'} \neq v_\ell$.

If $k = 4$ we have $(k-2)/2 = 1$ so P just consists of the edge $w_{j'_0} w_{\ell'}$ and we may set $Q := v_{j_0} v_\ell$. The edge $v_{j_0} v_\ell$ indeed lies in $E(G_{r-1})$: As $\ell - j_0$ is odd by choice of j_0 and $\ell - j_0 \leq 3^{r-1} + 2$ due to (2.27), we can apply Proposition 2.5.9 with $\rho = r-1$ to the path $v_{j_0} v_{j_0+1} \dots v_\ell$.

For the rest of the proof we assume that $k \geq 6$, which guarantees that $F(k-2,k)$ is positive and thus both $\alpha + \beta$ and $\gamma + \delta$ are at least one. Choose

$$\alpha_1, \dots, \alpha_{\frac{k-2}{2}} \in \left\{ \left\lfloor \frac{\alpha}{(k-2)/2} \right\rfloor, \left\lceil \frac{\alpha}{(k-2)/2} \right\rceil \right\}$$

and

$$\beta_1, \dots, \beta_{\frac{k-2}{2}} \in \left\{ \left\lfloor \frac{\beta}{(k-2)/2} \right\rfloor, \left\lceil \frac{\beta}{(k-2)/2} \right\rceil \right\}$$

with $\alpha_1 + \ldots + \alpha_{(k-2)/2} = \alpha$ and $\beta_1 + \ldots + \beta_{(k-2)/2} = \beta$, and define j_s, $1 \leq s < (k-2)/2$, by

$$j_s := j_{s-1} + (k-1)^{r-1} - \alpha_s(k-2) - \beta_s k.$$

Then for $1 \leq s < \frac{k-2}{2}$, the fact that $v_{j_{s-1}} v_{j_s} \in E(G_{r-1})$ follows from Proposition 2.5.9 with $\rho = r-1$ and $n' = j_s - j_{s-1}$ applied to the path $v_{j_{s-1}} v_{j_{s-1}+1} \dots v_{j_s}$.

Due to (2.29) the difference $\ell - j_{(k-2)/2-1}$ can be written as

$$\begin{aligned}
\ell - j_{\frac{k-2}{2}-1} &= \ell - j_0 - (j_{\frac{k-2}{2}-1} - j_0) \\
&= \ell - j_0 - \left(\frac{k-2}{2} - 1\right)(k-1)^{r-1} + \left(\alpha - \alpha_{\frac{k-2}{2}}\right)(k-2) + \left(\beta - \beta_{\frac{k-2}{2}}\right)k \\
&= (k-1)^{r-1} - \alpha(k-2) - \beta k + \left(\alpha - \alpha_{\frac{k-2}{2}}\right)(k-2) + \left(\beta - \beta_{\frac{k-2}{2}}\right)k \\
&= (k-1)^{r-1} - \alpha_{\frac{k-2}{2}} \cdot (k-2) - \beta_{\frac{k-2}{2}} k
\end{aligned}$$

and is positive because (2.30) implies

$$\alpha_{\frac{k-2}{2}} \cdot (k-2) + \beta_{\frac{k-2}{2}} k \;\leq\; (\alpha + \beta) \cdot k \;<\; \left(\frac{k^2}{2} + 4\right) \cdot k \;<\; (k-1)^{r-1},$$

where the final inequality uses that r is sufficiently large. Invoking Proposition 2.5.9 again, we obtain that $v_{j_{\frac{k-2}{2}-1}} v_\ell \in E(G_{r-1})$. An analogous argument shows that $v_{j_{\frac{k-2}{2}-1}-(k-2)} v_\ell \in E(G_{r-1})$. Therefore,

$$v_{j_0} \dots v_{j_{\frac{k-2}{2}-1}} v_\ell \quad \text{and} \quad v_{j_0} v_{j_1-(k-2)} \dots v_{j_{\frac{k-2}{2}-1}-(k-2)} v_\ell$$

are two paths of length $k-1$ from v_{j_0} to v_ℓ. They are internally disjoint as

$$\begin{aligned} j_i-(j_s-(k-2)) &= (i-s)\cdot(k-1)^{r-1}-(\alpha_t+\ldots+\alpha_{s+1})(k-2)-(\beta_t+\ldots+\beta_{s+1})k+(k-2)\\ &\geq (k-1)^{r-1}-\alpha(k-2)-\beta k\\ &\geq (k-1)^{r-1}-\left(\frac{k^2}{2}+4\right)\cdot k\\ &> 0 \end{aligned}$$

for $1\leq s<i<(k-2)/2$ and r sufficiently large. This together with $j_0\neq \ell'$, which holds due to (2.25), assures that at least one of those paths does not contain $w_{\ell'}$. Define Q to be such a path. We chose j_0, j_0' such that $j_0\equiv j_0' \mod 2$, hence we obtain either $j_0=j_0'-(k-2)$ or $j_0=j_0'+k$. In both cases,

$$j_s\equiv j_0\not\equiv j_0'\equiv j_{s'}' \mod k-1$$

and

$$j_s-(k-2)\equiv j_0-(k-2)\not\equiv j_0'\equiv j_{s'}' \mod k-1$$

for $0\leq s<(k-2)/2$ and $0\leq s'<(k-2)/2$ so Q and P do not intersect. Therefore, $P\cup\{w_{j_0'}v_{j_0}\}\cup Q$ is the desired path from x to y in G_{r-1}. □

2.6 Graphs with cycle components

In this section we prove Theorem 1.2.4.

Let H be a graph with a k-cycle component, and let $\tilde{H}\subset H$ such that $H\cong\tilde{H}\sqcup C_k$. Our goal is to show that for a suitable constant $\kappa=\kappa(H)$ one has

$$M_H(n)\leq \log_{k-1}(n)\ +\ \kappa. \tag{2.31}$$

for all $n\in\mathbb{N}$. We assume that $\tilde{H}$ contains at least one edge, for otherwise the claim follows directly from our results on cycles. Let

$$\lambda=\lambda(H):=40(v(H)^3\cdot v(H)!\cdot 2^{v(H)})^2. \tag{2.32}$$

For the purpose of the proof of Theorem 1.2.4, whenever a graph is denoted by Γ we let $(\Gamma_t)_{t\geq 0}$ be the H-process on Γ. Similarly if a graph is denoted by G, $(G_t)_{t\geq 0}$ is the H-process on G. We further introduce the following shorthand notation. If G is a graph and $U\subseteq V(G)$ we write $G[\setminus U]:=G[V(G)\setminus U]$.

We reduce Theorem 1.2.4 to the three lemmas below. The first of them tells us that if an n-vertex starting graph has the property that after removing a certain exceptional set the graph has only a constant number of components and each of the components has a small diameter, then the H-process on that graph takes only a constant number of steps before it stabilises. The purpose of the other two lemmas is to reduce the case of general starting graphs to the situation of the first lemma. The logarithmic term in (2.31) will come from the latter reduction and is again attributed to the fact that in the C_k-process on a graph the diameter decreases roughly by a factor $k-1$ in each step.

In the following the number of components of a graph G is denoted by $c(G)$, and we set

$$\mathrm{cdiam}(G) := \max_{\mathscr{C}} \mathrm{diam}(\mathscr{C})$$

where the maximum ranges over all components of G.

Lemma 2.6.1. *There exists a constant $\mu = \mu(H)$ such that the following holds. Let G be a graph that contains a copy $\tilde{H}_0$ of $\tilde{H}$ such that $c(G[\setminus V(\tilde{H}_0)]) \le \lambda$ and $\mathrm{cdiam}(G[\setminus V(\tilde{H}_0)]) \le \lambda$. Then we have that $\tau_H(G) \le \mu$.*

We will prove Lemma 2.6.1 in Section 2.6.3.

Lemma 2.6.2. *Let G be a graph on n vertices. If $\tilde{H}_1$ is a copy of $\tilde{H}$ in G_1 then for $t \ge \log_{k-1} n + 1$, one has $\mathrm{cdiam}(G_t[\setminus V(H_1)]) \le \lambda$.*

Lemma 2.6.2 will be shown in Section 2.6.2.

Lemma 2.6.3. *Let G be a graph and suppose that $\tilde{H}_1$ is a copy of $\tilde{H}$ in G_1. For every $t \in [\tau_H(G)]$ there exists $\tilde{t} \ge t$ and a subgraph $\tilde{G} \subseteq G_{\tilde{t}}$ such that $\tilde{H}_1 \subseteq \tilde{G}$,*

$$\tau_H(\tilde{G}) \ge \frac{\tau_H(G_t)}{\lambda} \qquad \text{and} \qquad c\left(\tilde{G}[\setminus V(\tilde{H}_1)]\right) \le \lambda,$$

and $\mathrm{cdiam}(\tilde{G}[\setminus V(\tilde{H}_1)]) \le \mathrm{cdiam}(G_{\tilde{t}}[\setminus V(\tilde{H}_1)])$.

This lemma is a consequence of two technical auxiliary statements that we will introduce in Section 2.6.1. The proof of Lemma 2.6.3 can be found in the same section.

When combined, the three lemmas above allow us to deduce the main theorem.

Proof of Theorem 1.2.4. Let G be an n-vertex graph such that $\tau_H(G) = M_H(n)$. If $\tau_H(G) < \log_{k-1} n + 1$ we are done. Thus we suppose that $\tau_H(G) \ge \log_{k-1} n + 1$.

Choose a copy $\tilde{H}_1$ of $\tilde{H}$ in G_1. Let $t_1 := \lceil \log_{k-1} n \rceil + 1$. Apply Lemma 2.6.3 with $t = t_1$ to obtain $\tilde{t} \ge t_1$ and $\tilde{G} \subseteq G_{\tilde{t}}$ such that $\tilde{H}_1 \subseteq \tilde{G}$, $c(\tilde{G}[\setminus V(\tilde{H}_1)]) \le \lambda$,

$$\tau_H(\tilde{G}) \;\ge\; \frac{\tau_H(G_{t_1})}{\lambda}, \tag{2.33}$$

and $\mathrm{cdiam}(\tilde{G}[\setminus V(\tilde{H}_1)]) \le \mathrm{cdiam}(G_{\tilde{t}}[\setminus V(\tilde{H}_1)])$. Lemma 2.6.2 now tells us that

$$\mathrm{cdiam}(\tilde{G}[\setminus V(\tilde{H}_1)]) \le \lambda.$$

The graph $\tilde{G}$ fulfils the hypotheses of Lemma 2.6.1, which together with (2.33) then yields

$$\tau_H(G) \;=\; t_1 + \tau_H(G_{t_1}) \;\le t_1 \;+ \lambda \cdot \tau_H(\tilde{G}) \;\le\; \log_{k-1}(n) + 1 + \lambda\mu.$$

where μ is the constant from Lemma 2.6.1. Thus (2.31) holds with $\kappa = 1+\lambda\mu$. ■

2.6.1 Component reduction

The technical lemmas in this section do not make use of the assumption that H contains a cycle component, and thus might also be of independent interest. Their purpose is to establish that at the cost of a constant factor in running time we may restrict ourselves to starting graphs that have few components, even after a small exceptional set has been removed.

If G is a graph and $x \in V(G)$ we denote the vertex set of the component of G containing x by $\mathscr{V}(G,x)$. Furthermore if $W \subset V(G)$ we define $\mathscr{V}(G,W) := \bigcup_{x\in W}\mathscr{V}(G,x)$. In particular, $\mathscr{V}(G,e) = \mathscr{V}(G,x)\cup\mathscr{V}(G,y)$ for any edge $e = xy$ of G.

Lemma 2.6.4. *Let Γ be a graph, and let $U \subseteq V(\Gamma)$. There exists a non-empty set $\hat{W} \subseteq V(\Gamma)$ of size at most $v(H)^3\cdot|U|!\cdot 2^{v(H)}$ such that $U\cap\hat{W} = \varnothing$ and for all $t \geq 0$ the following hold:*

(1) If $x,y \in V(\Gamma)\setminus U$ and H_{xy} is a copy of H completed by xy at time t such that $\mathscr{V}(\Gamma_{t-1}[\setminus U],x) \neq \mathscr{V}(\Gamma_{t-1}[\setminus U],y)$ and $\mathscr{V}(H_{xy}[V(H_{xy})\setminus U]-xy,x)\cap\hat{W} = \varnothing$, then there exists $\hat{y}\in\hat{W}$ and a copy $\hat{H}$ of H such that $x\hat{y}\in E(\hat{H})$, $\hat{H}-x\hat{y}\subseteq\Gamma_{t-1}$ and $V(\hat{H})\subseteq U\cup\hat{W}\cup\mathscr{V}(\Gamma_{t-1}[\setminus U],x)$. In particular, $x\hat{y}\in E(\Gamma_t)$.

(2) If $e\in E(\Gamma_t)\setminus E(\Gamma_{t-1})$ with $e \not\subseteq U$ and $\mathscr{V}(\Gamma_{t-1}[\setminus U],e\setminus U)\cap\hat{W} = \varnothing$, there exists a copy H' of H with $V(H')\subseteq U\cup\hat{W}\cup\mathscr{V}(\Gamma_{t-1}[\setminus U],e\setminus U)$ such that H' is completed by e at time t.

Proof. For $t \geq 1$, define

$$\Phi_t := \{(\varphi_{|\varphi^{-1}(U)}, e) : e\in E(H), \varphi\in \mathrm{Hom}(H-e,\Gamma_{t-1}), \varphi \text{ injective}\}.$$

These sets satisfy $\Phi_t\subseteq\Phi_{t+1}$ for all $t\geq 1$. Let $\Phi := \bigcup_{t\geq 0}\Phi_t$. We pick a set of representatives $\mathscr{R}\subseteq\bigcup_{t\geq 1}\{(\varphi,e) : e\in E(H), \varphi\in\mathrm{Hom}(H-e,\Gamma_{t-1}), \varphi \text{ injective}\}$ such that

$$\Phi = \{(\varphi_{|\varphi^{-1}(U)}, e) : (\varphi,e)\in\mathscr{R}\}$$

and for every $t\geq 1$, $e\in E(H)$, and injective $\varphi\in\mathrm{Hom}(H-e,\Gamma_{t-1})$, there exists precisely one $\psi\in\mathrm{Hom}(H-e,\Gamma_{t-1})$ with $(\psi,e)\in\mathscr{R}$ and $\varphi_{|\varphi^{-1}(U)} = \psi_{\psi^{-1}(U)}$. Now we set

$$\hat{W} := \bigcup_{(\psi,e)\in\mathscr{R}} V(\psi(H))\setminus U.$$

The set $\hat{W}$ is clearly disjoint from U and satisfies

$$|\hat{W}| \leq |\mathscr{R}|\cdot v(H) \leq |\Phi|\cdot v(H) \leq v(H)^3\cdot|U|!\cdot 2^{v(H)}.$$

Let $t\geq 0$. We now verify the properties (1) and (2).

(1) Let $x,y\in V(\Gamma)\setminus U$ such that $\mathscr{V}(\Gamma_{t-1}[\setminus U],x)\neq\mathscr{V}(\Gamma_{t-1}[\setminus U],y)$ and xy completes a copy H_{xy} of H at time t satisfying $\mathscr{V}_x\cap\hat{W} = \varnothing$, where $\mathscr{V}_x := \mathscr{V}(H_{xy}[V(H_{xy})\setminus U]-xy,x)$. Let $\varphi\in\mathrm{Hom}(H,\Gamma_t)$

with $\varphi(H) = H_{xy}$, let $e := \varphi^{-1}(xy)$. There exists $\psi \in \mathrm{Hom}(H-e, \Gamma_{t-1})$ such that $(\psi, e) \in \mathscr{R}$ and $\psi_{|\psi^{-1}(U)} = \varphi_{|\varphi^{-1}(U)}$, and hence

$$\tilde{\varphi} : V(H) \to V(\Gamma) \quad , \quad v \mapsto \begin{cases} \varphi(v) & \text{, if } v \in \varphi^{-1}(\mathscr{V}_x); \\ \psi(v) & \text{, otherwise} \end{cases}$$

is an embedding of $H-e$ into Γ_{t-1}. Indeed, all edges of $\psi(H-e)$ are present in Γ_{t-1}, and $\tilde{\varphi}$ is injective because $\psi(V(H)) \subseteq U \cup \hat{W}$ whereas $\mathscr{V}_x \cap (U \cup \hat{W}) = \varnothing$. Moreover, $H-e$ does not have any edges between $\varphi^{-1}(\mathscr{V}_x)$ and $V(H) \setminus \varphi^{-1}(U \cup \mathscr{V}_x)$, so $\tilde{\varphi}$ sends edges of $H-e$ to edges of Γ_{t-1}. Recall that $y \notin \mathscr{V}(\Gamma_{t-1}[\setminus U], x)$ and hence $y \notin \mathscr{V}_x$. Let $\hat{y} := \psi(\varphi^{-1}(y))$ and $\hat{H} := \tilde{\varphi}(H)$. We have that $\tilde{\varphi}(e) = x\hat{y}$, which tells us that $x\hat{y} \in E(\Gamma_t)$. Since $\varphi^{-1}(U) = \psi^{-1}(U)$ and $y \notin U$, we obtain $\hat{y} \in \hat{W}$.

(2) Let $e \in E(\Gamma_t) \setminus E(\Gamma_{t-1})$ such that $e \nsubseteq U$ and $\mathscr{V}(\Gamma_t[\setminus U], e \setminus U) \cap \hat{W} = \varnothing$. Pick an embedding $\varphi : H \to \Gamma_t$ with $\varphi(H) - e \subseteq \Gamma_{t-1}$, and set

$$\mathscr{V}_e := \mathscr{V}(\varphi(H)[V(\varphi(H)) \setminus U], e \setminus U).$$

Take the unique $(\psi, f) \in \mathscr{R}$ such that $f = \varphi^{-1}(e)$ and $\psi_{|\psi^{-1}(U)} = \varphi_{|\varphi^{-1}(U)}$. As $\mathscr{V}_e \cap \psi(V(H)) = \varnothing$ and H does not have any edges between $\varphi^{-1}(\mathscr{V}_e)$ and $V(H) \setminus \varphi^{-1}(\mathscr{V}_e \cup U)$ the map

$$\tilde{\varphi} : V(H) \to V(\Gamma) \quad , \quad v \mapsto \begin{cases} \varphi(v) & \text{, if } v \in \varphi^{-1}(\mathscr{V}_e); \\ \psi(v) & \text{, otherwise} \end{cases}$$

is an embedding of $H-f$ into Γ_{t-1} with vertices in $U \cup \hat{W} \cup \mathscr{V}(\Gamma_t[\setminus U], e \setminus U)$. Now $H' := \tilde{\varphi}(H)$ is the desired copy of H that is completed by e. □

For $r \in \mathbb{N}$, $m \in \mathbb{N}_0$ define

$$\lambda_1(r,m) := 40(r^3 \cdot m! \cdot 2^r)^2 \qquad , \qquad \lambda_2(r,m) := r2^{2r+m+m^2} + r^3 \cdot m! \cdot 2^r + 1. \tag{2.34}$$

Lemma 2.6.5. *Let $\tau' > 0$. If Γ is a graph with $\tau_H(\Gamma) \geq \tau'$, and $U \subseteq V(\Gamma)$, then there exists $t_0 \in [0, \tau']$ and $W \subseteq V(\Gamma)$ such that the following hold:*

(i) $U \subseteq W$,

(ii) $\tau_H(\Gamma_{t_0}[W]) \geq \frac{\tau'}{\lambda_1(v(H), |U|)}$,

(iii) $\Gamma_{t_0}[W \setminus U]$ is the union of at most $\lambda_2(v(H), |U|)$ components of $\Gamma_{t_0}[\setminus U]$.

Proof. Let $\hat{W}$ be as given by Lemma 2.6.4 applied to Γ and U. There can be at most $\binom{|U \cup \hat{W}|}{2}$ steps in the process in which a new edge with both endpoints in $U \cup \hat{W}$ is added. For this reason we can pick a set $T \subseteq [\tau']$ of size at most $\binom{|U \cup \hat{W}|}{2}$ such that

$$\Gamma_t[U \cup \hat{W}] = \Gamma_{t-1}[U \cup \hat{W}] \tag{2.35}$$

for $t \in \mathbb{N} \setminus T$. We point out that

$$|T|+1 \leq \frac{1}{20}\lambda_1(v(H),|U|).$$

In the following we assume that

$$\frac{\tau'}{\lambda_1(v(H),|U|)} > 1$$

for otherwise we could just take $t_0 = 0$ and $W = U \cup \mathscr{V}(\Gamma[\backslash U], V(H_1))$ where H_1 is a copy of H completed at time 1. In particular, we have that $\tau' \geq 10|T|$. Any partition of $[\tau']$ into more than $|T|$ intervals admits an interval that does not intersect T. Therefore we can find $t_1, t_3 \in [\tau']$ such that

$$t_3 - t_1 \geq \frac{\tau'}{5(|T|+1)} \qquad \text{and} \qquad [t_1,t_3] \cap T = \varnothing.$$

Every edge that is added at some time $t \in [t_1,t_3]$ has at least one endpoint outside $U \cup \hat{W}$. We set

$$t_2 := \left\lfloor \frac{t_1+t_3}{2} \right\rfloor.$$

Suppose there exists $t \in [t_2,t_3]$ and $e_t \in E(\Gamma_t) \setminus E(\Gamma_{t-1})$ such that

$$\mathscr{V}(\Gamma_t[\backslash U], e_t \setminus U) \cap \hat{W} = \varnothing. \tag{2.36}$$

Then Lemma 2.6.4 (2) gives a copy H' of H with $V(H') \subseteq U \cup \hat{W} \cup \mathscr{V}(\Gamma_{t-1}[\backslash U], e \setminus U)$ that is completed by e_t. Due to (2.35) H' must contain an edge $e_{t-1} \in E(\Gamma_{t-1}) \setminus E(\Gamma_{t-2})$ that intersects $\mathscr{V}(\Gamma_t[\backslash U], e_t \setminus U)$ and hence satisfies $\mathscr{V}(\Gamma_{t-1}[\backslash U], e_{t-1} \setminus U) \cap \hat{W} = \varnothing$. By iterating we obtain a sequence $e_t, e_{t-1} \ldots, e_{t_1}$ of edges such that $e_i \in E(\Gamma_i) \setminus E(\Gamma_{i-1})$ and $e_i \in \mathscr{V}(\Gamma_t[\backslash U], e_t \setminus U)$ for $i \in [t_1,t]$. Note that we cannot iterate further because $t_1 - 1$ might lie in T.

The set $\mathscr{V}(\Gamma_t[\backslash U], e_t \setminus U)$ is also the vertex set of a component of $\Gamma_{t_1}[\backslash U]$. Indeed, if it was not, we could find vertices $x,y \in \mathscr{V}(\Gamma_t[\backslash U], e_t \setminus U)$ such that for some $t' \in [t_1+1,t]$ one has $\mathscr{V}(\Gamma_{t'-1}[\backslash U],x) \neq \mathscr{V}(\Gamma_{t'-1}[\backslash U],y)$. Then any copy H_{xy} of H with $H_{xy} - xy \subseteq \Gamma_{t'-1}$ satisfies

$$\mathscr{V}(H_{xy}[V(H_{xy}) \setminus U] - xy, x) \cap \hat{W} \;\subseteq\; \mathscr{V}(\Gamma_t[\backslash U], e_t \setminus U) \cap \hat{W} \;=\; \varnothing,$$

so by Lemma 2.6.4 there exists $\hat{y} \in \hat{W}$ such that $x\hat{y} \in E(\Gamma_{t'})$. This however contradicts the assumption $\mathscr{V}(\Gamma_t[\backslash U], e_t \setminus U) \cap \hat{W} \neq \varnothing$.

Therefore, for the choice $W := \hat{W} \cup U \cup \mathscr{V}(\Gamma_t[\backslash U], e_t \setminus U)$ we have that $U \subseteq W$ and

$$\begin{aligned} c(\Gamma_{t_1}[W \setminus U]) &\leq |\hat{W}|+1 \\ &\leq v(H)^3 \cdot 2^{v(H)} \cdot |U|! + 1 \\ &\leq \lambda_2(v(H),|U|). \end{aligned}$$

Moreover, by finding the sequence $e_t, \ldots, e_{t_1}$ we have seen that

$$\Gamma_i[W] \neq \Gamma_{i-1}[W]$$

for $i \in [t_1+1, t]$. This does not yet show that $\tau_H(\Gamma_{t_1}) \geq t - t_1$ because the edges added in $E(\Gamma_i[W]) \setminus E(\Gamma_{i-1}[W])$ might come from copies of H that do not fully lie in W.

For each $i \in [t_1+1, t]$ and $e \in E(\Gamma_i[W]) \setminus E(\Gamma_{i-1}[W])$ we have $e \notin U \cup \hat{W}$ because $\Gamma_i[U \cup \hat{W}] = \Gamma_{i-1}[U \cup \hat{W}]$, and thus

$$\mathscr{V}(\Gamma_i[\setminus U], e \setminus U) \cap \hat{W} \subseteq \mathscr{V}(\Gamma_t[\setminus U], e_t \setminus U) \cap \hat{W} = \varnothing.$$

Then, by Lemma 2.6.4 (2) there exists a copy H' of H with $V(H') \subseteq W$, $e \in E(H')$, and $H' - e \subseteq \Gamma_{i-1}$. This tells us that every edge in $E(\Gamma_i[W]) \setminus E(\Gamma_{i-1}[W])$ completes at least one copy of H whose vertex set is contained in W. An inductive application of this fact shows that all edges in $E(\Gamma_t[W]) \setminus E(\Gamma_{t_1}[W])$ appear in the H-process on $\Gamma_{t_1}[W]$.

Thus, by setting $t_0 := t_1$ we obtain

$$\tau_H(\Gamma_{t_0}[W]) \geq t - t_1 \geq t_2 - t_1 \geq \left\lfloor \frac{\tau'}{10(|T|+1)} \right\rfloor \geq \frac{\tau'}{\lambda_1(v(H), |U|)}.$$

We have found the desired W and t_0 provided that (2.36) holds for suitable $t \in [t_2, t_3]$ and $e_t \in E(\Gamma_t) \setminus E(\Gamma_{t-1})$.

Now suppose that no such t exists, that is, for every $t \in [t_2, t_3]$ and $e \in E(\Gamma_t) \setminus E(\Gamma_{t-1})$ one has that $\mathscr{V}(\Gamma_t[\setminus U], e \setminus U) \cap \hat{W} \neq \varnothing$, or equivalently,

$$\mathscr{V}(\Gamma_t[\setminus U], e \setminus U) = \mathscr{V}(\Gamma_t[\setminus U], \hat{W}). \tag{2.37}$$

Let $W' := U \cup \mathscr{V}(\Gamma_{t_3}[\setminus U], \hat{W})$ and note that $\Gamma_{t_3}[W' \setminus U]$ is a union of components of $\Gamma_{t_3}[\setminus U]$. Then for all $t \in [t_2+1, t_3]$,

$$\Gamma_{t-1}[\setminus W'] = \Gamma_t[\setminus W'] \quad \text{and} \quad E_{\Gamma_t}(W', V(\Gamma) \setminus W') = E_{\Gamma_{t-1}}(W', V(\Gamma) \setminus W') \tag{2.38}$$

because any edge e in $E(\Gamma_t) \setminus E(\Gamma_{t-1})$ with an endpoint outside W' must violate (2.37). Moreover,

$$c(\Gamma_{t_3}[W' \setminus U]) \leq |\hat{W}|.$$

If $c(\Gamma_{t_2}[W' \setminus U]) > |\hat{W}|$ there exists $t' \in [t_2+1, t_3]$, a copy H' of H, and $x', y' \in V(H'') \setminus U$ such that H' is completed by $x'y'$ at time t',

$$\mathscr{V}\left(\Gamma_{t'-1}[\setminus U], x'\right) \cap \mathscr{V}\left(\Gamma_{t'-1}[\setminus U], \hat{W}\right) = \varnothing, \tag{2.39}$$

and $\mathscr{V}(\Gamma_{t'-1}[\setminus U], y') \subseteq \mathscr{V}\left(\Gamma_{t'-1}[\setminus U], \hat{W}\right)$. Lemma 2.6.4 (1) applied to x', y', and H' yields $\hat{y} \in \hat{W}$ with $x'\hat{y} \in E(\Gamma_{t'})$ and a copy $\hat{H}$ of H such that $\hat{H} - x'\hat{y} \subseteq \Gamma_{t-1}$ and $V(\hat{H}) \subseteq U \cup \hat{W} \cup \mathscr{V}\left(\Gamma_{t'-1}[\setminus U], x'\right)$. Because of (2.39) the edge $x'\hat{y}$ cannot be present in $\Gamma_{t'-1}$. However, by combining (2.35), (2.37), and (2.39) we obtain that $\hat{H} - x'\hat{y}$ is contained in $\Gamma_{t'-2}$. Thus $x'\hat{y} \in E(\Gamma_{t'-1})$, so we have arrived at a contradiction.

Therefore,

$$c(\Gamma_{t_2}[W' \setminus U]) \;\le\; |\hat{W}|. \tag{2.40}$$

For every $V, \tilde{V} \subseteq V(H)$ and arbitrary subgraph $\tilde{\Gamma} \subseteq \Gamma_{t_2}[U]$ with

$$\Phi_{V,\tilde{V},\tilde{\Gamma}} := \left\{ \phi \in \mathrm{Hom}(H[V \cup \tilde{V}], \Gamma_{t_2}) : \phi \text{ injective }, \phi(H[\tilde{V}]) = \tilde{\Gamma}, \phi(V) \subseteq V(\Gamma) \setminus W' \right\} \neq \varnothing,$$

fix an embedding $\phi_{V,\tilde{V},\tilde{\Gamma}} \in \Phi_{V,\tilde{V},\tilde{\Gamma}}$. Define

$$W'' := \bigcup_{(V,\tilde{V},\tilde{\Gamma}) : \Phi_{V,\tilde{V},\tilde{\Gamma}} \neq \varnothing} \phi_{V,\tilde{V},\tilde{\Gamma}}(V) \qquad \text{and} \qquad W := W' \cup \mathscr{V}(\Gamma_{t_2}[\setminus U], W'').$$

From (2.40) we deduce

$$\begin{aligned} c(\Gamma_{t_2}[W \setminus U]) &\le c(\Gamma_{t_2}[W' \setminus U]) + |W''| \\ &\le 1 + |\hat{W}| + 2^{2v(H)+|U|+|U|^2} \cdot v(H) \\ &\le 1 + v(H)^3 \cdot |U|! \cdot 2^{v(H)} + 2^{2v(H)+|U|+|U|^2} \cdot v(H) \\ &= \lambda_2(v(H), |U|). \end{aligned}$$

We have already seen that every edge in $E(\Gamma_{t_3}) \setminus E(\Gamma_{t_2})$ has both its endpoints in W'. It remains to show that those edges appear in the H-process on $\Gamma_{t_2}[W]$. Recall that $\Gamma_{t_3}[U \cup (V(\Gamma) \setminus W')] = \Gamma_{t_2}[U \cup (V(\Gamma) \setminus W')]$ by (2.35) and (2.38). If $e \in E(\Gamma_t) \setminus E(\Gamma_{t-1})$ for some $t \in [t_2+1, t_3]$, and $\varphi \in \mathrm{Hom}(H, \Gamma_t)$ is an embedding such that $\varphi(H)$ is completed by e let

$$V := \varphi^{-1}(V(\Gamma) \setminus W') \quad , \quad \tilde{V} := \varphi^{-1}(U) \quad \text{and} \quad \tilde{\Gamma} := \varphi(H) \cap \Gamma_{t_2}[U].$$

Then $(\varphi(H) \cap \Gamma_t[W']) \cup \phi_{V,\tilde{V},\tilde{\Gamma}}(H[V \cup \tilde{V}])$ is a copy of H that is completed by e and lies in W. An induction on $t \in [t_2, t_3]$ now shows that the edges in $E(\Gamma_{t_3}) \setminus E(\Gamma_{t_2})$ are added during the H-process on $\Gamma_{t_2}[W]$. Thus, with the choice $t_0 := t_2$ we get

$$\tau_H(\Gamma_{t_0}[W]) \ge t_3 - t_2 \ge \frac{\tau'}{10(|T|+1)} \ge \frac{\tau'}{\lambda_1(v(H), |U|)},$$

which completes the proof of Lemma 2.6.5. □

We will use the last two lemmas on various occasions throughout the remaining proof of Theorem 1.2.4. Our first application is the proof of Lemma 2.6.3.

Proof of Lemma 2.6.3. Recall the notation introduced in the statement of Lemma 2.6.3. Let $t \in [\tau_H(G)]$. Employ Lemma 2.6.5 with $(\tau', \Gamma, U) = (\tau_H(G_t), G_t, V(\tilde{H}_1))$ to obtain a set $W \subset V(G)$ and $t_0 \in [\tau_H(G_t)]$ such that $V(\tilde{H}_1) \subseteq W$, and the graph $\tilde{G} := G_{\tilde{t}}[W]$, where $\tilde{t} := t + t_0$, satisfies

$$\tau_H(\tilde{G}) \ge \frac{\tau_H(G_t)}{\lambda_1(v(H), v(\tilde{H}_1))},$$

and $\tilde{G}[\backslash V(\tilde{H}_1)]$ is the union of at most $\lambda_2(v(H), v(\tilde{H}_1))$ components of $G_{\tilde{t}}[\backslash V(\tilde{H}_1)]$. Then

$$\mathrm{cdiam}(\tilde{G}[\backslash V(\tilde{H}_1)]) \leq \mathrm{cdiam}(G_{\tilde{t}}[\backslash V(\tilde{H}_1)]).$$

and the claim follows from the fact that

$$\lambda_1(v(H), v(\tilde{H}_1)) \;\leq\; \lambda_1(v(H), v(H)) \;\leq\; \lambda$$

and

$$\lambda_2(v(H), v(\tilde{H}_1)) \;\leq\; \lambda_2(v(H), v(H)) \;\leq\; \lambda.$$

□

2.6.2 Diameter reduction

In this section we use Lemmas 2.6.4 and 2.6.5 to show Lemma 2.6.2.

Let $n \in \mathbb{N}$, and let G be an n-vertex graph such that $\tau_H(G) \geq \log_{k-1} n + 1$. Fix a copy $\tilde{H}_1$ of $\tilde{H}$ at time 1 and let $U_1 := V(\tilde{H}_1)$. Any copy of P_k that occurs in $G_t[\backslash U_1]$ for some $t \geq 1$ can be extended to a copy of H minus an edge by $\tilde{H}_1$. Therefore any edge that occurs in the i^{th} graph of the C_k-process on $G[\backslash U_1]$ for some $i \geq 0$ is also present in G_{i+1}. Denote the vertex sets of the components of $G[\backslash U_1]$ by $V_1, \ldots, V_s$ where $s := c(G[\backslash U_1])$.

Recall Lemma 2.3.4, which described how the distance between two vertices decreases during a C_k-process. If $x, y \in V_j$ for some $j \in [s]$, we can find an xy-path Q in $G_1[V_j]$. Lemma 2.3.4 applied to the C_k-process $(Q_t)_{t \geq 0}$ on Q shows that for $t \geq \log_{k-1}(n)$ we have

$$\begin{aligned}
\mathrm{dist}_{G_{1+t}[\backslash U_1]}(x,y) &\leq \mathrm{dist}_{Q_t}(x,y) \\
&\leq \frac{\mathrm{dist}_Q(x,y)}{(k-1)^{\lceil \log_{k-1} n \rceil}} + k - 2 \\
&\leq \frac{n-1}{(k-1)^{\lceil \log_{k-1} n \rceil}} + k - 2 \\
&< k - 1.
\end{aligned} \tag{2.41}$$

This tells us that for $t \geq \log_{k-1} n + 1$ the diameter of any $G_t[V_j]$ is at most $k-2$. However, it is possible that during the process some of the sets $V_1, \ldots, V_s$ merge into larger components. The next claim shows that this only affects the diameter by a constant factor. Let $\hat{W}$ be as given by Lemma 2.6.4 with $(\Gamma, U) = (G_1, U_1)$. Then

$$(2k-1)|\hat{W}| \leq \lambda$$

where λ is given by (2.32).

Let $t \geq \log_{k-1} n + 1$. Any component of $G_t[\backslash U_1]$ with at most λ vertices clearly has diameter less than λ. Let $\mathscr{C}$ be a component of $G_t[\backslash U_1]$ with $v(\mathscr{C}) > \lambda$ and hence $v(\mathscr{C}) > (2k-1)|\hat{W}|$. If $V(\mathscr{C}) = V_j$ for some $j \in [s]$, then (2.41) implies $\mathrm{diam}(\mathscr{C}) \leq k - 2$.

Otherwise $V(\mathscr{C})$ is the union of at least two of the sets $V_1,\dots,V_s$. For $\hat{w}\in\hat{W}\cap V(\mathscr{C})$, we define

$$\mathscr{V}(\hat{w}) := \{\hat{w}\}\cup \bigcup_{j\in[s]:N_{G_t}(\hat{w})\cap V_j\neq\varnothing} V_j.$$

Note that $\mathscr{V}(\hat{w})\subseteq V(\mathscr{C})$ because $\mathscr{C}$ is a connected component of $G_t[\backslash U_1]$. Let $j\in[s]$. If $V_j\cap\hat{W}\neq\varnothing$ and $\hat{w}\in V_j\cap\hat{W}$ we have $V_j\subseteq\mathscr{V}(\hat{w})$ because either V_j is a singleton and thus equals $\{\hat{w}\}$, or one of the vertices in $V_j\setminus\{\hat{w}\}$ must be adjacent to $\hat{w}$ as $G[V_j]$ is connected. In the case that $V_j\cap\hat{W}=\varnothing$ we let

$$t'_j := \min\{t'\leq t : V_j\neq\mathscr{V}(G_{t'}[\backslash U_1],V_j)\}.$$

Then there exists $i\in[s]\setminus\{j\}$ such that $V_i\subset V(\mathscr{C})$ and there is an edge xy in $E(G_{t'_j})\setminus E(G_{t'_j-1})$ between some $x\in V_j$ and $y\in V_i$. We can see that $\mathscr{V}(G_{t'_j-1}[\backslash U_1],x)\neq\mathscr{V}(G_{t'_j-1}[\backslash U_1],y)$ by the definition of t'_j. Now, the hypotheses of Lemma 2.6.4 (1) are satisfied and hence x has a $G_{t'_j}$-neighbour $\hat{w}$ in $\hat{W}$. Therefore, $V_j\subseteq\mathscr{V}(\hat{w})$. Since j was arbitrary we arrive at

$$V(\mathscr{C}) = \bigcup_{w\in\hat{W}\cap V(\mathscr{C})}\mathscr{V}(w). \tag{2.42}$$

We have $\operatorname{diam}(G_t[\mathscr{V}(w)])\leq 2k-2$ for all $w\in\hat{W}\cap V(\mathscr{C})$ because

$$\operatorname{dist}_{G_t[\mathscr{V}(w)]}(w,x) \;\leq\; \operatorname{diam}(G_t[V_j])+1 \;\leq\; k-1$$

whenever $j\in[s]$ and $x\in V_j\subset\mathscr{V}(w)$. Recall that for any two vertex-disjoint, connected graphs Γ,Γ' and vertices $z\in V(\Gamma)$, $z'\in V(\Gamma')$ one has

$$\operatorname{diam}(\Gamma\cup\Gamma'\cup\{zz'\})\leq\operatorname{diam}(\Gamma)+\operatorname{diam}(\Gamma')+1. \tag{2.43}$$

Since $\mathscr{C}$ is connected, (2.42) allows us to find an ordering $w_1,\dots,w_r$ of $\hat{W}\cap V(\mathscr{C})$, where $r:=|\hat{W}\cap V(\mathscr{C})|$, such that

$$E_{\mathscr{C}}(\mathscr{V}(w_1)\cup\ldots\cup\mathscr{V}(w_{i-1}),\mathscr{V}(w_i))\neq\varnothing$$

for $2\leq i\leq r$. Thus by iterating (2.43) we obtain

$$\begin{aligned}\operatorname{diam}(\mathscr{C}) &\leq \sum_{\hat{w}\in\hat{W}\cap V(\mathscr{C})}\operatorname{diam}(G_t[\mathscr{V}(w)])+|\hat{W}\cap V(\mathscr{C})|\\ &\leq (2k-1)\cdot|\hat{W}|\\ &\leq \lambda.\end{aligned}$$

This proves Lemma 2.6.2.

2.6.3 Small diameter and few components

We will now prove Lemma 2.6.1. The rough idea is as follows: We want to fix a copy of $\tilde{H}$ in G and run the C_k-process on the remaining graph. Since every component has small diameter the

C_k-process should turn each of the few components into a clique or a complete bipartite graph within a constant number of steps. This requires Lemma 2.5.3 and the following result.

