11-29-2012

On Integer Sequences, Packings and Games on Graphs

Ago-Erik Riet

ON INTEGER SEQUENCES, PACKINGS AND GAMES ON GRAPHS

by

Ago-Erik Riet

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Mathematical Sciences

The University of Memphis

December, 2012

To my mother Ene and my brother Ivo

ACKNOWLEDGEMENTS

I am especially grateful to my supervisor, Professor Béla Bollobás for all his guidance and encouragement. His criticism has encouraged me to better myself in many aspects of life and mathematics. I would like to thank him for all the opportunities he provided me, for giving me many elegant and stimulating mathematical problems, introducing me to brilliant mathematicians and taking me to exciting places.

My examiners, Professors Paul Balister, Anna Kaminska, Vladimir Nikiforov and Cecil Rousseau, have given me their helpful comments and suggestions for which I am thankful.

I would like to thank my co-authors Fabricio Benevides, Zoltán Füredi, Jonathan Hulgan, Steven Kalikow, Jonathan Lee, Nathan Lemons, Cory Palmer, Mykhaylo Tyomkyn and Jeffrey P. Wheeler for their collaboration on some of the problems of my dissertation.

I am thankful to my colleagues who have helped, supported and encouraged me in so many ways: Alex Mezei, Andrew Uzzell, Dominik Vu, Fabricio Benevides, Karen Johannson, Neal Bushaw, Ping Kittipassorn, Richard Johnson, Scott Binski, Tomas Juskevicius, Vivek Shandilya, and many others. I am especially grateful to Richard Johnson for reading parts of my dissertation and his advice.

Mrs. Tricia Simmons has helped me in countless ways with everyday things and to learn about the American way of life. Without her help, often beyond the call of duty, things would not have gone half as smoothly. Mrs. Gabriella Bollobás has shown me that art is an essential part of life, and by extension, of mathematics.

I would like to say a special thanks to my family members, especially my mother Ene Riet and my brother Ivo Riet who have supported me unconditionally before and during my PhD studies.

I would like to thank the Ministry of Education and Science of Estonia, the Archimedes Foundation of Estonia and the University of Memphis for their financial support which made writing my PhD possible.

ABSTRACT

Ago-Erik Riet. Ph.D. The University of Memphis. December, 2012. On Integer Sequences, Packings and Games on Graphs. Major Professor: Béla Bollobás.

This dissertation concerns four problems in combinatorics.

In Chapter 2 we consider the Prolonger-Shortener game of $\mathcal{F}$ saturation that was introduced by Füredi, Reimer and Seress: Players take turns drawing edges on an initially edgeless vertex set of size n with the restriction that they do not complete a copy of a graph in $\mathcal{F}$. The game ends when no more edges can be drawn. Prolonger wants as many edges as possible at the end of the game and Shortener as few as possible. We ask what is the final number of edges with both players playing optimally when $\mathcal{F}$ is a fixed path, or a collection of trees. We also consider a directed version of the game.

In Chapter 3 we address a question about completing partial packings of copies of a bipartite graph H, asked by Füredi and Lehel. An H-design on n vertices is an edge-disjoint collection of copies of H whose edge sets partition the edge set of the complete graph on n vertices. Given a bipartite graph H and an integer n, let $f(n; H)$ be the smallest integer such that any set of edge disjoint copies of H on n vertices can be extended to an H-design on at most $n + f(n; H)$ vertices. We establish tight bounds for the growth of $f(n; H)$ as $n \to \infty$. In particular, we prove the conjecture of Füredi and Lehel that $f(n; H) = o(n)$.

Chapter 4 is dedicated to a particular integer sequence, the Slowgrow sequence, originally introduced by Steven Kalikow. It starts with 1, and having defined terms $s_1, \ldots, s_n$ the term s_{n+1} is the smallest positive integer such that the block $s_{n-s_{n+1}+2} \ldots s_{n+1}$ has not occurred in the sequence earlier. Our main result is that blocks which can potentially occur multiple times in the sequence actually occur infinitely often. We also prove bounds on the time of the first occurrence of n in the Slowgrow sequence and that the limiting density of every number in the sequence is 0.

Chapter 5 is motivated by a question of András Sárközy. We prove sufficient conditions for existence of infinite sets of natural numbers A and B such that the number of solutions of the equation $a + b = n$ where $a \in A$ and $b \in B$ is monotone increasing for $n > n_0$. We also examine a generalized notion of Sidon sets, that is, sets A, B with the property that, for every $n \geq 0$, the equation above has at most one solution, i.e., all pairwise sums are distinct.

Contents

List of Figures

Chapter 1

Introduction

1.1 Introduction

This dissertation concerns four problems in combinatorics. Chapter 1 gives a basic overview of the structure of the dissertation, and also establishes the bulk of the terminology and notation used, leaving some of the more specific terminology for the chapters to come. Chapters 2 and 3 are on graph theoretic problems while Chapter 4 is on a specific integer sequence and Chapter 5 explores the additive number theoretic properties of integer sequences.

In Chapter 2 we consider a particular 2-player game on a finite graph, introduced by Füredi, Reimer and Seress [14]. This chapter is based on a joint work with Jonathan Lee. Given a collection of graphs $\mathcal{F}$, we define the Prolonger-Shortener game $\mathcal{G}(K_n, \mathcal{F})$ as follows: players start with the empty graph E_n on n vertices. They take turns drawing edges of the complete graph K_n on this vertex set with the restriction that no edge completes, with all the edges drawn so far, a copy of a graph in $\mathcal{F}$ as a subgraph. The game ends when no more edges can be drawn. Prolonger wants as many edges as possible at the end of the game and Shortener as few as possible. We ask what is the final number of edges with both players playing optimally. The game $\mathcal{G}(K_n, \{K_3\})$ was investigated by Füredi, Reimer and Seress [14] and studied also by Biró, Horn and Wildstrom [4]. We look at the specific cases when $\mathcal{F}$ contains only a fixed path, $\mathcal{F}$ is a collection of trees and separately a directed version of the game. Let P_k be the path on k vertices. We prove in Theorem 2.3 that the number of edges at the end of the game $\mathcal{G}(K_n, \{P_4\})$ is between $\frac{4}{5}n - 1$ and $\frac{4}{5}n + 1$ regardless of who starts in the game. We have a similarly tight result for the game $\mathcal{G}(K_n, \{P_5\})$, and we characterize the final types of components in $\mathcal{G}(K_n, \{P_6\})$. We also have a tight result for the final number of edges in the game $\mathcal{G}(K_n, \mathcal{T}_k)$ where $\mathcal{T}_k$ is the collection of all trees on k vertices. Separately, we also consider a directed version of the game which is played on a directed graph under the restriction that players are not allowed to create a directed walk on k vertices. We prove that the number of directed edges at the end of the game is $\frac{1}{3}n^2 + \frac{1}{3}kn + O(n + k^2)$.

Chapter 3 is motivated by a conjecture of Füredi and Lehel [13] concerning completion of partial packings of copies of a bipartite graph H. An H-design on n vertices is an edge-disjoint collection of copies of H whose edge sets union up to the edge set of the complete graph K_n on n vertices. Given a bipartite graph H and an integer n, let $f(n; H)$ be the smallest integer such that any set of edge disjoint copies of H on n vertices can be extended to an H-design on at most $n + f(n; H)$ vertices. We establish tight bounds for the growth of $f(n; H)$ as $n \to \infty$. In particular, we prove the conjecture of Füredi and Lehel [13] that $f(n; H) = o(n)$ which settles a long-standing open problem. Chapter 3 is based on a joint work with Zoltán Füredi and Mykhaylo Tyomkyn [15].

Chapter 4 is dedicated to a particular integer sequence. A *block* in a sequence is a finite subsequence consisting of consecutive terms of the given sequence. The Slowgrow sequence, introduced by Steven Kalikow, starts with 1 and, having defined terms $s_1, \ldots, s_n$, has for s_{n+1} the smallest positive integer such that the block $s_{n-s_{n+1}+2} \ldots s_{n+1}$ has not occurred in the sequence earlier. We prove bounds on the time of the first occurrence of n in the Slowgrow sequence and that the limiting density of every number in the sequence is 0. We also prove the more surprising result that blocks which can potentially occur multiple times in the sequence actually occur infinitely often. This is an existential result, giving no bounds on the density of occurrences of the block and is the most non-trivial fact we know about the Slowgrow sequence so far. We also present a classification of such blocks. I am grateful to Steven Kalikow for many helpful conversations.

In Chapter 5 we study the additive number theoretic properties of a pair of general sequences of non-negative integers. This chapter is based on a joint work with Fabricio Benevides, Jonathan Hulgan, Nathan Lemons, Cory Palmer and Jeffrey P. Wheeler [3]. Motivated by a question of András Sárközy, we prove sufficient conditions for existence of infinite sets of natural numbers A and B such that the number of solutions of the equation $a + b = n$ where $a \in A$ and $b \in B$, which we refer to as the *representation function* $r(A, B, n)$, is monotone increasing for $n > n_0$. We also examine a generalized notion of Sidon sets, that is, sets A, B with the property that, for every $n \geq 0$, the equation above has at most one solution, i.e., all pairwise sums are distinct. Our main theorem here states that for all $0 \leq \alpha, \beta < 1$, $1/2 < c_1, c_2 \leq 1$, there exist sets $A, B \subset \mathbb{N}_0$ such that $r(A, B, n)$ is monotone increasing in n;

$$\limsup_{n\to\infty} \frac{A(n)}{n^{c_1}} = \alpha; \qquad \limsup_{n\to\infty} \frac{B(n)}{n^{c_2}} = \beta.$$

1.2 Notation

Throughout we follow the notation that is most widely accepted. For graph theoretic notation, we mainly follow Bollobás [6]. We review the most commonly used terminology and notation in this section.

We use $\mathbb{N}_0$ to denote the natural numbers with zero, $\{0, 1, 2, \ldots\}$, and for $n \in \mathbb{N}_0$, we denote $[n] = \{1, 2, \ldots, n\}$, the set of the first n positive integers. Given a set X, we use $X^{(k)}$ to denote the set of all the k-element subsets of X; we will call a set of size k a k-set. We use $\mathcal{P}(X)$ to denote the *power set* of X, i.e. the set of all subsets of X.

We use the standard "big-oh" notations $O(\cdot), o(\cdot), \theta(\cdot), \Omega(\cdot), \omega(\cdot)$ for growth of functions on natural numbers. That is, $f(n) = O(g(n))$ if and only if $g(n) = \Omega(f(n))$ if and only if there exist constants C and n_0 such that for each $n' > n_0$, $|f(n')| \leq C|g(n')|$. A function $f(n) = \theta(g(n))$ if and only if it is $O(g(n))$ and $\Omega(g(n))$. Also, $f(n) = o(g(n))$ if and only if $f(n) \ll g(n)$ if and only if $g(n) = \omega(f(n))$ if and only if $g(n) \gg f(n)$ if and only if for each $\varepsilon > 0$, there exists n_0 such that for every $n' > n_0$, $|f(n')| \leq \varepsilon |g(n')|$.

A *(simple) graph* G is an ordered pair $(V(G), E(G))$ where $V(G)$ and $E(G) \subseteq V(G)^{(2)}$ are the *vertex set* and *edge set* of G, respectively. Throughout, unless otherwise stated, a graph named G is assumed to have vertex set $V = [n]$ and edge set E. The graph G is said to have *order* $|G| := |V(G)|$ and *size* $e(G) := |E(G)|$. We say an edge $\{u, v\}$ is *incident* to each of the vertices u and v. We say two vertices u and v of G are *adjacent* if $\{u, v\} \in E(G)$. Following the standard convention, we denote an edge $\{u, v\}$ of a graph by uv. A graph is *connected* if for any two of its vertices u and v it has $i \geq 0$ vertices $v_1, \ldots, v_i$ such that uv_1, $v_1v_2, \ldots, v_{i-1}v_i$, $v_iv \in E(G)$. The *degree* $d(v)$ of a vertex v is the number of edges containing it. A graph G is called *k-regular* if all vertex degrees in it are precisely k. The *minimum degree* $\delta(G)$ and *maximum degree* $\Delta(G)$ of a graph G are the minimum, respectively maximum of the vertex degrees over all the vertex set. We say that a vertex is *isolated* if it has degree 0. We say that an edge is *isolated* if all other edges have an empty pairwise intersection with it. A *subgraph* of G is a graph G' such that $V(G') \subseteq V(G)$ and $E(G') \subseteq E(G) \cap V(G')^{(2)}$; we say that a graph *contains* each of its subgraphs. A *component* of G is a maximal connected subgraph with respect to inclusion as subgraphs. For a subset $U \subseteq V(G)$, we use $G[U]$ to denote the graph *induced* by the vertices of U; that is, $G[U]$ has vertex set U and edge set $E(G) \cap U^{(2)}$.

For a vertex v in a graph G, the *open neighborhood* of v, denoted $N_G(v) = N(v)$, is the set of all vertices which share an edge with v, i.e., $N_G(v) := \{u \in V(G) : u \neq v, \{u, v\} \in E(G)\}$.

The *complete* graph K_n has n vertices and $E(K_n) = V(K_n)^{(2)}$ and the *empty* graph E_n has n vertices and $E(E_n) = \emptyset$. A *path* P_k of *length* $k-1$ or on k vertices is a graph where $V(P_k) = \{v_1, \ldots, v_k\}$ are distinct vertices and $E(P_k) = \{\{v_i, v_{i+1}\} | i = 1, 2, \ldots, k-1\}$. A *walk* of length $k-1$ or on k vertices in a graph G is a sequence $v_1, v_2, \ldots, v_k$ of its (not necessarily distinct) vertices such that $v_i v_{i+1} \in E(G)$ for all $i \in [k-1]$. A *cycle* C_k on k vertices is a graph where $V(C_k) = \{v_1, \ldots, v_k\}$ and $E(C_k) = \{\{v_i, v_{i+1}\} | i = 1, 2, \ldots, k-1\} \cup \{\{v_k, v_1\}\}$. A graph G is *bipartite* with partite classes (or partition classes) U and W if $V(G) = U \cup W$, $U \cap W = \emptyset$ and every edge of G has one of its vertices in U and one in W. The *complete bipartite* graph $K_{m,n}$ is a bipartite graph with partite classes U, $|U| = m$ and V, $|V| = n$ such that for any $u \in U$ and any $v \in V$, $\{u, v\} \in E(K_{m,n})$. A *matching* in the graph G is a set $U \subseteq E(G)$ such that any two edges in U have an empty intersection. A *perfect matching* is a matching whose union is the whole vertex set. An *independent set* U of vertices of G is a set such that $G[U]$ is the empty graph, i.e. it has no edges included in it. A graph G on n vertices is said to be *hamiltonian* with *Hamilton cycle* $v_1 \ldots v_n$ if $v_1, \ldots, v_n$ is an enumeration of $V(G)$ and $v_i v_{i+1} \in E(G)$ for all $i \in [n-1]$ and $v_n v_1 \in E(G)$.

A *multigraph* G is an ordered pair $(V(G), E(G))$ where $V(G)$ is a set and $E(G)$ is a multiset with elements from $V(G)^{(2)} \cup V(G)$. Many of the notions defined for graphs can be defined analogously for multigraphs.

We can generalize the notion of an edge to define a hypergraph. A *hypergraph* G is an ordered pair $(V(G), E(G))$ where $V(G)$ and $E(G) \subseteq \mathcal{P}(V(G))$ are the *vertex set* and *(hyper)edge set* of G, respectively. We say that the hypergraph G is *k-uniform* if $E(G) \subseteq V(G)^{(k)}$. Note that a 2-uniform hypergraph is a graph. We refer to a

k-uniform hypergraph as a k-graph. The degree of a vertex and the minimum and maximum degree are defined analogously to the graph case. A *multihypergraph* is defined like a hypergraph except instead of the edge set we have an edge multiset.

A *directed graph* or *digraph* G is an ordered pair $(V(G), E(G))$ where $V(G)$ and $E(G) \subseteq V(G) \times V(G)$ are the *vertex set* and *edge set* of G, respectively. An edge $(u, v) \in V(G) \times V(G)$ of a directed graph is said to *go* or be *directed* from u to v. The edge (u, v) is often denoted by $\overrightarrow{uv}$ but we choose to denote it uv when no confusion arises. The *indegree* $d_-(v)$ and *outdegree* $d_+(v)$ of a vertex v is the number of edges directed to v, respectively the number of edges directed from v.

Now we define some terminology related to integer sequences in additive number theory. Let $A, B \subseteq \mathbb{N}_0$ be sets of non-negative integers. The *sumset* of A and B, denoted $A + B$ is the set $\{a + b | a \in A, b \in B\}$. We say that a set of non-negative integers $A = \{a_0, a_1, \ldots\}$ is a *Sidon set* if all pairwise sums $a_i + a_j$ where $i \leq j$ are different.

Chapter 2

$\mathcal{F}$-saturation Games

2.1 $\mathcal{F}$-saturation Game

We start with some definitions. Let $\mathcal{F}$ be a family of graphs. A graph G is *$\mathcal{F}$-saturated* if G contains no member of $\mathcal{F}$ as a subgraph, but for any nonadjacent vertices u and v in G the graph obtained by adding uv to G contains some member of $\mathcal{F}$. For a discussion of saturated graphs see for example Bollobás [5].

In this chapter, we shall consider different 2-player games that we shall call *$\mathcal{F}$-saturation games* following West [33].

One recent variant of a game that we shall not study here but that is similar to ours was introduced by Ferrera, Harris and Jacobson in 2010 [12]. Let H be an arbitrary graph on n vertices and let $\mathcal{F}$ be a family of graphs. Two players start with the empty graph E_n on n vertices. They take turns adding an edge to it from the edge set of H, so the resulting graph is always a subgraph of H. The first player who creates a member of $\mathcal{F}$ loses. Equivalently, the player who reaches an $\mathcal{F}$-saturated graph wins.

The game we shall concentrate on in this chapter was introduced by Füredi, Reimer and Seress in 1991 [14] who studied it in a special case. Let H be an arbitrary graph on n vertices and let $\mathcal{F}$ be a family of graphs. Two players start with the empty graph E_n on n vertices. They take turns adding an edge to it so that the resulting graph is always a subgraph of H. No player is allowed to create a member of $\mathcal{F}$ out of the edges drawn so far. The game finishes when the resulting subgraph of H is $\mathcal{F}$-saturated or equal to H. The objective of one player, Prolonger, is to have as many edges in the resulting graph at the end of the game as possible and the other player, Shortener, wants as few edges as possible. Let us denote this game by $\mathcal{G}(H, \mathcal{F})$. If $\mathcal{F} = \{F\}$ then denote the game by $\mathcal{G}(H, F)$. Let the *game saturation number* $\mathrm{Sat}_g(H, \mathcal{F})$ or *game score* be the length of the game under optimal play by both players, with Prolonger starting. Let the game score with Shortener starting be $\mathrm{Sat}_{g'}(H, \mathcal{F})$.

Füredi, Reimer and Seress [14] concentrate on the game $\mathcal{G}(K_n, K_3)$. They prove that Prolonger can ensure that the number of edges at the end of the game is at least $(\frac{1}{2} + o(1))n \lg n$ where $\lg n$ is the binary logarithm. They attribute to Erdős a lost proof that Shortener can ensure that the number of edges at the end of the game is at most $\frac{n^2}{5}$. Biró, Horn and Wildstrom [4] demonstrate the improved upper bound $\frac{9n^2}{50}$.

2.2 Game of avoiding P_k

Let us denote a path on k vertices by P_k for every positive integer k. Let us write $\mathcal{G}$ for the game $\mathcal{G}(K_n, P_k)$.

First of all, we shall consider a variant of the game where Prolonger can skip his turn. We shall modify the game $\mathcal{G}$ so that Prolonger can skip his go if he chooses and Shortener must, on her turn, still draw one edge. Let us call the modified game $\mathcal{G}_1$.

Theorem 2.1. *In the modified game $\mathcal{G}_1$, the game score will be at least $n(k-2)/4$ regardless of who starts.*

Proof. We need to give a strategy for Prolonger that realizes the lower bound. His strategy is to keep all components in the graph under construction in the game hamiltonian (or isolated edges or isolated vertices). For the purpose of this proof call components that consist of an isolated edge or an isolated vertex hamiltonian (with 'Hamilton cycle' consisting of the single edge or the single vertex, respectively). Suppose all components are hamiltonian before a move by Shortener. If Shortener connects a component of at least two vertices with Hamilton cycle $a_1a_2\ldots a_l$ to an isolated vertex v by drawing the edge a_1v then Prolonger draws the edge a_2v thus making the new component hamiltonian with Hamilton cycle $a_1va_2a_3\ldots a_l$, see Figure 2.1. Analogously, if Shortener draws the edge a_1b_1 to connect two hamiltonian components with Hamilton cycles $a_1a_2\ldots a_l$ and $b_1b_2\ldots b_m$ then Prolonger can draw the edge a_2b_2 for hamiltonicity, see Figure 2.2. If Shortener makes any other move all components will continue to be hamiltonian after that and Prolonger can skip his go. That is, Prolonger's strategy is as follows:

i) If there is a component with a Hamilton path but no Hamilton cycle, augment to a Hamilton cycle;

ii) Otherwise skip turn.

We will reach a point when the total number of vertices in any two components is at least k. Then no two of the (hamiltonian) components can be connected by an edge. After that clearly all components will be completed to a clique. Since any two cliques have at least k vertices in total, the average degree in the graph is at least $k/2-1$. Hence the total number of edges in the graph will be at least $n(k-2)/4$. $\square$

We hope that a similar result can be proved for the unmodified game.

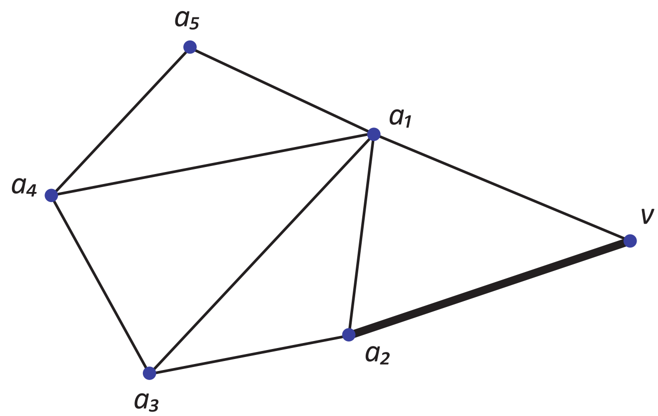

Figure 2.1: Unifying a hamiltonian component and an isolated vertex

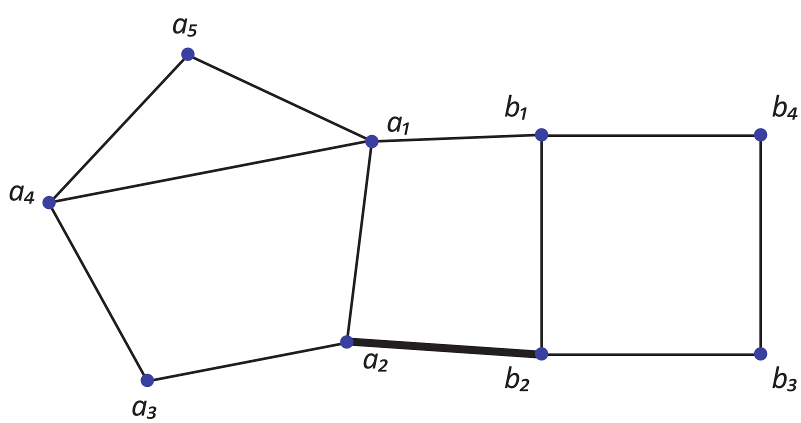

Figure 2.2: Unifying two hamiltonian components

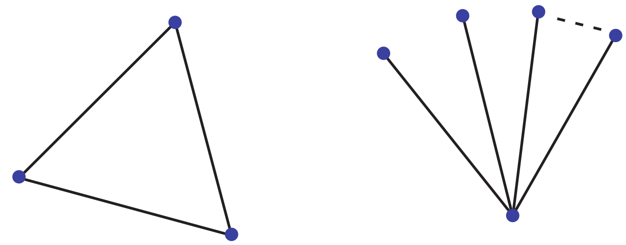

Figure 2.3: Types of components in a P_4 saturated graph

Let us introduce some notation which helps us to consider the game $\mathcal{G}(K_n, \mathcal{F})$ where $\mathcal{F}$ is either $\{P_4\}$ or $\{P_5\}$ or $\mathcal{T}_k$, the collection of all trees on k vertices. For concreteness we shall prove bounds on the game score in these cases with Prolonger starting the game. However, our proofs work unchanged for Shortener starting. Let the graph constructed during the game be initially $G_0 = E_n$. After i moves, i.e., after i edges have been drawn, let the graph constructed in the game be G_i. So after Prolonger's first move we have the graph G_1, after Shortener's first move G_2, after Prolonger's second move G_3 and so on.

2.2.1 Game of avoiding P_4

Let us consider the (unmodified) game $\mathcal{G}(K_n, P_4)$. We claim that the only types of components present in the P_4 saturated graph are those in Figure 2.3.

Lemma 2.2. *A P_4 saturated graph is either a vertex-disjoint union of triangles and stars with at least two vertices or a vertex-disjoint union of triangles and an isolated vertex.*

Proof. This follows by induction on the number of vertices. $\square$

We are going to bound the game score in $\mathcal{G}(K_n, P_4)$. Our bounds will not depend on who starts.

Theorem 2.3.

$$\frac{4}{5}n - 1 \leq \operatorname{Sat}_g(K_n, P_4), \operatorname{Sat}_{g'}(K_n, P_4) \leq \frac{4}{5}n + 1.$$

Proof. First we shall prove $\operatorname{Sat}_g(K_n, P_4), \operatorname{Sat}_{g'}(K_n, P_4) \leq \frac{4}{5}n + 1$ by fixing a strategy for Shortener. Let Shortener

i) extend a $K_{1,2}$ to a $K_{1,3}$ if possible, otherwise

ii) draw an isolated edge if possible, otherwise

iii) extend a star by attaching the central vertex to an isolated vertex if possible, otherwise

iv) extend a $K_{1,2}$ to a K_3.