Lemma 2.6.6. *For any connected graph $\tilde{G}$, $\tau_{C_k}(G) \leq \log_{k-1} \operatorname{diam}(\tilde{G}) + 9k^2$.*

Proof. The claim is obvious for $v(\tilde{G}) \leq k$, hence we assume $v(\tilde{G}) \geq k+1$. Let $x, y \in V(\tilde{G})$ with $xy \in E(\langle \tilde{G} \rangle_{C_k})$. In the C_k-process on $\tilde{G}$ the diameter at time $\lceil \log_{k-1}(\operatorname{diam}(\tilde{G})) \rceil$ is at most $k-2$. Let P be a shortest xy-path at time $\lceil \log_{k-1}(\operatorname{diam}(\tilde{G})) \rceil$ if k is odd, and a shortest odd xy-walk at time $\lceil \log_{k-1}(\operatorname{diam}(\tilde{G})) \rceil$ if k is even. In both cases P has at most $2k-1$ vertices (recall that one can choose a shortest odd xy-walk which is a union of two paths). By Lemma 2.5.1 there exists a k-cycle C at time 2 containing x. Choose an arbitrary vertex $z \in V(\tilde{G}) \setminus V(C)$ with a $\tilde{G}$-neighbour z' on C. This is possible because $v(\tilde{G}) \geq k+1$ and $\tilde{G}$ is connected. Then $xy \in E(\langle P \cup C \cup \{zz'\} \rangle_{C_k})$ by Lemma 2.5.3, so xy is present at time $\lceil \log_{k-1} \operatorname{diam}(\tilde{G}) \rceil + 2 + \binom{3k}{2} \leq \log_{k-1} \operatorname{diam}(\tilde{G}) + 9k^2$. □

In order to run the C_k-process we need paths of length $k-1$. In fact, we want to rule out the scenario that apart from a fixed k-cycle C' in the starting graph no path of length $k-1$ occurs throughout the process, that is, $G_t[\setminus V(C')]$ is P_k-free for all t, while the $\tilde{H}$-process on $G[\setminus V(C')]$ runs for $\omega(1)$ steps.

To exclude that situation we introduce the following lemma which shows that in any sufficiently long H-bootstrap process paths of arbitrary length occur, thereby giving the necessary room for the C_k-process to run as desired. The required duration of the process will only depend on the order of H and length of the desired path and not on the order of the starting graph.

Lemma 2.6.7. *Let $\ell, m \in \mathbb{N}$. There exists $\tau'(\ell, m) > 0$ such that if Γ is a graph with $\tau_H(\Gamma) \geq \tau'(\ell, m)$, and $U \subseteq V(\Gamma)$ with $|U| \leq m$, we can find a path of length ℓ in $\Gamma_{\tau'(\ell,m)}[\setminus U]$.*

Proof. Define $\tau'(0, m) := \binom{m}{2} + 1$ and for $\ell \geq 1$ set

$$\tau'(\ell, m) := 2\lambda_1(v(H), m) \cdot \tau'\left(\ell - 1, m + \lambda_2(v(H), m) \cdot \ell\right). \tag{2.44}$$

To show that this definition results in the claimed properties we will induct on ℓ.

Let Γ be a graph and $U \subseteq V(\Gamma)$ such that $\tau_H(\Gamma) \geq \tau'(\ell, m)$ and $|U| \leq m$.

For the base case $\ell = 0$ we only need that $V(\Gamma) \neq U$ because a path of length 0 is just a vertex. Since $\binom{m}{2} + 1 \leq \tau_H(\Gamma) \leq \binom{v(\Gamma)}{2}$ we can see that $m < v(\Gamma)$ and thus $V(\Gamma) \neq U$.

Now let $\ell \geq 1$. By Lemma 2.6.5 we can find $t_0 \in [0, \tau'(\ell, m)/2]$ and $W \subset V(\Gamma)$ such that

$$\tau_H(\Gamma_{t_0}[W]) \geq \frac{\tau'(\ell, m)}{2\lambda_1(v(H), m)} = \tau'\left(\ell - 1, m + \lambda_2(v(H), m) \cdot \ell\right)$$

and

$$c(\Gamma_{t_0}[W \setminus U]) \leq \lambda_2(v(H), m). \tag{2.45}$$

The induction hypothesis allows us to define paths Q_j, $j \in [\lambda_2(v(H),m)+1]$ as follows: Let

$$\tau^* := \tau'(\ell-1, m+\lambda_2(v(H),m)\cdot\ell)$$

and observe that $\tau^* \geq \tau'(\ell-1, m+(j-1)\cdot\ell)$ for $j \in [\lambda_2(v(H),m)+1]$. Pick a path $Q_1 \subset \Gamma_{t_0+\tau^*}[W\setminus U]$ of length $\ell-1$, and, given $Q_1,\dots,Q_{j-1}$, let Q_j be a path of length $\ell-1$ in $\Gamma_{t_0+\tau^*}[W\setminus(U\cup V(Q_1)\cup\ldots\cup V(Q_{j-1}))]$. Note that $|U\cup V(Q_1)\cup\ldots\cup V(Q_{j-1})| \leq m+(j-1)\ell$, and so we can indeed find Q_j. These paths are pairwise vertex-disjoint. By (2.45) and the Pigeonhole Principle there exist $j,j' \in [\lambda_2(v(H),m)+1]$, $j<j'$ such that Q_j and $Q_{j'}$ lie in the same component of $\Gamma_{t_0+\tau^*}[W\setminus U]$. We can then find $x_j \in V(Q_j)$ and $x_{j'} \in V(Q_{j'})$ such that $\Gamma_{t_0+\tau^*}[W\setminus U]$ contains an $x_jx_{j'}$-path $Q_{jj'}$ whose set of internal vertices is disjoint from $V(Q_j)\cup V(Q_{j'})\cup U$. Consider the two subpaths from the endpoints of Q_j to x_j. One of them has length at least $\lceil\frac{\ell-1}{2}\rceil$. Similarly there is a path of length at least $\lceil\frac{\ell-1}{2}\rceil$ with endpoint $x_{j'}$ in $V(Q_{j'})$. The two latter paths together with $Q_{jj'}$ form a path of length at least ℓ in $\Gamma_{t_0+\tau^*}[W\setminus U]$. Since

$$t_0+\tau^* \;\leq\; \frac{\tau'(\ell,m)}{2} + \tau'(\ell-1, m+\lambda_2(v(H),m)\cdot\ell) \;\leq\; \tau'(\ell,m)$$

the path of length ℓ we have found also lies in $\Gamma_{\tau'(\ell,m)}[\setminus U]$. This completes the induction. □

Remark 2.6.8. The definition in (2.44) yields a tower-type bound on $\tau'(\ell,m)$ where the height of the tower only depends on ℓ.

Next, we prove a useful little lemma that tells us that if at some point during an H-process there are many vertices with the same neighbourhoods, we can ignore most of them when it comes to determining the remaining number of steps in the process.

Lemma 2.6.9. *Let Γ be a graph, $r \geq 1$, and let $Z_1,\dots,Z_r \subseteq V(\Gamma)$ be pairwise disjoint such that $N_\Gamma(z)\setminus\{z'\} = N_\Gamma(z')\setminus\{z\}$ for all $z,z' \in Z_j$, $j \in [r]$. Then there exists $Z' \subseteq Z_1\cup\ldots\cup Z_r$ such that $|Z'| \leq r\cdot v(H)$ and*

$$\tau_H(\Gamma_t) = \tau_H(\Gamma_t[(V(\Gamma)\setminus Z)\cup Z'])$$

for $t \geq 0$.

Proof. For each $j \in [r]$ with $|Z_j| > v(H)$ choose a set $Z'_j \subset Z_j$ of size $v(H)$. For all other $j \in [r]$ set $Z'_j := Z_j$. Now let $Z := Z_1\cup\ldots\cup Z_r$, $Z' := Z'_1\cup\ldots\cup Z'_r$, and $V := (V(\Gamma)\setminus Z)\cup Z'$. Moreover, let

$$\mathscr{A} := \left\{\pi : V(\Gamma)\to V(\Gamma) | \pi \text{ bijective}, \pi_{|V(\Gamma)\setminus Z} = \mathrm{id}, \pi(Z_j) = Z_j \text{ for } j \in [r]\right\}.$$

Our hypotheses imply $\mathscr{A} \subseteq \mathrm{Aut}(\Gamma)$ and that for every $X \subseteq Z$ with $|X| \leq v(H)$, there exists $\pi \in \mathscr{A}$ such that $\pi(X) \subseteq Z'$ and $\pi_{|X\cap Z'} = \mathrm{id}$.

Let $\tilde\Gamma := \Gamma[V]$ with H-process $(\tilde\Gamma_t)_{t\geq 0}$. We will show that $\tilde\Gamma_t = \Gamma_t[V]$ for all $t \geq 0$. We clearly have $\tilde\Gamma_t \subseteq \Gamma_t[V]$ since $\tilde\Gamma \subseteq \Gamma$. If $t \geq 1$, $\tilde\Gamma_{t-1} = \Gamma_{t-1}[V]$, and $e \in E(\Gamma_t[V])\setminus E(\Gamma_{t-1})$ completes a copy H' of H with $V(H') \not\subseteq V$ we can choose $\pi \in \mathscr{A}$ with $\pi(V(H')\cap Z\setminus Z') \subseteq Z'$ and $\pi(e) = e$, and hence obtain

$$\pi(H') - e = \pi(H'-e) \subseteq \Gamma_{t-1} = \tilde\Gamma_{t-1}.$$

Thus, $\Gamma_t[V] \subseteq \tilde{\Gamma}_t$.

As $\Gamma_{\tau_H(\Gamma)}$ is H-stable, so is $\Gamma_{\tau_H(\Gamma)}[V]$. For this reason, $\tau_H(\Gamma) \geq \tau_H(\tilde{\Gamma})$.

If $e \in E(\Gamma_t) \setminus E(\Gamma_{t-1})$ for some $t \geq 0$ and H' is a copy of H in Γ_t completed by e we can find $\pi \in \mathscr{A}$ such that $\pi(V(H')) \subseteq V$. Then $\pi(e) \notin E(\tilde{\Gamma}_{t-1})$ because $\pi \in \mathrm{Aut}(\Gamma_{t-1})$ by 2.1.1, and $\pi(e) \in E(\tilde{\Gamma}_t)$ since $\pi \in \mathrm{Aut}(\Gamma_t)$. This holds in particular for $e \in E(\Gamma_{\tau_H(\Gamma)}) \setminus E(\Gamma_{\tau_H(\Gamma)-1})$, so $\tau_H(\Gamma) \leq \tau_H(\tilde{\Gamma})$. Now, for each $t \geq 0$, we obtain

$$\tau_H(\Gamma_t) \;=\; \tau_H(\Gamma) - t \;=\; \tau_H(\tilde{\Gamma}) - t \;=\; \tau_H(\tilde{\Gamma}_t) \;=\; \tau_H(\Gamma_t[V]).$$

$\square$

With Lemmas 2.6.7 and 2.6.9 at our disposal we are ready to complete the proof of Lemma 2.6.1.

For $\ell, m \in \mathbb{N}$ let $\tau'(\ell, m)$ as given by Lemma 2.6.7. Fix a copy $\tilde{H}_0$ of $\tilde{H}$ in G such that

$$c\left(G[\setminus V(\tilde{H}_0)]\right) \leq \lambda \qquad \text{and} \qquad \mathrm{cdiam}\left(G[\setminus V(\tilde{H}_0)]\right) \leq \lambda.$$

We set

$$U_0 := V(\tilde{H}_0) \quad , \quad W_0 := V(G) \quad , \quad \tau_0 := 0 \quad \text{and} \quad s := c(G[W_0 \setminus U_0]).$$

For $j \in [s]$, we are going to define $\tau_j \in \mathbb{N}_0$, $W_j \subseteq V(G)$, and $U_j \subseteq W_j$ with the following properties:

- $c(G_{\tau_j}[W_j \setminus U_j]) \leq s - j$,
- $\tau_H(G_t[W_j]) = \tau_H(G_t)$ for $t \geq \tau_j$,
- $U_{j-1} \subseteq U_j$,
- $|U_j| \leq |U_{j-1}| + 2^{|U_{j-1}|+1} \cdot v(H)$.

Suppose that $\tau_1, \ldots, \tau_{j-1}, W_1, \ldots, W_{j-1}, U_1, \ldots, U_{j-1}$ with the properties above are given. Lemma 2.6.7 with $(\ell, m, \Gamma, U) = (v(H), |U_{j-1}|, G_{\tau_{j-1}}[W_{j-1}], U_{j-1})$ shows that $G_{\tau_{j-1}+\tau'(v(H),|U_{j-1}|)}[W_{j-1} \setminus U_{j-1}]$ has a component $\mathscr{C}^j$ that contains a path of length $v(H)$. Define

$$\tau_j := \tau_{j-1} + \tau'(v(H), |U_{j-1}|) + \tau_{C_k}(\mathscr{C}^j) + 2.$$

If $c(G_{\tau_j}[W_{j-1} \setminus U_{j-1}]) < c(G_{\tau_{j-1}}[W_{j-1} \setminus U_{j-1}])$ we can simply set $W_j := W_{j-1}$, $\tau_j := \tau_{j-1}$, and $U_j := U_{j-1}$. For this reason we now assume that

$$c(G_{\tau_j}[W_{j-1} \setminus U_{j-1}]) = c(G_{\tau_{j-1}}[W_{j-1} \setminus U_{j-1}]). \tag{2.46}$$

Recall that every k-cycle with vertices in $V(\mathscr{C}^j)$ that occurs during the process can be extended to a copy of H by $\tilde{H}_0$. Therefore $\langle \mathscr{C}^j \rangle_{C_k} \subseteq G_{\tau_j - 2}[V(\mathscr{C}^j)]$. Since $\mathscr{C}^j$ contains a copy of $P_{v(H)}$, we can see that $\langle \mathscr{C}^j \rangle_{C_k}$ must be a complete or a complete bipartite graph by Lemma 2.5.3.

If $\langle \mathscr{C}^j \rangle_{C_k}$ is a complete graph or $\langle \mathscr{C}^j \rangle_{C_k} \neq G_{\tau_j - 2}[V(\mathscr{C}^j)]$, Lemma 2.5.2 tells us that $V(\mathscr{C}^j)$ is a clique at time τ_j. Then for any $x \in V(\mathscr{C}^j)$, the closed G_{τ_j}-neighbourhood of x has the form

$V(\mathscr{C}^j) \cup U'$, where $U' \subseteq U_{j-1}$, so there exists a partition $V(\mathscr{C}^j) = X_1 \cup \ldots \cup X_{2^{|U_{j-1}|}}$ such that the hypotheses of Lemma 2.6.9 are satisfied for $\Gamma = G_{\tau_j}[W_{j-1}]$ and $(Z_1, \ldots, Z_r) = (X_1, \ldots, X_{2^{|U_{j-1}|}})$. This gives us a set $Z' \subset V(\mathscr{C}^j)$ such that for $t \geq \tau_j$,

$$\tau_H(G_t[W_{j-1}]) = \tau_H(G_t[(W_{j-1} \setminus V(\mathscr{C}^j)) \cup Z']).$$

If $\langle \mathscr{C}^j \rangle_{C_k}$ is complete bipartite and $\langle \mathscr{C}^j \rangle_{C_k} = G_{\tau_j - 2}[V(\mathscr{C}^j)]$, let $X, Y \subset V(\mathscr{C}^j)$ be the partite sets of $\langle \mathscr{C}^j \rangle_{C_k}$. The $G_{\tau_j - 2}$-neighbourhood of any $x \in X$ is of the form $Y \cup U'$ for some $U' \subseteq U_{j-1}$, and similarly, the $G_{\tau_j - 2}$-neighbourhood of $y \in Y$ can be written as $X \cup U'$ for a suitable $U' \subseteq U_{j-1}$. We can find partitions $X = X_1 \cup \ldots \cup X_{2^{|U_{j-1}|}}$ and $Y = Y_1 \cup \ldots \cup Y_{2^{|U_{j-1}|}}$ such that the hypotheses of Lemma 2.6.9 are satisfied for $\Gamma = G_{\tau_j}[W_{j-1}]$ and $(Z_1, \ldots, Z_r) = (X_1, \ldots, X_{2^{|U_{j-1}|}}, Y_1, \ldots, Y_{2^{|U_{j-1}|}})$. As in the earlier case we obtain a set $Z' \subseteq V(\mathscr{C}^j)$ such that for $t \geq \tau_2$ one has

$$\tau_H\left(G_t[W_{j-1}]\right) = \tau_H\left(G_t[(W_{j-1} \setminus V(\mathscr{C}^j)) \cup Z']\right).$$

In both cases we define

$$U_j := U_{j-1} \cup Z' \qquad , \qquad W_j := (W_{j-1} \setminus V(\mathscr{C}^j)) \cup Z'.$$

With these definitions and the fact that $\tau_j \geq \tau_{j-1}$ we obtain both

$$\tau_H(G_t[W_j]) = \tau_H(G_t[W_{j-1}]) = \tau_H(G_t)$$

for any $t \geq \tau_j$ and

$$W_j \setminus U_j = W_{j-1} \setminus (U_{j-1} \cup V(\mathscr{C}^j)).$$

By construction, $G_{\tau_{j-1}}[W_{j-1} \setminus U_{j-1}]$ is a union of at most $s - (j-1)$ components of $G_{\tau_{j-1}}[\setminus U_0]$, and by assumption, (2.46) holds. So,

$$\begin{aligned}
c(G_{\tau_j}[W_j \setminus U_j]) &= c(G_{\tau_j}[W_{j-1} \setminus U_{j-1}]) - 1 \\
&= c(G_{\tau_{j-1}}[W_{j-1} \setminus U_{j-1}]) - 1 \\
&\leq s - (j-1) - 1 \\
&= s - j.
\end{aligned}$$

Moreover

$$|U_j| = |U_{j-1}| + |Z'| \leq |U_{j-1}| + 2^{|U_{j-1}|+1} \cdot v(H).$$

This completes the construction of $W_1, \ldots, W_s$, $\tau_1, \ldots, \tau_s$, and $U_1, \ldots, U_s$.

Now, $c(G_{\tau_s}[W_s \setminus U_s]) = 0$ yields $W_s = U_s$ and hence $\tau_H(G_{\tau_s}[W_s]) \leq \binom{|U_s|}{2}$. We thus arrive at

$$\begin{aligned}
\tau_H(G) &= \tau_s + \tau_H(G_{\tau_s}) \\
&= \tau_s + \tau_H(G_{\tau_s}[W_s]) \\
&= \tau_0 \,+\, \sum_{j=1}^{s} (\tau_j - \tau_{j-1}) \,+\, \binom{|U_s|}{2} \\
&= \sum_{j=1}^{s} (\tau'(v(H), |U_{j-1}|) + \tau_{C_k}(\mathscr{C}^j) + 2) \,+\, \binom{|U_s|}{2} \\
&\leq \sum_{j=1}^{s} \tau'(v(H), |U_{j-1}|) + s \cdot (\log_{k-1} \lambda + 9k^2) + 2s \,+\, \binom{|U_s|}{2}
\end{aligned}$$

where the last inequality uses Lemma 2.6.6. The claim now follows with

$$\mu := \sum_{j=1}^{s} \tau'(v(H), |U_{j-1}|) + s \cdot (\log_{k-1} \lambda + 9k^2) + 2s \,+\, \binom{|U_s|}{2}$$

and by the observation that μ depends only on λ, $v(H)$, s, $|U_0|, \ldots, |U_s|$, and k, while each of the latter depends solely on H. The dependence of s on H is given by the hypotheses of Lemma 2.6.1. This finishes the proof of Lemma 2.6.1.

2.7 The necessity of small degrees for sublinear running time

Let us recall the first part of the statement of Theorem 1.2.6. We are given a graph H such that every component H' of H satisfies $\delta(H') \geq 2$ and $\Delta(H') \geq 3$. We will show that the maximum running time $M_H(n)$ is asymptotically at least linear, i.e.

$$M_H(n) = \Omega(n).$$

The second part on bipartite H shall be handled separately in Section 2.7.1 as its proof will essentially consist of going through the proof of the first part and replacing all cliques by complete bipartite graphs. Let $r := |V(H)| - 1$ and write δ for $\delta(H)$.

Claim 2.7.1. *Let G be an n-vertex graph with an ordering $v_1, \ldots, v_n$ of its vertices such that $\{v_1, \ldots, v_r\}$ is a clique and for every $i \in [r+1, n]$, $|\{j \in [i-1] : v_j v_i \in E(G)\}| \geq \delta - 1$, i.e. every vertex but the first $\delta - 1$ is adjacent to at least $\delta - 1$ vertices preceding it. Then in the H-process $(G_t)_{t \geq 0}$ on G, for every $t \geq 0$, $\{v_1, \ldots, v_{r+t}\}$ is a clique in G_t. In particular, $G_{n-r} = K_n$.*

Proof. Observe that a vertex $v \in V(G)$ with $\delta - 1$ neighbours in an r-clique at time $t \geq 0$ will be adjacent to every vertex of the clique at time $t+1$ because every map from $V(H)$ to $V(G)$ that sends a minimum-degree vertex to v and all the other vertices to the clique in an arbitrary way is an embedding of some H^-. Since $\{v_1, \ldots, v_r\}$ is a clique at time 0, the claim follows by a standard inductive argument. □

The next step is to find a graph G that satisfies the hypotheses of Claim 2.7.1 and for which the clique number increases by at most a fixed constant in each step of the H-process.

The starting graph G will be the union of a $K_{\delta-1}$, a graph G' which we will define in a moment and a complete bipartite graph between the vertices of K_r and some vertices of G'. Let $\ell \in \mathbb{N}$ and denote the standard basis vectors of $\mathbb{Z}^r_{\delta-1}$ by $e_1, \ldots, e_r$. For $r < j \leq \ell$, let $e_j := e_{j \bmod r}$. Let G' be given by

$$
\begin{aligned}
V(G') &= [\ell] \times \mathbb{Z}^r_{\delta-1} \\
E(G') &= \left\{ \{(j,x), (j+1, x+\lambda e_j)\} : j \in [\ell-1], x \in \mathbb{Z}^r_{\delta-1}, \lambda \in \mathbb{Z}_{\delta-1} \right\}
\end{aligned}
$$

and write $W_j := \{j\} \times \mathbb{Z}^r_{\delta-1}$ for $j \in [\ell-1]$. The graph G' is the union of $\ell-1$ pairwise isomorphic $(\delta-1)$-regular bipartite graphs with partite sets W_j and W_{j+1} for $j \in [\ell-1]$.

Claim 2.7.2. *For any $e \in E(H)$ and any connected component H' of $H-e$, G' is H'-free.*

Proof. Let $e \in E(H)$, and let H' be a connected component of $H-e$. There must be a cycle in H'. If not, H' would be a tree and thus have at least two leaves. Since $H-e$ can have at most two vertices of degree $\delta-1$, H' would have precisely two leaves, that is, H' would be a path. The endpoints of H' would be the endpoints of e because $\delta = 2$. But then $H' \cup \{e\}$ would be a cycle, so H would have a component with maximum degree two, which contradicts our assumptions $\Delta(H) \geq 3$.

If $\delta = 2$, G' is just a path of length $\ell-1$ so in this case it is clearly H'-free. For the rest of the proof of Claim 2.7.2 we assume that $\delta \geq 3$. Suppose there was a copy of H' in G'. For the sake of notational simplicity we denote that copy by H', too. Let

$$
\begin{aligned}
j_0 &:= \min \left\{ j \in [\ell] : V(H') \cap W_j \neq \varnothing \right\} \\
j_1 &:= \max \left\{ j \in [\ell] : V(H') \cap W_j \neq \varnothing \right\}
\end{aligned}
$$

and for $j \geq j_0$, define

$$
H'[j_0, j] := H' \left[V(H') \cap ([j_0, j] \times \mathbb{Z}^r_{\delta-1}) \right].
$$

There exists a path from a vertex in W_{j_0} to a vertex in W_{j-1} in H'. Such a path must intersect W_j for $j_0 \leq j \leq j_1$. Thus $j_1 - j_0 \leq r$. By definition of G' any vertex in W_{j_0} has precisely $\delta-1$ G'-neighbours in W_{j_0+1}. Similarly, each vertex in W_{j_1} has precisely $\delta-1$ G'-neighbours in W_{j_1-1}. Since $H-e$ has at most two vertices of degree $\delta-1$ we get

$$
\left| V(H') \cap W_{j_0} \right| = 1 = \left| V(H') \cap W_{j_1} \right|.
$$

Then the degree of any vertex in $V(H') \cap (W_{j_0+1} \cup \ldots \cup W_{j_1-1})$ is at least δ. Take the unique $x_0, x_1 \in \mathbb{Z}^r_{\delta-1}$ with $(j_0, x_0), (j_1, x_1) \in V(H')$.

Now we show that for every $j \in [j_0, j_1]$, $H'[j_0, j]$ must be a full $(\delta-1)$-ary tree with root (j_0, x_0) whose set of leaves is $V(H') \cap (\{j\} \times \mathbb{Z}^r_{\delta-1})$. We induct on $j \in [j_0, j_1]$. As $\left|V(H') \cap W_j\right| = 1$, $H'[j_0, j_0]$ is just an isolated vertex and $H'[j_0, j_0+1]$ is a copy of $K_{1,\delta-1}$. Let $j_0 + 1 < j \leq j_1$.

Every vertex $(j-1,x)$ in $V(H')\cap W_{j-1}$ has precisely one $H'[j_0,j-1]$-neighbour. Consequently, the remaining H'-neighbours of $(j-1,x)$ must lie in W_j. There are at least $\delta-1$ of them so the H'-neighbours of $(j-1,x)$ in W_j are precisely its $\delta-1$ G'-neighbours in W_j. It remains to prove that the neighbourhoods of any two distinct $(j-1,x),(j-1,y)\in V(H')$ are disjoint. The unique paths from $(j-1,x)$ and $(j-1,y)$ to the root (j_0,x_0) in $H'[j_0,j-1]$ amount to $\lambda_{j_0},\ldots,\lambda_{j-1},\mu_{j_0},\ldots,\mu_{j-1}\in\mathbb{Z}_{\delta-1}$ with $(\lambda_{j_0},\ldots,\lambda_{j-1})\neq(\mu_{j_0},\ldots,\mu_{j-1})$ and

$$x=x_0+\lambda_{j_0}e_{j_0}+\ldots+\lambda_{j-1}e_{j-1}\qquad,\qquad y=x_0+\mu_{j_0}e_{j_0}+\ldots+\mu_{j-1}e_{j-1}.$$

Thus for every $z\in\mathbb{Z}^r_{\delta-1}$ with $x-z=\lambda_je_j$ for some $\lambda_j\in\mathbb{Z}_{\delta-1}$ one has

$$y-z=y-x+x-z=(\mu_{j_0}-\lambda_{j_0})e_{j_0}+\ldots+(\mu_{j-1}-\lambda_{j-1})e_{j-1}+\lambda_je_j\notin\mathbb{Z}_{\delta-1}\cdot e_j.$$

Here we used that $j-j_0\leq j_1-j_0\leq r$ so $e_{j_0},\ldots,e_j$ are linearly independent. This completes the induction.

As $H'=H'[j_0,j_1]$, we have arrived at a contradiction since trees have minimum degree one while $\delta(H')\geq\delta-1\geq 2$. Therefore G' is H'-free. $\square$

Define the starting graph G by

$$\begin{aligned}V(G)&=\{v_1,\ldots,v_r\}\,\cup\,V(G'),\\ E(G)&=\binom{\{v_1,\ldots,v_r\}}{2}\cup E(G')\cup\left\{\{v_j,(1,x)\}:j\in[r],x\in\mathbb{Z}^r_{\delta-1}\right\},\end{aligned}$$

where $v_1,\ldots,v_r$ are r newly introduced vertices, and let $(G_t)_{t\geq0}$ be the H-process on G. Pick an ordering of the vertices of $V(G)$ such that $v_1,\ldots,v_r$ are the first r vertices and for $j\in[\ell-1]$ each vertex in W_j precedes each vertex in W_{j+1}. With such an ordering G satisfies the hypotheses of Claim 2.7.1. Write

$$\begin{aligned}G'[j,\ell]&:=G'\left[V(G')\cap([j,\ell]\times\mathbb{Z}^r_{\delta-1})\right]\quad\text{and}\\ G_t[j,\ell]&:=G_t\left[V(G)\cap([j,\ell]\times\mathbb{Z}^r_{\delta-1})\right]\end{aligned}$$

for any $j\in[\ell],t\geq0$, and let t_j be the smallest positive integer such that $E(G_{t_j})\setminus E(G')$ contains an edge touching $[j,\ell]\times\mathbb{Z}^r_{\delta-1}$. The t_j are non-decreasing and well-defined as Claim 2.7.1 guarantees that every vertex of G receives a new neighbour at some stage of the process. We claim that $t_j>t_{j-r}$ for all $j\in[r+1,\ell]$. Suppose there existed $j\in[r+1,\ell]$ with $t_j=t_{j-r}$. At time t_j-1 we can find a copy H' of $H-e$ for some $e\in E(H)$ with a vertex w in $[j,\ell]\times\mathbb{Z}^r_{\delta-1}$. Since $t_j=t_{j-r}$, there are no edges in $E(G_{t_j-1})\setminus E(G')$ with an endpoint in $[j-r,\ell]\times\mathbb{Z}^r_{\delta-1}$, hence $G_{t_j-1}[j-r,\ell]=G'[j-r,\ell]$. The r^{th} neighbourhood of $[j,\ell]\times\mathbb{Z}^r_{\delta-1}$ in G' is $[j-r,\ell]\times\mathbb{Z}^r_{\delta-1}$. But then $H'\cap G_{t_j-1}[j-r,\ell]$ contains the r^{th} H'-neighbourhood of w and thereby a copy of a connected component of $H-e$ in G'. This contradicts Claim 2.7.2.

The order of G is $r+\ell\cdot(\delta-1)^r$. Therefore

$$M_H(r+\ell\cdot(\delta-1)^r) \;\geq\; \tau_H(G) \;\geq\; t_\ell \;\geq\; t_{r\cdot\lfloor \ell/r\rfloor} \;\geq\; \sum_{s=1}^{\lfloor \ell/r\rfloor-1} t_{(s+1)r}-t_{sr} \;\geq\; \left\lfloor \frac{\ell}{r}\right\rfloor -1.$$

Given a sufficiently large $n\in\mathbb{N}$, let $\ell := \lfloor \frac{n-r}{(\delta-1)^r}\rfloor$. During the H-process on any graph no isolated vertex receives a new neighbour since $\delta\geq 2$. Therefore, $M_H(n)$ is non-decreasing in n and

$$M_H(n)\geq M_H(r+\ell\cdot(\delta-1)^r)\geq \left\lfloor \frac{\ell}{r}\right\rfloor -1=\Omega(n).$$

2.7.1 A variant for bipartite graphs.

The second part of Theorem 1.2.6 states that if H is bipartite we may obtain the linear lower bound by choosing a bipartite starting graph. Observe that the graph G' is already bipartite. In order to replace G by a bipartite starting graph it will be sufficient to replace the r-clique $\{v_1,\dots,v_r\}$ by a complete bipartite graph with partite sets of size r each. The following is the bipartite analogue of Claim 2.7.1.

Claim 2.7.3. *Let G be a bipartite graph with partite sets X,Y. If $v_1,\dots,v_n$ is an ordering of $V(G)$ such that $\{v_1,\dots,v_r\}\subseteq X$, $\{v_{r+1},\dots,v_{2r}\}\subseteq Y$ and for every $i>2r$, either*

$$|\{j\in[i-1]: v_j\in Y, v_iv_j\in E(G)\}|\geq \delta-1 \quad \textit{or} \quad |\{j\in[i-1]: v_j\in X, v_jv_i\in E(G)\}|\geq \delta-1$$

then at time t in the H-process on G, $\{v_1,\dots,v_{2r+t}\}$ is the vertex set of a (not necessarily induced) complete bipartite graph.

Claim 2.7.3 follows from an inductive application of the observation that in the H-process on G, any vertex with at least $\delta-1$ neighbours in a partite set S of a complete bipartite graph with at least r vertices in each part will be adjacent to all remaining vertices in S after one more step.

Define the bipartite starting graph G via

$$\begin{aligned}
&V(G)=\{v_1,\dots,v_{2r}\}\,\dot\cup\, V(G') \text{ and}\\
&E(G)=\big\{v_iv_j : i\in[r], j\in[r+1,2r]\big\}\cup E(G')\cup\big\{\{v_j,(1,x)\} : j\in[r+1,2r], x\in\mathbb{Z}_{\delta-1}^r\big\}.
\end{aligned}$$

Again, choose an ordering of $V(G)$ such that $v_1,\dots,v_{2r}$ are the first $2r$ vertices and the vertices in $\{j\}\times\mathbb{Z}_{\delta-1}^r$ precede the vertices in $\{j+1\}\times\mathbb{Z}_{\delta-1}^r$. Then the hypotheses of Claim 2.7.3 are satisfied with

$$X=\{v_1,\dots,v_r\}\cup\bigcup_{j\in[\ell]\text{ odd}}\{j\}\times\mathbb{Z}_{\delta-1}^r \qquad , \qquad Y=\{v_{r+1},\dots,v_{2r}\}\cup\bigcup_{j\in[\ell]\text{ even}}\{j\}\times\mathbb{Z}_{\delta-1}^r.$$

The rest of the proof does not differ from the proof of the first part of Theorem 1.2.6 after Claim 2.7.1 was invoked.

2.8 Minimum degree one does not imply constant running time

The proof of the constant running times for trees may suggest that any graph with a degree one vertex has constant maximum running time. However, this is not the case as the following construction points out. The idea is to start with a graph $\tilde{H}$ satisfying $M_{\tilde{H}}(n) = \Omega(n)$ and then modifying it without decreasing the asymptotic running time such that we end up with a graph H with minimum degree one. Let us first give an informal description of a construction that is slightly weaker than Proposition 1.2.8 but captures the essence of the proof. Choose a 2-edge-connected bipartite graph $\tilde{H}$ and a bipartite graph $\tilde{G}$ such that $\tau_{\tilde{H}}(\tilde{G}) = \Omega(n)$. Such graphs exist by Theorem 1.2.6. Consider the disjoint union of $\tilde{H}$ and a clique of size 6. The $(\tilde{H} \sqcup K_6)$-process on $\tilde{G} \sqcup K_6$ stabilises after precisely $\tau_{\tilde{H}}(\tilde{G})$ steps because for every copy of $\tilde{H} \sqcup K_6$ that occurs throughout the process, the role of the clique of size six is always played by the vertices of the unique K_6 in $\tilde{G} \sqcup K_6$. A more rigorous approach would involve an inductive proof of the statement that the i^{th} graph of the process is just $\tilde{G}_t \sqcup K_6$ where $(\tilde{G}_t)_{t\geq 0}$ is the $\tilde{H}$-process on $\tilde{G}$. The crucial ingredient here is that all the $\tilde{G}_t$ are bipartite and hence cannot intersect a clique in more than two vertices. Next we add universal vertices to our graphs: Let H^* be the graph obtained from $\tilde{H} \sqcup K_6$ by introducing a new vertex that is adjacent to every other vertex, and let G^* be obtained from $\tilde{G} \sqcup K_6$ in the same manner. The t^{th} graph in the H^*-process on G^* is $\tilde{G}_t \sqcup K_6$ together with the universal vertex of G^* because every embedding of H^* minus an edge must send the universal vertex of H^* to the universal vertex of G^*. Now build H by appending a new vertex to the universal vertex of H^*. In the H-process on G^* the universal vertex of H must still be sent to the universal vertex of G^*. This implies

$$\tau_H(G^*) = \tau_{H^*}(G^*) = \tau_{\tilde{H} \sqcup K_6}(\tilde{G} \sqcup K_6) = \tau_{\tilde{H}}(\tilde{G}) = \Omega(n)$$

so we have found a graph with minimum degree one and at least linear running time.

From the perspective of keeping $\Delta(H)$ as small as possible the choices above are wasteful. To prove Proposition 1.2.8 we have to adjust the construction. However, as pointed out above, the underlying idea of starting with bipartite $\tilde{H}$, $\tilde{G}$ and building H and G around them remains the same.

Definition 2.8.1. We say that a bootstrap process $(G_t)_{t\geq 0}$ *simulates* another bootstrap process $(\tilde{G}_t)_{t\geq 0}$ if

$$\tilde{G}_0 \subseteq G_0 \qquad \text{and} \qquad G_t[V(G_0)] = \tilde{G}_t, \quad t \geq 0.$$

It is a simple observation that the running time of a process is at least the running time of any other process it simulates. We now give a recipe for constructing the desired graph of minimum degree one and maximum degree three that allows us to simulate the bootstrap process of a given bipartite and 2-edge-connected graph on any bipartite starting graph. In the following we refer to the graph obtained by deleting an arbitrary edge from K_4 as the *diamond*.

Let H' be a graph with vertices u_i, $i \in [0,17]$, and

$$E(H') := \{u_i u_{i+1} : i \in [0,16]\} \cup \{u_{17}u_0, u_1u_3, u_2u_4, u_5u_7, u_6u_8, u_{10}u_{12}, u_{11}u_{13}, u_{14}u_{16}, u_{15}u_{17}\}.$$

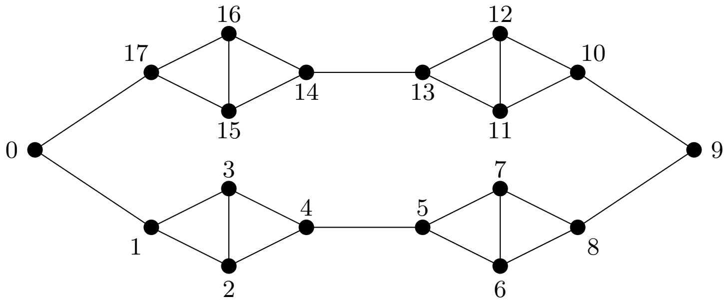

FIGURE 2.3: A drawing of H'. For clarity the vertex u_i is just labelled by i.

A visualisation of H' is shown in Figure 2.3. Note that the maps

$$\sigma : V(H') \to V(H'), u_i \mapsto u_{(i+9) \bmod 18} \quad \text{and} \quad \rho : V(H') \to V(H'), u_i \mapsto u_{(18-i) \bmod 18} \tag{2.47}$$

are automorphisms of H'. Let

$$U := \{u_0, \ldots, u_9\} \quad \text{and} \quad W := \{u_9, \ldots, u_{17}, u_0\}, \tag{2.48}$$

so $H'[U]$ and $H'[W]$ are isomorphic and contain two vertex-disjoint diamonds each.

Claim 2.8.2. *H' is itself H'-stable.*

Proof. The claim is equivalent to the assertion that for every $e_1 \in E(H')$ and $e_2 \in \binom{V(H')}{2} \setminus E(H')$, H' and $H'' := H' - e_1 \cup \{e_2\}$ are not isomorphic. As $H' = H'[U] \cup H'[W]$ and $H'[U] \cong H'[W]$ it suffices to consider the case $e_1 \subset U$. Let x, y be the endpoints of e_1. There are precisely two vertices of degree two in H' while all other vertices have degree three. If one of the two endpoints of e_2 does not lie in $\{x, y, u_0, u_9\}$ then $\Delta(H'') = 4$ and thus H'' cannot be isomorphic to H'. For this reason $e_2 \subset \{x, y, u_0, u_9\}$. As $e_2 \neq e_1$ we have that $e_2 \cap \{u_0, u_9\} \neq \varnothing$. Since $\sigma(u_0) = u_9$ and $\sigma \in \mathrm{Aut}(H')$ we may without loss of generality suppose that $u_0 \in e_2$. Denote the other endpoint of e_2 by z. Recall that $e_2 \notin E(H')$ so $z \neq u_1$. Observe that H' does not contain any ℓ-cycle for $\ell \in [5, 13]$. If $z = u_9$ then H'' contains an induced 8-cycle and thus is not isomorphic to H'. Now suppose that $z \in \{x, y\}$ and $z \neq u_9$. In that case $d_{H''}(u_9) = 2$. In H' the distance between the two vertices of degree two is seven. Hence if $xy \neq u_4u_5$ we have $\mathrm{dist}_{H''}(u_9, x) \leq 6$ and $\mathrm{dist}_{H''}(u_9, y) \leq 6$, so H'' and H' are not isomorphic. It remains to check the case $xy = u_4u_5$. Now H'' contains a 5-cycle when $z = u_4$, and an 11-cycle when $z = u_5$. Thus the claim follows. $\square$

Given a bipartite 2-edge-connected graph $\tilde{H}$ and $\tilde{v} \in V(\tilde{H})$ let H be defined by

$$\begin{aligned} V(H) &= V(\tilde{H}) \cup V(H') \cup \{z\}, \\ E(H) &= E(\tilde{H}) \cup E(H') \cup \{\{u_0, z\}, \{u_9, \tilde{v}\}\}, \end{aligned}$$

where z is a newly introduced vertex and, we assume that $V(\tilde{H}) \cap V(H') = \varnothing$. The H-degree of any $u \in V(H')$ is at most 3 while $d_{\tilde{H}}(\tilde{v}) + 1 \geq \delta(\tilde{H}) + 1 \geq 3$. Therefore $\delta(H) = 1$ and $\Delta(H) = \max\{\Delta(\tilde{H}), d_{\tilde{H}}(\tilde{v}) + 1\}$.

Let $\tilde{G}$ be a bipartite graph maximising $\tau_{\tilde{H}}(\tilde{G})$ among all bipartite starting graphs on $n-18$ vertices. We now construct a graph G on n vertices whose H-process simulates the $\tilde{H}$-process on $\tilde{G}$. Let H'_0 be isomorphic to H' and vertex-disjoint from $\tilde{G}$, and fix an isomorphism $\varphi' : H' \to H'_0$. Define G via

$$\begin{aligned} V(G) &= V(\tilde{G}) \cup V(H'_0), \\ E(G) &= E(\tilde{G}) \cup E(H'_0) \cup \left\{\varphi'(u_0)x : x \in V(\tilde{G})\right\} \cup \left\{\varphi'(u_9)x : x \in V(\tilde{G})\right\}, \end{aligned} \tag{2.49}$$

and set $u'_i := \varphi'(u_i)$ for $i = 0, \dots, 17$. We want to show that the H-process $(G_t)_{t\geq 0}$ on G simulates the $\tilde{H}$-process $(\tilde{G}_t)_{t\geq 0}$ on $\tilde{G}$. The condition $\tilde{G} \subseteq G$ is fulfilled by construction and the inclusion $\tilde{G}_t \subseteq G_t[V(\tilde{G})]$ follows from Observation 2.1.1 and $\tilde{G} \subseteq G$. It remains to verify $G_t[V(\tilde{G})] \subseteq \tilde{G}_t$ for $t \geq 0$. To do so we aim for the following slightly stronger statement:

$$E(G_t) \setminus E(G) \subseteq E(\tilde{G}_t) \setminus E(\tilde{G}) \qquad \text{for each } t \geq 0. \tag{2.50}$$

If (2.50) holds, then

$$\begin{aligned} G_t[V(\tilde{G})] &= \tilde{G} \cup \left\{ e \in E(G_t) \setminus E(G) : e \subset \binom{V(\tilde{G})}{2} \right\} \\ &\subseteq \tilde{G} \cup E(\tilde{G}_t) \setminus E(\tilde{G}) \\ &= \tilde{G}_t. \end{aligned}$$

To prove (2.50) we induct on t. The base case $t = 0$ is vacuously true. Now suppose that $t \geq 1$ and $E(G_{t-1}) \setminus E(G) \subseteq E(\tilde{G}_{t-1}) \setminus E(\tilde{G})$. Any copy of $H - e$, where $e \in E(H)$, contains a copy of H' or $H' - e$. For this reason the following claim will be helpful.

Claim 2.8.3. *Let $e' \in E(H')$. Every embedding $\phi : H' - e' \to G_{t-1}$ satisfies $\phi(V(H')) = V(H'_0)$ and $\{\phi(u_0), \phi(u_9)\} = \{u'_0, u'_9\}$.*

Proof. Let $\phi : H' - e' \to G_{t-1}$ be an embedding, and let $H'' := \phi(H' - e')$. Depending on whether e' lies in a diamond of H', there are either four or precisely three vertex-disjoint diamonds in H''. In the latter case H'' is 2-connected. We will show that the only diamonds that can occur in H'' are those from H'_0.

Due to the induction hypothesis the only vertices in $V(H'_0)$ with G_{t-1}-neighbours in $V(\tilde{G})$ are u'_0 and u'_9. Since $\tilde{G}_{t-1}$ is bipartite any triangle in G_{t-1} involving vertices of $\tilde{G}_{t-1}$ must use u'_0 or u'_9. Thus any diamond in G_{t-1} apart from those in H'_0 contains u'_0 or u'_9.

It is not possible that u'_0 and u'_9 lie in distinct diamonds of H'' for otherwise both had all three H''-neighbours in $V(\tilde{G}_{t-1})$ and none outside $V(\tilde{G}_{t-1})$. This would then contradict the connectedness of H'' since the remaining diamonds of H'', of which there is at least one, would have to lie in $V(H'_0) \setminus \{u'_0, u'_9\}$.

We have to rule out two more situations: The first one is when u'_0 occurs in a diamond of H'' while u'_9 does not (by symmetry this also covers the case that u'_9 lies in a diamond of H'' and u'_0 does not). Second, both u'_0 and u'_9 might lie in a common diamond of H''.

Suppose that $K \subset H''$ is a diamond with $V(K) \cap \{u'_0, u'_9\} \neq \varnothing$. If $u'_0 \in V(K)$ and u'_9 is not part of any diamond of H'', then $u'_0u'_1, u'_0u'_{17} \notin E(H'')$ because the three H''-neigbours of u'_0 must lie in $V(\tilde{G}_{t-1})$. The only diamonds of G_{t-1} that are vertex-disjoint from K and do not contain u'_9 are those in H'_0. Since H'' has at least two vertex-disjoint diamonds, $V(H'')$ intersects one of the sets $\{u'_1, \dots, u'_8\}$ and $\{u'_{10}, \dots, u'_{17}\}$, and hence $u'_9 \in V(H'')$ by connectedness of H''. Moreover, u'_9 is a cut-vertex of H''. Recall that H'' either is 2-connected or contains four diamonds. In the former case we have arrived at a contradiction. If there are four vertex-disjoint diamonds in H'' then $V(H'')$ meets both $\{u'_1, \dots, u'_8\}$ and $\{u'_{10}, \dots, u'_{17}\}$. But then $H'' - u'_9$ has at least three connected components, which contradicts the fact that H' is 2-connected.

Finally, we must exclude that both $u'_0 \in V(K)$ and $u'_9 \in V(K)$. Suppose for a contradiction that $u'_0, u'_9 \in V(K)$. The two common K-neighbours of u'_0 and u'_9 lie in $V(\tilde{G}_{t-1})$ and are adjacent in K as $u'_0u'_9 \notin E(G_{t-1})$ by the induction hypothesis. Both u'_0 and u'_9 have at most one H''-neighbour outside $V(K)$, so $|\{u'_0u'_1, u'_0u'_{17}\} \cap E(H'')| \leq 1$ and $|\{u'_9u'_8, u'_9u'_{10}\} \cap E(H'')| \leq 1$. At least one of the edges $u'_0u'_1$, $u'_0u'_{17}$, $u'_8u'_9$, $u'_9u'_{10}$ lies in $E(H'')$ because there are at least two vertex-disjoint diamonds in the connected graph H''. We have neither $\{u'_0u'_1, u'_8u'_9\} \subset E(H'')$ nor $\{u'_0u'_{17}, u'_9u'_{10}\} \subset E(H'')$ for otherwise $V(H'')$ would be contained in one of the twelve-element sets $V(K) \cup \{u'_1, \dots, u'_8\}$ and $V(K) \cup \{u'_{10}, \dots, u'_{17}\}$ due to connectedness of H'' and $\Delta(H'') = 3$. Thus, one of u'_0 and u'_9 is a cut-vertex of H''. If H'' is 2-connected we have again arrived at a contradiction. If H'' has four vertex-disjoint diamonds then three of the four sets $\{u'_1, \dots, u'_4\}$, $\{u'_5, \dots, u'_8\}$, $\{u'_{10}, \dots, u'_{13}\}$, and $\{u'_{14}, \dots, u'_{17}\}$ are subsets of $V(H'')$. The symmetries of H' given by (2.47) allow us to assume that $\{u'_1, \dots, u'_8\} \subset V(H'')$. Connectedness of H'' now implies that $u'_4u'_5$ and one of the two edges $u'_0u'_1$ and $u'_8u'_9$ lie in $E(H'')$. Therefore, $H''[V(K) \cup \{u'_1, \dots, u'_8\}]$ is connected. This, however, contradicts the fact that for any three distinct diamonds L_1, L_2, L_3 in H', the induced subgraph $H'[V(L_1) \cup V(L_2) \cup V(L_3)]$ is disconnected.

We have seen that all diamonds in H'' come from H'_0. Recall the definition of U and W in (2.48). Due to the symmetries of H' we may assume that e' lies in $H'[W]$. and is not incident to u_9. Then $\phi(U \setminus \{u_0, u_9\}) = \{u'_1, \dots, u'_8\}$ or $\phi(U \setminus \{u_0, u_9\}) = \{u'_{10}, \dots, u'_{17}\}$. In both cases the connectedness of H implies $\{\phi(u_0), \phi(u_9)\} = \{u'_0, u'_9\}$. If H'' has four diamonds then $H''[\phi(W \setminus \{u_0, u_9\})]$ contains two diamonds, so $\phi(W \setminus \{u_0, u_9\}) = V(H'_0) \setminus \phi(U)$. Now assume that H'' has exactly three diamonds, which happens only if e' lies in a diamond of H'. Observe that H'' intersects both $\{u'_1, \dots, u'_8\}$ and $\{u'_{10}, \dots, u'_{17}\}$. Furthermore, $V(H'') \cap \{u'_1, \dots, u'_8\}$ and $V(H'') \cap \{u'_{10}, \dots, u'_{17}\}$ are the vertex sets of two distinct components of $H'' - u'_0 - u'_9$ Removing u_0 and u_9 from $H' - e'$ results in a graph with at most two connected components. For this reason H'' cannot contain any vertices from $V(\tilde{G}_{t-1})$.

□

Let $\tilde{e} \in E(G_t) \setminus E(G)$. Our goal is to show that $\tilde{e} \in E(\tilde{G}_t) \setminus E(\tilde{G})$. We can assume $\tilde{e} \in E(G_t) \setminus$

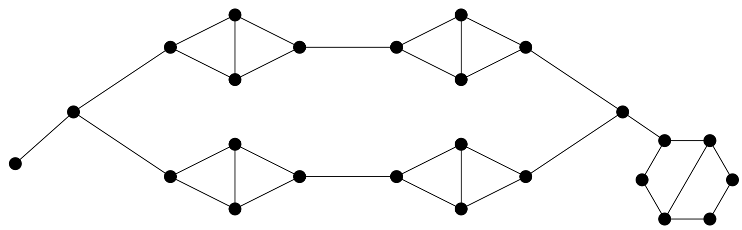

FIGURE 2.4: A graph with minimum degree one, maximum degree three and at least linear maximum running time.

$E(G_{i-1})$ as otherwise we would be done by the induction hypothesis. There exists an edge $e \in E(H)$ and an embedding $\varphi : H \to G_t$ such that $\varphi(e) = \tilde{e}$ and $\varphi(H - e) \subseteq G_{t-1}$. Let $e' := e$ if $e \in E(H')$ and an arbitrary edge of H' otherwise. Since $\varphi_{|V(H')}$ can be regarded as an embedding of $H' - e'$ into G_{t-1} we may apply Claim 2.8.3 to obtain $\varphi(V(H')) = V(H'_0)$ as well as $\{\varphi(u_0), \varphi(u_9)\} = \{u'_0, u'_9\}$. Since H' is H'-stable we conclude that $e \notin E(H')$. The only edges of H with precisely one endpoint in $V(H')$ are $u_0 z$ and $u_9 \tilde{v}$. However, both u'_0 and v'_0 are already adjacent to every vertex in $V(\tilde{G})$ at time 0, so the endpoints of e must lie in $V(\tilde{H})$. Then $\varphi_{|V(\tilde{H})}$ is an embedding of $\tilde{H}$ in $\tilde{G}_t$ with $\varphi_{|V(\tilde{H})}(H - e) \subset \tilde{G}_{t-1}$. This implies $\tilde{e} \in E(\tilde{G}_t)$, which completes the induction.

Now we simply need to choose a suitable candidate for $\tilde{H}$. Pick a cycle of length six with a chord between two fixed opposite vertices, e.g.

$$V(\tilde{H}) = \{v_1, \ldots, v_6\} \qquad , \qquad E(\tilde{H}) = \{v_1v_2, \ldots, v_5v_6, v_6v_1, v_1v_4\},$$

and choose $\tilde{v}$ to be a degree-two vertex of $\tilde{H}$. The graph H obtained for this choice of $\tilde{H}$ is depicted in Figure 2.4. Theorem 1.2.6 guarantees that the above defined $\tilde{G}$ satisfies

$$\tau_{\tilde{H}}(\tilde{G}) = \Omega(n)$$

and thus

$$M_H(n) \geq \tau_H(G) \geq \tau_{\tilde{H}}(\tilde{G}) = \Omega(n).$$

This finishes the proof.

In the introduction we have claimed that the construction above can be used to find a graph H with $\delta(H) = 1$ and $M_H(n) = \Omega(n^2)$. We now give a short sketch of how this can be achieved. The details are left to the reader. The starting graph $\tilde{G}$ used in [13] to show that $M_{K_6}(n) = \Omega(n^2)$, and its final graph $\langle \tilde{G} \rangle_{K_6}$ are K_7-free. We can now redo the proof of Proposition 1.2.8 with $\tilde{H} = K_6$, $H' = K_9$ and the $\tilde{G}$ introduced above. Since $\langle \tilde{G} \rangle_{K_6}$ is K_7-free, the only way to place a copy of K_9 in the starting graph G defined in (2.49) is to use the initial copy H'_0 of K_9. This locks K_9 in place throughout the process, resulting in the H-process on G simulating the K_6-process on $\tilde{G}$.

2.9 Open problems and further directions

The general open question that remains is how the influence of a degree-one vertex on the maximum running time can be characterised. Recall that Theorems 1.2.6 and 1.2.4 together yield an easily checkable necessary and sufficient condition for sublinear running time when the minimum degree is at least two.

In all cases known to the author appending a degree-one vertex to a given H either makes the maximum running time constant (cf. for example a clique with a pendent vertex) or does not affect it at all (cf. the proof of Proposition 1.2.8). While we could not show that this distinction applies to all H, it seems reasonable, as emphasised by the following informal considerations. Given a graph H and an H-process on some starting graph G, there are either many (in this context, at least $v(H)$) almost universal vertices that occur during the process or few of them (less than $v(H)$). In the former case the almost universal vertices contain a clique of size $v(H)$, which causes the process to stabilise within a constant number of steps (of course, the almost universal vertices have to appear early enough in the process). If there are few almost universal vertices then in all copies of H that occur during the process the roles of the unique neighbours of degree-one vertices are always played by vertices from the same small subset of $V(G)$. In that case the situation seems to be close to the simulation-type argument of Proposition 1.2.8, where the influence of degree-one vertices is restricted by forcing them to be fixed throughout the process. With these considerations in mind we ask if there is a graph with minimum degree whose maximum running time is non-constant and does not come from a simulation argument.

Question 2.9.1. Is there a graph H and $v \in V(H)$ with $d(v) = 1$ such that $M_{H-v}(n) = o(M_H(n))$ while $M_{H-v}(n) = \omega(1)$?

Our examples of graphs with minimum degree one and non-constant running time all have in common that they mimic the bootstrap process of a graph without a degree-one vertex. We believe that apart from constant and logarithmic there are no other types of running time below linear.

Conjecture 2.9.2. Every graph H with $M_H(n) = o(n)$ satisfies either $M_H(n) = O(1)$ or $M_H(n) = \Theta(\log n)$.

When H is the disjoint union of two arbitrary graphs H_1 and H_2 we do not know to what extent $M_H(n)$ depends on the individual running times $M_{H_1}(n)$ and $M_{H_2}(n)$. We have encountered examples for which the asymptotic running time of H matches one of the individual running times as well as examples where $M_{H_1}(n)$ and $M_{H_2}(n)$ were at least logarithmic while $M_H(n)$ was bounded by a constant. However, we have not seen whether $M_H(n)$ can be much bigger than $M_{H_1}(n)$ and $M_{H_2}(n)$.

Question 2.9.3. Are there graphs H_1, H_2 such that $M_{H_1 \sqcup H_2}(n) = \omega\left(M_{H_1}(n) + M_{H_2}(n)\right)$?

Finally we focus on a problem that came up in the proof of Theorem 1.2.1 (running time for forests). In both the two initial examples and the main result of Section 2.2 we made use of the fact that neighbours of degree-one vertices become adjacent to every vertex outside a fixed copy of T in a single step. However, this does not yield anything if there are no vertices outside that copy of T. In the proof of Theorem 1.2.1 this forced us to use a wasteful estimate on μ that dominates the upper bound in (1.4). Unfortunately we do not know of any bound which gives a significant improvement over the trivial estimate $k^2/4$ and works for all forests. There are families of forests for which we can do better than the trivial estimate. For example, it is an exercise to show that whenever T is a binary tree, that is, a rooted tree in which every vertex has at most two children, then $T[U]$ is a disjoint union of paths and hence $M_{T[U]}(|U|) \leq 3$. Using this improved estimate in (2.8), and recalling the definition (2.2) of μ, tells us that whenever T is a binary tree, one has $M_T(n) \leq (3k+50)/4$. This discussion motivates the following two questions.

Question 2.9.4. Is there an upper bound on $M_F(k)$ that is subquadratic in k, where $k = v(F)$?

Question 2.9.5. What is the smallest constant c_k such that $M_F(n) \leq c_k$ for every tree F on k vertices and $n \in \mathbb{N}$?

From an asymptotic (in n) viewpoint this question does not matter because the trivial bound $\binom{k}{2}$ is constant. However if the bound on μ could be replaced by $c \cdot v(F)$ for some constant $0 < c < \frac{1}{4}$ whenever F does not have a pendent path of length two then the star would be the unique forest maximising $M_F(n)$. The examples of F known to the author all have in common that $M_F(k)$ is at most linear in k with a leading constant below 1. We thus conclude this chapter with the question below.

Question 2.9.6. Is there a tree T on k vertices such that $M_T(n) > M_{K_{1,k-1}}(n)$ for infinitely many n?

Chapter 3

Linear and superlinear running times

This chapter focusses on the proofs of the results stated in Section 1.3. We begin with Theorems 1.3.1 and 1.3.5 in Section 3.1. In Section 3.2 we describe the two general constructions that yield all of our lower bounds for superlinear running times. Theorem 1.3.8 is treated in Section 3.3. Section 3.4 contains the proof of the lower bound for random graphs (Theorem 1.3.9). In Section 3.5 we link lower bounds on $M_H(n)$ for certain H to additive combinatorial constructions and use this link to prove Theorem 1.3.6. The bounds on $M_{Q_3}(n)$ (Theorem 1.3.7) are established in Section 3.6. We show the bounds for wheel graphs (Theorem 1.3.10) in Section 3.7. The proof of Theorem 1.3.12 is presented in Section 3.8. Lastly, in Section 3.9 we discuss open problems related to superlinear running times.

3.1 Upper bounds for bipartite graphs and for $K_{2,s}$

Here we present the proofs of our two upper bounds on $M_H(n)$ for bipartite H. The first is the general bound in terms of $\mathrm{ex}(n,H)$.

Proof of Theorem 1.3.1. Set $h := e(H)$. Let G be a graph on n vertices with $\tau_H(G) = M_H(n)$, and let $(G_t)_{t\geq 0}$ be the H-bootstrap process on G. The assumption $h \geq 2$ guarantees that G is a non-empty graph. Pick an arbitrary new edge from every second step of the process, that is, choose $e_0 \in E(G)$ and for $1 \leq t \leq \lfloor M_H(n)/2 \rfloor$ choose $e_t \in E(G_{2t}) \setminus E(G_{2t-1})$. The graph with vertex set $V(G)$ and edge set $\{e_0, \ldots, e_{\lfloor M_H(n)/2 \rfloor}\}$ must be H-free. Indeed, suppose it contained a copy H' of H with edges $e_{t_1}, \ldots, e_{t_h}$ where $t_1 < \ldots < t_h$. Note that $h-1 \geq 1$. Then $H' - e_{t_h}$ is a copy of H minus an edge in $G_{2t_{h-1}}$, so by the definition of the H-process, e_{t_h} would be present in $G_{2t_{h-1}+1}$, which contradicts the choice of e_{t_h}. Therefore

$$\left\lfloor \frac{M_H(n)}{2} \right\rfloor + 1 \leq \mathrm{ex}(n,H)$$

and hence the claim follows. ■

We now show that $K_{2,s}$ has linear running time for all $s \geq 3$.

Since $\delta(K_{2,s}) = 2$ and $\Delta(K_{2,s}) = s \geq 3$ the lower bound $M_{K_{2,s}} = \Omega(n)$ follows directly from Theorem 1.2.6. Therefore it remains to prove the linear upper bound $M_{K_{2,s}}(n) = O(n)$. Let G be an n-vertex graph and let $(G_t)_{t\geq 0}$ be the $K_{2,s}$-process on G. For simplicity we will denote the G_t-neighbourhood of a vertex x by $N_t(x)$. A copy of $K_{2,s}^-$ is just a pair of distinct vertices x,y together with $s-1$ common neighbours and another vertex z that is adjacent to either x or y. Therefore if at time $t\geq 0$, x and y have $s-1$ common neighbours and z is adjacent to x but not y then yz will complete a copy of $K_{2,s}$ at time $t+1$ and hence z will be a neighbour of y at time $t+1$. Let us summarise this observation:

Observation 3.1.1. For every $t\geq 0$ and any two distinct $x,y\in V(G)$ with $|N_t(x)\cap N_t(y)|\geq s-1$ we have $N_t(x)\setminus\{y\}\subseteq N_{t+1}(y)$, and similarly, $N_t(y)\setminus\{x\}\subseteq N_{t+1}(x)$.

The idea for the rest of the proof is to define partitions of $V(G)$ in which any two distinct vertices from the same block have at least $s-1$ common neighbours and show that unless the process has stabilised, one can coarsen these partitions every couple of steps. The claim then follows because any initial partition can have at most n blocks.

Lemma 3.1.2. *Let $t\geq 0$, and let $A,B\subset V(G)$ such that $|N_t(x_1)\cap N_t(x_2)|\geq s-1$ for any $x_1,x_2\in A$ and similarly $|N_t(y_1)\cap N_t(y_2)|\geq s-1$ for any $y_1,y_2\in B$. If $E_{G_t}(A,B)\neq\varnothing$, then $E_{G_{t+2}}(A,B)=\{xy : x\in A, y\in B\}$.*

Proof. We first look at the case $A\neq B$. Fix $x_0\in A$ and $y_0\in B$ such that $x_0y_0\in E(G_t)$. By assumption any $x\in A$ satisfies $|N_t(x)\cap N_t(x_0)|\geq s-1$, and hence $xy_0\in E(G_{t+1})$ by Observation 3.1.1. Similarly, we have $x_0y\in E(G_{t+1})$ for every $y\in B$. Applying Observation 3.1.1 again we obtain $N_{t+2}(x)\supseteq N_{t+1}(x_0)\supseteq B$ for every $x\in A$ and $N_{t+2}(y)\supseteq N_{t+1}(y_0)\supseteq A$ for every $y\in B$. Now suppose that $A=B$. Let $x_0,x_1\in A$ with $x_0x_1\in E(G_t)$. Observation 3.1.1 tells us that $xx_0\in E(G_{t+1})$ for all $x\in A\setminus\{x_0\}$. This yields $N_{t+2}(x)\supseteq\{x_0\}\cup N_{t+1}(x_0)\setminus\{x\}\supseteq A\setminus\{x\}$ for all $x\in A\setminus\{x_0\}$. □

We will use Lemma 3.1.2 to coarsen the aforementioned partitions of $V(G)$.

Lemma 3.1.3. *Let $t\geq 0$ with $t+4\leq\tau_{K_{2,s}}(G)$. If $\mathscr{P}$ is a partition of the vertex set of G such that any two disjoint vertices from the same block have at least $s-1$ common neighbours at time t, then there exist distinct $A,B\in\mathscr{P}$ such that in the coarser partition $\mathscr{P}\setminus\{A,B\}\cup\{A\cup B\}$ any two vertices from the same block have at least $s-1$ common neighbours at time $t+4$.*

Proof. Let $e\in E(G_{t+3})\setminus E(G_{t+2})$ and let A',B' be the (not necessarily distinct) blocks of $\mathscr{P}$ containing the endpoints of e. Then $E_{G_t}(A',B')=\varnothing$ for otherwise $e\in E(G_{t+2})$ by Lemma 3.1.2. This tells us that

$$t^* := \min\{t'\geq 0 : E_{G_{t'}}(A',B')\neq\varnothing\}\in\{t+1,t+2,t+3\}$$

Fix a copy of $K_{2,s}$ that was completed by an edge $e^*\in E_{G_{t^*}}(A',B')$ at time t^*. Let x^*,y^* be the two vertices forming that copy's partite set of size two. These two vertices cannot lie in the same block since one of them is an endpoint of e^* and if both of them lay in A' or B' there would be an edge between A' and B' at time t^*-1. Denote the block of $\mathscr{P}$ containing x^* by A and the block containing y^* by B. For every $x\in A$, $y\in B$ we have $N_{t+4}(x)\supseteq N_{t+3}(x^*)$ and $N_{t+4}(y)\supseteq N_{t+3}(y^*)$

by Observation 3.1.1, so

$$|N_{t+4}(x) \cap N_{t+4}(y)| \;\geq\; |N_{t+3}(x^*) \cap N_{t+3}(y^*)| \;\geq\; |N_{t^*}(x^*) \cap N_{t^*}(y^*)| \;>\; s-1.$$

Therefore the partition $\mathscr{P} \setminus \{A,B\} \cup \{A \cup B\}$ has the desired property. $\square$

Set $\mathscr{P}_0 := \{\{v\} : v \in V(G)\}$. This partition trivially fulfils the condition that any two distinct vertices from the same block have $s-1$ common neighbours at time 0. We may inductively apply Lemma 3.1.3 to obtain a finite sequence $\mathscr{P}_0, \mathscr{P}_4, \ldots, \mathscr{P}_{4\ell}$ of partitions of $V(G)$ where $\ell \in \mathbb{N}_0$,

$$4\ell \geq \tau_{K_{2,s}}(G) - 3,$$

and for all $t \in \{4,8,\ldots,4\ell\}$, $\mathscr{P}_t$ is a partition of $V(G)$ in which any two distinct vertices from the same block have at least $s-1$ common neighbours at time t. Moreover $\mathscr{P}_0$ has n blocks while for $t \in \{4,8,\ldots,4\ell\}$, $\mathscr{P}_t$ has one block less than $\mathscr{P}_{t-4}$ by construction. This implies

$$n \;=\; |\mathscr{P}_0| \;=\; |\mathscr{P}_{4\ell}| + \ell \;\geq\; \ell + 1 \;\geq\; \frac{\tau_{K_{2,s}}(G) - 3}{4} + 1$$

and thus $\tau_{K_{2,s}}(G) \leq 4n-1$. $\blacksquare$

3.2 Chain constructions: The formal setup and a recipe

The purpose of this section is to provide a unifying theme to the constructions that follow as well as the constructions in [13], on which ours are based. In the latter article the construction is phrased in terms of k-uniform hypergraphs where each hyperedge corresponds to a copy of K_k. This is possible because for complete graphs specifying the vertex set is equivalent to specifying the whole graph. While hypergraphs provide a suitable framework for cliques, we cannot use that framework for general H. Our next definition is that of a chain of copies of H in a graph. This concept already appeared (though in slightly different forms) in the articles [25, 13].

Definition 3.2.1. Let H be a fixed connected graph, and let $\ell \geq 1$. An *H-chain* of *length* ℓ is a sequence $(H_i, e_i)_{i \in [\ell]}$ of copies H_i of H and edges $e_i \in E(H_i)$ such that $e_i \in E(H_i \cap H_{i+1})$ for $i \in [\ell-1]$. We call $\bigcup_{i=1}^{\ell} H_i$ the *underlying graph* of the H-chain. The chain is called *proper* if additionally one has that

(1) $e_i \notin E(\langle H_j \rangle_H)$ for $1 \leq j < i \leq \ell$,

(2) for each $e \in E(H)$, every copy of $H-e$ in $\bigcup_{i=1}^{\ell} \langle H_i \rangle_H$ lies in $\langle H_j \rangle_H$ for some $j \in [\ell]$,

(3) for $e \in E(H)$ and $2 \leq j \leq \ell$, there are no copies of $H-e$ in $\langle H_j \rangle_H$ that are contained in $\left(\bigcup_{i=1}^{j-1} \langle H_i \rangle_H \cup \bigcup_{i=j}^{\ell} H_i\right) \setminus \{e_{j-1}, e_j\}$.

Remark 3.2.2. If H is *self-stable*, that is, if $\langle H \rangle_H = H$, then Condition (2) of the definition above simplifies to the assertion that the vertex set of every copy of $H-e$ in $\bigcup_{i=1}^{\ell} H_i$ lies in some H_j. Moreover, in that case Condition (3) follows from (2). In particular, this is true in the case when H is a complete graph, which is why (3) was not necessary in previous definitions of chains,

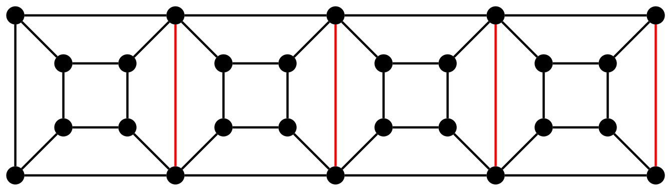

FIGURE 3.1: A proper Q_3-chain of length four. Removing the red edges results in a graph G with $\tau_{Q_3}(G) = 4$.

which were used to study cliques. We further note that we do not require any of the copies of H to be edge-disjoint. The underlying graph of a proper H-chain is H-stable if H is self-stable.

Removing the common edges e_i of neighbouring copies of H as well as the edge e_ℓ results in a graph whose H-process has running time equal to the chain's length:

Claim 3.2.3. *Let $(H_i, e_i)_{i\in[\ell]}$ be a proper H-chain. The graph $G := \left(\bigcup_{i=1}^{\ell} H_i\right) \setminus \{e_1, \ldots, e_\ell\}$ satisfies $\tau_H(G) \geq \ell$, and therefore*

$$M_H\left(\left|\bigcup_{i=1}^{\ell} V(H_i)\right|\right) \geq \ell.$$

Proof. The claim follows from the fact that in the H-process $(G_t)_{t\geq 0}$ on G we have

$$E(G_t) \cap \{e_1,, \ldots, e_\ell\} = \{e_1, \ldots, e_t\}$$

for every $0 \leq t \leq \ell$. To prove this fact we induct on t to show the following stronger statement

$$E(G_t) \cap \{e_1,, \ldots, e_\ell\} = \{e_1, \ldots, e_t\} \quad \text{and} \quad G_t \subseteq \bigcup_{i=1}^{t} \langle H_i \rangle_H \cup \bigcup_{i=t+1}^{\ell} H_i. \tag{3.1}$$

The case $t = 0$ follows immediately from the definition of G. Let $t \geq 1$. The induction hypothesis yields

$$E(G_{t-1}) \cap \{e_1, \ldots, e_\ell\} = \{e_1, \ldots, e_{t-1}\} \tag{3.2}$$

and

$$G_{t-1} \subseteq \bigcup_{i=1}^{t-1} \langle H_i \rangle_H \cup \bigcup_{i=t}^{\ell} H_i. \tag{3.3}$$

Due to Condition (1) of Definition 3.2.1 and (3.2) we can see that $H_t - e_t \subseteq G_{t-1}$, so $e_t \in E(G_t)$. Condition (2) and (3.3) guarantee that every copy of H completed at time t lies in $\langle H_i \rangle_H$ for some $i \in [\ell]$, while (3.2) and Condition (3) imply $i \leq t$. Therefore, $G_t \subseteq \bigcup_{i=1}^{t} \langle H_i \rangle_H \cup \bigcup_{i=t+1}^{\ell} H_i$.

We have to make sure that $e_j \notin E(G_t)$ for $t < j \leq \ell$. Suppose for a contradiction there was some $j \in [t+1, \ell]$ with $e_j \in E(G_t)$. Then e_j lies in $E(G_t) \setminus E(G_{t-1})$ because of (3.2). Let H' be a copy of H completed by e_j at time t. From (3.3) we obtain that $H' - e_j$ lies in $\bigcup_{i=1}^{\ell} \langle H_i \rangle_H$, and hence by Condition (2) there exists $i \in [t]$ such that $H' - e_j \subseteq \langle H_i \rangle_H$. However, (1) gives $e_j \notin E(\langle H_i \rangle_H)$ so we have arrived at the desired contradiction. Thus $\{e_{t+1}, \ldots, e_\ell\} \cap E(G_t) = \varnothing$. $\square$

Proper chains of arbitrary length do not necessarily exist. For example there are no such chains if $\delta(H) = 1$ because during the H-process the unique neighbour of a degree-one vertex in a copy of H becomes adjacent to every vertex outside that copy after one step. The following construction is a generalisation of the one used in the proof of Theorem 1.1.7.

Construction 3.2.4. *Given a quintuple (H, U, e, f, n) consisting of*

- *a graph H,*
- *a proper non-empty subset $U \subset V(H)$,*
- *two non-incident edges $e, f \in E_H(U, V(H) \setminus U)$,*
- *a natural number $n \geq 9v(H)$,*

the following construction yields an H-chain $(H_i, e_i)_{i \in [\ell]}$ of length $\ell := \lfloor \frac{n^2}{4v(H)^2} - \frac{3n}{v(H)} \rfloor + 7$ satisfying $e_i \notin E(\langle H_j \rangle_H)$ for $j < i$. Let

$$r := \left\lfloor \frac{n}{v(H)} \right\rfloor - \left(\left\lfloor \frac{n}{v(H)} \right\rfloor \mod 4 \right),$$

that is, r is the largest integer that is at most $\lfloor n/v(H) \rfloor$ and divisible by 4. Furthermore, let $s := r - 4$ and $W := V(H) \setminus U$. Note that $\ell < r \cdot s/4$. The condition $n \geq 9v(H)$ assures that r, s, and ℓ are positive. Pick a set of size $r \cdot (|U| - 1) + s \cdot (|W| - 1)$ and label its elements by $u_0, \dots, u_{r \cdot (|U|-1)-1}, w_0, \dots, w_{s \cdot (|W|-1)-1}$. Moreover, for convenience, we set $u_{r \cdot (|U|-1)} := u_0$ and $w_{s \cdot (|W|-1)} := w_0$. For $j \in [r]$ and $k \in [s]$, let

$$U_j := \{u_{(j-1) \cdot (|U|-1)}, \dots, u_{j \cdot (|U|-1)}\} \quad \text{and} \quad W_k := \{w_{(k-1) \cdot (|W|-1)}, \dots, w_{k \cdot (|W|-1)}\}.$$

Choose a bijection $\varphi : V(H) \to U_1 \cup W_1$ such that $\varphi(e) = u_0 w_0$ and $\varphi(f) = u_{|U|-1} w_{|W|-1}$. For $0 \leq i \leq \ell - 1$, let $\sigma_i : U_1 \cup W_1 \to U_{1+(i \bmod r)} \cup W_{1+(i \bmod s)}$ be the map that sends u_x to $u_{x+(i \mod r) \cdot (|U|-1)}$ and w_y to $w_{y+(i \mod s) \cdot (|W|-1)}$. Now, for $i \in [\ell]$, define H_i via

$$V(H_i) = \sigma_{i-1}(U_1 \cup W_1) \quad \text{and} \quad E(H_i) = \sigma_{i-1}(\varphi(E(H))), \tag{3.4}$$

and let

$$e_i := \sigma_{i-1}(\varphi(f)) = u_{(i \bmod r)(|U|-1)} w_{(i \bmod s)(|W|-1)}. \tag{3.5}$$

Proof. We have that $H_i \cong H$ and $e_i \in E(H_i)$ for all $i \in [\ell]$ by (3.4) and (3.5). Our choice of φ together with the definition of σ_i yields $\sigma_{i-1}(\varphi(f)) = \sigma_i(\varphi(e))$ for all $i \in [\ell - 1]$ and so $e_i \in E(H_{i+1})$.

As to Condition (1) of Definition 3.2.1 let $j, i \in [\ell]$ with $j < i$ and recall the definition of e_i and e_j in (3.4). Since $\ell < r \cdot s/4$ and $\gcd(r, s) = 4$, any $k \in [\ell]$ is uniquely determined by the pair $(k \bmod r, k \bmod s)$. Moreover, $(\ell \bmod r, \ell \bmod s) \neq (0, 0)$. Thus the assumption $j < i$ implies

$$(i \bmod r, i \bmod s) \notin \{(j - 1 \bmod r, j - 1 \bmod s), (j \bmod r, j \bmod s)\}. \tag{3.6}$$

As $4 \mid r$ and $s = r - 4$, $(i \bmod r)$ and $(i \bmod s)$, and similarly $(j \bmod r)$ and $(j \bmod s)$ have the same remainder modulo 4. Therefore

$$(i \bmod r, i \bmod s) \notin \{(j-1 \bmod r, j \bmod s), (j \bmod r, j-1 \bmod s)\}. \tag{3.7}$$

Combining (3.6) and (3.7) gives us $(i \bmod r) \notin \{j-1 \bmod r, j \bmod r\}$ or $(i \bmod s) \notin \{j-1 \bmod s, j \bmod s\}$ and thus $|e_i \cap V(H_j)| \le 1$, which tells us that $e_i \notin E(\langle H_j \rangle_H)$. □

Observe that the underlying graph $\bigcup_{i=1}^{\ell} H_i$ of the chain in Construction 3.2.4 has at most n vertices. Whether the chain itself is proper depends on H and the choice of the remaining parameters. In fact the main part of the proofs of Theorems 1.3.8 and 1.3.9 is to establish that the H-chains obtained in these sections by invoking Construction 3.2.4 are indeed proper. One of the tools we will use to establish properness is the following technical lemma.

Lemma 3.2.5. *Let the notation be as in Construction 3.2.4, and define $G := \bigcup_{i=1}^{\ell} \langle H_i \rangle_H$. If $|E_G(U_j, W_k)| > \nu(H)$ for some $j \in [r]$, $k \in [s]$, then $U_j \cup W_k = V(H_i)$ for some $i \in [\ell]$.*

Proof. Set $U_{r+1} := U_1$ and $W_{s+1} := W_1$. Let $j \in [r]$, $k \in [s]$ such that $|E_G(U_j, W_k)| > \nu(H)$. Let $I := \{(i \bmod r, i \bmod s) : i \in [\ell]\}$, so $U_j \cup W_k = V(H_i)$ for some $i \in [\ell]$ if and only if $(j,k) \in I$. Suppose for a contradiction that $(j,k) \notin I$. Recall that for $j' \in [r]$, $u_{j'(|U|-1)}$ is the unique vertex in $U_{j'} \cap U_{j'+1}$, and for $k' \in [s]$, $w_{k'(|W|-1)}$ is the only vertex in $W_{k'} \cap W_{k'+1}$. Define

$$\mathring{U}_j := U_j \setminus \{u_{(j-1)(|U|-1)}, u_{j(|U|-1)}\} \quad \text{and} \quad \mathring{W}_k := W_k \setminus \{w_{(k-1)(|W|-1)}, w_{k(|W|-1)}\}$$

Every edge in G comes from some $\langle H_i \rangle_H$ where $i \in [\ell]$. Any $i \in [\ell]$ for which both $V(H_i) \cap \mathring{U}_j \neq \varnothing$ and $V(H_i) \cap \mathring{W}_k \neq \varnothing$ satisfies

$$(i \bmod r,\ i \bmod s) = (j,k).$$

Since we assumed that no such i exists every edge in $E_G(U_j, W_k)$ has one of the four endpoints $\{u_{(j-1)(|U|-1)}, u_{j(|U|-1)}, w_{(k-1)(|W|-1)}, w_{k(|W|-1)}\}$. As $(j,k) \notin I$ there exists an edge between $u_{(j-1)(|U|-1)}$ and $\mathring{W}_k$ only if $(j-1,k) \in I$. In that case $(j-1) \bmod 4 = k \bmod 4$, which implies $j \bmod 4 \neq k \bmod 4$ and hence $E_G(\{u_{j(|U|-1)}\}, \mathring{W}_k) = \varnothing$. Similarly, $E_G(\{u_{j(|U|-1)}\}, \mathring{W}_k) \neq \varnothing$ would force $E_G(\{u_{(j-1)(|U|-1)}\}, \mathring{W}_k)$ to be empty. Therefore,

$$|E_G(\{u_{(j-1)(|U|-1)}\}, \mathring{W}_k)| + |E_G(\{u_{j(|U|-1)}\}, \mathring{W}_k)| \le |\mathring{W}_k| = |W| - 2. \tag{3.8}$$

An analogous relation holds between $E_G(\{w_{(k-1)(|W|-1)}\}, \mathring{U}_j)$ and $E_G(\{w_{k(|W|-1)}\}, \mathring{U}_j)$, that is,

$$|E_G(\{w_{(k-1)(|W|-1)}\}, \mathring{U}_j)| + |E_G(\{w_{k(|W|-1)}\}, \mathring{U}_j)| \le |\mathring{U}_j| = |U| - 2. \tag{3.9}$$

Together (3.8) and (3.9) give

$$|E_G(U_j, W_k)| \le 4 + |U| - 2 + |W| - 2 = \nu(H)$$

where the constant term 4 comes from the potential edges between the four vertices $u_{(j-1)(|U|-1)}$, $u_{j(|U|-1)}, w_{(k-1)(|W|-1)}$, and $w_{k(|W|-1)}$. This contradicts the assumption $|E_G(U_j,W_k)| > v(H)$. $\square$

Construction 3.2.4 is designed to provide quadratic lower bounds on the maximum running time. In certain situations, for example if H is bipartite, we already know that $M_H(n)$ is asymptotically subquadratic and so the construction above cannot yield a proper H-chain. Later we will see how the chain-based approach can be implemented to provide a larger class of running times.