Claim 2.4. *In graphs G_{2i+1}, $i \in \mathbb{N}_0$, there is at most one $K_{1,2}$ component. Unless all vertices of the graph have degree at least one the $K_{1,2}$ will not become a K_3; if a $K_{1,2}$ becomes a K_3, however, the game is finished. In graphs G_{2i}, $i \in \mathbb{N}_0$, there is at most one $K_{1,2}$ component; if in a graph G_{2i}, $i \in \mathbb{N}_0$, there is a $K_{1,2}$ component it will be extended into a K_3 and this finishes the game.*

Proof. Let us look at two cases and proceed by induction on i.

1. Suppose that in the graph G_{2i+1} there is a $K_{1,2}$ component where $i \in \mathbb{N}_0$.

   1) If there is an isolated vertex in the graph G_{2i+1}, Shortener will extend the $K_{1,2}$ to a $K_{1,3}$. Since then there will be no $K_{1,2}$ component in G_{2i+2}, there can be at most one of them in G_{2i+3}.

   2) If there is no isolated vertex in the graph G_{2i+1}, Shortener will extend the $K_{1,2}$ to a K_3. This finishes the game: Let us say that a component is *non-trivial* if it is not an isolated vertex. We can not unify two non-trivial components without

creating a P_4; hence when all vertices are non-isolated the only possible moves are filling in $K_{1,2}$'s to K_3's.

2. Suppose that there is no $K_{1,2}$ component in the graph G_{2i+1}.

   1) If there is an isolated vertex and another star of at least one vertex, Shortener will join the isolated vertex to a star (which may be an isolated vertex) creating at most one $K_{1,2}$ component. If he creates a $K_{1,2}$ component that means there was only one isolated vertex available. This means that on the next move Prolonger will have to extend the $K_{1,2}$ to a K_3 and thus finish the game with the graph G_{2i+3}.

   2) Otherwise the game has finished by Lemma 2.2.

This finishes the proof. □

By the Claim, until the last move of the game, the graph is a collection of vertex disjoint stars. Let λ be the number of components at the end of the game. Since there is at most 1 triangle, the score is bounded above by $n + 1 - \lambda$, with $n - \lambda$ moves producing non-trivial components (i.e. creating isolated edges) or extending stars. To prevent Shortener making a new non-trivial component, Prolonger must make a $K_{1,2}$, which occurs at most once in any component. Hence at most λ of Shortener's moves fail to make a non-trivial component. Hence there are at least $\frac{1}{2}(n - \lambda) - \lambda$ components. So $\lambda \geq \frac{1}{5}n$, and the score is at most $\frac{4}{5}n + 1$.

Now we shall prove $\frac{4}{5}n - 1 \leq \mathrm{Sat}_g(K_n, P_4), \mathrm{Sat}_{g'}(K_n, P_4)$ by fixing a strategy for Prolonger. Let Prolonger in G_{2i}

i) complete a triangle component if possible, otherwise

ii) complete a $K_{1,2}$ component if possible, otherwise

iii) extend a star component if possible, otherwise

iv) draw an isolated edge.

Note that Prolonger is forced to play an isolated edge only as the first move or after Shortener completes a triangle.

Let us say a move in the game *uses k new vertices* if the number of isolated vertices is reduced by k as a result of that move.

Suppose Prolonger has created a $K_{1,2}$ component in a graph G_{2i+1}, $i \in \mathbb{N}_0$. We shall show that the next move by Shortener and the next move by Prolonger will use at most two new vertices in total. If Shortener plays elsewhere, Prolonger will extend the $K_{1,2}$ to a K_3. If Shortener extends the $K_{1,2}$ to a K_3, Prolonger can make an arbitrary move. If Shortener extends the $K_{1,2}$ to a $K_{1,3}$ then Prolonger can extend that to a $K_{1,4}$. In any case at most two new vertices are used to obtain G_{2i+3} from G_{2i+1}.

Suppose Prolonger can not create a $K_{1,2}$ component in G_{2i+1} on his move that creates G_{2i+1} from G_{2i}. We will show that to get G_{2i+1} from G_{2i-1} (where $i > 0$) at most 2 new vertices are used in total. If Prolonger cannot create a $K_{1,2}$ component then that means Shortener did not play an isolated edge into G_{2i-1} (or there are no isolated vertices left in G_{2i}). Prolonger can be forced to play an isolated edge into G_{2i} only if Shortener completed a triangle into G_{2i-1}.

Suppose Prolonger can create a $K_{1,2}$ component when creating G_{2i+1} but does something else. That means Prolonger completes a triangle into G_{2i}. So when creating G_{2i+1} from G_{2i-1} at most 2 new vertices are used in total.

Note that with this strategy of Prolonger to obtain G_{2i+1} from G_{2i-1} we never use 4 new vertices. To obtain G_{2i+1} from G_{2i-1} we can use 3 new vertices only if in G_{2i-1} there was no $K_{1,2}$ component and in G_{2i+1} there is. As a consequence, no two consecutive pairs of moves by Shortener and then Prolonger both use 3 new vertices.

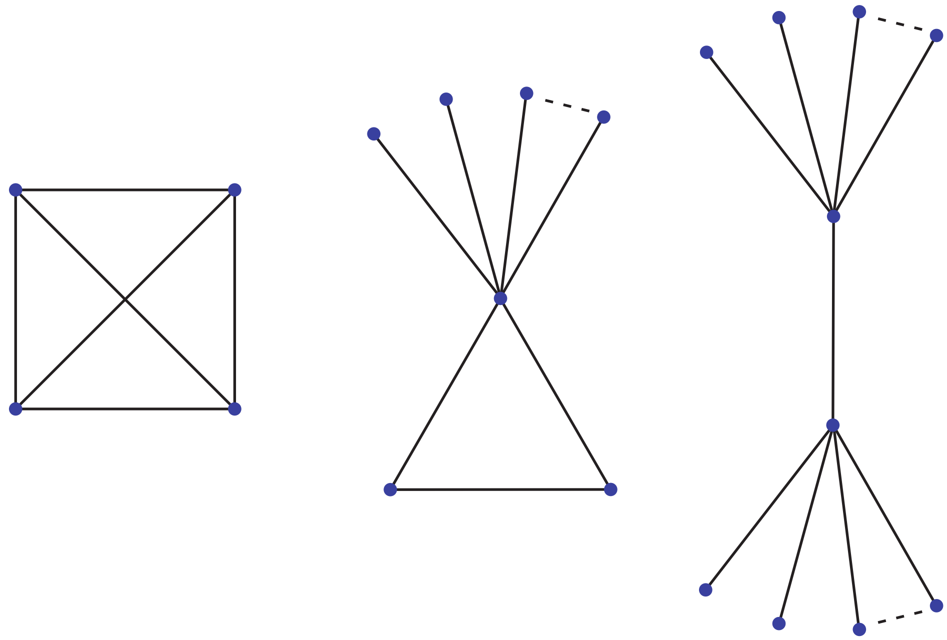

Figure 2.4: Types of components in a P_5 saturated graph

Therefore, per 4 consecutive moves (of Shortener, Prolonger, Shortener and then Prolonger), at most 5 new vertices are used. So the score is at least $\frac{4}{5}n - 1$. $\square$

This gives bounds for the game score in the game of avoiding a P_4 which are tight up to additive constants. We think that with more effort the exact value of the game score can be found, for each of the players starting.

2.2.2 Game of avoiding P_5

Let us consider the game $\mathcal{G}(K_n, P_5)$. We claim that the only types of components present in the P_5 saturated graph are those in Figure 2.4.

Lemma 2.5. *A P_5 saturated graph is either a vertex-disjoint union of K_4's, triangles with some (maybe zero) pendant edges at one vertex and double stars (more precisely, edges with at least two pendant edges on both ends) and at most one isolated edge or a vertex-disjoint union of one isolated vertex and some K_4's.*

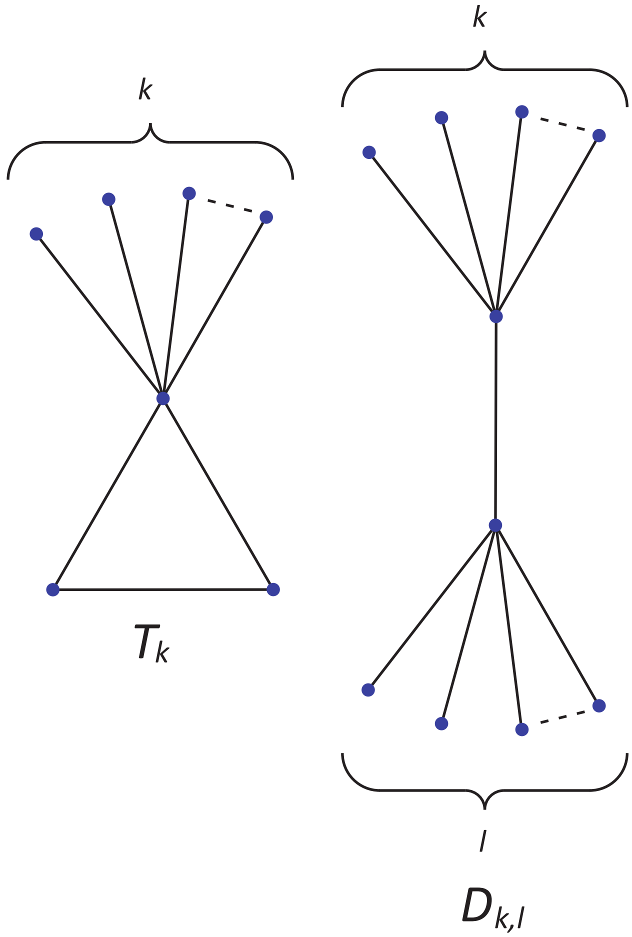

Figure 2.5: T_k and $D_{k,l}$

Proof. This follows by induction on the number of vertices. □

Let us denote a double star with k pendant edges at one end and l at the other end of the middle edge by $D_{k,l}$. Let us denote a triangle with k pendant edges at one vertex by T_k. See Figure 2.5 for these definitions in pictures.

We are going to prove bounds on the game score in $\mathcal{G}(K_n, P_5)$. The bounds will not depend on who starts.

Theorem 2.6. *For every positive integer n,*

$$n - 1 \leq \mathrm{Sat}_{\,g}(K_n, P_5), \mathrm{Sat}_{\,g'}(K_n, P_5) \leq n + 2.$$

Proof. Let us fix a strategy for Shortener to prove the upper bound. She will in G_{2i+1}, $i \in \mathbb{N}_0$,

i) extend a P_4 to a $D_{1,2}$ or extend a $K_{1,3}$ to a $D_{1,2}$ or extend a T_1 to a T_2 if possible, otherwise

ii) extend an isolated edge to a $K_{1,2}$ if possible, otherwise

iii) extend a component of 5 or more vertices by attaching to it an isolated vertex if possible, otherwise

iv) draw an isolated edge if possible, otherwise

v) play arbitrarily.

Lemma 2.7. *In any graph G_{2i+1}, $i \in \mathbb{N}_0$, there is at most one component of 4 vertices and at most one isolated edge, or, if there are no components of 4 vertices then there are at most two isolated edges. The same condition is satisfied in any graph G_{2i}, $i \in \mathbb{N}_0$.*

Proof. Let us look at two cases.

First, assume there are some isolated vertices in G_{2i+1}. We shall prove that then there is at most one component of 4 vertices which is a P_4 or a $K_{1,3}$ or a T_1 and at most one isolated edge, or, if there is no component of 4 vertices then there are at most 2 isolated edges. It is certainly true in G_1.

Now assume it is true in G_{2i+1}. If there is a P_4 or a $K_{1,3}$ or a T_1 component in G_{2i+1} then Shortener will extend it to a $D_{1,2}$, resp. $D_{1,2}$, resp. T_2; otherwise, if there are two isolated edges in G_{2i+1} then Shortener will extend one of them to a $K_{1,2}$. If in G_{2i+1} there is no P_4 or $K_{1,3}$ or T_1 or isolated edge then Shortener will unify a ≥ 5-vertex component with an isolated vertex (possible for every ≥ 5 vertex component by Lemma 2.5); if not possible draw an isolated edge; if not possible that means there is only one isolated vertex and possibly some 3-vertex components in G_{2i+1}, so Shortener plays arbitrarily. So in G_{2i+2} there is no 4-vertex component and at most one isolated edge, or, any one of P_4, $K_{1,3}$ or T_1 and no isolated edges and no isolated vertices. In the former case, into G_{2i+2} Prolonger can create at most one new 4-vertex component, or otherwise draw at most one isolated edge; in the latter case (if there was a 4-vertex component after Shortener's move), Prolonger can create into G_{2i+2} no new 4-vertex components and no isolated edges. When Prolonger creates a 4-vertex component (to form G_{2i+3} from G_{2i+2}) from two smaller components it will necessarily be a P_4 or a $K_{1,3}$ or a T_1. So the condition is satisfied both in G_{2i+2} and G_{2i+3}.

Second, assume there are no isolated vertices in G_{2i+1}. Before we run out of isolated vertices the condition of the lemma holds by the previous paragraph. We have also proved in the previous paragraph that it holds the first time when there are no isolated vertices left after a move by Prolonger; also, it holds in G_1. We shall prove that the condition of the lemma holds in G_{2i+2} and G_{2i+3}. When there are no isolated vertices left the only way to create new 4-vertex components is to join two isolated edges to create a P_4. But the condition of the lemma says that if we have a 4-vertex component then there is at most one isolated edge, so we cannot make a new 4-vertex component; otherwise we have at most two isolated edges, so we can create at most one 4-vertex component. Hence the condition of the lemma is satisfied both in G_{2i+2} and G_{2i+3}.

The proof is complete since the condition of the lemma is clearly satisfied in G_0. □

So at the end of the game there is at most one K_4. By Lemma 2.5 the number of edges does not exceed the number of vertices in all other components. Hence the game score is at most $n + 2$.

Let us fix a strategy for Prolonger to prove the lower bound. Let us call a component *trivial* if it consists of an isolated vertex. Let us call a non-trivial component *standalone* if it can not be connected to another non-trivial component without completing a P_5, otherwise call it *non-standalone*. Note that for a component to be non-standalone it has to have a vertex rooting no P_3. So the only non-standalone components are stars. Let Prolonger's strategy be, in the graph G_{2i}, $i \in \mathbb{N}_0$, to

(i) complete a triangle in a $D_{1,2}$ component to make it a T_2 or in a $K_{1,3}$ component to make it a T_1, or, if not possible

(ii) complete a triangle in a component without a triangle, or, if not possible

(iii) connect two isolated edges to form a P_4, or, if not possible

(iv) complete a $K_{1,2}$ component, or, if not possible

(v) draw an isolated edge, or, if not possible

(vi) play arbitrarily.

Lemma 2.8. *In any graph G_{2i+1}, $i \in \mathbb{N}_0$, the set of non-standalone components may be: empty; or one isolated edge; or one $K_{1,2}$. In any graph G_{2i}, $i \in \mathbb{N}_0$, the set of non-standalone components may be: empty; or $K_{1,2}$; or P_4; or $K_{1,3}$; or $K_{1,2}$ and an isolated edge; or two isolated edges; or one isolated edge.*

Proof. The result holds in G_0 and G_1. One can check that if in G_{2i+1} the set of non-standalone components was empty, an isolated edge or a $K_{1,2}$ then whatever Shortener does she can only create sets of non-standalone components in G_{2i+2} described in the statement of the lemma. Prolonger will

(i) complete a triangle in the $K_{1,2}$ component to create a K_3, in the P_4 component to create a T_1, in the $K_{1,3}$ component to create a T_1 all of which are standalone;

(ii) if not possible, he will complete a P_4 from two isolated edges which is standalone;

(iii) if not possible, he will complete a $K_{1,2}$ component from one isolated edge;

(iv) if not possible, he will draw an isolated edge;

(v) this failing, he will play arbitrarily but his move will not extend a star into a larger star (otherwise he could have completed a triangle in it),

so his edge will be a part of a standalone component. In any case the set of non-standalone components will be as described in the statement of the lemma. □

Lemma 2.9. *At the end of the game all standalone components will contain a triangle. The set of non-standalone components will consist of an isolated vertex or an isolated edge.*

Proof. In any graph G_{2i+1}, $i \in \mathbb{N}_0$, the set of non-trivial components without a triangle will be empty or consist of one component which will be an isolated edge, a $K_{1,2}$ or a P_4: This is clearly true in G_1. Assume it is true in G_{2i+1}; we shall show it is true in G_{2i+3}. By Lemma 2.8 there is at most one standalone component in G_{2i+1}, so if Shortener wants to connect two components one of them has to be an isolated vertex. So in G_{2i+2} the set of non-trivial components without a triangle will be empty, consist of an isolated edge or a $K_{1,2}$ or a P_4 or consist of any one of the preceding and an isolated edge or consist of a $K_{1,3}$ or a $D_{1,2}$. In each case, to form G_{2i+3} from G_{2i+2}, Prolonger will either

(i) complete a triangle in them to create a T_2 component or a T_1 component or a K_3 component or

(ii) connect two isolated edges to form a P_4 component or

(iii) connect an isolated edge to an isolated vertex to form a $K_{1,2}$ component or

(iv) create an isolated edge or

(v) else there is at most one non-trivial component without a triangle which can only be an isolated edge and he can play arbitrarily

– so again the set of non-trivial components without a triangle consists of an isolated edge, a $K_{1,2}$ or a P_4.

Shortener can never create components $D_{k,l}$ with $k, l \geq 2$: Indeed, she would have to go about building them by adding one isolated vertex at a time to the component since by Lemma 2.8 there will be at most one non-standalone component in G_{2i+1}; she would first have to build a $D_{1,2}$ component or a $K_{1,3}$ component which will immediately be completed into a T_2, resp. T_1 by Prolonger to form G_{2i+3}. At the end of the game the set of non-trivial components without a triangle will be an isolated vertex or an isolated edge, since all other components – none of which can be a $D_{k,l}$ with $k, l \geq 2$ in G_{2i+2} as we have proved – will get a triangle by Lemma 2.5. □

We are now ready to prove that $\mathrm{Sat}_g(K_n, P_5) \geq n - 1$ and $\mathrm{Sat}_{g'}(K_n, P_5) \geq n - 1$. By Lemma 2.9 all components at the end of the game will contain a triangle except for at most one component which can be either an isolated edge or an isolated vertex. That means that the number of edges in them is greater or equal to the number of vertices. Thus the game score will be at least $n - 1$. □

2.2.3 Game of avoiding P_6

Let us consider the game $\mathcal{G}(K_n, P_6)$. We claim that the only types of components present in a P_6 saturated graph are those in Figure 2.6.

Lemma 2.10. *A P_6 saturated graph is a vertex-disjoint union of components of the following types: K_5; K_4 with some pendant edges at one vertex; components consisting of a central vertex on which some triangles are built by identifying one of*

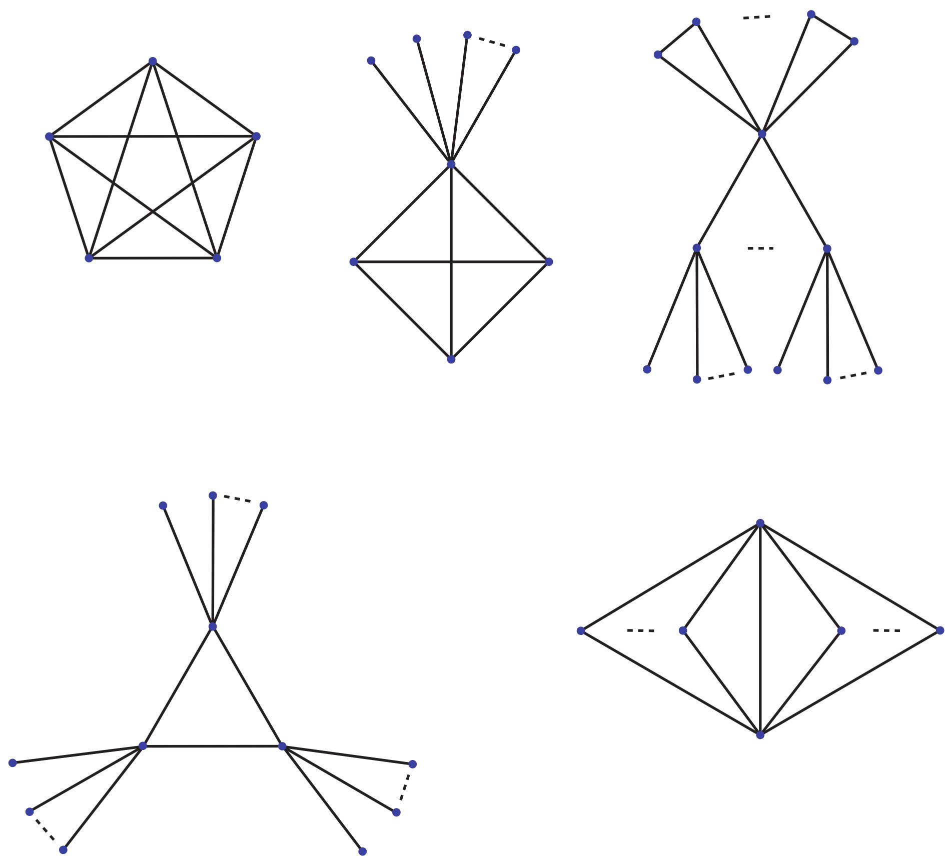

Figure 2.6: Types of components in a P_6 saturated graph

its vertices with the central vertex and some stars are built by identifying a leaf with the central vertex; triangles with some pendant edges on each vertex; components consisting of a central edge with some triangles built on it by identifying one of its edges with the central edge; some types of substructures of the components described.

Proof. This follows by induction on the number of vertices in the graph. □

That is all we have proved about this game. The case analysis to find good bounds on the game score seems hard.

2.3 Game of avoiding all trees on k vertices

Let us consider the Prolonger-Shortener game $\mathcal{G}(K_n, \mathcal{T}_k)$ where no player is allowed to complete any tree on k vertices. So we are considering the $\mathcal{F}$-saturation game where $\mathcal{F} = \mathcal{T}_k$ is the collection of all trees on k vertices. We are looking for the game scores $\mathrm{Sat}_g(K_n, \mathcal{T}_k)$ and $\mathrm{Sat}_{g'}(K_n, \mathcal{T}_k)$.

The condition in the game is equivalent to saying that all components have less than k vertices. Clearly at the end of the game all components will be cliques of at most $k-1$ vertices such that any two components have at least k vertices in total.

It is easy to see by computing the average degree that for Prolonger it is beneficial if in the graph created by the end of the game there are as many components of size $k-1$ as possible and one component left over. We shall prove that Prolonger can achieve almost that.

Theorem 2.11. *In the game $\mathcal{G}(K_n, \mathcal{T}_k)$ Prolonger can force*

$$\begin{aligned}
&\lfloor \tfrac{n}{k-1} \rfloor \tbinom{k-1}{2} + \tbinom{n-(k-1)\lfloor \frac{n}{k-1} \rfloor}{2} && \text{edges if} \quad n \not\equiv 1 \mod (k-1) \\
&\quad \text{and at least} \\
&\lfloor \tfrac{n}{k-1} \rfloor \tbinom{k-1}{2} + \tbinom{n-(k-1)\lfloor \frac{n}{k-1} \rfloor}{2} - (k-3) && \text{edges if} \quad n \equiv 1 \mod (k-1).
\end{aligned}$$

In other words, we prove that Prolonger can force the maximal possible number of edges if $n \not\equiv 1 \mod (k-1)$ and at least the maximal possible number minus $(k-3)$ edges otherwise.

Recall that $G_0 = E_n$ is the graph formed in the game in beginning, G_{2i-1} is the graph formed in the game after the ith move by Prolonger and G_{2i} the graph formed in the game after the ith move by Shortener where $i \in \mathbb{N}_0$.

Proof. Let the strategy for Prolonger be to choose two components with the greatest total number of vertices such that this number is at most $k-1$ and connect them by an edge.

Lemma 2.12. *In any graph G_{2i+1}, $i \in \mathbb{N}_0$, there are, either,*

1) some components of $k-1$ vertices, some isolated vertices and at most one more component, or,

2) some components of $k-1$ vertices, a component of $k-2$ vertices and an isolated edge.

Proof. It is clearly true in G_1. Let us assume it is true in G_{2i+1}. We shall prove it is true in G_{2i+3}. Shortener on her move has three choices:

a) connect the biggest component of less than $k-1$ vertices to an isolated vertex or

b) play an edge inside of a component or

c) connect two isolated vertices to make an isolated edge.

In the first and second case, on his next move Prolonger connects the two largest components of size less than $k-1$ (one of which is an isolated vertex by our induction hypothesis) to obtain Possibility 1) in G_{2i+3}. In the third case,

(i) if the second largest size component has $k-3$ or fewer vertices, he connects it to the isolated edge to obtain Possibility 1) in G_{2i+3};

(ii) if the second largest size component has size $k-2$ vertices and there are isolated vertices he connects an isolated vertex to the size $k-2$ component to obtain Possibility 1) in G_{2i+3} (with the middle size component being the isolated edge);

(iii) if the second largest size component has size $k-2$ vertices and there are no isolated vertices he plays arbitrarily inside a component (if the game has not ended because all components are cliques) to obtain Possibility 2) in G_{2i+3}.

This finishes the proof. $\square$

Clearly when Possibility 2) occurs in the game, the only possible moves are to play inside existing components. Then we end up with cliques of $k-1$ vertices, a clique of $k-2$ vertices and an isolated edge and Prolonger has forced the maximal possible number minus $\binom{k-1}{2} - (\binom{k-2}{2} + 1) = k-3$ of edges.

If Possibility 1) occurs towards the end of the game when the only possible moves are to play inside existing components then there are some components of $k-1$ vertices and another component. Then we end up with cliques of $k-1$ vertices and possibly another clique and Prolonger has forced the maximal possible number of edges. $\square$

2.4 Game of avoiding a directed walk on k vertices

Let us play the following game, starting on an empty directed graph on n vertices. Let us consider the Prolonger-Shortener game $\mathcal{G}_{\mathrm{dir}}$ where players put down directed edges and at no time there should be a directed walk on k vertices, i.e. there is no graph homomorphism from the directed path P_k into the graph.