3.2.1 Superimposing chains

In this section we describe the construction which underlies the proofs of Theorem 1.3.6, 1.3.12 and 1.3.10. It is a generalisation of the construction used by Balogh et al. to obtain their lower bound on $M_{K_5}(n)$.

Before we can dive into the construction, we have to introduce the concept of a simple chain. Perhaps the easiest way of creating H-chains of some length ℓ is to start with ℓ disjoint copies of H and glue them together along specified edges such that there are no intersections apart from the glued edges. This idea is formalised in the definition below.

Definition 3.2.6. Let H be a graph, and $e,f \in E(H)$ be non-incident. An H-chain $(H_i,e_i)_{i\in[\ell]}$ is called (e,f)*-simple* if for all $1 \le i < j \le \ell$,

$$V(H_i) \cap V(H_j) = \begin{cases} e_i & \text{, if } j = i+1; \\ \varnothing & \text{, otherwise;} \end{cases} \tag{3.10}$$

and there exist isomorphisms $\varphi_i : H \to H_i$, $i \in [\ell]$, such that $\varphi_i(f) = \varphi_{i+1}(e)$ for $i < \ell$.

Remark 3.2.7. In an (e,f)-simple chain every contiguous subchain $(H_i,e_i)_{i\in I}$, where $I \subset [\ell]$ is an interval of a given length, looks the same in the sense that one can pass from one to the other by a suitable isomorphism between the underlying graphs. This property will allow us to combine multiple (e,f)-simple chains on a common set of vertices into a single longer chain.

In the following we will call a chain just *simple* instead of (e,f)-simple if the edges e and f exist but there is no need to specify them. The example in Figure 3.1 is a simple chain. Building arbitrarily long (e,f)-simple H-chains is straightforward.

Observation 3.2.8. For any graph H and any non-incident $e,f \in E(H)$ there exists (e,f)-simple chains of arbitrary length. The underlying graph of a simple chain of length ℓ always has $2+\ell(v(H)-2)$ vertices.

Proof. Given H, e, f, and an n-element set V with an ordering $v_0,\dots,v_{n-1}$ of its elements, where $n := 2+\ell(v(H)-2)$, we can place an H-chain on V as follows. For every $i \in [\ell]$ choose a bijection $\varphi_i : V(H) \to \{v_{(i-1)\cdot(v(H)-2)},\dots,v_{i(v(H)-2)+1}\}$ such that

$$\varphi_i(e) = v_{(i-1)\cdot(v(H)-2)}v_{(i-1)\cdot(v(H)-2)+1} \quad \text{and} \quad \varphi_i(f) = v_{i\cdot(v(H)-2)}v_{i\cdot(v(H)-2)+1}.$$

Then $\varphi_i(f) = \varphi_{i+1}(e)$ for all $i \in [\ell-1]$. Define

$$H_i := \varphi_i(H) \quad \text{and} \quad e_i := \varphi_i(f)$$

for $i \in [\ell]$. With these choices $(H_i, e_i)_{i\in[\ell]}$ satisfies (3.10) as well as $e_i \in E(H_i \cup H_{i+1})$ for $i \in [\ell-1]$ and $e_\ell \in E(H_\ell)$. The second part on the number of vertices of the underlying graph follows from (3.10). □

While it is easy to build long simple chains, it is not guaranteed that those chains are proper. By definition, in a simple chain $(H_i, e_i)_{i\in[\ell]}$ any H_j, $j \geq 2$, intersects $\bigcup_{i\neq j}\langle H_i\rangle_H$ precisely in the edges e_{j-1} and e_j, and so $e_i \notin E(\langle H_i\rangle_H)$ for $j < i$. Additionally, $\left(\bigcup_{i=1}^{j-1}\langle H_i\rangle_H \cup \bigcup_{i=j}^{\ell} H_i\right) \setminus \{e_{j-1}, e_j\}$ has just $|E(H)|-2$ edges with both endpoints in $V(H_j)$. For this reason, Conditions (3) of Definition 3.2.1 always holds for simple chains, so showing that a given simple chain is proper amounts to checking Condition (2). If H is disconnected or has a cut-edge e, there is no hope of constructing proper simple chains of length at least two since we can place the components of $H-e$ in different copies of the chain. For 2-edge-connected H it suffices to consider chains of length at most $v(H)$. Indeed, any copy of H minus an edge is connected and hence if it lies in $\bigcup_{i=1}^{\ell}\langle H_i\rangle_H$ is must lie in a subchain of length at most $v(H)$. If the subchain is proper then the copy of H minus an edge must lie in $\langle H_j\rangle_H$ for some $j \in [\ell]$. We summarise the relation between short and long proper simple chains in the following observation.

Observation 3.2.9. If H is 2-edge-connected and there exists a proper (e,f)-simple H-chain of length $v(H)$ then every (e,f)-simple H-chain of length more than $v(H)$ is proper, too.

The discussion above does not tell us for which graphs we can obtain short proper H-chains in the first place. Fortunately, the desired chains exist as long as H is inseparable or bipartite-inseparable.

Proposition 3.2.10. *Let H be a graph that is inseparable or bipartite-inseparable. Then any simple H-chain is proper.*

Proof. We begin by establishing the following consequences of inseparability.

- If H is inseparable then for any two graphs G,G' with $v(G\cap G') = 2$ and $e(G\cap G') = 1$, each copy of $H-e'$ in $G\cup G'$ is fully contained in either G or G'.
- If H is bipartite-inseparable, and G, G' are two bipartite graphs with $v(G\cap G') = 2$ and $e(G\cap G') = 1$, then any $e' \in E(H')$ every copy of $H-e'$ in $G\cup G'$ lies in either G or G'.

Suppose that $H-e'$ is 3-connected for every $e' \in E(H)$. Let G, G' be graphs with precisely two common vertices x,y, and let H' be a copy of $H-e'$ in $G\cup G'$. If both $V(G)\cap V(H')\setminus\{x,y\} \neq \varnothing$ and $V(G')\cap V(H')\setminus\{x,y\} \neq \varnothing$, then $\{x,y\}\cap V(H')$ is a vertex-cut of H' of size at most two. This however contradicts the assumption that $H-e'$ is 3-connected. Therefore $V(H') \subseteq G$ or $V(H') \subseteq G'$. Since xy is an edge of both G and G' we arrive at either $H' \subseteq G$ or $H' \subseteq G'$.

Now assume that H is bipartite, and let G, G' be bipartite such that $V(G\cap G') = \{x,y\}$ for some vertices x,y and $E(G\cap G') = \{xy\}$. Recall that any embedding of a connected bipartite graph into another bipartite graph sends partite sets to partite sets. Thus if H' is a copy of $H-e'$ for some $e' \in E(H)$ that lies in neither $V(G)$ nor $V(G')$ then $V(H') \cap \{x,y\}$ would be a vertex cut of H' with at most one vertex from each partite set of H'. Again, we have arrived at a contradiction, so $V(H') \subseteq V(G)$ or $V(H') \subseteq V(G')$, and, since $xy \in E(G \cap G')$, $H' \subseteq G$ or $H' \subseteq G'$.

It remains to check that any (e,f)-simple H-chain $\mathscr{H} = (H_i, e_i)_{i\in[\ell]}$ is proper. Let $G := \bigcup_{i=1}^{\ell} \langle H_i \rangle_H$ and note that, due to our connectivity assumptions, G is bipartite if H is. Let $e' \in E(H)$, and let H' be a copy of $H-e'$ in G. Suppose that $V(H')$ is not contained in a single $\langle H_j \rangle_H$. Pick the smallest $j \in [\ell-1]$ such that $V(H') \cap V(H_j) \setminus (V(H_{j+1}) \cup \ldots \cup V(H_\ell)) \neq \varnothing$. Since the intersection of $\bigcup_{i=1}^{j} \langle H_i \rangle_H$ and $\bigcup_{i=j+1}^{\ell} \langle H_i \rangle_H$ is precisely the edge e_j, removing the endpoints of e_j disconnects G. However, removing them does not disconnect H' because of our assumptions on H. For this reason we must have that $V(H') \subseteq V(H_j)$. All edges in $G[V(H_j)]$ come from $\langle H_j \rangle_H$ because $\mathscr{H}$ is simple, so $H' \subseteq \langle H_j \rangle_H$.

We have shown that Condition (2) of Definition 3.2.1 is satisfied. Finally, Conditions (1) and (3) hold because the chain is simple. ☐

The number of vertices of (the underlying graph of) a simple chain is linear in the length of the chain. Therefore, single chains on their own do not yield superlinear running times. The next lemma formalises the idea of taking a family of simple chains on a common vertex set and combining them into a longer chain.

Lemma 3.2.11. *Let H be a 2-edge-connected graph on at least five vertices, let $e,f \in E(H)$ be non-incident, and let A be a finite ordered set. Let $(H_i^a, e_i^a)_{i\in[\ell]}$, $a \in A$, be a collection of proper, (e,f)-simple H-chains of length at least $2v(H)$ with underlying graphs G^a such that*

(i) $e_i^a \notin E(G^b)$ for $i \in [\ell]$, $a > b$,

(ii) for each $e' \in E(H)$, every copy of $H-e'$ in $\bigcup_{a\in A} \langle G^a \rangle_H$ is contained in exactly one of the $\langle G^a \rangle_H$,

If H is either inseparable or bipartite-inseparable and $\bigcup_{a\in A} \langle G^a \rangle_H$ does not contain any odd cycles of length at most $v(H)$, then

$$M_H\left(\left|\bigcup_{a\in A} V(G^a)\right| + 2|A| \cdot v(H)^2\right) \geq |A| \cdot \ell. \tag{3.11}$$

If H is not necessarily (bipartite-)inseparable but the chains above satisfy the additional condition

$$\bigcup_{i=1}^{v(H)} V(H_i^a) \cap V(G^b) = \varnothing \text{ and } \bigcup_{i=\ell-v(H)+1}^{\ell} V(H_i^a) \cap V(G^b) = \varnothing \text{ for } a \neq b \tag{3.12}$$

then we have that

$$M_H\left(\left|\bigcup_{a\in A} V(G^a)\right| + |A| \cdot v(H)\right) \geq |A| \cdot \ell. \tag{3.13}$$

Proof. We begin with the part when H is neither inseparable nor bipartite-inseparable. Once we have established that part, we reduce the first part to it.

We are only interested in the ordering and the cardinality of A, so we assume that $A = [s]$ for some $s \in \mathbb{N}$. Let $k := v(H) - 2$ and $G := \bigcup_{a\in A} \langle G^a \rangle_H$. For every $a \in A$ we introduce a set W^a of $v(H) - 4$ new vertices. Furthermore, let $\varphi^a : H \to H^a_1$ be a graph isomorphism with $\varphi^a(f) = e^a_1$ and define $e^a_0 := \varphi^a(e)$. Now, if $a \neq s$, choose an injective map $\psi^a : V(H) \to e^a_\ell \cup W^a \cup e^{a+1}_0$ with $\psi^a(e) = e^a_\ell$ and $\psi^a(f) = e^{a+1}_0$. Note that by (iii), e^a_ℓ and e^{a+1}_0 are disjoint, and $e^a_0 \notin E(G^b)$ for $b < a$. With these choices, the sequence

$$\left(H^a_{\ell-v(H)+1}, e^a_{\ell-v(H)+1}\right), \dots, \left(H^a_\ell, e^a_\ell\right), \left(\psi^a(H), e^{a+1}_0\right), \left(H^{a+1}_1, e^a_1\right), \dots, \left(H^{a+1}_{v(H)}, e^{a+1}_{v(H)}\right) \quad (3.14)$$

is an (e, f)-simple H-chain (up to relabelling since the indices above do not start with 1). It is also proper due to Remark 3.2.7, Observation 3.2.9, and the assumption that the subchain $\left(H^{a+1}_1, e^a_1\right), \dots, \left(H^{a+1}_{v(H)}, e^{a+1}_{v(H)}\right)$ is proper.

We use the pairs $(\psi^a(H), e^{a+1}_0)$, $a \in A$, to concatenate our chains. Set $\bar{\ell} := |A| \cdot (\ell + 1) - 1$ and define $\mathscr{H} := (\bar{H}_i, \bar{e}_i)_{i\in[\bar{\ell}]}$ by

$$\bar{H}_i := \begin{cases} H^a_{i-(a-1)(\ell+1)} & , \text{ if } (a-1)(\ell+1) < i < a(\ell+1); \\ \psi^a(H) & , \text{ if } i = a(\ell+1) \text{ and } a \neq s; \end{cases}$$

and

$$\bar{e}_i := e^a_{i-(a-1)(\ell+1)} \text{ for the unique } a \text{ with } (a-1)(\ell+1) \leq i < a(\ell+1).$$

By construction of $\psi^a(H)$, e^a_ℓ is the unique common edge of H^a_ℓ and $\psi^a(H)$ while H^{a+1}_1 and $\psi^a(H)$ intersect precisely in e^{a+1}_0. This together with the assumption that the $(H^a_i, e^a_i)_{i\in[\ell]}$ are chains tells us that $\mathscr{H}$ is an H-chain. It remains to check that $\mathscr{H}$ is proper.

Recall that $e^a_0 \notin E(G^b)$ for $b < a$ and $e^a_0 \cap W^b = \varnothing$ for all $b \in A$. Thus, $\bar{e}_{(a-1)(\ell+1)} \notin E(\bar{H}_j)$ for $j < (a-1)(\ell+1)$. If $i \in [\bar{\ell}]$ and $a \in A$ with $(a-1)(\ell+1) < i < a(\ell+1)$, then $\bar{e}_i \notin E(H^b_j)$ for $b < a$ and $j \in [\ell]$ by (i). Further, $\bar{e}_i \notin E(H^a_j)$ for $j < i$ because $(H^a_i, e^a_i)_{i\in[\ell]}$ is proper. Finally, $\bar{e}_i \cap W^b = \varnothing$ for $b \in A$, so $\bar{e}_i \notin E(\psi^b(H))$ for $b < a$. We have shown that $\bar{e}_i \notin E(\langle \bar{H}_j \rangle_H)$ for $j < i$.

Let H' be a copy of H minus an edge in $\bar{G} := G \cup \bigcup_{a\in[s-1]} \langle \psi^a(H) \rangle_H$. If $H' \subseteq G$ then by (ii) there exists $a \in A$ with $H' \subseteq \langle G^a \rangle_H$. Since $(H^a_i, e^a_i)_{i\in[\ell]}$ is a proper chain we find an $j \in [\ell]$ such that $H' \subseteq \langle H^a_j \rangle_H$. If H' is not contained in G there exists $a \in A$ such that $V(H') \cap W^a \neq \varnothing$. By (iii) the $v(H)^{\text{th}}$ $\bar{G}$-neighbourhood of any $w \in W^a$ is a subset of $\bigcup_{i=\ell-v(H)+1}^{v(H)} V(H^a_i) \cup W^a \cup \bigcup_{i=1}^{v(H)} V(H^{a+1}_i)$, that is, it is contained in the vertex set of the underlying graph of the H-chain (3.14). Therefore the connected graph H' lies in $\bigcup_{i=a(\ell+1)-v(H)}^{a(\ell+1)+v(H)} \langle \bar{H}_i \rangle_H$. Since (3.14) is a proper H-chain we obtain that $H' \subseteq \langle \bar{H}_j \rangle_H$ for some $j \in [a(\ell+1) - v(H), a(\ell+1) + v(H)]$.

Thus $\mathscr{H}$ is a proper H-chain of length at least $|A| \cdot (\ell+1) - 1$. The underlying graph of $\mathscr{H}$ has order at most $v(G) + (|A| - 1) \cdot (v(H) - 4)$. The claim now follows from Claim 3.2.3.

It remains to reduce the parts for inseparable or bipartitely inseparable H to the one we have just shown. The main idea is to use the inseparability of H to extend the chains $(H^a_i, e^a_i)_{i\in[\ell]}$ to longer chains that satisfy condition (3.12) and then apply the second part.

For every $a \in A$ we define e_0^a as before and introduce two disjoint sets U^a, V^a of new vertices, each of size $(v(H)+1)(v(H)-2)$. We place an (e,f)-simple chain $(\tilde{H}_i^a, \tilde{e}_i^a)_{i\in[v(H)+1]}$ on $U^a \cup e_0^a$ such that $\tilde{e}_{v(H)+1}^a = e_0^a$. For $i \in [v(H)+2, \ell+v(H)+1]$ let

$$\tilde{H}_i^a := H_{i-v(H)-1}^a \quad \text{and} \quad \tilde{e}_i^a := e_{i-v(H)-1}^a.$$

We now place an (e,f)-simple chain $(\tilde{H}_i^a, \tilde{e}_i^a)_{i\in[\ell+v(H)+2,\ell+2(v)+2]}$ on $V^a \cup e_\ell^a$ such that $\mathscr{H}^a := (\tilde{H}_i^a, \tilde{e}_i^a)_{i\in[\ell+2v(H)+2]}$ is an (e,f)-simple H-chain of length $\ell+2v(H)+2$. The latter chain is proper by Proposition 3.2.10. It remains to show that our new chains $\mathscr{H}^a$ satisfy the conditions (i) and (ii).

Denote the underlying graph of $\mathscr{H}^a$ by $\tilde{G}^a$, that is, let $\tilde{G}^a := \bigcup_{i\in[\ell+2v(H)+2]} \tilde{H}_i^a$. Since U^a and V^a were newly introduced sets of vertices, we have

$$\bigcup_{i\in[v(H)]\cup[\ell+v(H)+3,\ell+2v(H)+2]} V(\tilde{H}_i^a) \cap V(\tilde{G}^b) = \varnothing \tag{3.15}$$

for $a \neq b$. Therefore (3.12) is satisfied. Since (i) holds for the chains $(H_i^a, e_i^a)_{i\in[\ell]}$, $a \in A$, we have that

$$\tilde{e}_i^a \notin V(\tilde{G}^b) \tag{3.16}$$

for $v(H)+1 \le i \le \ell+v(H)+1$ and $a > b$. Because of (3.15) we can see that (3.16) also holds for $i \le v(H)$ and $i \ge \ell+v(H)+2$.

Lastly we have to verify condition (ii). We note that $\langle G^a\rangle_H$ is bipartite for all a if H is bipartitely inseparable since bipartitely inseparable graphs are 2-edge-connected. Suppose that H' is a copy of $H-e'$ for some $e' \in E(H)$ in $\tilde{G} := \bigcup_{a\in A}\langle \tilde{G}^a\rangle_H$. If $V(H') \cap U^a \neq \varnothing$ for some $a \in A$ then $V(H') \subseteq U^a \cup e_0^a$ as H is (bipartitely) inseparable. Here it is important that in the bipartite case $\bigcup_{a\in A}\langle G^a\rangle_H$ does not contain odd cycles of length at most $v(H)$ so $\tilde{G}[V(H')\setminus U_a]$ is bipartite with the endpoints of e_0^a lying in different parts. Similarly we obtain that $V(H') \subseteq V_a \cup e_\ell^a$ whenever $V(H') \cap V^a \neq \varnothing$ for some $a \in A$. In both cases H' is contained in $\langle \tilde{G}^a\rangle_H$. If $V(H') \cap U^a = V(H') \cap V^a = \varnothing$ for all $a \in A$ we are in the situation of (ii) for the original chains, so $H' \subseteq \langle G^a\rangle_H \subseteq \langle \tilde{G}^a\rangle_H$ for some $a \in A$.

Now the second part of this lemma applied to $\mathscr{H}^a$, $a \in A$, tells us that

$$M_H\left(v(\tilde{G}) + |A| \cdot v(H)\right) \ge |A| \cdot (\ell + 2v(H) + 2),$$

which implies (3.11). □

3.3 Dense graphs

In this section we give the proof of Theorem 1.3.8. Let $n \in \mathbb{N}$ and apply Construction 3.2.4 with the following choice of parameters:

- H the graph from the statement of Theorem 1.3.8

- U a subset of $V(H)$ of size $\lfloor v(H)/2 \rfloor$ such that $|E_H(U, V(H) \setminus U)|$ is minimised,
- e and f are arbitrary non-incident edges in $E_H(U,W)$,
- n as given.

The hypotheses $v(H) \geq 6$ and $\delta(H) > 3v(H)/4$ guarantee $E_H(U,W)$ contains two non-incident edges. We have to show that the chain $(H_i, e_i)_{i \in [\ell]}$ is proper. Condition (1) of Definition 3.2.1 is always satisfied by Construction 3.2.4.

As to Condition (2) let $e \in E(H)$, and let $\varphi \in \mathrm{Hom}(H-e, \bigcup_{i=1}^{\ell} \langle H_i \rangle_H)$ be injective. Write $H' := \varphi(H-e)$ and $G := \bigcup_{i=1}^{\ell} \langle H_i \rangle_H$. We will show that there exist $j \in [r]$ and $k \in [s]$ such that $V(H') = U_j \cup W_k$. By averaging we have

$$|V(H') \cap (U_1 \cup \ldots \cup U_r)| \geq |U| \quad \text{or} \quad |V(H') \cap (W_1 \cup \ldots \cup W_s)| \geq |W|.$$

We assume that $|V(H') \cap (U_1 \cup \ldots \cup U_r)| \geq |U| = \lfloor v(H)/2 \rfloor$. The case when $|V(H') \cap (W_1 \cup \ldots \cup W_s)| \geq |W| = \lceil v(H)/2 \rceil$ follows from analogous arguments together with the obvious inequality $\lceil v(H)/2 \rceil \geq \lfloor v(H)/2 \rfloor$.

Suppose for a contradiction that there is no $j \in [r]$ with $V(H') \cap (U_1 \cup \ldots \cup U_r) = U_j$. Then we can pick the smallest $j' \in [r-1]$ such that $V(H') \cap U_{j'} \setminus U_{j'+1} \neq \varnothing$. Let $U' := V(H') \cap U_{j'}$. Because $\delta(H') \geq \delta(H) - 1 \geq \lfloor 3v(H)/4 \rfloor$, any two distinct vertices of H' have an edge in their common H'-neighbourhood. This together with Lemma 3.2.5 gives $V(H') \cap U_{j'+1} \setminus U_{j'} = \varnothing$ and therefore $j' \leq \ell - 2$. Define $U'' := V(H') \cap (U_{j'+2} \cup \ldots \cup U_r)$ and observe that $U' \cup U'' = V(H') \cap (U_1 \cup \ldots \cup U_r)$. We have $|U'| + |U''| \geq \lfloor v(H)/2 \rfloor$ and thus

$$\max\left\{|U'|, |U''|\right\} \geq \frac{\lfloor v(H)/2 \rfloor}{2}.$$

There are no G-edges and hence no H'-edges between U' and U'' by Construction 3.2.4. Thus, if $|U''| > |U'|$ any $u \in U'$ has H'-degree

$$d_{H'}(u) \leq (v(H)-1) - |U''| < v(H) - 1 - \frac{\lfloor v(H)/2 \rfloor}{2} = \frac{\lceil 3v(H)/2 \rceil}{2} - 1,$$

which contradicts $\delta(H') \geq \lfloor 3v(H)/4 \rfloor$. We arrive at a similar contradiction if $|U'| > |U''|$. In the case $|U'| = |U''|$ every $u \in U' \cup U''$ has $d_{H'}(u) \leq \frac{\lceil 3v(H)/2 \rceil}{2} - 1$. Since $|U' \cup U''| \geq 3$ we can pick $u \in U' \cup U''$ with $u \notin \varphi(e)$ and thus $d_{H'}(u) \geq \delta(H)$ to arrive at the desired contradiction.

Therefore we can find $j \in [r]$ with $V(H') \cap (U_1 \cup \ldots \cup U_r) = U_j$. This gives us

$$|V(H') \cap (W_1 \cup \ldots \cup W_s)| = v(H) - |U_j| = \left\lceil \frac{v(H)}{2} \right\rceil. \tag{3.17}$$

Arguments analogous to the above yield $k \in [s]$ satisfying $V(H') \cap (W_1 \cup \ldots \cup W_s) = W_k$, so

$$V(H') = U_j \cup W_k.$$

Finally, Lemma 3.2.5 shows that $U_j \cup W_k = V(H_i)$ for some $i \in [\ell]$ since

$$E_{H'}(U_j, W_k) > |U_j| \cdot \left(\frac{3v(H)}{4} - |U_j| + 1\right) - 1 \geq \left\lfloor \frac{v(H)}{2} \right\rfloor \cdot \left(\frac{v(H)}{4} + 1\right) - 1 \geq v(H).$$

Now let $e' \in E(H')$. There exists $i' \in [\ell]$ with $e' \in E(\langle H_{i'} \rangle_H)$ and hence $V(H_i \cap H_{i'}) \neq \varnothing$. Let j', k' such that $V(H_{i'}) = U_{j'} \cup W_{k'}$. If $j' \neq j$ and $k' \neq k$ we have $(j', k') \in \{(j-1, k-1), (j+1, k+1)\}$ so $e' \in \{e_i, e_{i+1}\} \subset E(\langle H_i \rangle_H)$. If $j' = j$ or $k' = k$, then by Construction 3.2.4 there exists an isomorphism $\sigma : H_{i'} \to H_i$ with $\sigma_{|V(H_i \cap H_{i'})} = \mathrm{id}_{V(H_i \cap H_{i'})}$. This implies $e' = \sigma(e') \in \sigma(\langle H_{i'} \rangle_H) = \langle H_i \rangle_H$. Here we used that $\sigma \in \mathrm{Hom}(\langle H_{i'} \rangle_H, \langle H_i \rangle_H)$ due to Observation 2.1.1.

It remains to check Condition (3) of Definition 3.2.1. Let $i \in [2, \ell]$, $j \in [r]$, $k \in [s]$ such that $V(H_i) = U_j \cup W_k$. For every $i' \in [\ell] \setminus \{i\}$ and j', k' with $V(H_{i'}) = U_{j'} \cup W_{k'}$ we have

$$E_G(U_{j'}, W_{k'}) \cap E_G(U_j, W_k) \subseteq \{e_{i-1}, e_i\}.$$

Consequently, we can see that if H' is a copy of H minus an edge in $\left(\bigcup_{i'=1}^{i-1} \langle H_{i'} \rangle_H \cup \bigcup_{i'=i}^{\ell} H_{i'}\right) \setminus \{e_{i-1}, e_i\}$ with $H' \subseteq \langle H_i \rangle_H$, then $U_j \subseteq V(H')$, $W_k \subseteq V(H')$, and

$$|E_{H'}(U_j, W_k)| \leq |E_G(U_j, W_k) \setminus \{e_{i-1}, e_i\}| \leq |E_H(U, V(H) \setminus U)| - 2.$$

This contradicts the choice of U. Thus Condition (3) holds. We have shown that $(H_i, e_i)_{i \in [\ell]}$ is indeed an H-chain, and so Claim 3.2.3 yields

$$M_H(n) \geq \ell = (1 - o(1)) \frac{n^2}{4v(H)^2}.$$

■

3.4 Random graphs

In this section we prove Theorem 1.3.9.

As mentioned in the introduction, if $p = o\left(\frac{\log k}{k}\right)$ then H contains isolated edges with high probability. This shows the first part of the claim.

In the following we occasionally use *Chernoff's inequality* to show that a random variable is, with high probability, concentrated around its mean. There are several versions of that inequality. We use the one stated as Theorems 4.2 and 4.3 in [78].

Theorem 3.4.1 (Chernoff bound). *Let $X_1, \ldots, X_n$ be independent $\{0,1\}$-valued random variables. Then for $X := \sum_{i=1}^n X_i$, $\mu := \mathbb{E}(X) = \sum_{i=1}^n \mathbb{P}(X_i = 1)$, and $0 < \delta \leq 1$ one has*

$$\mathbb{P}(X < (1-\delta)\mu) < e^{-\frac{\mu\delta^2}{2}}$$

and

$$\mathbb{P}(X > (1+\delta)\mu) < e^{-\frac{\mu\delta^2}{4}}.$$

Let $p = p(k) = \omega(\frac{\log(k)}{k})$. We have already seen in Theorem 1.3.8 that k-vertex graphs of minimum degree greater than $3k/4$ have quadratic running time so we may suppose that $p \leq 4/5$ because for larger p, Chernoff's inequality implies that $\delta(H) > 3v(H)/4$ with high probability. This assumption is needed because our method will not work when $p(k) \to 1$ and so $G(k,p)$ is close to being a complete graph.

Again we are going to invoke Construction 3.2.4. However, before we specify the required parameters, we collect a few properties of $G(k,p)$ for our choice of p.

3.4.1 Preparation

Let H be a random graph on $[0,k-1]$ that is distributed as $G(k,p)$, and let $\varepsilon' := 1/100$ be a small constant and ε be a constant depending on ε' which we will choose in a moment. We want H to have the following properties with high probability as $k \to \infty$:

(i) Every vertex of H has at least $(1/2-\varepsilon')kp$ and at most $(1/2+\varepsilon')kp$ of its neighbours in $\{0,\ldots,\lceil k/2\rceil - 1\}$ and at least $(1/2-\varepsilon')kp$ and at most $(1/2+\varepsilon')kp$ neighbours in $\{\lceil k/2\rceil,\ldots,k-1\}$.

(ii) $\max\left\{\ell' : \text{ there exist disjoint } U,W \in \binom{[0,k-1]}{\ell'} \text{ with } E_H(U,W) = \varnothing\right\} < \varepsilon k/2$

(iii) There is no subgraph of order at most $3\varepsilon k$ and minimum degree at least $(1/4-3\varepsilon')kp$.

(iv) For every edge e of H there is no non-trivial embedding of $H-e$ into H.

All four properties hold asymptotically almost surely for an edge probability of $\omega(\log(k)/k)$.

Property (i): Let v be an arbitrary vertex of H, and for each $w \in V(H)\setminus\{v\}$ let X_{vw} be 1 if $vw \in E(H)$ and 0 otherwise. The random variables X_{vw} are independent with $\mathbb{P}(X_{vw}=1) = p$. We can write the number of neighbours of v in $\{0,\ldots,\lceil k/2\rceil-1\}$ as $\sum_{w\in\{0,\ldots,\lceil k/2\rceil-1\}\setminus\{v\}} X_{vw}$, and

$$\mu := \mathbb{E}\left(\sum_{w\in\{0,\ldots,\lceil k/2\rceil-1\}\setminus\{v\}} X_{vw}\right) \geq \left(\left\lceil\frac{k}{2}\right\rceil - 1\right)p \geq \left(\frac{1}{2}-\frac{1}{k}\right)kp. \tag{3.18}$$

Note that $(1/2-\varepsilon')kp \leq (1-\varepsilon')\mu$ when k is sufficiently large. Chernoff's inequality together with (3.18) implies

$$\mathbb{P}\left(\sum_{w\in\{0,\ldots,\lceil k/2\rceil-1\}\setminus\{v\}} X_{vw} < \left(\frac{1}{2}-\varepsilon'\right)kp\right) < e^{-\frac{\mu\varepsilon'^2}{2}} = e^{-\omega(1)\cdot\log k}$$

Taking a union bound over all k possible choices of v shows that w.h.p. every vertex of H has at least $(1/2-\varepsilon')kp$ neighbours in $\{0,\ldots,\lceil k/2\rceil-1\}$. Similar arguments show that the remaining three estimates of Property (i) hold with high probability, too.

Property (ii): This one follows from a union bound over all pairs of vertex sets of a given size. More precisely, for $\lceil \varepsilon k/2\rceil \leq \ell' \leq \lfloor k/2\rfloor$, let $F_{\ell'}$ be the event that there exist disjoint $U,W \in \binom{V(H)}{\ell'}$

with $E_H(U,W) = \varnothing$. We clearly have $F_{\ell'} \subseteq F_{\ell''}$ for $\ell'' \leq \ell'$. Therefore,

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{\ell'=\lceil \varepsilon k/2\rceil}^{\lfloor k/2\rfloor} F_{\ell'}\right) = \mathbb{P}(F_{\lceil \varepsilon k/2\rceil}) &\leq \binom{k}{\lceil \varepsilon k/2\rceil}^2 \cdot (1-p)^{\lceil \varepsilon k/2\rceil} \\
&\leq \left(\frac{k\cdot e}{\lceil \varepsilon k/2\rceil}\right)^{2\lceil \varepsilon k/2\rceil} \cdot e^{-p\lceil \varepsilon k/2\rceil^2} \\
&\leq \left(\frac{2e}{\varepsilon}\right)^{2\lceil \varepsilon k/2\rceil} \cdot e^{-\omega(1)\log(k)\cdot k\varepsilon^2/4} \\
&= e^{\log(2/\varepsilon)\cdot 2\lceil \varepsilon k/2\rceil + 2\lceil \varepsilon k/2\rceil - \omega(1)\log(k)\cdot k\varepsilon^2/4} \\
&\to 0 \quad \text{as } k \to \infty.
\end{aligned}$$

Property (iii): This property is a consequence of the following claim:

Claim 3.4.2. *Let $k \in \mathbb{N}$. For every $\delta > 0$ there exists $\varepsilon > 0$ such that with high probability $H \sim G(k,p)$ does not contain any subgraph of order at most $3\varepsilon k$ and minimum degree at least δkp.*

Proof. Choose $\varepsilon > 0$ such that $\frac{3\varepsilon e}{\delta} < e^{-1}$. Denote the bad event that H has a subgraph of order at most $3\varepsilon k$ and minimum degree at least δkp by B. For notational simplicity we assume that $\delta kp/2$ is an integer. For any fixed $1 \leq m \leq 3\varepsilon k$ we have

$$\begin{aligned}
\mathbb{P}(B) &\leq \sum_{Z\subset V(H):|Z|=m} \mathbb{P}\left(\delta(H[Z]) \geq \delta kp\right) \\
&\leq \sum_{Z\subset V(H):|Z|=m} \mathbb{P}\left(|E(H[Z])| \geq m\delta kp/2\right) \\
&\leq \sum_{Z\subset V(H):|Z|=m} \binom{\binom{m}{2}}{m\delta kp/2} \cdot p^{m\delta kp/2} \\
&\leq \binom{k}{m} \cdot \binom{\binom{m}{2}}{m\delta kp/2} \cdot p^{m\delta kp/2} \\
&\leq \left(\frac{ke}{m}\right)^m \cdot \left(\frac{m^2 e}{m\delta kp}\right)^{m\delta kp/2} \cdot p^{m\delta kp/2} \\
&< (k\cdot e)^m \cdot \left(\frac{3\varepsilon e}{\delta}\right)^{m\delta kp/2} \\
&\leq e^{m\cdot(1+\log k - \delta\cdot\omega(1)\cdot\log k)}
\end{aligned}$$

Summing over all $m \leq 3\varepsilon k$ shows that the probability of B is bounded from above by

$$\begin{aligned}
\sum_{m=1}^{\lfloor 3\varepsilon k\rfloor} e^{m\cdot(1+\log k-\delta\cdot\omega(1)\cdot\log k)} &< \frac{1}{1-e^{1+\log k-\delta\cdot\omega(1)\cdot\log k}} - 1 \\
&\to 0 \quad \text{as } k\to\infty
\end{aligned}$$

□

Now let ε be as given by the claim above when $\delta = 1/4 - 3\varepsilon'$ so w.h.p. (iii) holds.

Property (iv): The fourth property is similar to the statement that with high probability $G(k,p)$ has no non-trivial automorphisms. It is a direct consequence of the following result of Kim, Sudakov and Vu:

Definition 3.4.3 ([63] Definition 2.1)**.** Let G be a graph. The *defect* of G with respect to a permutation $\pi : V(G) \to V(G)$ is defined to be

$$D_\pi(G) = \max_{v \in V(G)} |N(\pi(v)) \triangle \pi(N(v))|$$

The *defect* of G is

$$D(G) := \min_{\pi \neq \mathrm{id}} D_\pi(G)$$

where the minimum ranges over all non-trivial permutations of the vertex set of G.

Lemma 3.4.4 ([63] Theorem 3.1)**.** *If p satisfies $p = \omega(\log(k)/k)$ and $1 - p = \omega(\log(k)/k)$ then*

$$D(G(k,p)) = (2 - o(1)) \cdot kp(1-p)$$

with high probability as $k \to \infty$.

With our value of p, Lemma 3.4.4 tells us that with high probability every bijection $\pi : V(H) \to V(H)$ apart from the identity comes with a vertex $v \in V(H)$ such that

$$|N_H(\pi(v)) \triangle \pi(N_H(v))| > \log k. \tag{3.19}$$

Let $e \in E(H)$, and let $\pi : V(H) \to V(H)$ be an embedding of $H - e$ into H, that is, π is a bijection and sends all edges of H but at most e to edges of H. Then π satisfies

$$|N_H(\pi(v)) \triangle \pi(N_H(v))| \leq 2$$

for every $v \in V(H)$. This contradicts (3.19) unless π is the identity map on $V(H)$. Thus Property (iv) is satisfied with high probability.

We have seen that the four properties above hold with high probability as $k \to \infty$. Now we are ready to construct a graph G that achieves $\tau_H(G) = \Omega(n^2)$.

3.4.2 Building the chain

Suppose that the properties (i)-(iv) hold. Apply Construction 3.2.4 with H the random graph above, n an arbitrary natural number, $U = \{0, \dots, \lceil k/2 \rceil - 1\}$, and e, f two arbitrary non-incident edges between U and $V(H) \setminus U$. A suitable choice of e and f exists by property (i). Due to Property (iv), H must be self-stable. Because of Construction 3.2.4 and Remark 3.2.2 it suffices to show that for each $e' \in E(H)$, every copy of $H - e'$ in $G := \bigcup_{i=1}^{\ell} H_i$ is contained in H_i for some

$i \in [\ell]$. Let $e' \in E(H)$, and let $H' \subseteq G$ be a copy of $H - e'$. Define

$$\mathscr{U} := U_1 \cup \ldots \cup U_r \quad \text{and} \quad \mathscr{W} := W_1 \cup \ldots \cup W_s.$$

By averaging we have $|\mathscr{U} \cap V(H')| \geq \lceil k/2 \rceil$ or $|\mathscr{W} \cap V(H')| \geq \lfloor k/2 \rfloor$. Assume the former (in the latter case the arguments are analogous). This allows us to fix the smallest $j \in [r]$ with

$$|V(H') \cap (U_1 \cup \ldots \cup U_j)| \geq \frac{\varepsilon k}{2} + 1.$$

Recall that $u_{j \cdot \lceil k/2-1 \rceil}$ is the unique common vertex of U_j and U_{j+1}. Then

$$|V(H') \cap (U_{j+1} \cup \ldots \cup U_n) \setminus \{u_{j \cdot \lceil k/2-1 \rceil}\}| \leq \frac{\varepsilon k}{2}$$

for otherwise this set and $V(H') \cap (U_1 \cup \ldots \cup U_j) \setminus \{u_{j \cdot \lceil k/2-1 \rceil}\}$, which by construction of G have no edges between each other, would violate property (ii). It follows that

$$\begin{aligned} |V(H') \cap \mathscr{U}| &\leq |V(H') \cap (U_1 \cup \ldots \cup U_{j-1})| + |U_j| + |V(H') \cap (U_{j+1} \cup \ldots \cup U_n) \setminus \{u_{j \cdot \lceil k/2-1 \rceil}\}| \\ &\leq \left\lceil \frac{k}{2} \right\rceil + \varepsilon k \end{aligned}$$

and that the vertices in $V(H') \cap \mathscr{U}$ are concentrated in U_j, i.e.

$$\begin{aligned} |V(H') \cap U_j| &\geq \left\lceil \frac{k}{2} \right\rceil - |V(H') \cap (U_1 \cup \ldots \cup U_{j-1})| - |V(H') \cap (U_{j+1} \cup \ldots \cup U_n) \setminus \{u_{j \cdot \lceil k/2-1 \rceil}\}| \\ &\geq \left\lceil \frac{k}{2} \right\rceil - \varepsilon k \end{aligned}$$

This implies $|V(H') \cap W| > \lfloor k/2 \rfloor - \varepsilon k$, hence analogous arguments give us

$$|V(H') \cap W_{j'}| \geq \left\lfloor \frac{k}{2} \right\rfloor - 2\varepsilon k$$

where j' is the smallest element of $[n]$ with $|V(H') \cap (W_1 \cup \ldots \cup W_{j'})| \geq 1 + \varepsilon k/2$.

Let $Z := V(H') \setminus (U_j \cup W_{j'})$. The order of $H'[Z]$ is at most $3\varepsilon k$. By construction of G, vertices in $\mathscr{U} \setminus U_j$ have at most one G-neighbour in U_j and by Property (i) vertices in $V(H') \cap \mathscr{U}$ have at most $(1/2 + \varepsilon')kp$ H'-neighbours in $W_{j'}$. Thus

$$\begin{aligned} d_{H'[Z]}(z) &= d_{H'}(z) - |N_{H'}(z) \cap U_j| - |N_{H'}(z) \cap W_{j'}| \\ &\geq (1 - 2\varepsilon')kp - 1 - \left(\frac{1}{2} + \varepsilon'\right) kp \\ &= \left(\frac{1}{2} - 3\varepsilon'\right) kp - 1 \\ &\geq \left(\frac{1}{4} - 3\varepsilon'\right) kp \end{aligned}$$

for every $z \in Z \cap \mathscr{U}$ and k sufficiently large. Similarly, $d_{H'[Z]}(z) \geq (1/4 - 3\varepsilon')kp$ for $z \in Z \cap \mathscr{W}$.

However, Property (iv) forbids subgraphs of minimum degree $(1/4-3\varepsilon')kp$ and order at most $3\varepsilon k$. Therefore $Z=\varnothing$, so $V(H')=U_j\cup W_{j'}$.

Because of property (i) every $u\in U_j$ has at most $(1/2-\varepsilon')kp$ G-neighbours in U_j, so

$$|E_{H'}(U_j,W_{j'})|\;\geq\;|U_j|\cdot\left(\delta(H')-\left(\frac{1}{2}-\varepsilon'\right)kp\right)\geq\left\lceil\frac{k}{2}\right\rceil\cdot\left(\frac{1}{2}-3\varepsilon'\right)kp>k$$

whenever k is sufficiently large. Now the second part of Lemma 3.2.5 implies $U_j\cup W_{j'}=V(H_i)$ for some $i\in[\ell]$. Since $G[V(H_i)]=H_i$ by construction of G we obtain $H'\subseteq H_i$. Thus $(H_i,e_i)_{i\in[\ell]}$ is indeed an H-chain, and

$$M_H(n)\geq\ell=(1-o(1))\frac{n^2}{4k^2}.$$

■

3.5 Chains via extremal additive problems

In this section we describe the two instances of chain constructions that underly Theorems 1.3.6 and 1.3.10. Both of them are consequences of Lemma 3.2.11 and use extremal sets from additive combinatorial problems to guarantee that Condition (2) of Definition 3.2.1 holds. The first is a generalisation of Theorem 1.1.8 that works for a class of graphs among which we can find K_5 and W_k for $k\geq 7$. The second is a translation of the first to a bipartite setting.

The underlying idea originally implemented in [13] was to associate undesired copies of K_5^- with non-trivial solutions to the equation $x+z=2y$ in the integers. We will generalise that approach by associating copies of H minus an edge (for H belonging to a certain class of graphs) with non-trivial solutions to a system of linear equations.

A solution $(a_1,\ldots,a_r)$ to a linear equation

$$k_1a_1+\ldots+k_ra_r=0$$

with coefficients $k_1,\ldots,k_r\in\mathbb{Z}$ is called *trivial* if the coefficients of coinciding values sum to zero, that is, there exists $s\leq r$ and a partition $[r]=I_1\cup\ldots\cup I_s$ such that $a_i=a_{i'}$ for $i,i'\in I_j$ and $\sum_{i\in I_j}k_i=0$ for all $1\leq j\leq s$.

3.5.1 Non-bipartite setting: Prerequisites

The 3-AP-free construction of Behrend can be used to forbid multiple linear equations in three variables at once. This requires only minor modifications in Behrend's original proof. The following two results (Theorem 3.5.1 and Corollary 3.5.2) are the mentioned modifications and are not new. They are merely included for completeness.

Theorem 3.5.1. *Let $k\geq 2$. There exists a constant $c>0$ such that for every $n\in\mathbb{N}$ one can find a set $S\subset[n]$ of size at least $n^{1-c/\sqrt{\log n}}$ such that the equation*

$$x_1+\ldots+x_k=kx_{k+1} \tag{3.20}$$

does not admit any non-trivial solution in S.

Proof. Given integral parameters $q \geq 2$, $d \geq 2$, $r \leq d(q-1)^2$, define the set

$$S_r(d,q) := \left\{a_1 + a_2(kq-1) + \ldots + a_d(kq-1)^{d-1} : a_1, \ldots, a_d \in [0, q-1], a_1^2 + \ldots + a_d^2 = r\right\}.$$

The upper bound on r is not a real restriction since $S_r(d,q)$ would be empty for larger values of r. The numbers $a_1, \ldots, a_d$ in the definition above are the digits in base $kq-1$, and hence they are uniquely determined for each $a \in S_r(d,q)$. For any $a \in \mathbb{N}$ with digits $a_1, \ldots, a_d$ in base $kq-1$, let $\|a\| := \sqrt{a_1^2 + \ldots + a_d^2}$. Note that if $a_i < q$ for all $i \in [d]$ then $\|ka\| = k\|a\|$.

Suppose there exist $a^{(1)}, \ldots, a^{(k)}, b \in S_r(d,q)$ such that $a^{(1)} + \ldots + a^{(k)} = kb$. We have that

$$\|a^{(1)} + \ldots + a^{(k)}\| = \|kb\| = k\sqrt{r} = \|a^{(1)}\| + \ldots + \|a^{(k)}\|.$$

Therefore the k vectors $(a_1^{(j)}, \ldots, a_d^{(j)}) \in \mathbb{R}^d$, $j \in [k]$, must be pairwise linearly dependent as this is the only case in which the triangle inequality is in fact an equality. The definition of $S_r(d,q)$ now implies $a_i^{(j)} = a_i^{(j)}$ for $i \in [d], j \in [d]$ and hence $a^{(1)} = \ldots = a^{(k)} = b$.

We have seen that $S_r(d,q)$ is free of non-trivial solutions to (3.20). Now it remains to choose parameters r, d, q such that $|S_r(d,q)|$ is as large as possible for $n \in \mathbb{N}$ with $S_r(d,q) \subset [n]$.

For a fixed choice of d and q,

$$\sum_{r=0}^{d(q-1)^2} |S_r(d,q)| = q^d.$$

By averaging we can pick $r = r(d,q)$ such that

$$|S_r(d,q)| \geq \frac{q^d}{d(q-1)^2+1} \geq \frac{q^{d-2}}{d}$$

With the choices $d := \lfloor\sqrt{\log n}\rfloor$ and $q := \lfloor\frac{n^{1/d}}{k}\rfloor$ we have both $S_r(d,q) \subseteq [n]$ and $|S| \geq n^{1-c/\sqrt{\log n}}$ for a suitable constant $c > 0$. □

Corollary 3.5.2. *Let $k \geq 2$. There exists a constant $c > 0$ such that for every $N \in \mathbb{N}$ one can find a set $S \subset [N]$ of size at least $N^{1-c/\sqrt{\log N}}$ such that none of the congruences*

$$\alpha x + \beta y + \gamma z \equiv 0 \mod N \qquad \textit{with} \qquad \alpha, \beta, \gamma \in [-k, k] \tag{3.21}$$

admit a non-trivial solution in S.

Proof. Apply Theorem 3.5.1 with k as given and $n = \lfloor N/(3k+1)\rfloor$. Call the obtained set S'. Let $r \in [3k+1]$ such that $|\{s' \in S' : s' \equiv r \mod (3k+1)\}|$ is maximised, and define

$$S := \{s' - (r-1) : s' \in S', s' \equiv r \mod (3k+1)\}.$$

We remark that $s \equiv 1 \mod (3k+1)$ for all $s \in S$. The size of S is at least $|S'|/(3k+1)$, which is still of the form $N^{1-c/\sqrt{\log N}}$. Let $\alpha, \beta, \gamma \in [-k,k]$, and suppose that at least two of them are non-negative and $\alpha \geq \beta \geq \gamma$. Any congruence of the form (3.21) in S becomes an equality since both sides are smaller than N in absolute value. Being free of non-trivial solutions to (3.20) is preserved under taking subsets and translates. Thus, if $\alpha+\beta+\gamma = 0$ the claim now follows from the observation that (3.21) is a special case of (3.20) with

$$x_1 = \ldots = x_\alpha, \qquad x_{\alpha+1} = \ldots = x_{\alpha+\beta}, \qquad x_{\alpha+\beta+1} = \ldots = x_k = x_{k+1}.$$

Now suppose that $\alpha+\beta+\gamma \neq 0$. We have

$$\alpha a_1 + \beta a_2 + \gamma a_3 \equiv \alpha+\beta+\gamma \not\equiv 0 \mod (3k+1)$$

for all $a_1, a_2, a_3 \in S$, and hence

$$\alpha a_1 + \beta a_2 + \gamma a_3 \not\equiv 0 \mod N$$

since $-N < \alpha a_1 + \beta a_2 + \gamma a_3 < N$. □

3.5.2 Non-biparte setting: Construction of the chains

Corollary 3.5.2 provides the extremal set we need to establish the following result, which is the aforementioned generalisation of Theorem 1.1.8

Theorem 3.5.3. *Let H be a fixed graph such that for some $e, f \in E(H)$ arbitrarily long proper (e,f)-simple H-chains exist, and that for each $e' \in E(H)$ every non-monochromatic colouring of the edges of $H - e'$ admits a cycle with at least two and at most three colours in which each colour class forms a path. Then $M_H(n) \geq n^{2-O(1/\sqrt{\log n})}$.*

Proof. Choose a prime p between $n/8$ and $n/4$. Corollary 3.5.2 allows us to pick a set $A \subseteq [p-1]$ of size $|A| \geq n^{1-O(1/\sqrt{\log n})}$ which is free of non-trivial solutions to any of the congruences

$$\alpha x + \beta y + \gamma z \equiv 0 \mod p \tag{3.22}$$

where $\alpha, \beta, \gamma \in [-v(H)^2, v(H)^2]$.

By assumption there exist $e, f \in E(H)$ such that there are arbitrarily long proper (e,f)-simple H-chains. Let V be a p-element set whose elements are labelled by v_x, $x \in [0, p-1]$, and let W be another p-element set with elements w_x, $x \in [0, p-1]$ that is disjoint from V. Let $\ell := \lfloor (p-3)/(v(H)-2) \rfloor$, and let $k := v(H)-2$. For $x \in [0, p-1]$ define

$$u_x := \begin{cases} v_x & \text{, if } 2+(v(H)+1)k \leq x < (\ell - v(H)-1)k; \\ w_x & \text{, otherwise.} \end{cases}$$

Place an (e,f)-simple H-chain $(H_i, e_i)_{i\in[\ell]}$ on the vertices $u_1, \ldots, u_{2+\ell k}$ such that the vertex set of the i^{th} copy of H is $V(H_i) = \{u_{1+(i-1)k}, \ldots, u_{2+ik}\}$. For $x, y \in [0, p-1]$ with $u_x u_y \in E(\bigcup_{i=1}^{\ell} \langle H_i \rangle_H)$

one has

$$x-y\in[-v(H)+1,v(H)-1]\setminus\{0\}.$$

For $a\in A$ and $i\in[\ell]$ define

$$H_i^a:=a\cdot H_i \quad \text{and} \quad e_i^a:=a\cdot e_i$$

where $a\cdot H_i$ and $a\cdot e_i$ are the images of H_i and e_i under the injective map $V\cup W\to V\cup W$, $v_x\mapsto v_{ax \bmod p}$, $w_y\mapsto w_{ay \bmod p}$. Let G^a be the underlying graph of the a^{th} chain.

We will now check the conditions of Lemma 3.2.11. From (3.22) with $\gamma=0$ we infer that

$$j\cdot a\not\equiv j'\cdot b \mod p \tag{3.23}$$

for $a,b\in A$, $a\neq b$, and $j,j'\in[-v(H)^2,v(H)^2]$ unless $j=j'=0$. Since any $x,y\in[0,p-1]$ with $u_xu_y\in E(G^a)$ satisfy $x-y=j\cdot a$ for some $j\in[-v(H)+1,v(H)-1]$, we can see that the G^a are pairwise edge-disjoint. In particular,

$$\{e_0^a,\dots,e_\ell^a\}\cap E(G^b)=\varnothing$$

for distinct $a,b\in A$ so the condition (i) is fulfilled.

The sets $W^a:=\{w_{j\cdot a \bmod p}:j\in[-v(H)^2,v(H)^2]\setminus\{0\}\}$, $a\in A$, are pairwise disjoint due to (3.23). We have that both $\bigcup_{i=1}^{v(H)}V(H_i^a)\subset W^a$ and $\bigcup_{i=\ell-v(H)+1}^{v(H)}V(H_i^a)\subset W^a$. Therefore (3.12) holds, too.

Next we verify that every copy of H minus an edge in $G:=\bigcup_{a\in A}G^a$ is fully contained in exactly one of the G^a. Suppose that for some $e'\in E(H)$ there was a copy of $H-e'$ with edges from distinct chains. We obtain a non-monochromatic edge-colouring of this copy by colouring each edge e'' with the unique $a\in A$ such that $e''\in E(G^a)$. Then for some $r\in[v(H)]$, there exists an r-cycle C with at least two and at most three colours whose colours classes form paths, that is, we can find $x_0,\dots,x_{r-1}\in[0,p-1]$, $a,b,c\in A$, $\alpha_1,\dots,\alpha_r\in[-v(H)+1,v(H)-1]$, and $0\le r'<r''\le r-1$ such that $a\notin\{b,c\}$ and

$$x_{j+1 \mod r}-x_j\equiv\begin{cases}\alpha_j a & ,\ 0\le j\le r';\\ \alpha_j b & ,\ r'<j\le r'';\\ \alpha_j c & ,\ r''<j\le r-1\end{cases} \mod p.$$

But now we have that

$$0\equiv\sum_{j=0}^{r-1}(x_{j+1 \mod r}-x_j)\equiv\sum_{j=0}^{r'}\alpha_j a+\sum_{j=r'}^{r''}\alpha_j b+\sum_{j=r''}^{r-1}\alpha_j c, \quad \mod p$$

where $\sum_{j=1}^{r'}\alpha_j\neq 0$ and $\sum_{j=r'+1}^{r''}\alpha_j\neq 0$ since the edges of C of colours a, b, respectively, form a path. This contradicts the definition of A.

All of our $|A|$ chains have length $\ell = \Theta(n)$. Lemma 3.2.11 yields

$$M_H(n) \geq M_H(2p + |A| \cdot v(H)) \geq |A| \cdot \ell \geq n^{2-O(1/\sqrt{\log n})},$$

where the leftmost inequality requires n to be sufficiently large. □

We note that the hypotheses of Theorem 3.5.3 hold for $H = K_5$, and thus we indeed recover Theorem 1.1.8. Verifying the colouring condition in Theorem 3.5.3 is not always straightforward. The following provides a condition that implies the necessary colouring condition.

Lemma 3.5.4. *Given a graph H, let $\mathscr{T}(H)$ be the graph whose vertices are the triangles in H and in which two vertices T_1, T_2 are adjacent if $e(T_1 \cap T_2) = 1$. If $e' \in E(H)$ such that every edge of $H - e'$ is contained in a triangle, and $\mathscr{T}(H - e')$ is connected, then in each non-m.c. edge-colouring of $H - e'$ there exists a non-m.c. triangle.*

Proof. Suppose for a contradiction that there exists $e' \in E(H)$ and an edge-colouring χ of $H - e'$ such that each edge of $H - e'$ lies in a triangle, $\mathscr{T}(H - e')$ is connected, and every triangle in $H - e'$ is monochromatic under χ. Since any two triangles with a common edge must have the same colour, we observe that the colour classes of χ induce a partition of $\mathscr{T}(H - e')$ into monochromatic connected components. But then $\mathscr{T}(H)$ must have at least two components because χ is non-monochromatic and each colour occurs in some triangle. This contradicts the assumption that $\mathscr{T}(H - e')$ is connected. □

Note that in a non-m.c. triangle the colour classes always form paths. We will invoke Lemma 3.5.4 in later sections when we apply Theorem 3.5.3.

3.5.3 Bipartite setting: K-fold Sidon sets

We translate the approach from the non-bipartite setting to the bipartite one by replacing the sets that are free of certain three-variable equations by so-called K-fold Sidon sets, where K denotes an arithmetic progression of natural numbers. These sets are a generalisation of k-fold Sidon sets, which were first defined by Lazebnik and Verstraëte in [73] to investigate hypergraphs of girth five.

Definition 3.5.5. Let $k, n \in \mathbb{N}$. A subset A of $[n]$ or $\mathbb{Z}_n$ is called a *k-fold Sidon set* if for any $a_1, a_2, a_3, a_4 \in A$ and $k_1, k_2, k_3, k_4 \in [-k, k]$ with $k_1 + k_2 + k_3 + k_4 = 0$ it does not contain any non-trivial solutions to the equation

$$k_1 x_1 + k_2 x_2 + k_3 x_3 + k_4 x_4 = 0. \tag{3.24}$$

We are interested in the maximum size of such sets. A simple counting argument shows that a largest k-fold Sidon set in $[n]$ has size $O(n^{1/2})$. In [30] it was pointed out that adapting a construction of Ruzsa [85] yields k-fold Sidon sets of order $\Omega(n^{1/2-o(1)})$. Unfortunately [30] does not contain any write-up of that adaption and there does not seem to be any other article containing one. However, for our application to graph bootstrap percolation it suffices if the absolute values of the coefficients in (3.24) are not taken from the whole interval $[0,k]$ but from a sufficiently long arithmetic progression. This motivates the following definition.

Definition 3.5.6. Let K be an arithmetic progression in the non-negative integers, and let $n \in \mathbb{N}$. A subset A of $[n]$ or $\mathbb{Z}_n$ is called a *K-fold Sidon set* if for any $k_1, k_2, k_3, k_4 \in K \cup (-K)$ with $k_1 + \ldots + k_4 = 0$ it does not contain any non-trivial solutions to the equation

$$k_1x_1 + k_2x_2 + k_3x_3 + k_4x_4 = 0. \tag{3.25}$$

We call a K-fold Sidon set *augmented* if (3.25) is also forbidden for $k_1, k_2, k_3, k_4 \in K \cup (-K)$ with $k_1 + \ldots + k_4 \neq 0$.

Recall that an interval can be interpreted as an arithmetic progression with common difference one. A k-fold Sidon set is the same as a $[0,k]$-fold Sidon set. Moreover if $k \geq \max K$ then any k-fold Sidon set is also a K-fold Sidon set. In the following we denote the size of a largest K-fold Sidon set in $[n]$ or $\mathbb{Z}_n$, respectively, by $s_K(n)$ or $s_K(\mathbb{Z}_n)$, respectively. We also write $s'_K(n)$ and $s'_K(\mathbb{Z}_n)$ for the sizes of largest augmented K-fold Sidon sets. Let us collect a few observations on those quantities.

Observation 3.5.7. Let K be an arithmetic progression of positive integers. The following hold.

1. $s_K(n)$ is non-decreasing in n.
2. $s_K(\lambda n) \leq \lambda s_K(n)$ for all $\lambda \in \mathbb{N}$.
3. $s'_K(\mathbb{Z}_n) \geq s'_K(\lfloor n/(4\max K + 1) \rfloor)$
4. $s'_K(n) \geq s_K(n)/(4\max K + 1)$.

Proof. (1) A K-fold Sidon set in $[n]$ is clearly one in $[n']$ for every $n' \geq n$.

(2) Being a K-fold Sidon set is preserved under taking subsets and translates. Thus if $A \subset [\lambda n]$ is a K-fold Sidon set so are the shifted subsets $(A \cap [(j-1)n+1, jn]) - (j-1)n$, $1 \leq j \leq \lambda$, each of which has size at most $s_K(n)$.

(3) An augmented K-fold Sidon set in $[n/(4\max K + 1)]$ can be interpreted as an augmented K-fold Sidon set in $\mathbb{Z}_n$ since for $k_1, \ldots, k_4 \in K \cup (-K)$ and $a_1, \ldots, a_4 \in [n/(4\max K + 1)]$ the term $|k_1a_1 + \ldots + k_4a_4|$ never exceeds $n-1$.

(4) Given a K-fold Sidon set $A \subset [n]$ we can pick $r \in [0, 4\max K]$ such that $|\{a \in A : a \equiv r \mod (4\max K + 1)\}|$ is maximised. Then the translated subset $A' := \{a - (r-1) : a \in A, a \equiv r \mod (4\max K + 1)\}$ is an augmented K-fold Sidon set since for any $a_1, \ldots, a_4 \in A'$ and $k_1, \ldots, k_4 \in K \cup (-K)$ with $k_1 + \ldots + k_4 \neq 0$ one has

$$k_1a_1 + \ldots + k_4a_4 \equiv k_1 + \ldots + k_4 \not\equiv 0 \mod (4\max K + 1).$$

The size of A' at least $|A|/(4\max K + 1)$ by our choice of r. $\square$

3.5.4 Bipartite setting: Construction of the chains

We are now able to formulate the following bipartite variant of Theorem 3.5.3.

Theorem 3.5.8. *Let H be a bipartite graph such that arbitrarily long proper simple H-chains exist, and assume that for each $e \in E(H)$, every non-m.c. edge-colouring of $H - e$ contains a non-m.c. copy of C_4. Then for each fixed arithmetic progression K of length $2v(H)$ and common difference at least 2 in the non-negative integers, there exists a constant $c_{H,K} > 0$ that depends only on H and K such that one has*

$$M_H(n) \geq c_{H,K} \cdot n \cdot s_K(n). \tag{3.26}$$

Proof. As in the non-bipartite setting we use Lemma 3.2.11. Let p be a prime with $n/16 \leq p \leq n/8$. Denote the sizes of the two partite sets of H by s and t, respectively, and suppose that $s \leq t$. Pick two disjoint p-element sets $U = \{u_x : x \in [0, p-1]\}$ and $W = \{w_y : y \in [0, p-1]\}$.

Denote the common difference of K by d and the smallest element of K by m. Since H is bipartite and allows simple chains we can choose suitable $e, f \in E(H)$ and place a proper, (e, f)-simple H-chain $(H_i, e_i)_{i\in[\ell]}$ of length $\ell := \left\lfloor \frac{p}{2d\cdot t} \right\rfloor$ on $V := U \cup W$ such that for any $i \in [\ell]$, one has $V(H_i) = U_i \cup W_i$ and $E(H_i) \subseteq \{uw : u \in U_i, w \in W_i\}$, where

$$\begin{aligned} U_i &:= \{u_{d\cdot x} : x \in [1+(i-1)(t-1), 1+(i-1)(t-1)+(s-2)]\} \cup \left\{u_{d\cdot(1+i(t-1))}\right\}, \\ W_i &:= \left\{w_{d\cdot y+m} : y \in [2+i(t-1), 1+(i+1)(t-1)]\right\} \cup \left\{w_{d\cdot(2+(i+1)(t-1))+m}\right\}. \end{aligned}$$

This is well-defined in the sense that for sufficiently large p all indices that occur in the definition of U_i and W_i are smaller than p. With the labelling above, one has $y - x \in K$ whenever $u_x \in U_i$ and $w_y \in W_i$ for some $x, y \in [0, p-1]$ and $i \in [\ell]$.

Let $A \subset [p-1]$ be a largest augmented K-fold Sidon set when interpreted as a subset of $\mathbb{Z}_p$. For $a \in A$ and $1 \leq i \leq \ell$ define

$$H_i^a := a \cdot H_i \qquad , \qquad e_i^a := a \cdot e_i$$

where $a \cdot H_i$ and $a \cdot e_i$ are the images of H_i and e_i, respectively, under the injective map

$$V \to V, \quad u_x \mapsto u_{ax \bmod p}, \; w_y \mapsto w_{ay \bmod p}.$$

The sequences $(H_i^a, e_i^a)_{i\in[\ell]}$, $a \in A$, are proper, (e, f)-simple H-chains. For $a \in A$, let $G^a := \bigcup_{i=1}^{\ell} H_i^a$ be the underlying graph of $(H_i^a, e_i^a)_{i\in[\ell]}$ and observe that every edge of G^a as well as every edge of $\langle G^a \rangle_H$ is of the form $u_x w_y$ where $y - x \equiv k \cdot a \mod p$ for some $k \in K$. Note that $ka \not\equiv k'b \mod p$ for $a \neq b$ and $k, k' \in K$ and hence $E(G^a) \cap E(G^b) = \varnothing$ and $E(\langle G^a \rangle_H) \cap E(\langle G^b \rangle_H) = \varnothing$.

The assumption that A is a K-fold Sidon set translates into the property that every C_4 in $G := \bigcup_{a\in A} \langle G^a \rangle_H$ lies in a single $\langle G^a \rangle_H$. Indeed, let C' be a cycle of length four in G and suppose for a contradiction that C' does not lie in a single $\langle G^a \rangle_H$. Let $x, y, x', y' \in [0, p-1]$ such that $u_x, w_y, u_{x'}, w_{y'}$ are the vertices of C'. There exist $a_1, a_2, a_3, a_4 \in A$, not all equal, and $k_1, k_2, k_3, k_4 \in$

K such that

$$y-x\equiv k_1a_1 \quad , \quad y-x'\equiv k_2a_2 \quad , \quad y'-x'\equiv k_3a_3 \quad , \quad y'-x\equiv k_4a_4 \mod p,$$

and thus

$$k_1a_1-k_2a_2+k_3a_3-k_4a_4\equiv 0 \mod p.$$

We cannot have $k_1a_1\equiv k_2a_2$ or $k_1a_1\equiv k_4a_4 \mod p$ for otherwise $x=x'$ or $y=y'$. Similarly $k_3a_3\not\equiv k_2a_2$ and $k_3a_3\not\equiv k_4a_4 \mod p$. Finally, $k_1+k_3\not\equiv 0$ and $k_2+k_4\not\equiv 0 \mod p$ whenever p is sufficiently large. Since $a_1,\dots,a_4$ are not all the same the solution must be non-trivial. But this contradicts the fact that A is an augmented K-fold Sidon set in $\mathbb{Z}_p$.

Unfortunately, the chains above are not guaranteed to satisfy (3.12). The latter is necessary to apply Lemma 3.2.11 because H was not assumed to be inseparable or bipartite-inseparable. For this reason we will make the initial and terminal segments of the chains artificially disjoint by introducing new vertices. For each $a\in A$ and $v\in V$ with $v\in V(H_i^a)$ for some $i\in[v(H)]\cup[\ell-v(H),\ell]$ introduce a new vertex v^a. Write $V^a:=\{v^a : v\in\bigcup_{i\in[v(H)]\cup[\ell-v(H)+1,\ell]}V(H_i^a)\}$ and define

$$\theta^a : V\to V\cup V^a,\quad \theta^a(v):=\begin{cases} v^a & \text{, if } v\in\bigcup_{i\in[v(H)]\cup[\ell-v(H)+1,\ell]}V(H_i);\\ v & \text{, otherwise.}\end{cases}$$

Consider the chains $\bar{\mathscr{H}}^a:=(\bar{H}_i^a,\bar{e}_i^a)_{i\in[\ell]}$, where $\bar{H}_i^a:=\theta^a(H_i^a)$ and $\bar{e}_i^a:=\theta^a(e_i^a)$, and let $\bar{G}^a:=\bigcup_{i=1}^{\ell}\bar{H}_i^a$. The sets V^a, $a\in A$, are pairwise disjoint by definition, so the new chains $\bar{\mathscr{H}}^a$ satisfy (3.12). Moreover, $\bar{e}_i^a\notin E(\bar{G}^b)$ for $a\neq b$ as G^a and G^b are edge-disjoint and $V(\bar{G}^a)\cap V(\bar{G}^b)\subseteq V(G^a)\cap V(G^b)$. It remains to show that (ii) of Lemma 3.2.11 holds for the new chains.

Suppose that $\bar{H}$ is a copy of H minus an edge in $\bigcup_{a\in A}\langle\bar{G}^a\rangle_H$ which is not fully contained in $\langle\bar{G}^a\rangle_H$ for any $a\in A$. We can colour the edges of $\bar{H}$ by assigning to $e\in E(\bar{H})$ the unique $a\in A$ with $e\in E(\langle\bar{G}^a\rangle_H)$. This colouring admits a non-monochromatic four-cycle $\bar{C}$. We now project $\bar{C}$ to the original chains. More precisely, we define $\pi : V\cup\bigcup_{a\in A}V^a\to V$ by setting $\pi(v):=v$ and $\pi(v^a):=v$ for all $v\in V$ and $a\in A$. For each $a\in A$, π is injective on $V(\bar{G}^a)$ and sends edges of $\langle\bar{G}^a\rangle_H$ to edges of $\langle G^a\rangle_H$. This and the fact that $\langle G^a\rangle_H$ and $\langle G^b\rangle_H$ are edge-disjoint for $a\neq b$ tell us that π cannot send distinct edges of $\bigcup_{a\in A}\langle\bar{G}^a\rangle_H$ to the same edge of G. Therefore, $\pi(\bar{C})$ is a four-cycle in G that does not lie in a single $\langle G^a\rangle_H$. However, we have already established that such a four-cycle cannot exist, and hence we have obtained a contradiction.

We can now apply Lemma 3.2.11 to the chains $\bar{\mathscr{H}}^a$, $a\in A$, to deduce

$$M_H(n)\geq M_H\left(\left|\bigcup_{a\in A}V(\bar{G}^a)\right|+|A|\cdot v(H)\right)\geq \ell\cdot|A|.$$

for sufficiently large n. The definitions of ℓ, A, and p together with Observation 3.5.7 lead to

$$M_H(n)\geq\left(\left\lceil\frac{p}{2dt}\right\rceil-2\right)\cdot s'_K(\mathbb{Z}_p)\geq c_{H,K}\cdot n\cdot s_K(n)$$

for a suitable $c_{H,K}>0$. $\square$

To make the estimate given in Theorem 3.5.8 useful we need lower bounds on $s_K(n)$.

Lemma 3.5.9. *Let $r \geq 1$ and $d \geq 2$ be fixed. There exists a function $f : \mathbb{N} \to \mathbb{R}_{\geq 0}$ satisfying $\lim_{m\to\infty} f(m) = 0$ such that if K is an arithmetic progression of length r with base point $m \in \mathbb{N}$ and common difference d, there exists a K-fold Sidon set $A \subset [n]$ with*

$$|A| \geq n^{\frac{1}{2}-f(m)-o(1)}. \tag{3.27}$$

The asymptotic variable of the $o(1)$ term is n.

We defer the proof of Lemma 3.5.9, which is independent of the application to bootstrap percolation, to Section 3.5.5.

A combination of Theorem 3.5.8 and Lemma 3.5.9 together with a standard asymptotic calculation gives the result below, where the provided lower bound does not depend on any choice of an arithmetic progression K.

Corollary 3.5.10. *Let H be a bipartite graph such that for some $e, f \in E(H)$ arbitrarily long proper (e, f)-simple H-chains exist, and assume that for each $e' \in E(H)$, every non-m.c. edge-colouring of $H - e'$ contains a non-m.c. copy of C_4. Then*

$$M_H(n) \geq n^{\frac{3}{2}-o(1)}.$$

Proof. Theorem 3.5.8 and Lemma 3.5.9 show that there exist functions $f : \mathbb{N} \to \mathbb{R}_{\geq 0}$ and $\varepsilon : \mathbb{N}\times\mathbb{N} \to \mathbb{R}_{\geq 0}$ with $\lim_{m\to\infty} f(m) = 0$ and $\lim_{n\to\infty} \varepsilon(m,n) = 0$ for each $m \in \mathbb{N}$ such that

$$M_H(n) \geq n^{3/2-f(m)-\varepsilon(m,n)}$$

for every $m \in \mathbb{N}$. Consider the sequence $(n'_m)_{m\in\mathbb{N}}$ defined by

$$n'_1 := 1 \qquad , \qquad n'_m := \max\left(n'_{m-1}+1, \min\left\{n' : \varepsilon(m,n) < \frac{1}{m} \text{ for all } n \geq n'\right\}\right).$$

Now define $g(n) := f(m) + \varepsilon(m,n)$ for the unique m with $n'_m \leq n < n'_{m+1}$. To complete the proof, note that both $\lim_{n\to\infty} g(n) = 0$ and $M_H(n) \geq n^{3/2-g(n)}$. □

We are finally able to prove Theorem 1.3.6. To do so it suffices to show that for $3 \leq r \leq s$, $K_{r,s}$ satisfies the conditions stated in Corollary 3.5.10.

Proof of Theorem 1.3.6. Removing one vertex from each partite set of $K_{r,s}$ yields a copy of $K_{r-1,s-1}$, which is 2-connected since $r,s \geq 3$. By Proposition 3.2.10 (ii), any (e, f)-simple $K_{s,t}$-chain is proper for arbitrary non-incident $e, f \in E(H)$.

Take a copy of $K_{r,s}$ and label its vertices by $v_1,\ldots,v_r,w_1,\ldots,w_s$. Since $K_{r,s}$ is edge-transitive it suffices to check the colouring hypothesis of Theorem 3.5.8 for $e = v_rw_s$. Let χ be a non-m.c. edge-colouring of $K_{r,s} - e$. Let $e', e'' \in E(K_{r,s})$ with $\chi(e') \neq \chi(e'')$. Pick a copy K of $K_{3,3}^-$ in $K_{r,s}$ such that $e', e'' \in E(K)$. Any two four-cycles in $K_{3,3}$ intersect in an edge, and any edge in

$K_{3,3}^-$ lies in a four-cycle. Let $C', C'' \subset K$ be four-cycles containing e' and e'', respectively. Since $\chi(e') \neq \chi(e'')$ and C shares an edge with C' it is not possible that both cycles are monochromatic. We can now apply Theorem 3.5.8 to get the desired bound $M_{K_{r,s}}(n) \geq n^{\frac{3}{2}-o(1)}$. ■

The exponent $3/2 - o(1)$ in Corollary 3.5.10 is a consequence of the lower bound on $s_K(n)$ provided in Lemma 3.5.9. Determining the maximum size of a K-fold (or k-fold) Sidon set is still an open problem. It has been conjectured in [73] that for any k and every $n \in \mathbb{N}$ there exists a k-fold Sidon set of size at least $c_k\sqrt{n}$, where $c_k > 0$ is a constant that only depends on k. An affirmative answer to this conjecture would immediately improve the lower bound of Theorem 3.5.8 to $\Theta(n^{3/2})$ since any k-fold Sidon set is also a K-fold Sidon set if $k \geq \max K$. As a final remark in this section we point out that Theorem 3.5.8 also applies to Q_3. Thus an improvement of the lower bound on the maximum size of a k-fold Sidon set in $[n]$ to $\Theta(\sqrt{n})$ would render the proof in Section 3.6 obsolete.

3.5.5 A lower bound on the maximum size of a K-fold Sidon set

We now prove Lemma 3.5.9 by constructing a K-fold Sidon of the size specified in (3.27). The construction we will use consists of slightly altering and glueing together the proofs of Theorems 2.3, 7.3, and 7.5 of [85]. We will not optimise the precise form of the $o(1)$-term in (3.27) as our aim is a short and clear exposition. The glueing part of the construction is described by the following two lemmas which formalise the well-known process of taking intersections of random translates of two or more sets.

Lemma 3.5.11. *Let $\mathscr{P}$ be a property of finite subsets of $\mathbb{Z}$ that is preserved under taking subsets and translates, that is, if $A \subset \mathbb{Z}$ has property $\mathscr{P}$, so do each $A' \subset A$ and $A+t$ for $t \in \mathbb{Z}$. Then for every $A \subset [n]$ with property $\mathscr{P}$ and $B \subset [n]$ there exists $B_0 \subseteq B$ of size at least $\frac{|A|}{2n} \cdot |B|$ for which $\mathscr{P}$ holds.*

Proof. Let $B_0 := B \cap (A+t)$ where $t \in [-n+1, n-1]$ is chosen uniformly at random, and apply the first moment method to $|B_0|$. □

We can apply Lemma 3.5.11 several times to get a set satisfying multiple desired properties:

Lemma 3.5.12. *Let $\mathscr{P}_1, \ldots, \mathscr{P}_r$ be properties preserved under taking subsets and translates. If $B \subset [n]$ and $A_1, \ldots, A_r \subset [n]$ such that for $i \in [r]$, A_i has property $\mathscr{P}_i$ then B contains a subset of size at least $\frac{|A_1| \cdot \ldots \cdot |A_r|}{(2n)^r} \cdot |B|$ for which $\mathscr{P}_1, \ldots, \mathscr{P}_r$ hold.*

The final preliminary result we need is another variant of Behrend's lower bound on the maximum size of a 3-AP-free set. Again, the result is well-known and its proof is included for completeness.

Corollary 3.5.13. *Let $k \geq 2$. There exists a constant $c > 0$ such that for every $N \in \mathbb{N}$ one can find a set $S \subset [N]$ of size at least $N^{1-c/\sqrt{\log N}}$ such that none of the congruences*

$$\alpha x_1 + \beta x_2 + \gamma x_3 \equiv \delta x_4 \mod N \qquad \textit{with} \qquad \alpha, \beta, \gamma, \delta \in [0,k], \alpha + \beta + \gamma = \delta \tag{3.28}$$

admit a non-trivial solution in S.

Proof. Apply Theorem 3.5.1 with k as given and $n = \lfloor N/(k+1) \rfloor$. Any congruence of the form (3.28) becomes an equality since both sides are smaller than N. The claim now follows from the observation that (3.28) is a special case of (3.20) with

$$x_1 = \ldots = x_\alpha, \qquad x_{\alpha+1} = \ldots = x_{\alpha+\beta},$$
$$x_{\alpha+\beta+1} = \ldots = x_{\alpha+\beta+\gamma}, \qquad x_{\alpha+\beta+\gamma+1} = \ldots = x_{k+1}.$$

□

With the tools above at hand we proceed with the actual construction. Let K be an arithmetic progression of length r with base point $m \in \mathbb{N}$ and common difference d. Our goal is to find a function $f : \mathbb{N} \to \mathbb{R}_{\geq 0}$ with $\lim_{m\to\infty} f(m) = 0$ and build a set $A \subseteq [n]$ with $|A| \geq n^{\frac{1}{2}-f(m)-o(1)}$ that does not admit non-trivial solutions to any of the equations

$$k_1x_1 + k_2x_2 + k_3x_3 + k_4x_4 = 0 \tag{3.29}$$

with $k_1, \ldots, k_4 \in K \cup (-K)$ and $k_1 + \ldots + k_4 = 0$. It suffices to restrict our attention to the case when at most two of $k_1, \ldots, k_4$ are negative as we can simply multiply all four of them by -1. Moreover we will assume that $k_1 \geq \ldots \geq k_4$. In the following we suppose that $m > 2r \cdot d$ since for smaller m we can take A to be a singleton and artificially set $f(m) = 1/2$. This assumption is necessary to guarantee that some of the sets that occur below are non-empty. We will introduce the precise form of f at a later stage when it is actually needed. We have to distinguish between the following special cases of (3.29).

(3.29.1) The only negative coefficient is k_4.

(3.29.2) Both k_3 and k_4 are negative and $(k_1, k_2) = (-k_4, -k_3)$.

(3.29.3) Both k_3 and k_4 are negative and $(k_1, k_2) \neq (-k_4, -k_3)$.

To obtain A we will take sets A_1, A_2, and $A_3(k_1, k_2, k_3, k_4)$ such that A_1 simultaneously forbids all instances of (3.29.1), A_2 is free of non-trivial solutions to all equations of the form (3.29.2), and $A_3(k_1, k_2, k_3, k_4)$ avoids (3.29.3) for the specified choice of $k_1, \ldots, k_4$. Finally those sets will be combined using Lemma 3.5.12. Let A_1 be given by Corollary 3.5.13 for $k = \max K$ and $N = n$, and let $c_1 > 0$ such that $|A_1| \geq n^{1-c_1/\sqrt{\log n}}$, and note that each subset as well as each translate of A_1 is free of non-trivial solutions to (3.29.1).

To construct A_2 we choose a prime p between $\sqrt{n}/2$ and $\sqrt{n}$ and let $S \subseteq [p]$ be as given by Corollary 3.5.13 with $k = \max K$ and $N = p$. Define

$$A_2 := \{s + (4k+1)p \cdot (s^2 \bmod p) : s \in S\}.$$

Suppose there exist $k_1, \ldots, k_4 \in K \cup (-K)$ with $(k_1, k_2) = (-k_4, -k_3)$ and $a_1, \ldots, a_4 \in A_2$ satisfying $k_1a_1 + \ldots + k_4a_4 = 0$. Let $s_1, \ldots, s_4 \in S$ such that $a_i = s_i + (4k+1)p \cdot (s^2 \bmod p)$ for

$1 \leq i \leq 4$. By construction of A_2 we obtain

$$k_1 s_1 + \ldots + k_4 s_4 = 0 \qquad \text{and} \qquad k_1 s_1^2 + \ldots + k_4 s_4^2 \equiv 0 \mod p. \tag{3.30}$$

We can rewrite (3.30) as

$$k_1(s_1 - s_4) = k_2(s_3 - s_2) \qquad \text{and} \qquad k_1(s_1^2 - s_4^2) \equiv k_2(s_3^2 - s_2^2) \mod p, \tag{3.31}$$

where $k_1 \geq k_2 > 0$. If $s_1 - s_4 = 0$ or $s_3 - s_2 = 0$ then $(s_1, s_2) = (s_4, s_3)$ so we have a trivial solution. If not we can divide the second part of (3.31) by the first part to obtain $s_1 + s_4 \equiv s_2 + s_3 \mod p$, which together with the first part of (3.31) yields

$$2k_1 s_1 \equiv (k_1 + k_2)s_3 + (k_1 - k_2)s_4 \mod p.$$

Note that (3.28) in Corollary 3.5.13 also covers equations in three variables (e.g. by considering $\gamma = 0$). Consequently, $s_1 = s_3$ if $k_1 = k_2$, and $s_1 = s_3 = s_4$ if $k_1 \neq k_2$. In both cases $(s_1, \ldots, s_4)$ is a trivial solution to (3.29.2). Thus, we have established that A_2 does not admit non-trivial solutions to (3.29.2). Moreover there exists a constant $c_2 > 0$ such that $|A_2| \geq n^{1/2 - c_2/\sqrt{\log n}}$.