Clearly the directed graph constructed in the game at any stage of the game must have at most one edge (i.e. one direction) between any pair of vertices, no loops (i.e. no edges starting and ending at the same vertex) and it is acyclic (i.e. has no directed cycle), otherwise there is a homomorphism from P_λ to the graph for every positive λ. Such a graph can always be *topologically sorted* to give a linear ordering of the vertices such that $uv \in E(G)$ implies $u < v$ in the ordering. Let us call $v_1 v_2 \dots v_m$ a *descending sequence* of vertices if $v_i v_{i+1}$ is a directed edge for every $i = 1, 2, \dots, m-1$. Suppose that a maximum descending sequence of vertices has l vertices. Let us define classes $G_1, G_2, \dots, G_l$ so that $v \in G_i$ if the longest descending sequence of vertices ending at v has i elements.

Assume $k \leq n$ – otherwise the game ends in the complete acyclic simple graph on n vertices. Clearly at the end of the game $l = k-1$ (otherwise we can take any class G_i with at least two vertices and put an edge uv between them, splitting G_i as $G_i \backslash \{v\}$ and $\{v\}$). Furthermore, all edges of the form wz exist where $w \in G_i$ and $z \in G_j$ and $i < j$ as the game cannot be continued.

So at the end of the game the game score is $s = \frac{1}{2}(n^2 - \sum_{i=1}^{k-1} c_i^2)$ where $c_i = |G_i|$.

Let us treat the case $k < 4$ separately.

The case $k = 1$ is not possible. If $k = 2$ then no edges can be created in the game and the game score is 0. If $k = 3$ then all edges created go from G_1 to G_2, then G_1 and G_2 end up being as equal as possible: Indeed, let us consider the classes as empty in the beginning. If an edge is created, let the initial vertex be put to class G_1 and the final vertex to G_2. Let us prove that after any move by Prolonger except at the end where all vertices have been assigned a class the number of vertices in G_1 and G_2 is the same. If Shortener plays an edge between vertices that have already places in the classes or creates an isolated edge, thus placing two vertices in the classes, one in each of G_i then Prolonger can keep the balance by either playing an edge between vertices already placed in G_1 and G_2 or creating an isolated edge. If Shortener connects a vertex already placed in G_i to an unassigned vertex then Prolonger can connect a vertex in G_{3-i} to an unassigned vertex, for every $i = 1, 2$, thus keeping the balance. There may end up to be a one-vertex difference in the size of classes only when all vertices have been assigned to classes. So the game score will be $\frac{1}{2}(n^2 - \lfloor \frac{n}{2} \rfloor^2 - \lceil \frac{n}{2} \rceil^2)$.

Now assume $k \geq 4$.

Theorem 2.13. *Let $k \geq 4$. The game score in $\mathcal{G}_{\text{dir}}$ is at most*
$\frac{1}{2}(n^2 - k + 4 - 2\lfloor \frac{n-k+4}{3} \rfloor^2 - \lceil \frac{n-k+4}{3} \rceil^2)$.

Proof. Let us give a strategy for Shortener.

We will be looking at the digraph that is empty in the beginning and gets edges added to it during the game. First let Shortener build a path of $k-1$ vertices. She can use the edges that Prolonger builds for that, so that all of Prolonger's edges except at most one both start and end on the path:

- if Prolonger puts down an isolated edge uv and the existing path $v_1v_2 \ldots v_l$ has $k-3$ vertices or less then Shortener will attach the isolated edge at the end of the path thus forming the path $v_1v_2 \ldots v_l uv$;
- if Prolonger puts down an isolated edge uv and if the existing path has $k-2$ edges then Shortener will attach the isolated edge to the second vertex from the end of the path thus forming the path $v_1v_2 \ldots v_{l-1}uv$ and leaving one vertex v_l in one of G_{k-1} or G_{k-2};
- if Prolonger connects two vertices already on the longest path then Shortener will attach an isolated vertex at the end of the path;
- if Prolonger puts down an edge v_iv to an isolated vertex v where $i < l$ then Shortener will create the edge v_lv thus extending the existing path to $v_1v_2 \ldots v_lv$;
- if Prolonger puts down an edge vv_i from an isolated vertex v where $1 < i$ then Shortener will create the edge vv_1 thus extending the existing path to $vv_1v_2 \ldots v_l$;

- if Prolonger extends the existing path then Shortener will also just increase the existing longest path (if it has less than $k-1$ vertices) by attaching an isolated vertex at the end.

Note that after the path of $k-1$ vertices is formed, there is one vertex in each class G_i and there may be one vertex that can belong to either G_{k-2} or G_{k-1}.

So now we have a path $v_1v_2\dots v_{k-1}$.

Suppose u_i is a vertex that has been forced to the class G_i (e.g. it could be v_i). If Prolonger creates the edge u_iv then Shortener will create the edge $v_{k-2}v$; if Prolonger creates the edge vu_i then Shortener will create the edge vv_2. So when Prolonger attaches any isolated vertex v to an already assigned vertex Shortener will force v to be in either G_1 or G_{k-1}. If Prolonger plays an edge between vertices already assigned to classes then Shortener will create the edge $v_{k-2}v$ where v is an isolated vertex which will thus be forced into G_{k-1}. If Prolonger plays an isolated edge uv then Shortener will create the edge $v_{k-3}u$ thus forcing u into G_{k-2} and v into G_{k-1}. See Figure 2.7 for an illustration.

It can thus be seen that after any move by Shortener all non-isolated vertices will be forced to classes – except there may be a vertex which belongs to G_{k-2} but may be forced to G_{k-1} at the end of the game. Also, note that all classes except G_1, G_{k-2} and G_{k-1} will contain only one vertex. Assume that at the end of the game this situation will be the worst possible for Shortener: each class G_i where $2 \le i \le k-3$ contains one vertex and the classes G_1, G_{k-2} and G_{k-1} contain an almost equal number of vertices (which will thus differ by at most 1 from $\frac{n-k+4}{3}$). Hence Shortener can force the game score to be at most $\frac{1}{2}(n^2 - k + 4 - 2\lfloor\frac{n-k+4}{3}\rfloor^2 - \lceil\frac{n-k+4}{3}\rceil^2)$. □

Theorem 2.14. *Let $k \ge 4$. The game score in $\mathcal{G}_{\mathrm{dir}}$ is at least* $\binom{k-1}{2} + (n-k+1)(k-2) + 1/2(n-k-9)^2(1-1/3)$.

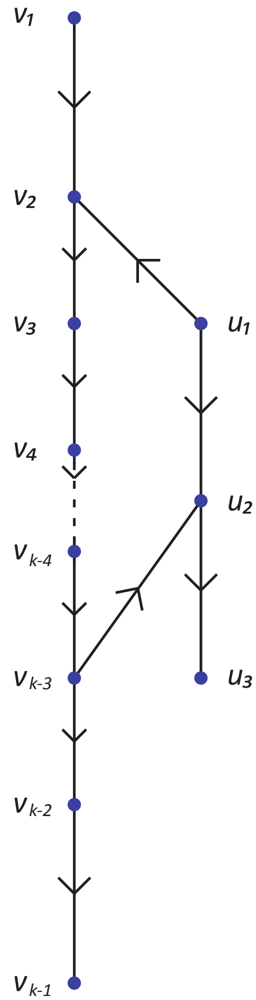

Figure 2.7: Shortener's strategy to force into classes the vertices of paths on 3 vertices

Proof. Let us give a strategy for Prolonger.

Suppose Prolonger likes to build long paths. The structures that arise in the graph we are building in the game as a result of some λ moves by Prolonger and the λ moves by Shortener after each of those are

A. λ vertices with in-edges and a path on $\lambda + 1$ vertices

B. $\lambda - 1$ vertices with in-edges and a path on $\lambda + 1$ vertices and a vertex on the path with an in-edge or an out-edge

C. $\lambda - 2$ vertices with in-edges and a path on $\lambda + 1$ vertices and two vertices of the path with an in-edge or an out-edge

Let us denote the structures by A_λ, B_λ and C_λ, respectively, see Figure 2.8 for illustration. Note that if Prolonger leaves these stuctures alone, Shortener can eventually force at most $\lambda + 1$, resp. λ, resp. $\lambda - 1$ vertices into one class and the rest of the vertices will be in other individual classes. Note that if Shortener hopes to stop the growth (by addition of isolated vertices) of the path $u_1 \dots u_l$ she must force the initial vertex u_1 to G_1 (for example by drawing the edge $u_1 v_2$) and the final vertex u_l to G_{k-1} (for example by drawing the edge $v_{k-2} u_l$) – hence the structures B_λ and C_λ.

Let a structure D consist of r vertices and force vertices to t different classes, with d_j vertices forced to the jth class ($\sum_{j=1}^{t} d_j = r$). We can define the *normalized score* for a structure as $s(D) = (\sum_{j=1}^{t} d_j^2)/r^2$. Let us choose the classes in the structure so that $d_1 \geq d_2 \geq \dots \geq d_t$. We can define $d_i = 0$ for every $i > t$.

Consider the graph at the end of the game. Let it consist of m different types of structures D_i with r_i vertices forced to at most t classes and class sizes $d_1^i \geq d_2^i \geq \dots \geq d_t^i$ for every $i = 1, 2, \dots, m$ (where the i is an upper index). Let the vertex set of the graph be partitioned into $g_i n / r_i$ copies of structure of type i for

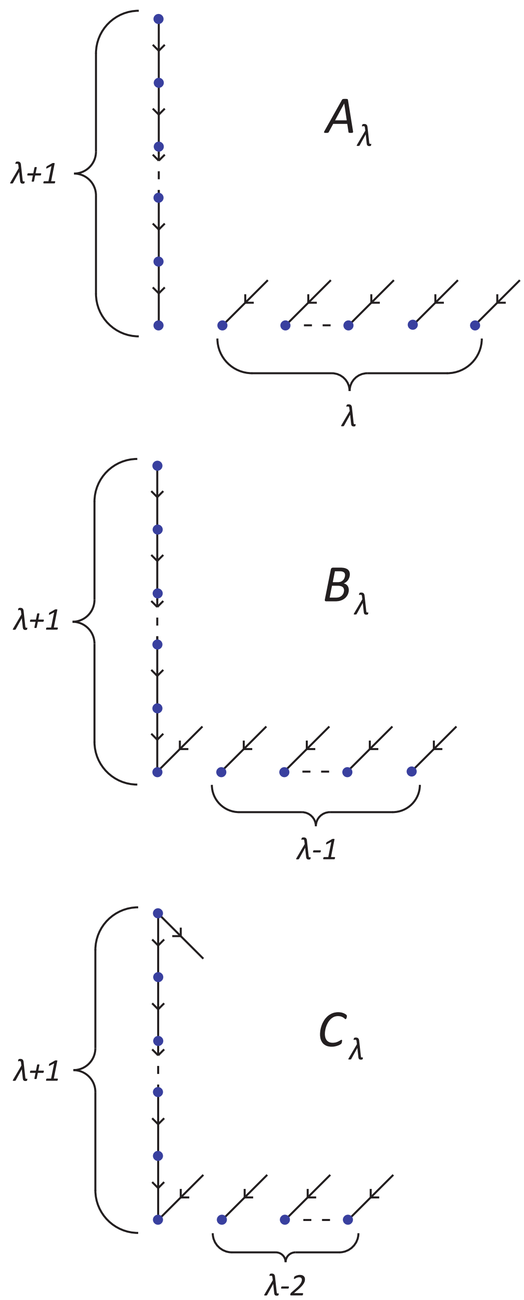

Figure 2.8: Structures A_λ, B_λ and C_λ

every i. Then the game score will be at least

$$\begin{aligned}
&\frac{1}{2}\Big(n^2 - \sum_{i=1}^{t}\big(\sum_{j=1}^{m} g_j n d_i^j / r_j\big)^2\Big) \\
=&\frac{n^2}{2}\Big(1 - \sum_{i=1}^{t}\big(\sum_{j=1}^{m} g_j d_i^j / r_j\big)^2\Big) \\
\geq&\frac{n^2}{2}\Big(1 - \sum_{i=1}^{t}\sum_{j=1}^{m} g_j \big(d_i^j / r_j\big)^2\Big) \\
=&\frac{n^2}{2}\Big(1 - \sum_{j=1}^{m} g_j \sum_{i=1}^{t} \big(d_i^j / r_j\big)^2\Big) \\
=&\frac{n^2}{2}\Big(1 - \sum_{j=1}^{m} g_j s(D_j)\Big)
\end{aligned}$$

where the inequality follows from the arithmetic mean-quadratic mean inequality with weights g_j.

We can compute

$$s(A_\lambda) = \frac{\lambda + (\lambda+1)^2}{(2\lambda+1)^2},$$

$$s(B_\lambda) = \frac{\lambda + \lambda^2}{(2\lambda)^2} = \frac{1}{4} + \frac{1}{4\lambda},$$

$$s(C_\lambda) = \frac{\lambda + (\lambda-1)^2}{(2\lambda-1)^2}.$$

It is beneficial for Prolonger to have structures of low normalized score. We shall prove that he can force only structures with normalized score $\frac{1}{3}$ or less. Note that the functions $s(A_\lambda)$, $s(B_\lambda)$ and $s(C_\lambda)$ are decreasing in λ when $\lambda \geq 1$. Note that $s(C_2) = \frac{1}{3}$, $s(B_3) = \frac{1}{3}$, $s(A_4) > \frac{1}{3}$ and $s(A_5) < \frac{1}{3}$. Also note that normalized score goes down if the path length increases and the number of non-path vertices stays the same.