Now consider (3.29.3). Given $k_1, \ldots, k_4$ with $k_1 \geq k_2 \geq 0 \geq k_3 \geq k_4$ and $(k_1, k_2) \neq (-k_4, -k_3)$, let $\tilde{m} := \min\{k_2, -k_3\} + 1$ and let $B \subset [\tilde{m}/(2r \cdot d + 1)]$ be given by Corollary 3.5.13 for $N = \lfloor \tilde{m}/(2r \cdot d + 1) \rfloor$ and $k = \max K - m = r \cdot d$. Define

$$A_3 = A_3(k_1, k_2, k_3, k_4) := \left\{ \sum_{j=0}^{\lfloor \log_{\tilde{m}} n \rfloor - 1} b_j \cdot \tilde{m}^j \;:\; b_j \in B \text{ for } 0 \leq j \leq \lfloor \log_{\tilde{m}} n \rfloor - 1 \right\}.$$

Suppose that A_3 contains $a_1, \ldots, a_4$ with $k_1 a_1 + \ldots + k_4 a_4 = 0$, and write

$$a_i = \sum_{j=0}^{\lfloor \log_{\tilde{m}} n \rfloor - 1} b_{ij} \cdot \tilde{m}^j$$

for $1 \leq i \leq 4$. Suppose for a contradiction that the solution $(a_1, \ldots, a_4)$ is not trivial. Pick the smallest j such that the digits b_{1j}, b_{2j}, b_{3j}, and b_{4j} are not all the same. Those digits satisfy

$$k_1 b_{1j} + k_2 b_{2j} \equiv (-k_3) b_{3j} + (-k_4) b_{4j} \mod \tilde{m}. \tag{3.32}$$

With $d_i := k_i - \tilde{m}$ for $i \in \{1, 2\}$ and $d_i := (-k_i) - \tilde{m}$ for $i \in \{3, 4\}$ the congruence (3.32) becomes

$$d_1 b_{1j} + d_2 b_{2j} \equiv d_3 b_{3j} + d_4 b_{4j} \mod \tilde{m}$$

The coefficients $d_1, \ldots, d_4$ satisfy $d_1 + d_2 = d_3 + d_4$ and precisely one of d_2 and d_3 is equal to -1. Here we used that $(k_1, k_2) \neq (-k_4, -k_3)$, $k_1 \geq \ldots \geq k_4$, and $k_1 + \ldots + k_4 = 0$, so $k_2 \neq -k_3$. Furthermore,

$$d_1, \ldots, d_4 \in \{-1, d - 1, 2d - 1, \ldots, (r-1)d - 1\},$$

which together with the estimate $\max B \le \tilde{m}/(2r\cdot d+1)$ implies

$$d_1b_{1j}+d_2b_{2j}=d_3b_{3j}+d_4b_{4j}.$$

Each of $d_1,\dots,d_4$ is smaller than $\max K-m$, so we have arrived at a contradiction to our choice of B. Therefore $(a_1,\dots,a_4)$ must be a trivial solution.

Corollary 3.5.13 provides a constant $c_3>0$ that does not depend on n such that

$$\begin{aligned}|A_3|=|B|^{\lfloor \log_m n\rfloor} &\ge \left(\frac{m}{2r\cdot d+1}\right)^{\left(1-c_3/\sqrt{\log(m/(2r\cdot d+1))}\right)\cdot\lfloor \log_m n\rfloor}\\ &= n^{\left(1-c_3/\sqrt{\log(m/(2r\cdot d+1))}\right)\cdot\left(1-\frac{\log(2r\cdot d+1)}{\log m}\right)\cdot\frac{\lfloor \log_m n\rfloor}{\log_m n}}\\ &= n^{1-f(m)/r^4-o(1)}\end{aligned}$$

when m is sufficiently large in terms of r and d. Here we define $f(m)$ such that

$$1-f(m)/r^4=\left(1-c_3/\sqrt{\log(m/(2r\cdot d+1))}\right)\cdot\left(1-\frac{\log(2r\cdot d+1)}{\log m}\right)$$

Note that $f(m)$ tends to 0 as $m\to\infty$.

We now apply Lemma 3.5.12 to A_1,A_2, and $A_3(k_1,\dots,k_4)$ for all $k_1,\dots,k_4$ that belong to (3.29.3) to obtain the desired set A that is free of non-trivial solutions to (3.29). There are at most r^4 choices of $k_1,\dots,k_4\in K\cup(-K)$ such that $k_1\ge k_2\ge 0\ge k_3\ge k_4$. Hence,

$$\begin{aligned}|A| &\ge \frac{1}{(2n)^{r^4+1}}\cdot n^{1-c_1/\sqrt{\log n}}\cdot n^{1/2-c_2/\sqrt{\log n}}\cdot\left(n^{1-f(m)/r^4-o(1)}\right)^{r^4}\\ &= n^{1/2-(c_1+c_2)/\sqrt{\log n}-f(m)-o(1)}.\end{aligned}$$

This completes the proof of Lemma 3.5.9.

3.6 The three-dimensional cube

The upper bound of Theorem 1.3.7 is a direct consequence of Theorem 1.3.1 and the well-known bound

$$\mathrm{ex}(n,Q_3)=O(n^{8/5})$$

introduced by Erdős and Simonovits [38] in 1970.

It remains to establish the lower bound $M_{Q_3}(n)=\Omega(n^{3/2})$. To do so we will once again employ Lemma 3.2.11. Since Q_3 cannot be disconnected by removing an edge and one vertex from each partite set, Proposition 3.2.10 implies that for any non-incident $e,f\in E(Q_3)$ there exist arbitrarily long proper, (e,f)-simple Q_3-chains. In our construction large C_4-free bipartite graphs play a crucial role. One way to obtain such graphs is via maximum Sidon sets in the integers, which, recall, are sets where all the sums $a+b$ are distinct up to the obvious permutation of the summands. Erdős and Turán [40] showed that there exist Sidon sets $A\subset[n]$ of size $|A|=$

$(\frac{1}{\sqrt{2}} - o(1)) \cdot \sqrt{n}$ and that the maximum size of such a set is less than $(1+\varepsilon)\sqrt{n}$ for any $\varepsilon > 0$ and sufficiently large n. In the following we assume that n is suffciently large so we do not run into degenerate cases. Choose a set $A \subset [n]$ such that

(1) $|A| \geq \left(\frac{1}{2\sqrt{30}} - o(1)\right) \cdot \sqrt{n}$,

(2) $\max A \leq n/12$,

(3) $|a_1 + a_2 - a_3 - a_4| \geq 5$ whenever $a_1, a_2, a_3, a_4 \in A$ satisfy $\{a_1, a_2\} \neq \{a_3, a_4\}$.

Such a set can be obtained by taking a maximum Sidon set $B \subset [n/60]$ and defining $A := 5 \cdot B$. Note that choosing $a_2 = a_3$ in (3) gives $|a_1 - a_4| \geq 5$ for $a_1 \neq a_4$.

We build our simple chains in the spirit of Construction 3.2.4. Let s be the largest positive integer with $n \geq 6s + 2 + 128|A|$, that is, $s := \lfloor (n - 2 - 128|A|)/6 \rfloor$. Let $\ell := s - \max A$, and for $a \in A$, define $(H_i^a, e_i^a)_{i \in [\ell]}$ as follows: Take two disjoint $(3s+1)$-element sets

$$W = \{u_1, \ldots, u_s, v_0, \ldots, v_s, w_1, \ldots, w_s\} \qquad W' = \{u'_1, \ldots, u'_s, v'_0, \ldots, v'_s, w'_1, \ldots, w'_s\}$$

and for $1 \leq i \leq s$, let

$$W_i := \{v_{i-1}, u_i, w_i, v_i\} \qquad \text{and} \qquad W'_i := \left\{v'_{i-1}, u'_i, w'_i, v'_i\right\}.$$

For $i \in [\ell]$, define H_i^a via $V(H_i^a) = W_i \cup W'_{i+a}$ and

$$\begin{aligned} E(H_i^a) = \{v_{i-1}u_i, v_{i-1}w_i, u_iv_i, w_iv_i\} &\cup \{v'_{i-1}u'_i, v'_{i-1}w'_i, u'_iv'_i, w'_iv'_i\} \\ &\cup \{v_{i-1}v'_{i-1+a}, u_iu'_{i+a}, w_iw'_{i+a}, v_iv'_{i+a}\} \end{aligned} \tag{3.33}$$

Each H_i^a is a copy of Q_3 and $e_i^a := v_iv'_{i+a}$ is the unique edge in $E(H_i^a) \cap E(H_{i+1}^a)$. Denote the underlying graph of $(H_i^a, e_i^a)_{i \in [\ell]}$ by G^a and write $G := \bigcup_{a \in A} G^a$. Observe that G is bipartite with partite sets $\{u_1, \ldots, u_s, w_1, \ldots, w_s, v'_0, \ldots, v'_s\}$ and $\{u'_1, \ldots, u'_s, w'_1, \ldots, w'_s, v_0, \ldots, v_s\}$.

By construction of G we have $N_G(v_i) \cap W' \subseteq \{v'_{i+a} : a \in A\}$ for all $i \in [\ell]$. Therefore, (3) implies that any two G-neighbours of v_i in W' have distance at least 5 in $G[W']$. Analogous statements hold for u_i and w_i. We will refer to this property as the *distance condition*.

Let us verify that the chains constructed above indeed satisfy the conditions of Lemma 3.2.11 for bipartite-inseparable graphs. As G is bipartite it does not contain any odd cycles. By (3.33) the sets $E_G(W_i, W'_{i+a})$ and $E_G(W_j, W'_{j+a'})$ are disjoint for distinct $a, a' \in A$ and arbitrary $i, j \in [\ell]$. Thus, each of the edges e_i^a lies in precisely one chain. It remains to show that every copy of Q_3^- in G is contained in exactly one of the chains. This follows from the third part of the claim below.

Claim 3.6.1. *1. Any 4-cycle in G has either all its vertices in W or all its vertices in W' or two vertices in W and two in W'.*

2. Any 4-cycle in G with two vertices in W and two vertices in W' has at least one edge in $G[W]$ and at least one in $G[W']$.

3. If Q is a copy of Q_3^- in G, then there exists $1 \leq i \leq j \leq s$ such that $V(Q) = W_i \cup W'_j$.

Proof. 1. Suppose there is a 4-cycle with one vertex in W and three vertices in W'. Then the three vertices in W' form a path of length two in W' whose endpoints have a common neighbour in W, which contradicts the distance condition. Similarly, there cannot be a 4-cycle that has three vertices in W and one in W'.

2. Suppose that $(\{x_1,x_2,x_3,x_4\},\{x_1x_2,x_2x_3,x_3x_4,x_4x_1\}) \subset G$ is a 4-cycle all of whose edges lie in $E_G(W,W')$. Without loss of generality, $x_1 \in W$. Then there exist $i_1, j_2, i_3, j_4 \in [s]$ such that $x_1 \in W_{i_1}$, $x_2 \in W'_{j_2}$, $x_3 \in W_{i_3}$, $x_4 \in W'_{j_4}$ and thus

$$j_2 - i_1, j_4 - i_1, j_2 - i_3, j_4 - i_3 \in A + \{-1,0,1\} \quad \text{and} \quad (j_2-i_1)+(j_4-i_3) = (j_4-i_1)+(j_2-i_3).$$

This, however, contradicts property (3) of A.

3. The copy Q contains two vertex-disjoint 4-cycles R_1, R_2 that cover $V(Q)$. If one of them is fully contained in $G[W]$ or $G[W']$, respectively, then the other one lies in $G[W']$ or $G[W]$, respectively, by the distance condition and the claim follows. Therefore we may assume that $|V(R_1)\cap W| = |V(R_1)\cap W'| = |V(R_2)\cap W| = |V(R_2)\cap W'| = 2$, so there exist i_1, j_1, i_2, j_2 with

$$V(R_1)\cap W \subset W_{i_1}\ ,\ V(R_1)\cap W' \subset W'_{j_1}\ ,\ V(R_2)\cap W \subset W_{i_2}\ ,\ V(R_2)\cap W' \subset W'_{j_2}$$

This implies

$$j_2 - i_1, j_1 - i_1, j_2 - i_2, j_1 - i_2 \in A + \{-1,0,1\} \qquad , \qquad (j_2-i_1)+(j_1-i_2) = (j_2-i_2)+(j_1-i_1),$$

which contradicts the property (3) of A unless $i_1 = i_2$ and $j_1 = j_2$. □

Let Q be a copy of Q_3^- in G. By Claim 3.6.1 we can find $i, j \in [s]$ such that $V(Q) = W_i \cup W'_j$. Because there are at least three edges between W_i and W'_j we obtain that $j - i \in A$ and $Q \subset H_i^{j-i}$, so Q is fully contained in precisely one chain. The conclusion of Lemma 3.2.11 now tells us that

$$M_{Q_3}(|V(G)| + 128|A|) \geq |A| \cdot (s - \max A).$$

From the definition of s and properties (1) and (2) of A we deduce

$$M_{Q_3}(n) \geq \left(\frac{1}{24\sqrt{30}} - o(1)\right) \cdot n^{3/2}.$$

3.7 Wheel graphs

This section is devoted to the proof of Theorem 1.3.10.

Let $k \geq 7$ be an odd integer. Recall that we refer to the unique universal vertex of W_k as the *hub* and call the cycle that remains after removing the hub the *outer cycle*. The proof is split into two parts. First, we will prove the lower bound $M_H(n) \geq n^{2-o(1)}$. Second, we will show that $M_{W_k}(n) = o(n^2)$.

3.7.1 Lower bound

The lower bound is a consequence of Theorem 3.5.3. To apply that theorem we have to verify that W_k allows proper, simple chains of arbitrary length, and for each $e \in E(W_k)$, every non-m.c. colouring of $W_k - e$ contains a cycle with at least two and at most three colours in which the colour classes are edge sets of paths.

Let v denote the hub of W_k. For each $e \in E(W_k)$, $\Delta(W_k - e) \geq k-1 \geq 6$ whereas the vertices on the outer cycle have degree three in W_k. Therefore every embedding $\varphi : W_k - e \to W_k$ fixes the hub, and hence φ restricted to $(W_k - e) - v$ is an embedding of P_k or C_k into C_k. In both cases $\varphi(e)$ must be an edge of W_k. We have thus shown that W_k is self-stable.

Given $\ell \in \mathbb{N}$ let us construct a proper simple W_k-chain of length ℓ. Fix two non-incident edges e and f on the outer cycle of W_k. Finding an (e,f)-simple W_k-chain $(H_i,e_i)_{i\in[\ell]}$ of length ℓ is immediate by Observation 3.2.8. Let $G := \bigcup_{i=1}^{\ell} H_i$ and denote the hub of H_i by v_i. We need to make sure that $(H_i,e_i)_{i\in[\ell]}$ is proper, so let us check that Condition (2) of Definition 3.2.1 holds. Let $e \in E(W_k)$, and let $\varphi : W_k - e \to G$ be an embedding. Recall that v denotes the hub of W_k, and $d(v) = k \geq 7$. We have $d_G(w) \leq 5$ for all $w \in V(G) \setminus \{v_1,\dots,v_\ell\}$. Therefore we can find $j \in [\ell]$ such that $\varphi(v) = v_j$. If e lies on the outer cycle of W_k we deduce

$$\varphi\left(V(W_k) \setminus \{v\}\right) \subseteq N_G(\varphi(v)) \subseteq V(H_j).$$

Since all edges in $G[V(H_j)]$ are contributed by H_j we arrive at

$$\varphi(W_k - e) \subset H_j.$$

Now assume that e is incident to v, and let $w \in V(W_k)$ such that $e = vw$. Then $\varphi(V(W_k) \setminus \{v,w\}) \subseteq N_G(v_j)$ and $d_{W_k - v}(w) = 2$. Moreover $\varphi(W_k) - v_j$ must be a k-cycle. The only vertices in G that have two neighbours in $V(H_j)$ but do not lie in H_j themselves are v_{j-1} and v_{j+1} (if $j \in \{1,\ell\}$ only one of them exists). However, for each $w' \in V(H_j - v_j)$ both $G[\{v_{j-1}\} \cup V(H_j - v_j) \setminus \{w'\}]$ and $G[\{v_{j+1}\} \cup V(H_j - v_j) \setminus \{w'\}]$ are paths of length $k-1$ and thus cannot contain a k-cycle. Therefore $\varphi(w)$ must lie in $V(H_j)$, which implies

$$\varphi(W_k - e) = H_j - v_j\varphi(w).$$

We have shown that Condition (2) of Definition 3.2.1 holds for the above defined W_k-chain. This together with the fact that W_k is W_k-stable shows that the chain is proper.

Let us verify the colouring condition of Theorem 3.5.3. There are two types of edges in W_k. Those incident to the hub and those lying on the outer cycle. If $e \in E(W_k)$ lies on the outer cycle, every edge of $W_k - e$ is contained in a triangle and the triangle graph $\mathscr{T}(W_k - e)$ is connected. In that case, by Lemma 3.5.4, every non-m.c. colouring of $W_k - e$ admits a non-m.c. triangle. Now suppose that e is incident to the hub. Let w be the endpoint of e other than the hub, or equivalently, the unique degree-two vertex of $W_k - e$. Let χ be an arbitrary non-m.c. colouring of $E(W_k)$. Note that $W_k - w$ is isomorphic to $W_{k-1} - f$ where f is an arbitrary edge on the outer cycle of W_{k-1}. Hence if $W_k - w$ is non-m.c. under χ we can proceed as above to find a non-m.c. triangle. If $W_k - w$ is monochromatic under χ then the 4-cycle C formed by w, the hub of W_k, and their two common neighbours has at least two (since χ is non-m.c.) and at most three colours (because the two edges of C in $W_k - w$ have the same colour). Moreover the colour classes form paths (again because the edges of C in $W_k - w$ share a colour). Now Theorem 3.5.3 implies

$$M_{C_k}(W_k) \geq n^{2-o(1)}.$$

3.7.2 Upper bound

Let $\varepsilon > 0$ be an arbitrary constant. We have to show that $M_{W_k}(n) < \varepsilon n^2$ for sufficiently large n. Our proof relies on the following consequence of Szemerédi's Regularity Lemma. Recall that an *induced matching* in a graph G' is a matching M such that M is the edge set of an induced subgraph of G'.

Theorem 3.7.1 ([71], Theorem 3.2)**.** *If a graph G' is the union of n induced matchings, then $e(G') = o(n^2)$.*

As pointed out in [71] the theorem above is equivalent to the $(6,3)$-theorem of Ruzsa and Szemerédi in [87] (see also [70] for a survey on the Regularity Lemma and its applications).

Choose a graph G on n vertices such that $\tau := \tau_{W_k}(G) = M_{W_k}(n)$, and let $(G_t)_{t\geq 0}$ be the W_k-process on G. For our arguments we require a certain type of H-chain as specified by the claim below.

Claim 3.7.2. *For any $\tau' \in [\tau_H(G)]$ and $e_{\tau'} \in E(G_{\tau'}) \setminus E(G_{\tau'-1})$ there exists an H-chain $(H_t, e_t)_{t\in[\tau']}$ such that $e_t \in E(G_t) \setminus E(G_{t-1})$ and $H_t - e_t \subseteq G_{t-1}$ for all $t \in [\tau']$.*

Proof. We induct on τ'. As to the case $\tau' = 1$ if suffices to pick any copy of H completed at time 1. Suppose that $\tau' > 1$. Pick a copy $H_{\tau'}$ of H completed by $e_{\tau'}$ at time τ', so $H_{\tau'} - e_{\tau'} \subseteq G_{\tau'-1}$ and $e_{\tau'} \notin E(G_{\tau'-1})$. One of the edges of $H_{\tau'} - e_{\tau'}$ must have been added at time $\tau' - 1$ for otherwise $H_{\tau'-1}$ would have been completed by time $\tau' - 1$. Let $e_{\tau'-1}$ be such an edge and let $H_{\tau'-1}$ be a copy of H completed by $e_{\tau'-1}$. Now apply the induction hypothesis to $e_{\tau'-1}$ to obtain an H-chain $\mathscr{H} = (H_t, e_t)_{t\in[\tau'-1]}$ with $e_t \in E(G_t) \setminus E(G_{t-1})$ and $H_t - e_t \subseteq G_{t-1}$ for $t \in [\tau' - 1]$. We want to extend $\mathscr{H}$ by $(H_{\tau'}, e_{\tau'})$ to obtain the desired chain. We have that $(H_t, e_t)_{t\in[\tau']}$ satisfies condition (2) of Definition 3.2.1 by the induction hypothesis and the choice of $H_{\tau'-1}$ and $e_{\tau'-1}$. Conditions (1) and (3) clearly hold, too. It remains to show that (4) holds, which due to the induction hypothesis and the fact that W_k is self-stable reduces to checking that $e_{\tau'} \notin E(H_j)$ for $j < t'$. The copies $H_1, \ldots, H_{\tau'-1}$ are all completed before time τ', and hence $e_{\tau'}$ cannot lie in any of them. □

Pick a W_k-chain $(H_t, e_t)_{t\in[\tau]}$ as given by Claim 3.7.2 for $\tau' = \tau$ and an arbitrary edge e_τ added at the final step of the process.

Suppose that $\tau > \frac{\varepsilon}{2}n^2$ (otherwise we are already done). We will proceed as follows: First we prove that there exist $1 \le t_0 \le t_1 \le \varepsilon n^2/2$ such that $V(H_{t_0})$ is a clique in G_{t_1}. Second, we show that once such a clique occurs the process will stabilise within at most $(n-k+1)$ more steps.

Without loss of generality we assume that $\varepsilon n^2/4$ is an integer. Denote the hub of H_t by v_t, and recall that $H_t - v_t$ is a k-cycle. For $v \in V(G)$, let

$$\tilde{N}(v) := \bigcup_{t\in[\varepsilon n^2/4]\,:\,v_t=v} V(H_t - v_t) \qquad , \qquad \tilde{G}(v) := G_{\varepsilon n^2/4}[\tilde{N}(v)].$$

If there exists $t_0 \le \varepsilon n^2/4$ such that $\tilde{G}(v_{t_0})$ is not a disjoint union of k-cycles, we can find a connected subgraph $\Gamma \subseteq \tilde{G}(v_{t_0})$ of order at most $k+1$ that contains a k-cycle. Any path of length $k-1$ in $\tilde{G}(v_{t_0})$ together with v_{t_0} forms a copy of W_k minus an edge. Therefore $\langle\Gamma\rangle_{C_k} \subseteq G_{t_1}$ where $t_1 := t_0 + M_{C_k}(v(\Gamma))$. We note that $t_1 \le \varepsilon n^2/2$ for large n since $M_{C_k}(v(\Gamma))$ does not depend on n. Recall from our results on cycles (Lemma 2.5.3) that $\langle\Gamma\rangle_{C_k}$ must be a complete graph, so we have found the desired t_0, t_1.

Claim 3.7.3. *There exists $t_0 \in [\varepsilon n^2/4]$ such that $\tilde{G}(v_{t_0})$ is not a disjoint union of k-cycles*

Proof. Suppose for a contradiction that for any $t \in [\varepsilon n^2/4]$, $\tilde{G}(v_t)$ is a disjoint union of k-cycles. In that case, $v_t \ne v_{t+1}$ for all $t \in [\frac{\varepsilon n^2}{4} - 1]$ because the k-cycles $H_t - v_t$ and $H_{t+1} - v_{t+1}$ intersect in at least one endpoint of e_t. We consider

$$\tilde{G} := \bigcup_{t\in[\varepsilon n^2/4]} \tilde{G}(v_t)$$

and estimate its number of edges. For $t \in [\frac{\varepsilon n^2}{4} - 1]$, we have $e_t \in E(\tilde{G}(v_t)) \cup E(\tilde{G}(v_{t+1}))$ unless $e_t = v_t v_{t+1}$. Let $T := \{t \in [\frac{\varepsilon n^2}{4} - 1] : v_t v_{t+1} = e_t\}$, and choose an H_{t+1}-neighbour x_t of v_t other than v_{t+1} for each $t \in T$. Observe that

$$E(\tilde{G}) \supseteq \left\{ e_t : t \in \left[\frac{\varepsilon n^2}{4} - 1\right] \setminus T \right\} \cup \{v_t x_t : t \in T\} \tag{3.34}$$

The edges $v_t x_t$, $t \in T$, are pairwise distinct. Indeed, if $v_{t'} x_{t'} = v_t x_t$ for some $t, t' \in T$ with $t' < t$, we have $v_{t+1} \in V(H_{t'})$ because $x_t v_{t+1}, v_t v_{t+1} \in E(G_{\varepsilon n^2/4})$ and $\tilde{G}(v_{t'})$ is a disjoint union of k-cycles. However, $v_t v_{t+1} \notin E(H_{t'})$ so $H_{t'} - v_{t'}$ cannot be an induced k-cycle in $G_{\varepsilon n^2/4}$, which contradicts our assumption on $\tilde{G}(v_{t'})$.

Now (3.34) implies

$$e(\tilde{G}) \ge \max\left\{ \frac{\varepsilon n^2}{4} - 1 - |T|, |T| \right\} \ge \frac{\varepsilon n^2}{8} - 1.$$

On the other hand, for every $v \in V(G)$, $\tilde{G}(v)$ can be written as the union of three induced matchings of $G_{\varepsilon n^2/4}$. Therefore $\tilde{G}$ is the union of $3n$ induced matchings. This contradicts Theorem 3.7.1 when n is sufficiently large. $\square$

Now that we have found the required t_0 and t_1 we show that the process can run for at most $(n-k+1)$ more steps.

Claim 3.7.4. *For every $s \in \mathbb{N}_0$ with $t_1 + 2s \le \tau$ there exists a clique of size $k+s$ at time $t_1 + 2s$ containing $V(H_{t_0})$.*

Proof. We induct on s. The case $s = 0$ is clear by choice of t_0 and t_1. Let $s \ge 1$. By the induction hypothesis we can find a clique W of size $k + (s-1)$ in $G_{t_1+2(s-1)}$ that contains $V(H_{t_0})$. Define

$$t' := \min\{t \in [t_0+1, t_1+2s-1] : V(H_t) \nsubseteq W\}.$$

This is well-defined since W is a clique at time $t_1 + 2(s-1)$ and the edge that completes H_{t_1+2s-1} is not present at time $t_1 + 2(s-1)$. The minimality of t' gives $V(H_{t'-1}) \subseteq W$ whenever $t' > t_0 + 1$. If $t' = t_0 + 1$ we have $V(H_{t'-1}) \subseteq W$ be assumption. In both cases $|V(H_{t'}) \cap W| \ge v(H_{t'-1} \cap H_{t'}) \ge 2$.

There exists $v \in V(H_{t'}) \setminus W$ with two $G_{t'}$-neighbours in W. Indeed, if $v_{t'} \notin W$ we can pick $v = v_{t'}$. If not, the cycle $H_{t'} - v_{t'}$ has an edge uv with $u \in W$, $v \notin W$. Now both u and $v_{t'}$ are $G_{t'}$-neighbours of v in W. Since $\delta(W_k) = 3$ and W is a clique in $G_{t'}$, we can, for any $w \in W \setminus N_{G_{t'}}(v)$ find a copy of W_k in $G_{t'+1}[W \cup \{v\}]$ that is completed by vw. Therefore $W \cup \{v\}$ is a clique of size $k+s$ in $G_{t'+1}$ and the claim follows from the observation that $t'+1 \le t_1 + 2s$. □

Clearly, a clique of size $n+1$ cannot occur in the W_k-process on an n-vertex graph. For this reason, by Claim 3.7.4 we cannot have $t_1 + 2(n-k+1) \le \tau$. This implies

$$M_{W_k}(n) < t_1 + 2(n-k+1) < \varepsilon n^2$$

provided that n is sufficiently large. ■

Remark 3.7.5. The reason why we have to restrict ourselves to odd k is that for even k the final graph of the C_k-process on a connected graph of size $k+1$ with a k-cycle can be a complete bipartite graph instead of a clique, and so the base case of Claim 3.7.4 might not be satisfied. However, we believe that by a careful analysis one can show Theorem 1.3.10 for even $k \ge 8$, too.

3.8 High connectivity guarantees superlinear running time

We turn our attention to the proof of Theorem 1.3.12.

In the previous sections undesired copies of H minus an edge gave rise to non-monochromatic cycles in the underlying graph of a chosen H-chain. By making the underlying graph free of such cycles to begin with we avoided the undesired copies. In this section we are going to use this approach by employing graphs with arbitrarily large girth that have as many edges as possible. Finding the maximum number of edges in a graph of a prescribed girth is a problem with a rich history in extremal combinatorics. An overview of that history can be found in [51]. We use the following type of graph found by Lazebnik-Ustimenko-Woldar [72], and further investigated by Lazebnik-Viglione [74].

Theorem 3.8.1 ([72] Theorem 3.2 and [74] Theorem 2)**.** *Let $k \geq 1$ be an odd integer, and let q be a prime power. There exists a connected, bipartite, q-regular graph of order $2q^{\lceil (3k+2)/4 \rceil}$ and girth at least $k+5$.*

The original theorem in [72] only provides an upper bound of $2q^{\lceil (3k+2)/4 \rceil}$ on the order. The sharpness result comes from [74]. In their original forms both theorems are much more elaborate than the reduced version cited here, which only contains the parts we need for our application, and include a precise construction of the graph that achieves the properties above.

Proof of the first part of Theorem 1.3.12. Let $\mathscr{G} = \mathscr{G}(q)$ be a graph as given by Theorem 3.8.1 when $k = 2v(H) - 3$ and q is an arbitrary prime power, and denote its two partite sets by X, Y. Further, denote the neighbourhood of $y \in Y$ by $\mathscr{N}(y)$. The girth of $\mathscr{G}$ is at least $2v(H)+1$.

Once again our lower bound comes from Lemma 3.2.11.We start by introducing suitable H-chains. Recall that a simple H-chain of length ℓ has $2 + \ell \cdot (v(H) - 2)$ vertices. Let e, f be arbitrary non-incident edges of H. Since $|\mathscr{N}(y)| = q$ for all $y \in Y$, we can place a simple (e, f)-chain $(H_i^y, e_i^y)_{i \in [\ell]}$ of length $\ell := \lfloor (q-2)/(v(H)-2) \rfloor$ on each of the $\mathscr{N}(y)$. By Proposition 3.2.10 those chains are proper. Moreover they are pairwise edge-disjoint because $\mathscr{G}$ is C_4-free and so $|\mathscr{N}(y) \cap \mathscr{N}(y')| \leq 1$ for all $y, y' \in Y$. Denote the underlying graphs of the chains by G^y, $y \in Y$. Fix an arbitrary ordering of Y. In this ordering we clearly have $e_j^y \notin E(G^{y'})$ for any $y > y'$ and $0 \leq j \leq \ell$.

Claim 3.8.2. *Every cycle of length at most $v(H)$ in G lies in $\langle G^y \rangle_H$ for some $y \in Y$.*

Proof. Let $x_1, \dots, x_s \in X$ and $x_1x_2, \dots, x_{s-1}x_s, x_sx_1$ be the vertices and edges of a cycle in G. For $1 \leq j < s$, let $y_j \in Y$ be the unique vertex in Y such that $x_j, x_{j+1} \in \mathscr{N}(y_j)$, and let $y_s \in Y$ with $x_sx_1 \in \mathscr{N}(y_s)$. Suppose that the cycle is not contained in $\langle G^y \rangle_H$ for all $y \in Y$. With that assumption $y_1, \dots, y_s$ are not all the same. Thus there exist $r \in [2, s]$ and $1 \leq j_1 < \dots < j_r \leq s$ such that $\{j_1, \dots, j_r\} = \{j \in [s] : y_j \neq y_{j-1 \bmod s}\}$, and consequently, $x_{j_1}y_{j_1} \dots x_{j_r}y_{j_r}x_{j_1}$ is a circuit of length at most $2s$ in $\mathscr{G}$. As a shortest circuit in a graph is always a cycle, we conclude that $\mathscr{G}$ contains a cycle of length at most $2s$. We have arrived at a contradiction since $2s \leq 2v(H)$ and $\mathscr{G}$ has girth larger than $2v(H)$. □

Recall that if H is bipartite so are the $\langle G^y \rangle_H$. Thus Claim 3.8.2 tells us that in the bipartite setting there are no odd cycles of length at most $v(H)$ in G. It remains to verify that copies of H minus an edge are restricted to individual chains. Let H' be a copy of $H - e$ in G for an arbitrary $e \in E(H)$. Suppose for a contradiction that there exist $y, y' \in Y$, $y \neq y'$, such that $E(H' \cap \langle G^y \rangle_H) \neq \varnothing$ and $E(H' \cap \langle G^{y'} \rangle_H) \neq \varnothing$. Since H' is 2-vertex-connected, any two of its edges lie on a common cycle. We can thus find a cycle $C \subseteq H'$ with $E(C \cap \langle G^y \rangle_H) \neq \varnothing$ and $E(C \cap \langle G^{y'} \rangle_H) \neq \varnothing$. This however contradicts Claim 3.8.2.

Now we can apply Lemma 3.2.11 to obtain

$$M_H\left(\left|\bigcup_{y \in Y} G^y\right| + 2|Y| \cdot v(H)^2\right) \geq |Y| \cdot \ell. \tag{3.35}$$

As $|X| = |Y| = q^{\lceil (6v(H)-7)/4 \rceil}$ and $\ell = \Theta(q)$, (3.35) turns into

$$M_H\left((4v(H)^2+1)\cdot q^{\lceil (6v(H)-7)/4 \rceil}\right) \;=\; \Omega\left(q^{1+\left\lceil \frac{6v(H)-7}{4} \right\rceil}\right).$$

To complete the proof observe that for every $n \in \mathbb{N}$ there exists a prime power q such that n is at most a fixed constant factor away from $q^{\lceil (6v(H)-7)/4 \rceil}$. ■

We now move to the second part of Theorem 1.3.12. At its heart lies the following observation that was also used in [25] to prove $M_{K_4}(n) = n-3$. The variant we present is slightly more general than the one in [25] but is the same in essence.

Observation 3.8.3. Let H be a 3-connected graph such that for some $e \in E(H)$, $H-e$ has a vertex-cut of size two. Let $t \geq 1$. In the H-process on any graph G, if two cliques $U,W \subset V(G)$ of size at least $v(H)$ at time t satisfy $|U \cap W| \geq 2$, then $U \cup W$ a clique at time $t+1$.

Proof of Observation 3.8.3. We can assume that $U \not\subseteq W$ and $W \not\subseteq U$ for otherwise the claim is trivial. Let $x,y,z,z' \in V(H)$ such that $\{z,z'\}$ is a vertex cut of $H-xy$. We have $\{x,y\}\cap\{z,z'\} = \varnothing$. If not, $\{z,z'\}$ would also be a vertex-cut of H. For any $u \in U \setminus W$ and $w \in W \setminus U$ we can find an embedding $\varphi : V(H) \to U \cup W$ of $H-xy$ at time t such that $\varphi(\{z,z'\}) \subseteq U \cap W$ and $\varphi(x) = u$, $\varphi(y) = w$. Therefore u and w are adjacent at time $t+1$. □

Proof of the second part of Theorem 1.3.12. Let G be an arbitrary n-vertex graph with $\tau_H(G) = M_H(n)$, and let $(G_t)_{t\geq 0}$ be its H-process. Write $\tau := \tau_H(G)$ and take a sequence $H_1,\ldots,H_\tau$ of copies of H such that H_t is completed at time t for $t \in [\tau]$, and $v(H_t \cap H_{t+1}) \geq 2$ for $t \in [\tau-1]$. We can build such a sequence as we did in Section 3.7.2 by starting from H_τ and iteratively defining $H_{\tau-1},\ldots,H_1$.

We now use Observation 3.8.3 inductively to show that for $t \in [\tau]$, $V(H_1 \cup \ldots \cup H_t)$ is a clique at time $t+c$, where $c := M_H(v(H))+1$. By our assumptions on H, $V(H_t)$ is a clique at time $t+c-1$. In particular this covers the base case $t=1$. For $t \geq 2$ we have $V(H_1 \cup \ldots \cup H_{t-1}) \cap V(H_t)| \geq v(H_{t-1} \cap H_t) \geq 2$. The induction hypothesis tells us that $V(H_1 \cup \ldots \cup H_{t-1})$ is a clique in G_{t-1+c}. Observation 3.8.3 implies that $V(H_1 \cup \ldots \cup H_t)$ is a clique in G_{t+c}. This completes the induction.

Since $V(H_1 \cup \ldots \cup H_t)$ is a clique at time $t+c$ and H_{t+c+1} is completed at time $t+c+1$ we obtain $v(H_1 \cup \ldots \cup H_t) < v(H_1 \cup \ldots \cup H_{t+c+1})$ whenever $t+c+1 \leq \tau$. Therefore

$$n \geq v(H_1 \cup \ldots \cup H_{1+r(c+1)}) \geq v(H) + r$$

for any $r \geq 0$ with $1+r(c+1) \leq \tau$. This yields

$$M_H(n) = \tau < 1 + (n+1-v(H))\cdot(c+1).$$

■

3.9 Open problems and further directions

So far we only know of three different types of asymptotic running times within the superlinear range:

- $n^{3/2}$ (the cube Q_3),
- $n^{2-o(1)}$ and $o(n^2)$ (Wheel graphs W_k where $k \geq 7$),
- n^2 (K_r for $r \geq 6$, dense H, $G(k,p)$ for $p = \omega(\log(k)/k)$)

The other superlinear bounds we found lack a complementing upper or lower bound. It would be interesting to know if there exists an infinite family of distinct asymptotic running times. The most promising candidates for such a family seem to be complete bipartite graphs. Theorem 1.3.1 together with the Kővari-Sós-Turán Theorem gives the upper bound

$$M_{K_{s,t}}(n) = O\left(n^{2-\frac{1}{s}}\right)$$

for $3 \leq s \leq t$. A construction involving extremal $K_{s-1,t-1}$-free graphs for suitable values of s and t could lead to a lower bound of $M_{K_{s,t}} = \Omega(\mathrm{ex}(n, K_{s-1,t-1}))$. With Theorems 1.3.5 and 1.3.6 in mind we ask the following question.

Question 3.9.1. Let $2 \leq s \leq t$. Does the maximum running time of $K_{s,t}$ satisfy

$$M_{K_{s,t}}(n) = \Theta(\mathrm{ex}(n, K_{s-1,t-1}))?$$

A positive answer to Question 3.9.1 would give running times of the form $\Theta(n^q)$ for infinitely many $3/2 \leq q < 2$. Indeed, the theorem of Kővari-Sós-Turán together with the constructions in [69] shows that complete bipartite graphs offer infinitely many distinct exponents. Intuitively there should be many bipartite graphs with running time $\Theta(n^q)$ and $3/2 < q < 2$, even if Question 3.9.1 is answered in the negative. As to the gap between linear running time and $\Theta(n^{3/2})$ there is currently no graph H known to satisfy $M_H(n) = \omega(n)$ and $M_H(n) = o(n^{3/2})$. While Theorem 1.3.12 provides superlinear lower bounds for many graphs, our current methods are not suitable to find upper bounds for those graphs in the desired asymptotic range. This leads us to the question below.

Question 3.9.2. For which real $1 < q < \frac{3}{2}$ does a graph H with $M_H(n) = n^{q+o(1)}$ exist?

In Theorem 1.3.8 we have seen that any H with $\delta(H) > 3v(H)/4$ has quadratic running time and the example in Figure 1.3 tells us that there are H with $\delta(H) = v(H)/2$ and linear running time. This motivates the problem of finding the smallest minimum degree that forces $M_H(n)$ to be quadratic, or equivalently, the largest minimum degree that allows subquadratic running time.

Problem 3.9.3. Determine

$$\bar{c} := \limsup_{v(H)\to\infty} \max_{M_H(n)=o(n^2)} \frac{\delta(H)}{v(H)}.$$

The discussion above yields $1/2 \leq \bar{c} \leq 3/4$. Theorem 1.3.8 involves splitting $V(H)$ into (almost) equally sized parts. Perhaps another partition of $V(H)$ could lead to an improved upper bound on $\bar{c}$. As to improving the lower bound Mantel's theorem [75] rules out bipartite H. Furthermore, a consequence of Theorem 1.3.9 is that almost all H with $\delta(H) \geq v(H)/2$ have quadratic running time. It seems that an H with $\delta(H) \geq v(H)/2$ and subquadratic running time is required to have a structure that we can exploit such as a cut-edge and many non-trivial automorphisms.

Our results on cycles suggest that in the maximum running times odd cycles and even cycles do not behave differently from an asymptotic point of view and the proofs for even cycles are just technically more involved because one has to distinguish between bipartite and non-bipartite starting graphs. We thus conjecture that Theorem 1.3.10 can be extended to even k.

Conjecture 3.9.4. The hypothesis that k be odd in Theorem 1.3.10 is not necessary.

Chapter 4

Strong B_h-sets of integers

In this chapter we present the proofs of the results stated in Section 1.4. We show Theorem 1.4.7 in Section 4.1, followed by Theorem 1.4.6 in Section 4.2. We then turn our attention to the proof of Theorem 1.4.8 in Section 4.3.

4.1 An upper bound on the growth of strong B_h-sets

In order to prove Theorem 1.4.7 we need to find $c(\alpha,h)$ such that $S(n) \le c(\alpha,h)n^{(1-\alpha)/h}$ for every α-strong B_h-set of integers. For this purpose we introduce a finite version of α-strong B_h-sets as a natural extension of the concept of (n,α)-strong Sidon sets established in [66].

Definition 4.1.1 (Finite strong B_h-sets)**.** A set $S \subset [n]$ is called an *n-finite α-strong B_h-set* if

$$|(x_1 + \ldots + x_h) - (y_1 + \ldots + y_h)| \ge n^{\alpha}$$

for any $x_1,\ldots,x_h,y_1,\ldots,y_h \in S$ with $\{x_1,\ldots,x_h\} \ne \{y_1,\ldots,y_h\}$.

Now the idea is to express a given infinite α-strong B_h-set as the countable union of suitable finite α-strong B_h-sets and bound the finite sets individually.

Proposition 4.1.2. *Let $h \ge 2$, $n \in \mathbb{N}$ and $0 \le \alpha < 1$. Any n-finite α-strong B_h-set $S \subset \mathbb{N}$ satisfies*

$$|S| < 2h \cdot n^{\frac{1-\alpha}{h}}.$$

Proof. Let $S \subset [n]$ be an n-finite α-strong B_h-set. The claim is obvious if $|S| < 2h$, so we suppose that $|S| \ge 2h$. Each interval of size at most n^{α} contains at most one element of hS. We can partition $[hn]$ as follows:

$$[hn] = \bigcup_{i=1}^{\lfloor h \cdot n^{1-\alpha} \rfloor} ((i-1)n^{\alpha}, i \cdot n^{\alpha}] \cup \left(\lfloor h \cdot n^{1-\alpha} \rfloor \cdot n^{\alpha}, hn\right]. \tag{4.1}$$

Since $hS \subset [hn]$ and each of the $\lfloor h \cdot n^{1-\alpha} \rfloor + 1$ intervals on the right hand side of (4.1) has size at most n^α, we obtain

$$|hS| = \sum_{i=1}^{\lfloor h \cdot n^{1-\alpha} \rfloor} |hS \cap ((i-1)n^\alpha, i \cdot n^\alpha]| + \left|hS \cap \left(\lfloor h \cdot n^{1-\alpha} \rfloor, hn\right]\right| \leq \lfloor h \cdot n^{1-\alpha} \rfloor + 1$$

Combining this with the lower bound

$$|hS| > \binom{|S|}{h} > \frac{|S|^h}{h^h}$$

and the simple estimate $h^{1/h} < 2$ yields

$$|S| < \left(h \cdot n^{1-\alpha} \cdot h^h\right)^{1/h} < 2h \cdot n^{(1-\alpha)/h}$$

□

Now, let $S \subset \mathbb{N}$ be an infinite α-strong B_h set. Observe that for every $n \in \mathbb{N}$ and $x_1, y_1, \ldots, x_h, y_h \in S \cap [n, \infty)$ we have

$$|(x_1 - n + \ldots + x_h - n) - (y_1 - n + \ldots + y_h - n)| = |(x_1 + \ldots + x_h) - (y_1 + \ldots + y_h)| \geq n^\alpha$$

so $(S-n) \cap [n]$ is an n-finite α-strong B_h set of size $|S \cap (n, 2n]|$. For this reason we partition S into the sets $S_i := S \cap (2^i, 2^{i+1}]$, $i \geq 0$, and $S_{-1} := S \cap \{1\}$. For each $i \geq 0$, the translated sets $S_i - 2^i$ is a 2^i-finite α-strong B_h-set. This implies

$$\begin{aligned}
S(n) &= \sum_{i=-1}^{\infty} |S_i \cap [n]| \\
&\leq 1 + \sum_{i=0}^{\lceil \log_2 n \rceil} |S_i| \\
&< 1 + \sum_{i=0}^{\lceil \log_2 n \rceil} 2h \cdot 2^{i \cdot \frac{1-\alpha}{h}} \\
&< 1 + 2h \cdot \frac{2^{\lceil \log_2(n)+1 \rceil \cdot (1-\alpha)/h} - 1}{2^{(1-\alpha)/h} - 1} \\
&\leq 1 + \frac{4h}{2^{(1-\alpha)/h} - 1} \cdot n^{\frac{1-\alpha}{h}}
\end{aligned}$$

and hence we can choose $c(\alpha, h) = 8h/(2^{(1-\alpha)/h} - 1)$ to obtain the desired bound

$$S(n) \leq c(\alpha, h) n^{\frac{1-\alpha}{h}}.$$

■

4.2 An infinite strong B_h-sets

We are going to prove a slightly stronger variant of Theorem 1.4.6 that holds not just for α-strong B_h-sets but also for the following.

Definition 4.2.1 ((α,γ)-strong B_h-set)**.** Let $h \geq 2$. Given real numbers $0 \leq \alpha < 1$ and $\gamma \geq 1$ we say that a set of integers $S \subset \mathbb{N}$ is an *(α,γ)-strong B_h-set* if

$$|(x_1 + \ldots + x_h) - (y_1 + \ldots + y_h)| \geq \gamma \cdot \max\{x_1^\alpha, y_1^\alpha, \ldots, x_h^\alpha, y_h^\alpha\} \tag{4.2}$$

for any $x_1, y_1, \ldots, x_h, y_h \in S$ with $\max\{x_1, \ldots, x_h\} \neq \max\{y_1, \ldots, y_h\}$.

Remark 4.2.2. This definition provides a natural generalisation of the concept of (α, c)-strong Sidon sets introduced in [66]. An $(\alpha, 1)$-strong B_h-set is just an α-strong B_h-set. The additional factor γ gives us more flexibility that is need when we apply the theorem to find B_h-set in random subsets of the integers. Also note that, given $\alpha' > \alpha \geq 0$ and $\gamma \geq 1$, we have $\gamma n^\alpha < n^{\alpha'}$ when n is sufficiently large, so any α'-strong B_h-set is (after removing some initial elements, if necessary) also an (α,γ)-strong B_h-set.

Let us state the aforementioned variant of Theorem 1.4.6.

Theorem 4.2.3. *For every $h \geq 2$ and reals $0 \leq \alpha < 1$ and $\gamma \geq 1$ there exists an (α,γ)-strong B_h-set $S \subset \mathbb{N}$ satisfying*

$$S(n) \geq n^{\sqrt{(h-1+\frac{\alpha}{2})^2+1-\alpha}-(h-1+\frac{\alpha}{2})+o(1)}. \tag{4.3}$$

This theorem clearly implies Theorem 1.4.6 by setting $\gamma = 1$. The size estimates rely on the well-known Prime Number Theorem, which is stated below. We refer to the classic text of Hardy and Wright [56] for a proof.

Theorem 4.2.4 (Prime Number Theorem)**.** *Let $\pi(n)$ denote the number of primes in $[n]$. Then*

$$\pi(n) \sim \frac{n}{\log n}.$$

In Section 4.2.1 we construct a candidate for an (α,γ)-strong B_h-set by following the construction of Cilleruelo [29] and adapting it to our setting. In the same section we provide several lemmas that describe the important properties of the candidate set. Those lemmas will be generalisations of their counterparts in [29]. In Section 4.2.2 we refine our candidate set by removing elements that cause a violation of (4.2), and show that this alteration does not affect the growth of the counting function. Our exposition will be self-contained, so although we encourage the reader to have a look at the original work of Cilleruelo, familiarity with that work is not required as we will reintroduce the necessary ideas and definitions.

4.2.1 The construction and its basic properties

We will use the same sets $A_{\bar{q},c} \subset \mathbb{N}$ that appear in [29], where $\bar{q}$ is a generalised basis (as defined below) and c is a real parameter c with $0 < c < \frac{1}{2}$. For our purposes we fix the choice

$$c := \sqrt{\left(h-1+\frac{\alpha}{2}\right)^2 + 1 - \alpha} - \left(h-1+\frac{\alpha}{2}\right). \tag{4.4}$$

The reasons behind this choice become apparent later in the proof. The difference to the original construction (i.e. the case $\alpha = 0$) lies in the choice of c and the elements $A_{\bar{q},c}$ that we have to remove later.

Given a sequence $\bar{q} = (\bar{q}_1, \bar{q}_2, \ldots)$ of positive integers, any $a \in \mathbb{N}_0$ can be uniquely expressed as

$$a = x_1 + x_2 \cdot \bar{q}_1 + x_3 \cdot \bar{q}_1\bar{q}_2 + x_4 \cdot \bar{q}_1\bar{q}_2\bar{q}_3 + \ldots + x_k \cdot \bar{q}_1 \ldots \bar{q}_{k-1}, \tag{4.5}$$

where $0 \le x_i < \bar{q}_i$ for $i \in [k]$ and $x_k \neq 0$. We refer to the numbers x_i (henceforth written as $x_i(a)$ to bring out the dependence on a) as the *digits* of a in the *generalised basis* $\bar{q}$. Given (4.5) re call $\operatorname{len}(a) := k$ the *length* of a, that is, the length is the number of digits in the generalised basis. For convenience we also set $x_i(a) := 0$ for $i > \operatorname{len}(a)$. In the following the basis $\bar{q}$ will always be clear from context, hence we did not explicitely indicate the basis in the notation above.

For $i \ge 1$, let q_i be a prime satisfying

$$2^{2i-1} < q_i \le 2^{2i+1}. \tag{4.6}$$

Such a choice is possible due to Bertrand's Postulate (for a proof of that postulate see for example [4]). Consider the generalised basis

$$\bar{q} := (h^2 q_1, h^2 q_2, h^2 q_3, \ldots).$$

The set $A_{\bar{q},c}$ will be constructed by defining its elements digitwise. Denote the set of primes by $\mathscr{P}$ and, for $k \ge 3$, let

$$\mathscr{P}_k := \left\{ p \in \mathscr{P} : 2^{c(k-1)^2 - f(c,k-1)} < p \le 2^{ck^2 - f(c,k)} \right\},$$

where $f(c,k) := ck^2/\sqrt{\log k}$. The sets $\mathscr{P}_k$ form a partition $\mathscr{P} = \bigcup_{k=3}^{\infty} \mathscr{P}_k$ of the primes. Any $a \in \mathbb{N}$ is uniquely determined by its digits $x_i(a)$, $i \ge 1$. For $p \in \mathscr{P}$ define the number a_p by its digits in the basis $\bar{q}$ as follows. Take the uniqe $k \ge 3$ with $p \in \mathscr{P}_k$. For $i \in [k]$, let the digit $x_i(a_p)$ be the unique solution to the congruence

$$g_i^{x_i(a_p)} \equiv p \mod q_i \quad , \quad (h-1)q_i + 1 \le x_i(a_p) \le hq_i - 1. \tag{4.7}$$

Here the condition $(h-1)q_i + 1 \le x_i(a_p) \le hq_i - 1$ is necessary to guarantee uniqueness since $g_i^{q_i - 1} \equiv 1 \mod q_i$. Moreover, let $x_0(a_p) = 0$ and $x_i(a_p) = 0$ for $i \ge k+1$. With these choices we have $\operatorname{len}(a_p) = k$. Note that a_p uniquely determines the residue class of p modulo q_i for each

$i \in [k]$, and hence, by the Chinese Remainder Theorem, the residue class modulo $q_1 \dots q_k$. This determines p uniquely as $p \leq 2^{ck^2 - f(c,k)} \leq 2^{k^2} < q_1 \dots q_k$. Therefore, $a_p \neq a_{p'}$ for $p \neq p'$.

By construction a_p has precisely k digits if $p \in \mathscr{P}_k$. Cilleruelo ([29], Section 3.1) showed that the sequence $A_{\bar{q},c}$ satisfies

$$A_{\bar{q},c} \geq n^{c+o(1)}. \tag{4.8}$$

By the two inequalities in (4.7) and the constant factor h^2 in the choice of $\bar{q}$ we can sum any h elements of $A_{\bar{q},c}$ by simply summing their digits, and the magnitude of any digit of the sum tells us how many of the h summands have a non-zero digit at the same position. We summarise these observation in the following remark.

Remark 4.2.5. For any $p_1, \dots, p_s \in \mathscr{P}$, where $s \in [h]$, and $i \geq 1$ we have

$$x_i(a_{p_1} + \dots + a_{p_s}) = x_i(a_{p_1}) + \dots + x_i(a_{p_s}),$$

and thus

$$\mathrm{len}(a_{p1} + \dots + a_{p_s}) = \max\{\mathrm{len}(a_{p_1}), \dots, \mathrm{len}(a_{p_s})\}.$$

Furthermore, the i^{th} digit satisfies

$$m_i(h-1)q_i + m_i \leq x_i(a_{p_1} + \dots + a_{p_s}) \leq m_i h q_i - m_i,$$

where $m_i := |\{j \in [s] : x_i(a_{p_j}) \neq 0\}|$. In particular, m_i is uniquely determined by the digit $x_i(a_{p_1} + \dots + a_{p_s})$.

As in the usual decimal system, the number of digits of a in the generalised basis gives bounds on the magnitude of a. Note, however, that since $h^2 q_i$ is increasing in i the upper upper and the lower bound on the magnitude are more than a constant factor apart from each other.

Lemma 4.2.6. *Let $a \in \mathbb{N}$ and $k := \mathrm{len}(a)$. We have that*

$$h^{2k-2} 2^{k^2 - 2k + 1} < a < h^{2k} 2^{k^2 + 2k}.$$

Proof. By the definition of $\bar{q}$ and the assumption that a has k digits we obtain

$$h^2 q_1 \cdot \ldots \cdot h^2 q_{k-1} \leq a < h^2 q_1 \cdot \ldots \cdot h^2 q_k.$$

These estimates combined with (4.6) yield

$$a > h^{2k-2} \prod_{i=1}^{k-1} 2^{2i-1} = h^{2k-2} 2^{k^2 - 2k + 1}$$

and

$$a < h^{2k} \prod_{i=1}^{k} 2^{2i+1} = h^{2k} 2^{k^2 + 2k}.$$

$\square$

Lemma 4.2.6 has the following consequence.

Lemma 4.2.7. *For any $a,k \in \mathbb{N}$ with $k = \mathrm{len}(a)$, and real numbers $0 \leq \alpha < 1$, $\gamma \geq 1$,*

$$\mathrm{len}(\lfloor \gamma a^\alpha \rfloor) \leq \left(\alpha k^2 + (\log_2 h + 1)2\alpha k + \log_2 \gamma\right)^{1/2}.$$

Proof. Write $k := \mathrm{len}(a)$ and $\ell := \left(\alpha k^2 + (\log_2 h + 1)2\alpha k + \log_2 \gamma\right)^{1/2}$. Then

$$\gamma a^\alpha < \gamma(h^2 q_1 \dots h^2 q_k)^\alpha \leq \gamma h^{\alpha 2k} 2^{\alpha(k^2+2k)} \leq h^{2(\ell+1)-2} 2^{(\ell+1)^2-2(\ell+1)+1}.$$

The lower bound in Lemma 4.2.6 now tells us that $\lfloor \gamma a^\alpha \rfloor$ has at most ℓ digits, and hence the claim follows. □

The next statement is the (α,γ)-strong analogue of Proposition 7 in [29].

Proposition 4.2.8. *Let $0 \leq \alpha < 1$ and $\gamma \geq 1$ be real parameters. Let $p_1, p'_1, \dots, p_h, p'_h$ be primes such that*

1. *$a_{p_1} \geq \dots \geq a_{p_h}$ and $a_{p'_1} \geq \dots \geq a_{p'_h}$ as well as $a_{p_1} > a_{p'_1}$,*
2. *$|(a_{p_1} + \dots + a_{p_h}) - (a_{p'_1} + \dots + a_{p'_h})| < \gamma a_{p_1}^\alpha$.*

Write $k_i := \mathrm{len}(a_{p_i})$ and $k'_i := \mathrm{len}(a_{p'_i})$ for $i \in [h]$. Then there exists $\ell \in [0, k_1]$ and $s \in [h]$ such that

(i) $k_i = k'_i \geq \ell + 1$ *for all* $i \in [s]$,

(ii) $\ell^2 \leq \alpha k_1^2 + (\log_2 h + 1)2\alpha k_1 + \log_2 \gamma + 1$,

(iii) $\ell^2 \geq (1-c)k_s^2 - c(k_1^2 + \dots + k_{s-1}^2)$,

(iv) $q_{\ell+1} \dots q_{k_1} \,|\, \prod_{j=1}^s (p_1 \dots p_j - p'_1 \dots p'_j)$

Proof. Note that $k_1 \geq \dots \geq k_h$ and $k'_1 \geq \dots \geq k'_h$ due to (1). We set

$$\ell := \max\left\{ i \in [0, k_1] : x_i(a_{p_1} + \dots + a_{p_h}) \neq x_i(a_{p'_1} + \dots + a_{p'_1}) \right\},$$
$$s := \max\left\{ j \in [h] : \max\{k_j, k'_j\} > \ell \right\}.$$

Recall that by Remark 4.2.5, $|\{j \in [s] : x_i(a_{p_j}) \neq 0\}|$ is determined by $x_i(a_{p_1} + \dots + a_{p_h})$, and similarly, $|\{j \in [s] : x_i(a_{p'_j}) \neq 0\}|$ is given by $x_i(a_{p'_1} + \dots + a_{p'_h})$ for any $i \geq 1$. From the definition of ℓ and the assumptions that $a_{p_1} \geq \dots \geq a_{p_h}$ and $a_{p'_1} \geq \dots \geq a_{p'_h}$ we deduce that for $i \geq \ell + 1$,

$$\{j \in [h] : x_i(a_{p_j}) \neq 0\} = \{j \in [h] : x_i(a_{p'_j}) \neq 0\}.$$

For all $j \in [s]$ with $k_j \geq k'_j$, we have $x_{k_j}(a_{p_j}) \neq 0$ by definition of k_j and $k_j \geq \ell + 1$ by definition of s, so $x_{k_j}(a_{p'_j}) \neq 0$ and thus $k'_j \geq k_j$. Similarly, $k_j \geq k'_j$ whenever $j \in [s]$ with $k'_j \geq k_j$. This verifies property (i). By definition of ℓ and the fact that the length function len is non-decreasing we can see that

$$\begin{aligned}\ell-1 &\le \operatorname{len}\left(\left|(a_{p_1}+\ldots+a_{p_h})-(a_{p'_1}+\ldots+a_{p'_h})\right|\right)\\ &\le \operatorname{len}(\lfloor\gamma a_{p_1}^{\alpha}\rfloor)\\ &\le \left(\alpha k_1^2+(\log_2 h+1)2\alpha k_1+\log_2\gamma\right)^{1/2}\end{aligned}$$

and hence (ii) holds. As to (iii) and (iv) observe that for $i\in[\ell+1,k_1]$ Remark 4.2.5 and the definiton of ℓ imply

$$x_i(a_{p_1})+\ldots+x_i(a_{p_h})=x_i(a_{p'_1})+\ldots+x_i(a_{p'_h}).$$

and thus

$$g_i^{x_i(a_{p_1})+\ldots+x_i(a_{p_h})}\equiv g_i^{x_i(a_{p'_1})+\ldots+x_i(a_{p'_h})} \mod q_i.$$

By (4.7) one then has

$$p_1\ldots p_{j(i)}\equiv p'_1\ldots p'_{j(i)} \mod q_i. \tag{4.9}$$

for $i\in[\ell+1,k_1]$ where $j(i)$ is the largest index in $[h]$ satisfying $i\le k_{j(i)}$. For every $j\in[h]$, we have $\{i\ge 1: j(i)=j\}=[\max\{\ell+1,k_{j+1}+1\},k_j]$, where for convenience we set $k_{h+1}:=0$. Furthermore the primes $q_{\max\{\ell,k_{j+1}\}+1},\ldots,q_{k_j}$ are pairwise distinct so the Chinese Remainder Theorem yields

$$p_1\ldots p_j\equiv p'_1\ldots p'_j \mod q_{\max\{\ell,k_{j+1}\}+1}\cdots q_{k_j}.$$

and consequently

$$q_{\max\{\ell,k_{j+1}\}+1}\cdots q_{k_j}|(p_1\ldots p_j-p'_1\ldots p'_j) \tag{4.10}$$

Property (iv) now follows by taking the products of both sides of (4.10) over all $j\in[s]$. We have $p_1\ldots p_s\ne p'_1\ldots p'_s$ because of assumption (1). Hence, by (4.9) for $j=s$, we can see that $p_1\ldots p_s-p'_1\ldots p'_s$ must be a positive multiple of $q_{\ell+1}\ldots q_{k_s}$. Combining this with (4.6) gives us

$$2^{c(k_1^2+\ldots+k_s^2)}\ge|p_1\ldots p_s-p'_1\ldots p'_s|\ge q_{\ell+1}\ldots q_{k_s}>2^{k_s^2-\ell^2},$$

which by taking logarithms and rearranging summands implies property (iii). □

4.2.2 Proof of Theorem 4.2.3

We have seen that for any choice of $q_1,q_2,\ldots$ the counting function of $A_{\bar q,c}$ satisfies

$$A_{\bar q,c}(n)\ge n^{c+o(1)}.$$

Yet, $A_{\bar q,c,h}$ is not necessarily an (α,γ)-strong B_h-set. From now on we suppose that c is fixed and has the value given in (4.4). Call a tuple $(p_1,p'_1,\ldots,p_h,p'_h)$ of primes *bad* for $\bar q$ if

$$a_{p_1}\ge\ldots\ge a_{p_h}\quad,\quad a_{p'_1}\ge\ldots\ge a_{p'_h}\quad,\quad a_{p_1}>a_{p'_1},$$

and

$$|(a_{p_1}+\ldots+a_{p_h})-(a_{p'_1}+\ldots+a_{p'_h})|\geq \gamma a_{p_1}^{\alpha}.$$

Denote the set of bad tuples by $\mathscr{B}(\bar{q})$. Recall that the basis $\bar{q}$ was used to define the a_{p_i}, so whether a tuple is bad indeed depends on $\bar{q}$. To arrive at the desired (α,γ)-strong B_h-set we pass to a subset of $A_{\bar{q},c}$ by removing a_{p_1} for every bad tuple $(p_1,p'_1,\ldots,p_h,p'_h)$. For $k\geq 1$ let $\mathscr{B}_k(\bar{q})$ be the set of primes in $\mathscr{P}_k$ that appear as the first entry of a bad tuple. Define

$$S_{\bar{q}} := \left\{ a_p : p \in \bigcup_{k=3}^{\infty} \mathscr{P}_k \setminus \mathscr{B}_k(\bar{q}) \right\}$$

While $S_{\bar{q}}$ is clearly an (α,γ)-strong B_h-set we need to make sure that the removal above does not affect the asymptotics of the counting function, that is, we want to show that

$$S_{\bar{q}}(n)=\Omega(A_{\bar{q},c}(n)) \tag{4.11}$$

We cannot guarantee this for every basis $\bar{q}$. However, in order to find the desired B_h-set it suffices to have some choice of $\bar{q}$ satisfying (4.11). For this reason we will choose a basis randomly and show that (4.11) holds with positive probability.

Consider the probability space of sequences $(q_i)_{i\in\mathbb{N}}$ with $2^{2i-1}<q_i\leq 2^{2i+1}$ and q_i prime for all $i\in\mathbb{N}$, where each q_i is chosen independently and uniformly at random from $\mathscr{P}\cap\left(2^{2i-1},2^{2i+1}\right]$, and set $\bar{q}:=(h^2q_i)_{i\in\mathbb{N}}$. The construction of such a probability space is standard though not obvious and can be found in most modern textbooks on probability theory (e.g. [52]).

In the following, the labels (i), (ii), (iii), (iv) always refer to the conclusions (i)-(iv) of Proposition 4.2.8. If $k\in\mathbb{N}$, and $p_1\in\mathscr{P}_k$ is the first entry of a bad tuple $(p_1,p'_1,\ldots,p_h,p'_h)$, then by Proposition 4.2.8 there exist $k_1,k'_1,\ldots,k_h,k'_h\in\mathbb{N}$, $s\in[h]$, and $\ell\in[k_1]$ satisfying $k_1=k$ and (i)-(iv). We will use this fact to estimate the size of $|\mathscr{B}_k(\bar{q})|$.

We define the following sets. For $s\in[h]$, $k\geq 3$, and $\ell\in[0,k]$, let

$$\mathscr{K}_{s,k,\ell} := \{(k_1,\ldots,k_s) : 0\leq \ell\leq k_s\leq\ldots\leq k_1=k,\ \text{(ii) and (iii) hold}\}.$$

If $k_1\geq\ldots\geq k_s$, write

$$\mathscr{P}_{k_1,\ldots,k_s} := \left\{(p_1,p'_1,\ldots,p_s,p'_s) : p_i,p'_i\in\mathscr{P}_{k_i} \text{ for } i\in[s]\right\}.$$

Finally, given $s\in[h]$, $k\geq 3$, $\ell\in[0,k]$ define

$$\mathscr{Q}_{s,k,\ell} := \left\{ (p_1,p'_1,\ldots,p_s,p'_s)\in\mathscr{P}^{2s}\ :\ q_{\ell+1}\ldots q_k \text{ divides } \prod_{i=1}^{s}(p_1\ldots p_i-p'_1\ldots p'_i) \right\}.$$

Note that $\mathscr{K}_{s,k,\ell}$ and $\mathscr{P}_{k_1,\dots,k_s}$ are deterministic whereas $\mathscr{Q}_{s,k,\ell}$ is random. These sets allows us to bound $|\mathscr{B}_k(\bar{q})|$ as follows.

$$|\mathscr{B}_k(\bar{q})| \leq \left| \bigcup_{s=1}^{h} \bigcup_{\ell=1}^{k} \bigcup_{(k_1,\dots,k_s)\in\mathscr{K}_{s,k,\ell}} \mathscr{P}_{k_1,\dots,k_s} \cap \mathscr{Q}_{s,k,\ell} \right| \tag{4.12}$$

To see why this holds, observe that $\mathscr{B}_k(\bar{q})$ is contained in the set obtained by projecting the elements of the set whose cardinality we take on the right hand side of (4.12) to their first coordinates. From (4.12) we infer

$$\begin{aligned}\mathbb{E}(|\mathscr{B}_k(\bar{q})|) &\leq \sum_{s=1}^{h}\sum_{\ell=1}^{k}\sum_{\mathbf{k}\in\mathscr{K}_{s,k,\ell}} \mathbb{E}\left(|\mathscr{P}_{\mathbf{k}}\cap\mathscr{Q}_{s,k,\ell}|\right)\\ &= \sum_{s=1}^{h}\sum_{\ell=1}^{k}\sum_{\mathbf{k}\in\mathscr{K}_{s,k,\ell}}\ \sum_{(p_1,p_1',\dots,p_s,p_s')\in\mathscr{P}_{\mathbf{k}}} \mathbb{P}\left(q_{\ell+1}\dots q_k \text{ divides } \prod_{i=1}^{s}(p_1\dots p_i - p_1'\dots p_i')\right).\end{aligned}$$

We have that

$$\mathbb{P}\left(q_{\ell+1}\dots q_k \text{ divides } \prod_{i=1}^{s}(p_1\dots p_i - p_1'\dots p_i')\right) \leq \frac{d\left(\prod_{i=1}^{s}(p_1\dots p_i - p_1'\dots p_i')\right)}{\prod_{i=\ell+1}^{k}\left(\pi(2^{2i+1})-\pi(2^{2i-1})\right)}$$

where d is the divisor function. The Prime Number Theorem implies

$$\pi(2^{2i+1})-\pi(2^{2i-1}) = 2^{2i-1+O(\log i)}$$

and the divisor function obeys the bound $d(n) = 2^{O(\log(n)/\log\log(n))}$ (cf. [56] Theorem 317). Since $\prod_{i=1}^{s}(p_1\dots p_i - p_1'\dots p_i') \leq 2^{s\cdot c(k_1^2+\dots+k_s^2)} \leq 2^{O(k^2)}$, we arrive at

$$\frac{d\left(\prod_{i=1}^{s}(p_1\dots p_i - p_1'\dots p_i')\right)}{\prod_{i=\ell+1}^{k}\left(\pi(2^{2i+1})-\pi(2^{2i-1})\right)} \leq \frac{2^{O(k^2/\log k)}}{2^{k^2-\ell^2+O(k\log k)}} \leq 2^{-k^2+\ell^2+O(k^2/\log k)}. \tag{4.13}$$

For $s\in[h]$, $k\geq 3$, $\ell\in[k]$, and $(k_1,\dots,k_s)\in\mathscr{K}_{s,k,\ell}$ we have the bounds

$$|\mathscr{K}_{s,k,\ell}| \leq k^h \qquad \text{and} \qquad |\mathscr{P}_{k_1,\dots,k_s}| = \prod_{j=1}^{s}|\mathscr{P}_{k_j}| =\leq 2^{2c(k_1^2+\dots+k_s^2)-2f(c,k)}, \tag{4.14}$$

and property (iii) of Proposition 4.2.8 yields

$$\begin{aligned}2c(k_1^2+\dots+k_s^2) &\leq 2c\left(k_1^2+\dots+k_{s-1}^2+\frac{\ell^2+c(k_1^2+\dots+k_{s-1}^2)}{1-c}\right)\\ &= \frac{2c}{1-c}\left(k_1^2+\dots+k_{s-1}^2+\ell^2\right)\\ &\leq \frac{2c}{1-c}(s-1)k^2+\frac{2c}{1-c}\ell^2.\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{E}(|\mathscr{B}_k(\bar{q})|) &\le h \cdot k \cdot k^h \cdot 2^{2c(k_1^2+\ldots+k_s^2)-2f(c,k)} \cdot 2^{-k^2+\ell^2+O(k^2/\log k)} \\ &\le 2^{(\frac{2c}{1-c}(s-1)-1)k^2+(\frac{2c}{1-c}+1)\ell^2-2f(c,k)+O(k^2/\log k)} \\ &\le 2^{(\frac{2c(s-1)+(1+c)\alpha}{1-c}-1)k^2-2f(c,k)+O(k^2/\log k)}\end{aligned}$$

where the last inequality uses that $\ell^2 \le \alpha k^2 + O(k)$ by (ii) in Proposition 4.2.8. Recall 4.4. With that definition c satisfies

$$\frac{2c(h-1)+(1c)\alpha}{1-c} - 1 = c.$$

Hence we obtain

$$\mathbb{E}(|\mathscr{B}_k(\bar{q})|) \le 2^{ck^2-2f(c,k)+O(k^2/\log k)}.$$

As $|\mathscr{P}_k| = \Omega(2^{ck^2-f(c,k)-\log_2(ck^2-f(c,k))})$ for every $k \ge 3$, we get the estimate

$$\mathbb{E}\left(\sum_{k\ge 3}\frac{|\mathscr{B}_k(\bar{q})|}{|\mathscr{P}_k|}\right) \le \sum_{k\ge 3} 2^{-f(c,k)+O(k^2/\log k)} \le \sum_{k\ge 3} 2^{-(c+o(1))k^2/\sqrt{\log k}}. \tag{4.15}$$

The series on the right hand side of (4.15) converges, so by Markov's inequality $\sum_{k\ge 3}\frac{|\mathscr{B}_k(\bar{q})|}{|\mathscr{P}_k|}$ converges with probability one. For this reason $|\mathscr{B}_k(\bar{q})| = o(|\mathscr{P}_k|)$ with probability one for all $k \ge 3$. Given $n \in \mathbb{N}$ let $k(n)$ be the largest integer with $h^{2k(n)}2^{k(n)^2+2k(n)} \le n$. Then Lemma 4.2.6 implies $\{a_p : p \in \mathscr{P}_k\} \subseteq [n]$ for $k \le k(n)$, and

$$\begin{aligned}S_{\bar{q}}(n) &\ge \sum_{k=3}^{k(n)}(|\mathscr{P}_k| - |\mathscr{B}_k(\bar{q})|) \\ &= \sum_{k=3}^{k(n)}(1-o(1))|\mathscr{P}_{k(n)}| \\ &\ge (1-o(1))\pi\left(2^{ck(n)^2-ck(n)^2/\sqrt{\log k(n)}}\right) \\ &\ge n^{c+o(1)}.\end{aligned}$$

This completes the proof of Theorem 4.2.3.

■

4.3 B_h-sets in random sets of integers

We now prove Theorem 1.4.8. Let us recall the necessary definitions and notation. We are given a real parameter $0 < \delta \le 1$ and consider the random subset R_δ of $\mathbb{N}$ given by picking each positive integer m independently with probability $p_m := 1/m^{1-\delta}$. We now want to show that for any integer $h \ge 2$ the random set R_δ contains with probability 1 a B_h-set S satisfying

$$S(n) \ge n^{\sqrt{(h-1+\frac{1-\delta}{2})^2+\delta}\ -\ (h-1+\frac{1-\delta}{2})+o(1)}.$$

To do so we use the approach introduced in [66] to prove the analogous claim for Sidon sets. We start with a strong B_h-set S' and show that with high probability R_δ contains a subset S such that every element of S is close to S' and no two elements of S are near the same element of S'. This will yield the desired set since slightly displacing each element of S' results in a B_h-set.

We partition the set of positive integers into intervals as follows

$$\mathbb{N} = \bigcup_{j\geq 1} I_j \qquad \text{, where} \quad I_j := \left[j^{1/\delta}, (j+1)^{1/\delta}\right). \tag{4.16}$$

Note that the sizes of the intervals I_j are monotone in j. Given two positive integers $a, b \in \mathbb{N}$ we write

$$a \sim b$$

if they lie in the same part of the partition we just defined, that is, if $a, b \in I_j$ for some $j \geq 1$. Before we choose the set S' we have to establish a few auxiliary results.

4.3.1 A couple of auxiliary statements

The choice of partition (4.16) guarantees that the random set R_δ intersects each part with at least a fixed constant probability. This fact already appeared in [66] and we refer to that article for a proof.

Lemma 4.3.1 ([66], Lemma 13)**.** *There exists $j_0 \geq 1$ such that for any $j \geq j_0$,*

$$\mathbb{P}(R_\delta \cap I_j \neq \varnothing) \geq \frac{1}{3}. \tag{4.17}$$

The size of the interval I_j, $j \geq 1$, obeys the following upper bound that is easier to handle than the obvious precise term $(j+1)^{1/\delta} - j^{1/\delta}$.

Lemma 4.3.2. *For any $j \geq 1$ we have $|I_j| < 2^{\frac{1}{\delta}} j^{\frac{1}{\delta}-1}$.*

Proof. The claim follows from the fact that for any real $x, \alpha \geq 1$ one has

$$(x+1)^\alpha - x^\alpha \leq 2^\alpha \cdot x^{\alpha-1},$$

which is equivalent to

$$x \cdot \left(\frac{(x+1)^\alpha}{x^\alpha} - 1\right) \leq 2^\alpha. \tag{4.18}$$

Suppose that $\alpha \geq 1$ is fixed. The real function

$$f : \mathbb{R} \to \mathbb{R} \quad , \quad x \mapsto x \cdot \left(\frac{(x+1)^\alpha}{x^\alpha} - 1\right)$$

satisfies $f(1) = 2^\alpha - 1 < 2^\alpha$ and

$$f'(x) = \frac{(x+1)^\alpha - x^\alpha - \alpha(x+1)^{\alpha-1}}{x^\alpha}.$$

We need that $f'(x) \le 0$ for all $x \ge 1$. Since the denominator x^α is positive we have to show that

$$(x+1)^\alpha - x^\alpha - \alpha(x+1)^{\alpha-1} \le 0$$

for $x \ge 1$. This follows directly from the convexity of the function $x \mapsto x^\alpha$. Now (4.18) with $\alpha = 1/\delta$ and $x = j$ yields the claim. $\square$

Our next statement tells us that given a sufficiently strong B_h-set, we can move around its elements within each of the intervals I_j without destroying the B_h-property.

Lemma 4.3.3. *If $S \subset \mathbb{N}$ is a $(1-\delta, h2^{1+\frac{1}{\delta}})$-strong B_h-set, and $\sigma : S \to \mathbb{N}$ is an injection such that $\sigma(s) \sim s$ for all $s \in S$, then $\sigma(S)$ is a B_h-set.*

Proof. Let S and σ be as in the hypotheses. Assume for a contradiction that one can find $\tilde{x}_1, \tilde{y}_1, \ldots, \tilde{x}_h, \tilde{y}_h \in \sigma(S)$ such that $\max\{\tilde{x}_1, \ldots, \tilde{x}_h\} \neq \max\{\tilde{y}_1, \ldots, \tilde{y}_h\}$ and $\tilde{x}_1 + \ldots + \tilde{x}_h = \tilde{y}_1 + \ldots + \tilde{y}_h$. For $i \in [h]$, let $x_i := \sigma^{-1}(\tilde{x}_i)$ and $y_i := \sigma^{-1}(\tilde{y}_i)$ be the corresponding elements in S. Without loss of generality we may assume that $x_1 \ge \ldots \ge x_h$ and $y_1 \ge \ldots \ge y_h$ as well as $x_1 > y_1$. Let $j_1 \ge 1$ such that $x_1 \in I_{j_1}$. Lemma 4.3.2 implies

$$\begin{aligned} |(x_1 + \ldots + x_h) - (y_1 + \ldots + y_h)| &\le h2|I_{j_1}| \\ &< h2^{1+\frac{1}{\delta}} j_1^{\frac{1}{\delta}-1} \\ &\le h2^{1+\frac{1}{\delta}} x^{1-\delta}, \end{aligned}$$

which contradicts the hypotheses that S is a $(1-\delta, h2^{1+\frac{1}{\delta}})$-strong B_h-set. $\square$

Lemma 4.3.3 has an immediate consequence.

Corollary 4.3.4. *Any $(1-\delta, 2^{1+\frac{1}{\delta}})$-strong B_h-set S satisfies*

$$|S \cap I_j| \le 2$$

for all $j \ge 1$.

Proof. If there was $j \ge 1$ with $|S \cap I_j| \ge 3$ we could simply pick distinct $x, y, z \in S \cap I_j$ and choose a bijection $\sigma_j : I_j \to I_j$ such that $\sigma_j(x)$, $\sigma_j(y)$, and $\sigma_j(z)$ lie on an arithmetic progression of length three. But then the map $\sigma : S \to \mathbb{N}$ that coincides with σ_j on $S \cap I_j$ and is the identity otherwise leads to a contradiction to Lemma 4.3.3 because B_h-sets do not contain 3-APs. $\square$

4.3.2 Proof of Theorem 1.4.8

Let $S' \subset \mathbb{N}$ be a $(1-\delta, h2^{1+\frac{1}{\delta}})$-strong B_h-set with

$$S'(n) \ge n^{u(\delta)-o(1)} \qquad \text{where} \qquad u(\delta) = \sqrt{\left(h-1+\frac{1-\delta}{2}\right)^2 + \delta} - \left(h-1+\frac{1-\delta}{2}\right).$$

Such a set exists by Theorem 4.2.3. By Corollary 4.3.4 we have $|S' \cap I_j| \leq 2$ for each $j \geq 1$. Let S'' be the subset of S' obtained by removing the larger element of $S' \cap I_j$ for all $j \geq 1$ with $|S' \cap I_j| = 2$. The set S'' is clearly a $(1-\delta, h2^{1+\frac{1}{\delta}})$-strong B_h-set and its counting function satisifies

$$S''(n) \geq \frac{S'(n)}{2}$$

for all $n \in \mathbb{N}$. For each $s \in S''$, let $j(s)$ be the unique index with $s \in I_{j(s)}$. By construction, $j(s) \neq j(s')$ for distinct $s, s' \in S''$. Let j_0 be as given by Lemma 4.3.1 and define the random set

$$S''' := \{s \in S'' : j(s) \geq j_0 \text{ and } R_\delta \cap I_{j(s)} \neq \varnothing\}.$$

Take the unique injection $\sigma : S''' \to R_\delta$ such that $\sigma(s) \sim s$ for $s \in S'''$. By construction, $S := \sigma(S''')$ is a subset of R_δ and by Lemma 4.3.3 it is a B_h-set. Observe that S is a random set that depends on R_δ.