At the end of the game the graph will consist of many copies of the structures A_λ, B_λ and C_λ (for different λ), a path of $k-1$ vertices and possibly one more vertex in either G_{k-2} or G_{k-1} and up to 9 more vertices. The 9 leftover vertices are the final ones touched and the only ones Prolonger cannot force to structures of normalized score $\leq \frac{1}{3}$ (the structure with the biggest number of vertices with normalized score greater than $\frac{1}{3}$ is A_4; it has 9 vertices).

Note that we can move between structures using the next move by Prolonger and then Shortener but only from A_λ through B_λ to C_λ and towards increasing λ.

Prolonger may create the first directed path $p := v_1 v_2 \ldots v_{k-1}$ of length $k-1$ exactly as described in the beginning of Shortener's strategy, so that when done there will be one vertex in each class G_i, $i \in \{1, 2, \ldots, k-1\}$ and possibly one extra vertex in either G_{k-2} or G_{k-1}.

We shall describe the structures that arise as Prolonger goes about building long paths.

If from some point on Shortener is allowed to play until the end without interference from Prolonger (or, technically, Shortener will make moves instead of Prolonger), then we say we are in the *endgame*. At this point it is the most beneficial to Shortener if, for example, he forces the endvertices of the paths to G_{k-1}, the second to last vertices to G_{k-2} etc. So if Prolonger can create the disjoint paths p_1, p_2, ..., p_m involving all vertices then the game score will be at least
$g := \frac{1}{2}(n^2 - \sum_{i=1}^{k-1} c_i^2)$ where $c_i = |G_i| = |\{j : |p_j| >= k - i\}|$.

Note that in the endgame, structure A_λ will contribute $\lambda + 1$ vertices to one class (for example G_{k-1}) and the other λ vertices to other individual classes; structure B_λ will contribute λ vertices to one class and the other λ vertices to other individual classes; structure C_λ will contribute $\lambda - 1$ vertices to one class and the other λ vertices to other individual classes.

Prolonger's strategy is to build isolated edges. When Shortener increases the in- or out-degree of a vertex of an isolated edge and

- if there is another isolated edge available Prolonger will attach that to the non-assigned vertex of the first isolated edge to form a path of 4 vertices and obtain structure B_3;
- otherwise, Prolonger will attach an isolated vertex to the non-assigned vertex of the first isolated edge to form a path of 3 vertices, and after the next move by Shortener obtain structure $C_{\geq 2}$ if Shortener touches this path again, or after the next moves by Shortener, Prolonger and Shortener obtain a structure at least as good as $B_{\geq 3}$.

If before a move by Prolonger there are two isolated edges, both with both vertices non-assigned then Prolonger attaches one edge to the other to form a path of 4 vertices; whatever Shortener does Prolonger can on the next move attach a path (maybe just a vertex) to the path to form a path on ≥ 5 vertices: if Shortener assigned a vertex of the path to a class we are in situation $B_{\geq 4}$ which is good for Prolonger (normalized score $\leq \frac{1}{3}$), otherwise if it is an A-structure, on the next step Prolonger can attach another vertex to the path to obtain an $A_{\geq 5}$ or join to some other vertex to obtain a $B_{\geq 5}$ if there are no isolated vertices which is again beneficial for Prolonger in terms of the score.

So at the end of the game there is a path on $k-1$ vertices whose vertices become joined to all other vertices on the path, every other vertex is joined to $k-2$ vertices of the path. All vertices not on the path are on structures of normalized score $\leq \frac{1}{3}$, except there may be one vertex left over to G_{k-2} or G_{k-1} and up to 9 more vertices on structures of normalized score $> \frac{1}{3}$. So the game score is at least

$$\binom{k-1}{2} + (n-k+1)(k-2) + 1/2(n-k-9)^2(1-1/3).$$

$\square$

Note that both the lower and upper bounds on the game score in the game $\mathcal{G}_{\mathrm{dir}}$ are

$$\frac{1}{3}n^2 + \frac{1}{3}kn + O(n + k^2).$$

So we have determined the game score asymptotically if $n \gg k$.

Chapter 3

Completing Partial Packings of Bipartite Graphs

3.1 Introduction

Let H be a simple graph. A *partial H-packing of order n*, or simply *H-packing*, is a set $\mathcal{P} := \{H_1, H_2, \dots, H_m\}$ of edge-disjoint copies of H whose union forms a simple graph on n vertices. We say that an H-packing of order n is *complete* or an *H-design* if the edge sets of $H_i,\ i = 1, \dots, m$ partition the edge set of the complete graph on n vertices. More generally, we say that a graph G can be *edge-decomposed* into copies of H if G is the union of some H-packing.

A long-standing problem in design theory is to find a way of completing an H-packing into an H-design of a larger size, using as few new vertices as possible. We define $f(n; H)$ to be the smallest integer such that any H-packing on n vertices, can be extended to an H-design on at most $n + f(n; H)$ vertices.

The existence of $f(n; H)$ for any n and H follows from Wilson's theorem [34], see Section 3.3 for details. Many bounds of the type of $f(n; H) \leq c(H)n$ have been proved for various graphs H by explicit constructions. A (by no means complete) list of references includes Hoffman, Küçükçifçi, Lindner, Roger, Stinson [20], [23], [24], [25], [26], [27], Jenkins [21], Bryant, Khodkar and El-Zanati [7]. See also Füredi and Lehel [13] for a survey of these results.

Hilton and Lindner [19] achieved a breakthrough, having proved a sub-linear bound on $f(n; H)$ for a particular H. More precisely, they showed that a C_4-packing can be completed by adding $O(n^{3/4})$ new vertices.

Füredi and Lehel [13] applied methods from extremal graph theory and managed to find the right order of magnitude for $f(n; C_4)$. They proved that

$$f(n; C_4) = \Theta(\sqrt{n}).$$

This settled the case $H = C_4$ and solved a problem proposed decades ago (see [27]). Based on their theorem, Füredi and Lehel [13] conjectured that for *any* bipartite graph H the packing can be completed by adding $o(n)$ new vertices. Our aim in this article is to give a proof of their conjecture.

Theorem 3.1. *For every bipartite graph H there is a function $f(n; H) = o(n)$ such that every H-packing of order n can be completed to an H-design on at most $n + f(n; H)$ vertices.*

In fact we determine the asymptotic growth of the function $f(n; H)$ *exactly.*

To present our main result, we need to define a new property of graphs. We say that a (not necessarily bipartite) graph H is *matching-friendly* if its vertex set $V(H)$ can be partitioned into V_1 and V_2 such that V_2 is an independent set of vertices and the induced graph $H[V_1]$ consists of a non-empty matching and a set of isolated vertices. For example, C_4 is not matching-friendly, but every other cycle is. The choice of the name 'matching-friendly' should become clear in the course of the proof.

Theorem 3.2. *If H is matching-friendly, then*

$$f(n; H) = \Theta(\mathrm{ex}\,(n, H)/n).$$

If H is not matching-friendly, then

$$f(n; H) = \Theta\left(\max\left\{\mathrm{ex}\,(n, H)/n, \sqrt{n}\right\}\right).$$

Here, as usual, $\mathrm{ex}\,(n, H)$ stands for the extremal number of H, see next section for its definition.

Theorem 3.2 applies to all graphs H, not just bipartite ones. However if H is not bipartite, it just states that $f(n; H) = \Theta(n)$. This is rather easy to deduce: take a packing $\mathcal{P}_n$, whose union consists of two complete graphs on $n/2$ vertices each. Such a packing exists for infinitely many values of n by Wilson's theorem, to be stated in Section 3.3. It is not hard to check that $\mathcal{P}_n$ needs $\Omega(n)$ vertices in order to be extended to an H-design. On the other hand, every H-packing can be extended to an H-design by adding $O(n)$ new vertices; this is a consequence of Gustavsson's theorem, to be stated in Section 3.3.

Thus from now on we shall assume that H is bipartite. Note that Theorem 3.2 implies Theorem 3.1.

3.2 Notation and basic Tools

As usual, we write $|G|$, $e(G)$, $\delta(G)$ and $\Delta(G)$ for the the number of vertices, number of edges, minimum degree and maximum degree of a graph G. These quantities will also be used for multigraphs and (multi)-hypergraphs. Denote by $N(v)$ the neighbourhood of v, excluding v.

Let K_n and $K_{m,n}$ denote the complete graph on n vertices and the complete bipartite graph with bipartition classes of size m and n. The graph $K_{1,k}$ is also called a *k-star*. It has a *central vertex* of degree k and k *endvertices* or *leaves* of degree 1.

The *degeneracy* of G is $\mathrm{dg}\,(G) := \max\,(\delta(G'))$, where the maximum is taken over all induced non-empty subgraphs G' of G. Suppose that the vertices of G are numbered $v_1, v_2, \ldots, v_n$, starting the numbering from v_n backwards, so that v_i is a minimum degree vertex of $G_{(i)} := G[v_1, \ldots, v_i]$, the subgraph of G induced by the vertices v_1 through v_i, for every $i = 1, 2, \ldots, n$. It is easy to see that $\mathrm{dg}\,(G) = \max \delta(G_{(i)})$.

A *transversal* of a graph G is a subset U of its vertices such that every edge of G has at least one endpoint in U. In other words, transversals are complements of independent sets. The *transversal number* $\tau(G)$ is the size of the smallest transversal of the graph G.

A graph G not containing H as a (not necessarily induced) subgraph is called *H-free*. Let us denote by $\mathrm{ex}\,(n, H)$ the *extremal number* for H, i.e. the maximum number of edges of an H-free graph on n vertices. More generally, let $\mathrm{ex}\,(G, H)$ be the maximum number of edges in an H-free subgraph of G. Then $\mathrm{ex}\,(n, H) = \mathrm{ex}\,(K_n, H)$. Also, if $F \subset H$ then $\mathrm{ex}\,(n, F) \leq \mathrm{ex}\,(n, H)$.

In our proof of Theorem 3.2 we shall use the following crude bound on *symmetric Zarankiewicz numbers* $z = z(m, n, s, s) = \mathrm{ex}\,(K_{m,n}, K_{s,s})$, see for instance [5].

Theorem 3.3. *For all $m, n \geq s$, and $s \geq 1$ we have*

$$z(m, n, s, s) \leq 2nm^{1-1/s} + sm.$$

It is a well-known fact that $z(n, n, s, s) \geq 2\mathrm{ex}\,(n, K_{s,s})$, see [5]. Since every bipartite graph H is a subgraph of $K_{s,s}$ for some s, it follows that an H-free graph G

on n vertices has at most $c(H)n^{2-\varepsilon(H)}$ edges, where $\varepsilon = \varepsilon(H)$ is a small positive number. Therefore $\delta(G) \leq cn^{1-\varepsilon}$. Furthermore, since a subgraph of an H-free graph is also H-free, we may conclude that $\mathrm{dg}\,(G) \leq cn^{1-\varepsilon}$. A more careful estimate on the degeneracy of an H-free graph is given by the following lemma.

Lemma 3.4. *For every H-free graph G,*

$$\mathrm{dg}\,(G) \leq \frac{4\mathrm{ex}\,(n,H)}{n} + 2|H| \leq C_H \frac{\mathrm{ex}\,(n,H)}{n}.$$

In other words, every H-free graph G of order $m \leq n$ has a vertex of degree at most $C_H \mathrm{ex}\,(n,H)/n$, where C_H is a constant that depends only on H.

Proof. The second inequality follows from the fact $\mathrm{ex}\,(n,H) \geq n/2$ for all graphs H containing more than one edge (the case $e(H) \leq 1$ is trivial), thus we can take $C_H = 4 + 4|H|$.

To prove the first inequality, notice that every H-free graph G on m vertices contains a vertex of degree at most $2e(G)/m \leq 2\mathrm{ex}\,(m,H)/m$. Hence, it suffices to show that

$$\frac{\mathrm{ex}\,(m,H)}{m} \leq \frac{2\mathrm{ex}\,(n,H)}{n} + |H| \tag{3.1}$$

for all $1 \leq m \leq n$. We claim that

$$\lfloor n/m \rfloor \cdot \mathrm{ex}\,(m,H) - |H| \cdot m \cdot \lfloor n/m \rfloor \leq \mathrm{ex}\,(n,H) \tag{3.2}$$

for all $1 \leq m \leq n$. It is easy to check that (3.2) implies (3.1), no matter if the left hand side is positive or not.

To see that (3.2) holds, consider an m-vertex H-free graph G with the maximum number of edges, and take $\lfloor n/m \rfloor$ of its vertex disjoint copies $G_1, G_2, \ldots$. If their union is H-free (e.g., in the case when H is connected) then

$$\lfloor n/m \rfloor \mathrm{ex}\,(m, H) \leq \mathrm{ex}\,(m\lfloor n/m \rfloor, H) \leq \mathrm{ex}\,(n, H).$$

If H is disconnected with components $C_1, ..., C_t$ then let s be the maximum integer that the graph F_s with components $C_1, ..., C_s$ appears in G; by assumption that G is H-free we have $s < t$. Let $G' = G \setminus V(F_s)$, that is remove F_s and all edges adjacent to it from G; we have deleted at most $m|H|$ edges. Then the graph comprising $\lfloor n/m \rfloor$ vertex disjoint copies of G' is F_{s+1}-free, and therefore H-free as well, which implies (3.2). $\square$

We shall need two basic facts about graph colouring. Their proofs can be found in any standard textbook on graph theory e.g. [6]. One is the fact that a graph of maximal degree Δ can be $\Delta + 1$-coloured by a greedy algorithm. The other theorem we need is Vizing's theorem: a graph of maximal degree Δ can be edge-coloured using $\Delta + 1$-colours or, equivalently, can be decomposed into $\Delta + 1$ matchings.