It remains to show that the counting function of S attains the claimed asymptotic growth with probability 1. We have that $S(n) \geq S'''(n) - 1$ for all $n \geq 1$. The -1 comes from the fact that the unique $s \in S'''$ with $s \sim \max(S \cap [n])$ could be larger than n. The counting function $S'''(n)$ satisfies

$$S'''(n) = \left| \{s \in S'' : j_0^{1/\delta} \leq s \leq n \text{ and } R_\delta \cap I_{j(s)} \neq \varnothing\} \right|,$$

so Lemma 4.3.1 and linearity of expectation give us

$$\begin{aligned}
\mathbb{E}(S'''(n)) &\geq \frac{1}{3}\left(S''(n) - S''\left(\left\lceil j_0^{1/\delta} \right\rceil - 1\right)\right) \\
&\geq \frac{1}{6} S'(n) - \frac{1}{3} S'\left(\left\lceil j_0^{1/\delta} \right\rceil - 1\right) \\
&= n^{u(\delta) - \varepsilon(n)}
\end{aligned}$$

for some $\varepsilon(n) = o(1)$. This together with Chernoff's inequality yields $n_0 \in \mathbb{N}$ such that for $n \geq n_0$,

$$\mathbb{P}\left(S'''(n) \leq \frac{1}{2} n^{u(\delta)-\varepsilon(n)}\right) \leq e^{-n^{u(\delta)-\varepsilon(n)}/8} \leq \frac{1}{n^2}. \tag{4.19}$$

Since $\sum_{n=1}^{\infty} \frac{1}{n^2} < \infty$, the Borel-Cantelli Lemma (see for example [64]) tells us that with probability 1 there exists $n_1 \in \mathbb{N}$ such that $S'''(n) \geq \frac{1}{2} n^{u(\delta)-\varepsilon(n)}$ for all $n \geq n_1$. Therefore, with probability 1,

$$S(n) \geq n^{u(\delta) - o(1)}$$

for all $n \in \mathbb{N}$, which completes the proof.

∎

4.4 Remarks and open questions

Our result on the density infinite α-strong B_h-sets (Theorem 1.4.6) was obtained by generalising the construction of a (non-strong) B_h-set which in turn generalised the construction of an infinite

Sidon set. It seems difficult to obtain an improvement on the density in the strong setting for B_h set without also improving on the density of the densest known infinite Sidon set.

As to the question of maximising the growth of infinite Sidon or B_h-sets in random infinite subsets of $\mathbb{Z}$, Theorem 1.4.6 provides the currently best lower bound for $h = 2$ and $\frac{5}{6} < \delta < 1$. In the case $h > 2$ no other bounds apart from Theorem 1.4.8 are known, though it seems likely that the results of Dellamonica-Kohayakawa-Lee-Rödl-Samotij [32, 33] on B_h-sets in random subsets of $[n]$ (see also [67] by the almost same set of authors) can be extended to the infinite setting to obtain exact exponents for the case $0 < \delta < \frac{h}{2h-1}$. We think that if δ is sufficiently large Theorem 1.4.8 yields the best lower bound that does not require new insights into the non-random, non-strong setting.

Kohayakawa et al. asked if their upper bound on the size of infinite α-strong Sidon sets (Theorem 6 in [66]) can be improved to an analogue of the bound (1.9) introduced by Erdős. We extend their question to the setting of α-strong B_h-sets.

Question 4.4.1. Let $h \geq 2$, and let $0 \leq \alpha < 1$ be a real. Does every α-strong B_h-set S satisfy

$$\liminf_{n\to\infty} \frac{S(n)}{n^{(1-\alpha)/h}} = 0?$$

Chapter 5

Splitting matchings

5.1 Non-realisable splits

We now prove Proposition 1.5.6.

Let $k \leq n$, and let $a_1, \dots, a_k \in \mathbb{N}_0$ such that $a_1 + \dots + a_k = n$. We proceed as follows: In the first step we show that if there exist pairwise distinct $x_1, \dots, x_k \in \mathbb{Z}_n$ such that $a_1x_1 + \dots + a_kx_k \neq \mathbf{0}$ then we can construct pairwise disjoint perfect matchings $M_1, \dots, M_k$ on a common vertex set of size $2n$ such that there exists no matching M with $|M \cap M_i| = a_i$ for $1 \leq i \leq k$. In the second step we prove that unless n is odd, $k = n$, and $a_1 = \dots = a_n = 1$, we can always find suitable $x_1, \dots, x_k$.

Given $x_1, \dots, x_k$ as above we define the desired matchings on two copies of $\mathbb{Z}_n$. To distinguish between the copies we write edges as ordered pairs, that is, we consider the matchings as subsets of $\mathbb{Z}_n^2$. For $i \in [k]$ let

$$M_i := \{(y, y + x_k) : y \in \mathbb{Z}_n\}.$$

Suppose there exists a perfect matching $M \subset \mathbb{Z}_n^2$ such that $|M \cap M_i| = a_i$ for all $i \in [k]$. We can now sum the elements of $\mathbb{Z}_n$ in two different ways to obtain

$$\sum_{y \in \mathbb{Z}_n} y \;=\; \sum_{y \in \mathbb{Z}_n} M(y)$$

where $M(y)$ is the unique vertex with $(y, M(y)) \in M$. This implies

$$\mathbf{0} \;=\; \sum_{y \in \mathbb{Z}_n} (M(y) - y) = a_1x_1 + \dots + a_kx_k,$$

which contradicts our choice of $x_1, \dots, x_k$.

As mentioned above a suitable choice of $x_1, \dots, x_k$ does not always exist. If n is odd, $k = n$, and $a_1 = \dots = a_n = 1$, we have $\{x_1, \dots, x_k\} = \mathbb{Z}_n$ and thus

$$a_1x_1 + \dots + a_kx_k = \sum_{x \in \mathbb{Z}_n} x = \mathbf{0}.$$

In the remaining cases, that is, if $n > k$, or n is even, or $\min_{i \in [k]} a_i = 0$ we can find $x_1, \dots, x_k$ as follows: Fix $i \in [k]$ with $a_i \neq 0$. If $k < n$, choose arbitrary pairwise distinct $x_1, \dots, x_k \in \mathbb{Z}_n \setminus \{\mathbf{0}\}$

such that $x_i = \mathbf{1}$. Then

$$(a_1x_1 + \ldots + a_kx_k) - (a_1x_1 + \ldots + a_{i-1}x_{i-1} + a_i \cdot \mathbf{0} + a_{i+1}x_{i+1} + \ldots + a_kx_k) = a_i \cdot \mathbf{1} \neq \mathbf{0}.$$

Therefore at least one of $a_1x_1 + \ldots + a_kx_k$ and $(a_1x_1 + \ldots + a_{i-1}x_{i-1} + a_i \cdot \mathbf{0} + a_{i+1}x_{i+1} + \ldots + a_kx_k)$ is not zero. If $k = n$ and $\min\{a_1, \ldots, a_k\} = 0$ take $j \in [k]$ such that $a_j = 0$. Choose arbitrary pairwise distinct $x_\ell \in \mathbb{Z}_n \setminus \{\mathbf{0}, \mathbf{1}\}$ for $\ell \in [k] \setminus \{i, j\}$. We have that

$$\mathbf{1} \cdot a_i + \mathbf{0} \cdot a_j + \sum_{\ell \in [k] \setminus \{i,j\}} a_\ell x_\ell \neq \mathbf{0} \cdot a_i + \mathbf{1} \cdot a_j + \sum_{\ell \in [k] \setminus \{i,j\}} a_\ell x_\ell,$$

so at least one of the choices $(x_i, x_j) = (\mathbf{1}, \mathbf{0})$ and $(x_i, x_j) = (\mathbf{0}, \mathbf{1})$ leads to $a_1x_1 + \ldots + a_kx_k \neq \mathbf{0}$. In the case that $k = n$, $\min\{a_1, \ldots, a_k\} = 1$, and n is even there is nothing to choose because $a_1 = \ldots = a_n = 1$ and $\{x_1, \ldots, x_n\} = \mathbb{Z}_n$ whenever $x_1, \ldots, x_n$ are pairwise distinct. So,

$$a_1x_1 + \ldots + a_kx_k = \sum_{x \in \mathbb{Z}} x = \frac{n}{2} \cdot \mathbf{1} \neq \mathbf{0}.$$

■

5.2 Almost arbitrary splits of three matchings

In this section we give the proof Theorem 1.5.7.

We say that a matching $M \subset E(G)$ is *distributed as* (a_1, a_2, a_3) if it satisfies $|M \cap M_1| = a_1$, $|M \cap M_2| = a_2$, and $|M \cap M_3| = a_3$. It suffices to prove the claim for triples (a_1, a_2, a_3) with $a_1 = \max\{a_1, a_2, a_3\}$ as the roles of the matchings are interchangeable. We will show that given an M that is distributed as (a_1, a_2, a_3) with $a_1 + a_2 + a_3 = n - 2$ we can find a matching M' that is distributed as $(a_1 - 1, a_2 + 1, a_3)$. This also implies the existence of matchings distributed as $(a_1 - 1, a_2, a_3 + 1)$. Starting from M_1 minus two arbitrary edges we can then find a matching distributed as (a_1, a_2, a_3) for any such triple satisfying $a_1 + a_2 + a_3 = n - 2$.

For any matching $M \subset E(G)$ of size $n - 2$ and any vertex x that is unmatched by M, let $P_{23}(M, x)$ be the maximum $(M_2 \setminus M)$-$(M_3 \cap M)$-alternating path starting at x, and let $\ell_{23}(M, x)$ be its length. Let

$$\ell_{23}(M) := \min_{x \text{ unmatched by } M} \ell_{23}(M, x).$$

For a matching M of G and $v \in V(G)$, denote by $M(v)$ the vertex u that is matched by M to v i.e. $M(v) = u$ if and only if $\{v, u\} \in M$. Choose M such that $\ell_{23}(M)$ is minimised over all matchings that are distributed as (a_1, a_2, a_3). Pick an unmatched vertex x with $\ell_{23}(M, x) = \ell_{23}(M)$ and an unmatched vertex z that is distinct from the endpoints of $P_{23}(M, x)$ and from $M_3(x)$. We can choose such vertices because there are four unmatched vertices in total. If $M_2(x)$ is incident to an edge of $M \cap M_1$ or unmatched we are done since in the former case the matching

$$M \setminus \{M_2(x)M_1(M_2(x))\} \cup \{xM_2(x)\}$$

is distributed as $(a_1 - 1, a_2 + 1, a_3)$ while in the latter we can pick

$$M \setminus \{e\} \cup \{xM_2(x)\}$$

for any $e \in M \cap M_1$. Hence we assume that $M_2(x)$ is incident to an edge of $M \cap M_3$. Now $M_3(z)$ cannot be incident to an edge of $M \cap M_2$ because

$$M' := M \setminus \{M_2(x)M_3(M_2(x)), M_3(z)M_2(M_3(z))\} \cup \{xM_2(x), zM_3(z)\}$$

would be a matching that is distributed as (a_1, a_2, a_3) and in which $P_{23}(M', M_3(M_2(x)))$ would be a path of length $\ell_{23}(M,x) - 2$, which contradicts our choice of M. Here it was important that z is different from the endpoints of $P_{23}(M,x)$ so $P_{23}(M', M_3(M_2(x)))$ is a subpath of $P_{23}(M,x)$ not containing x and therefore $P_{23}(M', M_3(M_2(x)))$ has smaller length than $P_{23}(M,x)$. Therefore $M_3(z)$ is unmatched or incident to an edge of $M \cap M_1$. If $M_3(z)$ is incident to $M \cap M_1$ then

$$M'' := M \setminus \{M_2(x)M_3(M_2(x)), M_3(z)M_1(M_3(z))\} \cup \{xM_2(x), zM_3(z)\}$$

is the desired matching. Should $M_3(z)$ be unmatched then for any $e \in M \cap M_1$,

$$M''' := M \setminus \{M_2(x)M_3(M_2(x)), e\} \cup \{xM_2(x), zM_3(z)\}$$

is distributed as $(a_1 - 1, a_2 + 1, a_3)$. Here we used that $M_3(x) \neq z$, or equivalently that $M_3(z) \neq x$. So under the previous assumption that $M_2(x)$ is incident to an edge in $M \cap M_3$, we have that the edges $xM_2(x), zM_3(z)$ are disjoint. Hence M'' and M''' are indeed matchings of G.

■

5.3 Further directions of research

The bound $n-2$ in Theorem 1.5.7 cannot be improved to $n-1$ without further assumptions. For any even $n \geq 4$ consider the unique (up to relabelling) decomposition of $n/2$ disjoint copies of K_4 into three perfect matchings M_1, M_2, and M_3. We depict the situation in Figure 5.1. The only matchings of size two in a copy K of K_4 are $M_i \cap E(K)$ for $i \in [3]$. Therefore for each matching $M \subset M_1 \cup M_2 \cup M_3$ we have that if $|M \cap M_i|$ is odd for some $i \in [3]$ there exists a copy K of K_4 such that $M \cap E(K) \subset M_i$ and $|M \cap E(K)| = 1$. This implies $|M| \leq n - |\{i \in [3] : |M \cap M_i| \text{ odd}\}|$. Now we can see that if $a_1 + a_2 + a_3 = n - 1$ and at least two of the a_i are odd (and hence all of them are odd by parity) we cannot find any matching M satisfying $|M \cap M_i| = a_i$ for $i \in [3]$.

We conjecture that this is the only obstruction.

Conjecture 5.3.1. Let G be a graph on $2n$ vertices whose edge set is decomposed into three perfect matchings M_1, M_2, M_3. Let a_1, a_2, a_3 be non-negative integers such that $a_1 + a_2 + a_3 = n - 1$. If G has a component that is not isomorphic to K_4, or at least one of a_1, a_2, a_3 is even then there exists a matching M in G such that $|M \cap M_i| = a_i$ for $i \in [3]$.

FIGURE 5.1: A union of two copies of K_4 in the case $n = 4$ with the matchings indicated by colours. It is not possible to find a split with multiplicities $(a_1, a_2, a_3) = (1, 1, 1)$ because in each K_4 any two edges of distinct colours intersect.

A simple parity argument shows that if n is odd then G has a component that is not isomorphic to K_4. Another way to exclude any K_4 in G is to assume that G is bipartite. Conjecture 5.3.1 would resolve the case $k = 3$ of Question 1.5.1 since then the always realisable triples (a_1, a_2, a_3) would be those for which either $a_1 + a_2 + a_3 \leq n - 2$, or $\max\{a_1, a_2, a_3\} = n$, or $a_1 + a_2 + a_3 = n - 1$ and one of the a_i is even.

As to fair splits, an immediate consequence of Theorem 1.5.7 is that for any initial matchings M_1, M_2, M_3 we can find a split with $|M \cap M_i| = \left\lfloor \frac{n-2}{3} \right\rfloor \geq \frac{n}{3} - 1$. Proposition 1.5.6 shows that even if n is divisible by 3 a split with multiplicities $(n/3, n/3, n/3)$ is not possible in general. This motivates the problem of how close to a perfect split one can always get. Arman, Rödl, and Sales note in the concluding section of their article that by a modification of their results one can show that given positive rational numbers $\alpha_1, \ldots, \alpha_k$ with $\sum_{i=1}^{k} \alpha_i \leq 1$ there exists an integer κ such that one can always find a split M with $|M \cap M_i| \geq \alpha_i n - \kappa$ for $i \in [k]$ whenever n is sufficiently large. In particuar this holds in the case $\alpha_1 = \ldots = \alpha_k = \frac{1}{k}$, and leads us to the following question.

Question 5.3.2. Given $k \in \mathbb{N}$ what is the smallest integer c_k such that for every graph G on $2n$ vertices whose edge set is the union of k pairwise edge-disjoint perfect matchings $M_1, \ldots, M_k$ there exists a matching M in G satisfying $|M \cap M_i| \geq \frac{n}{k} - c_k$ for each $i \in [k]$?

By an application of Alon's Necklace Splitting Theorem (cf. [8, 9]) similar to the one in [11] we can show that $c_k \leq 4k - 6$ for all k. However, a proof of that estimate is not included here for it does not provide anything substantially new compared to the application of Necklace Splitting just mentioned. Moreover we conjecture that $c_k = 1$ and in fact the following more general statement holds.

Conjecture 5.3.3. Let G be a complete bipartite graph on $2n$ vertices whose edge set is decomposed into n perfect matchings $M_1, \ldots, M_n$. If $(a_1, \ldots, a_n)$ is a tuple of non-negative integers such that $a_1 + \ldots + a_n = n - 1$ then there exists a matching M in G such that $|M \cap M_i| = a_i$ for each $i \in [n]$.

This conjecture was independently formulated by Noga Alon as a question [6]. It encompasses the situation of Question 1.5.1 when $k < n$. Indeed, by König's Theorem or alternatively Hall's Theorem (see [34] for an exposition of these two theorems) any three pairwise disjoint perfect matchings between two sets of size n can be extended to n pairwise disjoint matchings between the same vertex sets. Thus, Question 1.5.1 for $k < n$ and multiplicities $(a_1, \ldots, a_k)$ is equivalent to the same question for $k = n$ and the n-tuple $(a_1, \ldots, a_k, 0, \ldots, 0)$.

Conjecture 5.3.3 is very optimistic as it implies the Ryser-Brualdi-Stein conjecture by setting $a_1 = \ldots = a_{n-1} = 1$ and $a_n = 0$. It is also related to the Aharoni-Berger conjecture [82, 31] and other generalisations of the Ryser-Brualdi-Stein conjecture such as Conjecture 1.9 in [3]. A similar line of study for fair splits that has a geometric flavour was initiated in [21].

The counterexample in Proposition 1.5.6 comes from addition tables of Abelian groups. As shown by Hall [55] and, independently, Salzborn and Szekeres [91] there cannot be such a counterexample that falsifies Conjecture 5.3.3. A modern exposition of their results is given in [99]. Another direction of research would be to generalise these results to the setting of non-commutative groups. Such a generalisation would give further evidence for Conjecture 5.3.3.

Lastly, it would be interesting to understand an unordered variant of Question 1.5.1 where the multiplicities a_i for $i \in [k]$ and the intersections $|M \cap M_i|$ for $i \in [k]$ are equal as multisets.

Question 5.3.4. Let G be a graph on $2n$ vertices whose edge set is the union of k pairwise disjoint perfect matchings $M_1, \ldots, M_k$. For which tuples $(a_1, \ldots, a_k)$ of non-negative integers with $a_1 + \ldots + a_k \leq n$ can we always find a new matching M in G and a permutation $\pi : [k] \to [k]$ such that $|M \cap M_i| \geq a_{\pi(i)}$ for all $i \in [k]$?

List of Symbols

$\lceil \cdot \rceil$	ceiling function
$\lfloor \cdot \rfloor$	floor function
$\mathbf{0}$, $\mathbf{1}$, $\mathbf{2}$	Elements of a cyclic group corresponding to the integers $0, 1, 2$
$(a_1, \dots, a_k)$	multiplicity tuple for splitting matchings
A_t	set of integers describing differences occuring during the C_k-process on P_n
A'_t	subset of A_t
$Aut(G)$	set of automorphisms of G
$[a,b]_1$	odd integers between a and b
$[a,b]$	set of integers between a and b
$A+B$	Minkowki sum of the sets A and B
c_1, c_2, $\dots$	positive constants
$c(G)$	number of components of G
C_k	cycle on k vertices
C'	copy of a k-cycle
$\mathscr{C}(G,x)$	component of G containing x
D_i	set of differences occuring in the C_k-process on P_n
$\delta(H)$	minimum degree of H
$\Delta(H)$	maximum degree of H
$\mathrm{diam}(G)$	diameter of G
$\mathrm{dist}_G(x,y)$	distance between vertices x, y in G
e,f, e'	edges of a graph
$e(H)$	number of edges of H
$E(H)$	edge set of H
$E_G(X,Y)$	edges between X and Y in G
F	a forest
$F(x,y)$	Frobenius number of two integers x, y
$F'(x,y)$	Frobenius number of even integers x, y with greates common divisor 2
$G, G', \tilde{G}$	graphs on which we run a bootstrap process
$G^{(1)}, \dots, G^{(s)}$	connected components of G
G^a	underlying graph of an H-chain indexed by a
$G[\setminus U]$	induced subgraph of G on the vertex set $V(G) \setminus U$
$\mathscr{G}$	an auxiliary graph
$(G_t)_{t \geq 0}$	bootstrap process on G
$\langle G \rangle_H$	final graph of the H-bootstrap process on G

hA	h-fold sum set of A
$h \cdot A$	dilate of A
H	the graph that is replicated during a bootstrap process
$H[U]$	induced subgraph of H on the vertex set U.
$H_1 \sqcup H_2$	disjoint union of H_1 and H_2
H'	a subgraph of H or a copy of H minus an edge
$\tilde{H}$	a subgraph of H
H^-	graph obtained by removing an edge from the edge-transitive graph H
$H - v$	graph obtained by removing the vertex v from H
H_e	graph obtained by removing the edge e from H
$\mathscr{H}$, $\bar{\mathscr{H}}$, $\mathscr{H}^a$	H-chains.
$\mathrm{ht}_z(T)$	height of T with respect to the root z
$\mathrm{ht}_{\mathbf{z}}(F)$	height of the forest F with respect to the tuple of roots $\mathbf{z}$
$\mathrm{Hom}(H,G)$	set of graph homomorphisms from H to G
i, j	integers used for indexing purposes
κ	a constant integer
k	a natural number, mostly used as a parameter for a family of sets or graphs
K_n	complete graph on n vertices
$K_{s,t}$	complete bipartite graph with partite sets of sizes s and t
ℓ	natural number that usually indicates the length of a path or a chain
$\ell_{23}(M,x)$	length of the path $P_{23}(M,x)$
$\ell_{23}(M)$	smallest value of $\ell_{23}(M,x)$ over all vertices unmatched by M
λ, μ	constant integers
$M_1, M_2, \ldots$	initial matchings in the context of splitting matchings
M, M', M'', M'''	matchings intersecting the initial matchings with given multiplicities
$M_H(n)$	maximum running of the H-bootstrap process
n	number of vertices in the starting graph
$[n]$	the set of the first n natural numbers
$[n]_1$	odd integers among the first n positive integers
$n_H(G)$	number of labelled copies of H in G
$N_G(x)$	neighbourhood of the vertex x in G
$\mathbb{N}$	set of natural numbers (excluding zero)
$\mathbb{N}_0$	set of non-negative integers
O, Θ, o, ω	asymptotic notation
P,Q	paths in a graph
P_n	path on n vertices
$P_{23}(M,x)$	maximum $(M_2 \setminus M)$-$(M_3 \cap M)$-alternating path starting at x
P^Δ	graph constructed to prove the presented lower bound on $M_{C_k}(n)$
$(P^{\Delta,t})_{t \geq 0}$	C_k-process on P^Δ
ϕ, φ, ψ	graph homomorphisms
Q_3	three-dimensional hypercube
r	parameter indicating the maximum running time for cycles; a natural number

ρ	alternate time variable if t is already in use; an automorphism of H
s	a natural number, mostly used to define ranges of indices
S	a B_h-set
$S(n)$	counting function of the set S
σ	an automorphism of H
τ	a natural number used to indicate a fixed time during a bootstrap proces
$\tau_H(G)$	running time of the H-bootstrap process on G.
t	a non-negative integer; used as time variable for bootstrap processes
T	a tree; a set of times in a bootstrap process
$T_0, T_1, T_2, \ldots$	copies of a forest T
$T^{(x \to y)}$	copy of T obtained by replacing $x \in V(T)$ by $y \notin V(T)$
u, v, w	vertices
U, V, W	sets of vertices
$U_j, \tilde{U}_j$	vertex sets indexed by j
$\mathscr{U}$	union of U_j for certain indices j
$v(H)$	number of vertices of H
$V(H)$	vertex set of H
$\mathscr{V}(G,x)$	vertex set of the component of G containing x
W_k	wheel graph on $k+1$ vertices; vertex set indexed by k
$\tilde{W}_k$	vertex set indexed by k
x, y	vertices of a graph; integers
X, Y	subsets of the vertices of a graph (usually partite sets of a biparite graph)
$X \times Y$	Cartesian product of X and Y
$\mathbb{Z}$	set of integers
$\mathbb{Z}_k$	additive group of integers modulo k

Bibliography

[1] J. Adler and U. Lev. "Bootstrap Percolation: visualizations and applications". *Brazilian Journal of Physics* 33 (2003), pp. 641–644.

[2] R. Aharoni, N. Alon, and E. Berger. "Eigenvalues of K1,k-Free Graphs and the Connectivity of Their Independence Complexes". *Journal of Graph Theory* 83 (2016), pp. 384–391.

[3] R. Aharoni, N. Alon, E. Berger, M. Chudnovsky, D. Kotlar, M. Loebl, and R. Ziv. "Fair representation by independent sets". *A Journey Through Discrete Mathematics: A Tribute to Jiří Matoušek* (2017), pp. 31–58.

[4] M. Aigner and G. M. Ziegler. *Proofs from THE BOOK*. Springer Berlin Heidelberg, 2010.

[5] J. L. R. Alfonsín. *The diophantine Frobenius problem*. OUP Oxford.

[6] N. Alon. *Personal Communication*.

[7] N. Alon. "An extremal problem for sets with applications to graph theory". *Journal of Combinatorial Theory, Series A* 40 (1985), pp. 82–89.

[8] N. Alon. "Splitting necklaces". *Advances in Mathematics* 63 (1987), pp. 247–253.

[9] N. Alon, D. Moshkovitz, and S. Safra. "Algorithmic construction of sets for k-restrictions". *ACM Transactions on Algorithms* 2 (2006), pp. 153–177.

[10] M. Anastos, D. Fabian, A. Müyesser, and T. Szabó. "Splitting matchings and the Ryser-Brualdi-Stein conjecture for multisets". *The Electronic Journal of Combinatorics* (2023).

[11] A. Arman, V. Rödl, and M. T. Sales. "Colorful Matchings". *SIAM Journal on Discrete Mathematics* 37 (2023), pp. 925–950.

[12] P. Balister, B. Bollobás, and P. Smith. "The time of bootstrap percolation in two dimensions". *Probability Theory and Related Fields* 166 (2016), pp. 321–364.

[13] J. Balogh, G. Kronenberg, A. Pokrovskiy, and T. Szabó. "The maximum length of K_r-Bootstrap Percolation". *Proceedings of the American Mathematical Society* (To appear).

[14] J. Balogh, B. Bollobás, H. Duminil-Copin, and R. Morris. "The sharp threshold for bootstrap percolation in all dimensions". *Transactions of the American Mathematical Society* 364 (2011), pp. 2667–2701.

[15] J. Balogh, B. Bollobás, and R. Morris. "Graph bootstrap percolation". *Random Structures & Algorithms* 41 (2012), pp. 413–440.

[16] J. Balogh, B. Bollobás, R. Morris, and O. Riordan. "Linear algebra and bootstrap percolation". *Journal of Combinatorial Theory, Series A* 119 (2012), pp. 1328–1335.

[17] E. Bayraktar and S. Chakraborty. "$K_{r,s}$ graph bootstrap percolation". *Preprint* (2019), arXiv:1904.12764.

[18] F. A. Behrend. "On sets of integers which contain no three terms in arithmetical progression". *Proceedings of the National Academy of Sciences* 32 (1946), pp. 331–332.

[19] F. Benevides and M. Przykucki. "Maximum Percolation Time in Two-Dimensional Bootstrap Percolation". *SIAM Journal on Discrete Mathematics* 29 (2015), pp. 224–251.

[20] M. Bidgoli, A. Mohammadian, and B. Tayfeh-Rezaie. "On $K_{2,t}$-Bootstrap Percolation". *Graphs and Combinatorics* 37 (2021), pp. 731–741.

[21] A. Black, U. Cetin, F. Frick, A. Pacun, and L. Setiabrata. "Fair splittings by independent sets in sparse graphs". *Israel Journal of Mathematics* 236 (2020), pp. 603–627.

[22] T. F. Bloom and O. Sisask. "The Kelley–Meka bounds for sets free of three-term arithmetic progressions". *Essential Number Theory* 2 (2023), pp. 15–44.

[23] B. Bollobás. "Weakly k-saturated graphs". In: *Beiträge zur Graphentheorie (Kolloquium, Manebach)*. Vol. 25. 1968, pp. 25–31.

[24] B. Bollobás. *Random Graphs*. Second edition. Cambridge University Press, 2001.

[25] B. Bollobás, M. Przykucki, O. Riordan, and J. Sahasrabudhe. "On the Maximum Running Time in Graph Bootstrap Percolation". *The Electronic Journal of Combinatorics* 24 (2017), P2.16.

[26] S. R. Broadbent and J. M. Hammersley. "Percolation processes: I. Crystals and mazes". *Mathematical Proceedings of the Cambridge Philosophical Society* 53 (1957), pp. 629–641.

[27] R. A. Brualdi and H. J. Ryser. *Combinatorial matrix theory*. Cambridge University Press, 1991.

[28] J Chalupa, P. L. Leath, and G. R. Reich. "Bootstrap percolation on a Bethe lattice". *Journal of Physics C: Solid State Physics* 12 (1979), pp. L31–L35.

[29] J. Cilleruelo. "Infinite sidon sequences". *Advances in Mathematics* 255 (2014), pp. 474–486.

[30] J. Cilleruelo and C. Timmons. "k-Fold Sidon Sets". *The Electronic Journal of Combinatorics* 21 (2014).

[31] D. M. Correia, A. Pokrovskiy, and B. Sudakov. "Short Proofs of Rainbow Matchings Results". *International Mathematics Research Notices* 2023 (2023), pp. 12441–12476.

[32] D. Dellamonica, Y. Kohayakawa, S. J. Lee, V. Rödl, and W. Samotij. "On the Number of B_h -Sets". *Combinatorics, Probability and Computing* 25 (2016), pp. 108–129.

[33] D. Dellamonica, Y. Kohayakawa, S. J. Lee, V. Rödl, and W. Samotij. "The number of Bh-sets of a given cardinality". *Proceedings of the London Mathematical Society* 116 (2018), pp. 629–669.

[34] R. Diestel. *Graph Theory*. Springer Berlin Heidelberg, 2017.

[35] S. Eberhard and F. Manners. "The apparent structure of dense Sidon sets". *Preprint* (2022), arXiv:2107.05744.

[36] M. Elkin. "An Improved Construction of Progression-Free Sets". In: *Proceedings of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms*. Society for Industrial and Applied Mathematics, 2010, pp. 886–905.

[37] P. Erdős and M. Simonovits. "A limit theorem in graph theory". In: *Studia Scientiarum Mathematicarum Hungarica 1*. 1966.

[38] P Erdős and M Simonovits. “Some extremal problems in graph theory”. In: *Combinatorial Theory and its Applications I*. North-Holland Amsterdam, 1970, pp. 378–392.

[39] P. Erdős and A. H. Stone. “On the structure of linear graphs”. *Bulletin of the American Mathematical Society* 52 (1946), pp. 1087–1091.

[40] P. Erdos and P. Turán. “On a problem of Sidon in additive number theory, and on some related problems”. *J. London Math. Soc* 16 (1941), pp. 212–215.

[41] A. Espuny Díaz, B. Janzer, G. Kronenberg, and J. Lada. “Long running times for hypergraph bootstrap percolation”. *European Journal of Combinatorics* 115 (2024).

[42] D. Fabian, P. Morris, and T. Szabó. “Slow graph boostrap percolation I: Cycles”. *Preprint* (2023), arXiv:2308.00498.

[43] D. Fabian, P. Morris, and T. Szabó. “Slow graph boostrap percolation II: Accelerating properties”. *Preprint* (2023), arXiv:2311.18786.

[44] D. Fabian, P. Morris, and T. Szabó. “Slow graph boostrap percolation III: Chain constructions” (2023). In preparation.

[45] D. Fabian, P. Morris, and T. Szabó. “Maximum running times for graph bootstrap percolation processes”. *Discrete Mathematics Days 2022* 263 (2022), pp. 134–139.

[46] D. Fabian, J. Rué, and C. Spiegel. “On strong infinite Sidon and Bh sets and random sets of integers”. *Journal of Combinatorial Theory, Series A* 182 (2021).

[47] J. R. Faudree, R. J. Faudree, and J. R. Schmitt. “A survey of minimum saturated graphs”. *The Electronic Journal of Combinatorics* 1000 (2011), DS19.

[48] J. Fox. “A new proof of the graph removal lemma”. *Annals of Mathematics* 174 (2011), pp. 561–579.

[49] P. Frankl. “An Extremal Problem for two Families of Sets”. *European Journal of Combinatorics* 3 (1982), pp. 125–127.

[50] A. Frieze and M. Karoński. *Introduction to random graphs*. Cambridge University Press.

[51] Z. Füredi and M. Simonovits. “The history of degenerate (bipartite) extremal graph problems”. In: *Erdős Centennial*. Springer, 2013, pp. 169–264.

[52] H.-O. Georgii. *Stochastics: Introduction to Probability and Statistics*. De Gruyter, 2012.

[53] B. Green and J. Wolf. “A Note on Elkin’s Improvement of Behrend’s Construction”. In: *Additive Number Theory*. Springer New York, 2010, pp. 141–144.

[54] K. Gunderson, S. Koch, and M. Przykucki. “The time of graph bootstrap percolation”. *Random Structures & Algorithms* 51 (2017), pp. 143–168.

[55] M. Hall. “A combinatorial problem on abelian groups”. *Proceedings of the American Mathematical Society* 3 (1952), pp. 584–587.

[56] G. H. Hardy and E. M. Wright. *An Introduction to the Theory of Numbers*. Sixth edition. Oxford University Press, 2008.

[57] I. Hartarsky and L. Lichev. “The Maximal Running Time of Hypergraph Bootstrap Percolation”. *SIAM Journal on Discrete Mathematics* 38 (2024), pp. 1462–1471.

[58] S. Janson, T. Łuczak, T. Turova, and T. Vallier. “Bootstrap percolation on the random graph $G_{n,p}$”. *The Annals of Applied Probability* 22 (2012).

[59] G. Kalai. “Weakly Saturated Graphs are Rigid”. In: *North-Holland Mathematics Studies*. Vol. 87. Elsevier, 1984, pp. 189–190.

[60] P. Keevash, A. Pokrovskiy, B. Sudakov, and L. Yepremyan. “New bounds for Ryser’s conjecture and related problems”. *Transactions of the American Mathematical Society, Series B* 9 (2022), pp. 288–321.

[61] Z. Kelley and R. Meka. “Strong Bounds for 3-Progressions”. In: *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*. 2023, pp. 933–973.

[62] P Kővári, V. T Sós, and P. Turán. “On a problem of Zarankiewicz”. In: *Colloquium Mathematicum*. Vol. 3. Polska Akademia Nauk, 1954, pp. 50–57.

[63] J. H. Kim, B. Sudakov, and V. H. Vu. “On the asymmetry of random regular graphs and random graphs”. *Random Structures & Algorithms* 21 (2002), pp. 216–224.

[64] A. Klenke. *Wahrscheinlichkeitstheorie*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2020.

[65] Y. Kohayakawa, S. J. Lee, C. G. Moreira, and V. Rödl. “Infinite Sidon Sets Contained in Sparse Random Sets of Integers”. *SIAM Journal on Discrete Mathematics* 32 (2018), pp. 410–449.

[66] Y. Kohayakawa, S. J. Lee, C. G. Moreira, and V. Rödl. “On strong Sidon sets of integers”. *Journal of Combinatorial Theory, Series A* 183 (2021).

[67] Y. Kohayakawa, S. J. Lee, V. Rödl, and W. Samotij. “The number of Sidon sets and the maximum size of Sidon sets contained in a sparse random set of integers”. *Random Structures & Algorithms* 46 (2015), pp. 1–25.

[68] B. Kolesnik. “The sharp K_4-percolation threshold on the Erdős–Rényi random graph”. *Electronic Journal of Probability* 27 (2022).

[69] J. Kollár, L. Rónyai, and T. Szabó. “Norm-graphs and bipartite turán numbers”. *Combinatorica* 16 (1996), pp. 399–406.

[70] J. Komlós, A. Shokoufandeh, M. Simonovits, and E. Szemerédi. “The regularity lemma and its applications in graph theory”. *Theoretical Aspects of Computer Science: Advanced Lectures* (2002), pp. 84–112.

[71] J. Komlós and M. Simonovits. “Szemeredi”s Regularity Lemma and its applications in graph theory” (1995).

[72] F. Lazebnik, V. A. Ustimenko, and A. J. Woldar. “A new series of dense graphs of high girth”. *Bulletin of the American mathematical society* 32 (1995), pp. 73–79.

[73] F. Lazebnik and J. Verstraëte. “On hypergraphs of girth five”. *The Electronic Journal of Combinatorics* 10 (2003).

[74] F. Lazebnik and R. Viglione. “On the connectivity of certain graphs of high girth”. *Discrete Mathematics* 277 (2004), pp. 309–319.

[75] W. Mantel. “"Problem 28 (Solution by H. Gouwentak, W. Mantel, J. Teixeira de Mattes, F. Schuh and W. A. Wythoff)"”. *Wiskundige Opgaven* (1907), pp. 60–61.

[76] K. Matzke. “The saturation time of graph bootstrap percolation”. *Preprint* (2015), arXiv:1510.06156.

[77] R. Morris. “Bootstrap percolation, and other automata”. *European Journal of Combinatorics* 66 (2017), pp. 250–263.

[78] R. Motwani and P. Raghavan. *Randomized Algorithms*. First edition. Cambridge University Press, 1995.

[79] J. von Neumann. “Theory of self-reproducing automata”. *Edited by Arthur W. Burks, University of Illinois Press* (1966).

[80] J. A. Noel and A. Ranganathan. “On the Running Time of Hypergraph Bootstrap Percolation”. *Preprint* (2022), arXiv:2206.02940.

[81] K. O’Bryant. “A Complete Annotated Bibliography of Work Related to Sidon Sequences”. *The Electronic Journal of Combinatorics* 1000 (2004), DS11.

[82] A. Pokrovskiy. “An approximate version of a conjecture of Aharoni and Berger”. *Advances in Mathematics* 333 (2018), pp. 1197–1241.

[83] M. Przykucki. “Maximal Percolation Time in Hypercubes Under 2-Bootstrap Percolation”. *The Electronic Journal of Combinatorics* 19 (2012), P41.

[84] K. F. Roth. “On certain sets of integers”. *J. London Math. Soc* 28 (1953).

[85] I. Ruzsa. “Solving a linear equation in a set of integers I”. *Acta Arithmetica* 65 (1993).

[86] I. Z. Ruzsa. “An infinite Sidon sequence”. *Journal of Number Theory* 68 (1998), pp. 63–71.

[87] I. Z. Ruzsa and E. Szemerédi. “Triple systems with no six points carrying three triangles”. *Combinatorics (Keszthely, 1976), Coll. Math. Soc. J. Bolyai* 18 (1978), pp. 939–945.

[88] H. J. Ryser. “Neuere Probleme der Kombinatorik”. *Vorträge über Kombinatorik, Oberwolfach* 69 (1967), p. 91.

[89] A. A. Saberi. “Recent advances in percolation theory and its applications”. *Physics Reports* 578 (2015), pp. 1–32.

[90] M. Sahimi. *Applications of percolation theory*. Springer Nature, 2023.

[91] F. Salzborn and G. Szekeres. “A problem in combinatorial group theory”. *Ars Combinatoria* 7 (1979), pp. 3–5.

[92] T. Sanders. “On Roth’s theorem on progressions”. *Annals of Mathematics* 174 (2011), pp. 619–636.

[93] S Sidon. “Bemerkungen über Fourier-und Potenzreihen”. *Acta Univ. Szeged Sect. Sci. Math* 7 (1935).

[94] C. Spiegel. “Additive structures and randomness in combinatorics”. 2020.

[95] D. Stauffer and A. Aharony. *Introduction to percolation theory*. CRC press, 2018.

[96] S. K. Stein. “Transversals of Latin squares and their generalizations.” *Pacific Journal of Mathematics* 59 (1975), pp. 567–575.

[97] A. Stöhr. “Gelöste und ungelöste Fragen über Basen der natürlichen Zahlenreihe. I.” *Journal für die reine und angewandte Mathematik* 194 (1955), pp. 40–65.

[98] S. Ulam and others. “Random processes and transformations”. In: *Proceedings of the International Congress on Mathematics*. Vol. 2. Citeseer, 1952, pp. 264–275.

[99] D. H. Ullman and D. J. Velleman. “Differences of Bijections”. *The American Mathematical Monthly* 126 (2019), pp. 199–216.