3.3 A Primer on Graph Decompositions

In this section we shall state various theorems on graph decompositions that we shall use in the proof of Theorem 3.2.

Let H be a bipartite simple graph of order d with vertices $v_1, v_2, \ldots, v_d$ and let $\deg(v_i)$ denote the degree of v_i. Denote $\gcd(H) = \gcd(\deg(v_1), \ldots, \deg(v_d))$. For an H-design of order n to exist we need the following obvious conditions:

$$e(H) | \binom{n}{2} \qquad \text{and} \qquad \gcd(H) | (n-1).$$

If these conditions hold we say that n is *H-divisible*. If n admits an H-design, we call it *H-admissible*. Wilson [34] proved the following fundamental theorem.

Theorem 3.5. *There exists an integer n_0, depending on H, such that every $n > n_0$ that is H-divisible is also H-admissible.*

Wilson's theorem implies that $f(n; H)$ exists for every H and n. Indeed, the union of an H-packing $\mathcal{P}$ on n vertices can be considered as our new 'building block' H'. By Theorem 3.5 there exists an H'-design $\mathcal{P}'$ for a sufficiently large H'-divisible number n'. By decomposing each copy of H' in $\mathcal{P}'$ into copies of H, we obtain an H-design on n' vertices. Since for a given n there are only finitely many H-packings on n vertices, and each of them can be completed to an H-design as above, $f(n; H)$ is well-defined.

More generally, let us say a graph G is *H-divisible* if all degrees of G are multiples of $\gcd(H)$ and $e(H)|e(G)$.

A very deep and powerful extension of Wilson's theorem was proved by Gustavsson [16].

Theorem 3.6. *For any digraph D there exist $\varepsilon_D > 0$ and $N_D > 0$ such that if G is a digraph satisfying:*

(a) $e(G)$ is divisible by $e(D)$;

(b) there exist non-negative integers a_{ij} such that

$$\sum_{v_i \in V(D)} a_{ij} d_D^+(v_i) = d_G^+(u_j), \qquad \sum_{v_i \in V(D)} a_{ij} d_D^-(v_i) = d_G^-(u_j)$$

for every $u_j \in V(G)$;

(c) if there exists $u_1\vec{u}_2 \in E(G)$ such that $u_2\vec{u}_1 \notin E(G)$ then there exists $v_1\vec{v}_2 \in E(D)$ such that $v_2\vec{v}_1 \notin E(D)$;

(d) $|V(G)| \geq N_D$;

(e) $\delta^+, \delta^- > (1 - \varepsilon_D)|V(G)|$

then G can be written as an edge-disjoint union of copies of D.

Viewing simple graphs G and H as digraphs, by orienting each edge in both directions, the above theorem translates to

Theorem 3.7. *For every H there exist m_0 and ε_0 such that every H-divisible graph G on $m > m_0$ vertices with minimum degree at least $(1 - \varepsilon_0)m$ can be edge-decomposed into copies of H.*

In the proof of Theorem 3.2 we shall need the analogue of Wilson's theorem for H-packings into complete bipartite graphs $K_{m,n}$, in which case the obvious divisibility conditions are

$$e(H)|mn \quad , \quad \gcd(H)|m \quad \text{and} \quad \gcd(H)|n.$$

Theorem 3.8. *Let H be a bipartite graph. There exists an integer n_0, depending on H, such that every H-divisible $K_{m,n}$ with $m, n > n_0$ can be edge-decomposed into copies of H.*

This was proved by Häggkvist [17] for the case when H is regular, $m = n$, and under stronger divisibility assumptions. However, Häggkvist's proof was before Gustavsson's theorem. With Theorem 3.6 at our disposal, we can give a proof of Theorem 3.8. While it is almost certain that its statement has been well-known, we could not find any explicit reference. Thus, we shall give a proof sketch, skipping some technical details.

Proof. First suppose that $m = n$. The graph $K_{n,n}$ on vertices $[n]$ and $\{1', ..., n'\}$ can be thought of as a directed graph with loops on $[n]$ by replacing each edge ab' with a directed edge a to b. By embedding H so that the bipartite classes of H are sent to disjoint subsets of $[n]$ we can regard H as a directed graph H' without loops. By removing n copies of H from $K_{n,n}$ first, where each copy has exactly one 'vertical' edge, we reduce to the case of decomposing a dense digraph G (without loops) into copies of the digraph H'. Here 'dense' means that we must ensure that $\delta^{\pm}(G) > (1-\varepsilon)n$. The packing of G can be done provided (a) n is large enough; (b) the number of edges is divisible by $e(H)$; and (c) the in- and out-degrees of any vertex of G are representable as a non-negative linear combination of the in- and out-degrees of vertices of H'. This last condition should translate to the assumption than n is divisible by both the gcd of the degrees of the vertices in A and the gcd of the degrees of the vertices in B, where (A, B) is the bipartition of H. (This assumes one wants to pack all the copies of H the same way round. If not, pack $H \cup H^r$ where H^r is H with the bipartition reversed, and possibly remove one extra copy of H initially to ensure that $2e(H)$ divides $e(G)$. Then n needs only be divisible by the gcd(H).)

So there is an integer n'_0 such that the theorem holds for all $K_{n,n}$ with $n > n'_0$. In fact, the same construction works for $K_{m,n}$ if $n \leq m \leq (1+\varepsilon'(H))n$. To see this, remove some copies of H in order to isolate $m-n$ vertices in the larger partition class, making sure that we do not reduce the degrees of the remaining vertices too much. Having done that, apply the above digraph reduction to the remaining graph, which can be viewed as a subgraph of $K_{n,n}$. Then apply Theorem 3.6 as above.

Given $m, n \geq n_0 = (n'_0)^2$, we can partition both sets $\{1, ..., m\}$ and $\{1, ..., n\}$ into subsets of size about n_0 each and such that each complete bipartite graph (X, Y) induced on two partition classes $X \subset \{1, ..., m\}$ and $Y \subset \{1, ..., n\}$ is H-divisible. Pack every such graph with copies of H as described above. $\square$

This theorem does not have the full strength of Gustavsson's theorem but it is enough for our purposes to decompose complete bipartite graphs.

3.4 Upper bound: Outline of the Proof

In this section we would like to describe our strategy for proving the upper bound in Theorem 3.2.

Consider an H-packing $\mathcal{P} = \{H_1, H_2, \dots, H_m\}$ on n vertices. We want to complete it to an H-design by adding few vertices. We consider the *uncovered graph* $G_0 = (K_n) \setminus \cup_{i=1,\dots,m} E(H_i)$ i.e. the graph consisting of edges that are not covered by copies of H.

We proceed in three steps:

Step 1: Reducing the transversal. We add some new vertices and all possible edges from those to other vertices. Now we delete an edge-disjoint collection of copies of H from the resulting graph, so that the resulting graph has a smaller transversal than the graph we started with. This step constitutes a major part of the proof of Theorem 3.2 and will be carried out in Sections 3.5 through 3.7.

More precisely, in Section 3.5 we shall construct a 'nice' collection of disjoint k-stars on the edges of any given graph G. This construction will be applied in Section 3.6 to G_0 in order to construct a hypergraph M with a small edge-chromatic number, related to G_0. Then in Section 3.7 we shall use M and its edge-colouring in order to extend $\mathcal{P}$ to a packing on a larger vertex set, such that the uncovered graph has a small transversal.

In Section 3.8 we shall describe how we iterate Step 1 in order to obtain further packings with yet smaller transversals of the uncovered graphs.

Step 2: Decreasing the number of uncovered edges. Starting with an uncovered graph G_1 that has a small transversal we extend the new packing to obtain a new uncovered graph G_2 with very few edges. This will be established in Section 3.9.

Step 3: Completing the packing. This will be done by applying Theorem 3.7 and Theorem 3.8 in Section 3.10.

3.5 Degeneracy

The aim of this section is to prove Proposition 3.9: this will be our main tool for reducing the transversal of the uncovered graph. We also believe that the statement of Proposition 3.9 is interesting in its own right; see Section 3.12 for related questions.

Recall that a k-star is a copy of $K_{1,k}$.

Proposition 3.9. *For every integer k and a graph G of degeneracy d there is a maximal collection $\mathcal{C}$ of edge disjoint k-stars on G such that each vertex of G is an endvertex to at most $d + k - 1$ stars in $\mathcal{C}$.*

Case $k = 2$ was proved by Füredi and Lehel [13]. We are following their approach, using downdegree instead of updegree since this feels more natural to us. Let us choose an ordering $v_1, v_2, \ldots, v_n$ of vertices of G_0 such that the *(maximum) downdegree* $\overleftarrow{\Delta}(G_0)$, defined as the maximum of the number of edges from a vertex v_i to vertices $v_j,\ j < i$, over all $i = 1, 2, \ldots, n$, equals $d = \mathrm{dg}\,(G)$.

Let us construct $\mathcal{C}$ as follows: take a maximal collection of edge-disjoint k-stars whose central vertex is smaller in the given ordering than any of its endvertices, and then extend it to a maximal collection of edge-disjoint k-stars. Then $u \in G$ appears as an endvertex of a star of the first kind, or as such endvertex of a star of the second kind which is greater than its centre at most $\overleftarrow{\Delta}(G)$ times. It appears as an endvertex smaller than the centre of a star of the second kind at most $k - 1$ times since otherwise we could form a star of the first kind with u at its centre – this is a contradiction as we started taking stars of the second kind in a graph containing no stars of the first kind.

It follows that u can appear at most $\overleftarrow{\Delta}(G) + k - 1 = d + k - 1$ times as an endvertex of a star in $\mathcal{C}$, which proves Proposition 3.9. Note that the maximality of $\mathcal{C}$ implies $\Delta(G \setminus \bigcup \mathcal{C}) \le k - 1$.

3.6 A Hypergraph and its Colouring

We continue carrying out the plan outlined in Section 3.4. Recall that we are given an H-packing $\mathcal{P} = \{H_1, H_2, \dots, H_m\}$ on n vertices and
$G_0 = (K_n) \setminus \cup_{i=1,\dots,m} E(H_i)$ is our uncovered graph.

In this section we shall give a construction of a certain hypergraph M on a vertex set of G_0 along with its edge-colouring; we shall need it in order to extend $\mathcal{P}$ to a packing on a larger set of vertices, in which the uncovered graph will have a small transversal.

First of all, we can assume without loss of generality that G_0 is H-free (by removing a maximal set of edge-disjoint copies of H from G_0). By Lemma 3.4 we know that $\mathrm{dg}\,(G_0) = O(\mathrm{ex}\,(n, H)/n)$.

For a fixed vertex v of H, let $k = \deg(v)$ and $W_1 = N(v)$. Let (U, W) be a bipartition of H such that $v \in U$ and $W_1 \subset W$. Denote by R the ratio $|W|/|W_1|$, rounded up to the nearest integer. Let $s = |U|$ and $t = |W|$ be the sizes of the bipartition classes. For convenience we can assume that $s \ge t$, perhaps choosing another v.

By Proposition 3.9 there is a collection $\mathcal{C}$ of disjoint k-stars on G_0 with the property that each vertex of G_0 is an endvertex to at most $\mathrm{dg}\,(G_0) + k - 1$ stars in $\mathcal{C}$. Define a multi-k-graph (k-uniform hypergraph with several edges on the same set of vertices allowed) called M as follows: for every star of $\mathcal{C}$ there is a k-edge containing precisely the leaves of the star. The maximum degree $\Delta(M)$ (i.e. the maximum number of edges containing any given vertex) is bounded by $\mathrm{dg}\,(G_0) + k - 1 \leq c_3 * \mathrm{ex}\,(n, H)/n$, where c_3 is a positive constant depending only on H. We shall denote edges of M by (c, e) where $c \in G_0$ is the centre of the respective star and e is the hyperedge consisting precisely of the leaves of the star.

Let us introduce an edge-colouring on M so that each colour class forms a vertex-disjoint collection of hyperedges. Since every hyperedge intersects at most $k(\Delta(M) - 1)$ other hyperedges, it can be done, using at most $k(\Delta(M) - 1) + 1 = c_4 * \mathrm{ex}\,(n, H)/n$ colours: let us colour greedily as many hyperedges with colour 1 as we can, then with colour 2 and so on (again c_4 is a positive constant depending only on H).

Split every colour class i into $R = \lceil |W|/|W_1| \rceil$ (almost) equal parts $i.1$ through $i.R$. For every colour class i, fix a map σ_i which, for every j, takes hyperedges coloured $i.j$ to disjoint $|W_1|(R-1)$-subsets of vertices inside the union of hyperedges coloured with one of the colours $i.l$, $l \neq j$. Note that this mapping takes hyperedges into sets which are disjoint from the hyperedge itself.

Now we are ready to extend $\mathcal{P}$ in order to reduce G_0 to a new uncovered graph G_1 that has a new transversal.

3.7 Construction of a transversal

We shall prove that, by adding a small set of new vertices Q, we can use up all the edges inside G_0 in edge-disjoint copies of H and end up with a graph G_1 on the vertex set $V \cup Q$ with no edges inside V (i.e. with transversal Q).

The following construction decreases the degrees of the vertices in V below k.

Construction 1. Covering all k-stars. Write $V = V(G_0)$. Consider $v \in H$, $k = \deg(v)$, the bipartition $H = (U, W)$ and the colouring of the multihypergraph M

as before. For every colour $i.j$ add to G_0 a set $Q^{i.j} = q_1^{i.j}, \ldots, q_{|U|-1}^{i.j}$ of $|U|-1$ new vertices and place a copy of $H = (U, W)$ in the obvious way on every star (c, e) of colour $i.j$ such that $U = \{c, q_1^{i.j}, \ldots, q_{|U|-1}^{i.j}\}$ and $W \subset e \cup \sigma_i(e)$ (if $|W|$ is divisible by $|W_1|$ then we have $W = e \cup \sigma_i(e)$). Note that the sets $e \cup \sigma_i(e)$ for different hyperedges e of colour $i.j$ are pairwise disjoint and so the copies of H are placed edge-disjointly. We needed $O(\mathrm{ex}\,(n, H)/n)$ new vertices.

The following construction takes care of all the edges from within V.

Construction 2. Covering the remaining edges. By Vizing's theorem, the set of remaining edges inside V can be partitioned into (at most) k matchings $L_1, \ldots, L_k$. Consider the smallest r such that $\binom{r}{2} \geq e(H)\frac{n}{2}$ and K_r can be packed completely with copies of H. By Theorem 3.5 we can pick $r = O(\sqrt{n})$. For each matching L_i, add to G_0 a set Q^{L_i} of r new vertices, and pack the copies of H into $K_r \cup L_i$ so that the packing is almost like the complete packing of K_r, except with all edges in L_i covered by an edge from different copies of H. This way we clearly pack copies of H edge-disjointly. Note that $|Q^{L_i}| = O(n^{1/2})$ for every i, so we need $O(n^{1/2})$ new vertices for this construction.

However, if H is matching-friendly, we can do much better. Recall, H is matching-friendly if $V(H)$ can be partitioned into V_1 and V_2, where V_2 is independent and V_1 is 'almost' independent, i.e. the V_1-induced subgraph of H is a non-empty matching and some isolated vertices. This implies that we can cover at least one edge of an uncovered matching L_i by adding $|V_2|$ new vertices such that no edge between the new vertices will be used. It follows easily that the whole L_i can be covered using at most a constant number of $c(H)$ new vertices.

Let $Q = \cup_{i,j} Q^{i.j} \bigcup \cup_i Q^{L_i}$. We have constructed a graph G_1 on vertex set $V \cup Q$ with transversal Q. By removing copies of H, we can assume that G_1 is H-free. For the number of added vertices we have the bound $|Q| \leq c_5 * \max\,\{\mathrm{ex}\,(n, H)/n, \sqrt{n}\}$. If H is matching-friendly, we obtain $|Q| \leq c_6 * \mathrm{ex}\,(n, H)/n$.

3.8 Further transversals

We can add some more vertices to G_1 to reduce the transversal number of the resulting graph even further. This procedure can be repeated many times.

It suffices to prove the following lemma.

Lemma 3.10. *Let G be an H-free graph on n vertices, containing a transversal Q of size $q = o(n)$. Then there is an ordering $v_1, \dots, v_n$ of the vertices of G such that $\overleftarrow{\Delta}(G) \le Cq^{1-\varepsilon}$, where C and $0 < \varepsilon = \varepsilon(H) < 1$ are constants depending only on H. In particular,* $\mathrm{dg}\,(G) \le Cq^{1-\varepsilon}$.

Proof. Let us write $Y = V(G)\backslash Q$ and consider the bipartite graph G' with bipartition (Y, Q), whose edges are the edges of G having precisely one vertex in each of Q and Y. Let $G'' = G[Q]$ be the subgraph of G induced by Q. Then the edge sets of G' and G'' partition the edge set of G.

Since G'' is an H-free graph on q vertices, its degeneracy is at most $c''q^{1-\varepsilon}$ for a positive constant c'' depending only on H. Let us fix an ordering $u_1, u_2, \dots, u_q$ of the vertices in Q such that $\overleftarrow{\Delta}(G'') = \mathrm{dg}\,(G'')$.

Select s and t with $s \ge t$ such that $H \subset K_{s,t} \subset K_{s,s}$ and s is chosen as small as possible. By Theorem 3.3 we have that

$$z(|Q|, |Y|, s, s) \le 2|Y||Q|^{1-1/s} + s|Q|.$$

Let $\varepsilon \le 1/s$. We find that

$$\mathrm{ex}\,(K_{|Q|,|Y|}, H) \le \mathrm{ex}\,(K_{|Q|,|Y|}, K_{s,s}) = z(|Q|, |Y|, s, s) \le 2|Y||Q|^{1-1/s} + s|Q|.$$

Therefore, as long as $|Y| \ge q^{1/s}$, the minimal degree in Y satisfies $\delta(Y) = O(q^{1-1/s})$.

Let v_1 be a vertex of Y of smallest possible degree in the graph G', let v_2 be a vertex of Y of minimal degree in $G'[V(G)\backslash\{v_1\}]$, take v_3 to be a vertex of Y of minimal degree in $G'[V(G)\backslash\{v_1, v_2\}]$ and so on, until v_r, where $r = |Y| - q^{1/s}$. Each of those degrees is $O(q^{1-\varepsilon})$, by the previous paragraph. Let $v_{r+1}, v_{r+2} \ldots v_{n-q}$ be the remaining vertices in Y.

Define the ordering $v_{r+1}, v_{r+2} \ldots v_{n-q}, u_1, u_2, \ldots, u_q, v_1, v_2, \ldots, v_r$. It follows from the construction that $\overleftarrow{\Delta}(G') \leq c' q^{1-\varepsilon(H)}$. $\square$

The lemma allows us to iterate the construction of Sections 3.6 and 3.7. An H-free uncovered graph with a transversal of size q has by Lemma 3.10 degeneracy $c'q^{1-\varepsilon}$. Hence we can define a hypergraph as in Section 3.6 and use it in order to construct a new packing as in Section 3.7. The number of new vertices needed in Construction 1 will be $O(q^{1-\varepsilon})$ and in Construction 2 of Section 3.7 each matching has cardinality at most q, so we need to add a set Q^{L_i} of $O(q^{1/2})$ additional vertices for every matching L_i. Hence, the total number of new vertices will be at most $C(H)q^{1-\varepsilon(H)}$. By construction, this set of vertices will be a transversal of the new packing, so we can just repeat the procedure, using the new transversal. We iterate as long as $Cq^{1-\varepsilon} \leq q/2$, that is $q \geq C'(H) = (2C)^{1/\varepsilon}$. The number of new vertices halves after each step, thus by adding $O(\max\{\mathrm{ex}(n, H)/n, \sqrt{n}\})$ new vertices, or $O(\mathrm{ex}(n, H)/n)$ if H is matching-friendly, we can make the transversal smaller than the constant $C'(H)$.

3.9 Decreasing the number of uncovered edges

Our next objective is to reduce the number of uncovered edges. Furthermore we shall make sure that the number of vertices in the uncovered graph is congruent 1 modulo $e(H)$. This will be needed later for completing the packing.

Write G_2 for the uncovered graph with $Q \subset V(G_2)$ a transversal and $Y = V(G_2) \backslash Q$. As we know from Section 3.8, we may assume that $|Q| < C'(H)$. Define $g = \gcd(H)$. By adding a few new vertices to Q we may also assume that $|G_2| \equiv 1 \mod e(H)$. Since G_2 is the complement of a partial packing and $g|e(H)$ (because H is bipartite), all degrees in G_2 must be multiples of g. This implies that every vertex in Y is either isolated or has at least g neighbours in Q. We shall add a set Z of new vertices of size $m|Q|^g$ in order to reduce G_2 to a graph G_3 in which every subset of vertices of Q of size g has at most m common neighbours in Y and every vertex in Y has either none or at least g neighbours in Q. That would bound the number of edges between Y and Q by $m|Q|^g$. In addition every vertex from Z will have at most m uncovered edges in Y incident with it. Then G_3 would have at most $m|Q|^g + m|Z| + 1/2(|Z| + |Q|)^2 = C''(H)$ edges.

Let $m = 2n_0$, where n_0 a multiple of $e(H)$ that satisfies Theorem 3.8 for H, that is any H-divisible complete bipartite graph with at least n_0 vertices in each partition class can be edge-decomposed into copies of H.

Let us pick a set $K = \{q_1, q_2, \ldots, q_g\}$ of some g vertices in Q and write N for their common neighbourhood in Y: $N = N(q_1) \cap N(q_2) \cap \ldots \cap N(q_g) \cap Y$. If $|N| > m$, we are going to add to G_2 an additional set $Q^*_{q_1,\ldots,q_g} = Q^* = \{q_1^*, q_2^*, \ldots, q_m^*\}$ of m vertices. If $|N| \leq m$, we just pick the next K.

We are going to cover almost all the edges in the complete bipartite graphs $(K \cup Q^*, N)$ and $(Q^*, Y \backslash N)$. Since $|Q^*|$ and $|K \cup Q^*|$ are both divisible by g, to make those graphs H-divisible, it suffices to omit less than $e(H)$ vertices from each of the sets N and $Y \backslash N$ — so that we obtain respectively sets N' and Y'. By Theorem 3.8 it follows that both complete bipartite graphs $(K \cup Q^*, N')$ and (Q^*, Y') can be packed completely with edge-disjoint copies of H.

The uncovered graph has obtained m new vertices, each of which has at most m (in fact at most $2e(H)$) uncovered edges into Y and the vertices in Q^* have now at most m common neighbours inside Y. Also, for each vertex in Y, the number of its remaining neighbours in Q is a multiple of g.

If we repeat the procedure for all possible sets $K \subset Q$ of size g, we obtain the desired graph G_3, taking Z to be the union over all K. Notice also that by adding m vertices at a time, we make sure that $|G_3| \equiv 1 \mod e(H)$.

3.10 Completing the packing

We shall now apply Theorems 3.7 and 3.8 to complete the packing. Since the uncovered graph G_3 has a constant number of edges, the number of non-isolated vertices in it is also constant. Let Q be a set of vertices of size $C_3(H)$ such that all vertices in $Y = G_3 \setminus Q$ are isolated and $|Y| \equiv 0 \mod e(H)$; hence also $|Y| \equiv 0 \mod g$, where g is the greatest common divisor of all degrees in H, as before. By the construction in the previous section we may assume that $|Q| + |Y| = |G_3| \equiv 1 \mod e(H)$, thus $|Q| \equiv 1 \mod e(H)$.

We now apply Theorem 3.7 to $G_3[Q]$ to extend the packing by adding a set X of few new vertices. More precisely, we pick X to be a set of new vertices of size $\max\{m_0, (1/\varepsilon_0)|Q|\}$, where m_0 and ε_0 are as in Theorem 3.7, this is a constant of H. Also let $|X| \equiv |Y| + |Q| - 1 \mod 2e(H)$. To complete the packing it suffices to make sure that the uncovered graph on $Q \cup X$ and the complete bipartite graph $K_{X,Y}$ are H-divisible.

One divisibility condition requires $|X| + |Q| \equiv 1 \mod g$ for the former graph and $|X|, |Y| \equiv 0 \mod g$ for the latter. Both conditions are satisfied since $|X| \equiv 0 \mod g$.

The other divisibility condition requires the number of edges in each graph to be divisible by $e(H)$. This is certainly true for $K_{X,Y}$, by the choice of Y. So we only need to make sure that $e(H)$ divides the number of edges of the uncovered graph on $Q \cup X$, in other words

$$e(H) \Big| \left(\binom{|X| + |Q|}{2} - \binom{|Q|}{2} + e(G_3) \right).$$

Since G_3 is the complement of an H-packing, we know that

$$e(G_3) \equiv \binom{|Q| + |Y|}{2} \mod e(H).$$

Therefore we need $e(H)$ to divide

$$\binom{|X| + |Q|}{2} - \binom{|Q|}{2} + \binom{|Q| + |Y|}{2} = \binom{|X| + |Y| + |Q|}{2} - |X||Y|.$$

This is true whenever $|X| \equiv |Y| + |Q| - 1 \mod 2e(H)$.

Hence we can satisfy all divisibility conditions in order to apply Theorems 3.7 and 3.8 to complete the packing. This finishes the proof of the upper bound in Theorem 3.2.

3.11 Lower bound

In this section we want to show the existence of H-packings that need $\Omega(\mathrm{ex}\,(n, H)/n)$ vertices in order to be completed. If H is not matching-friendly, there exist also packings that need $\Omega(\sqrt{n})$ new vertices.

Let us start with the second claim. If H is not matching-friendly, we need $\Omega(\sqrt{n})$ new vertices in order to cover the edges of a complete matching L on n vertices. Indeed, any time we place a copy of H that covers at least one edge of L, we must use an edge between two new vertices (otherwise H would be matching-friendly). Hence, in order to cover $n/2$ edges of L we need about $\sqrt{n}$ new vertices.

Now we have to make sure that the complement of a perfect matching *is* the union of an H-packing for infinitely many n. Take two disjoint copies of H and view their union H' as a bipartite graph with equal partition classes, i.e. one copy of H is 'upside down'. Let s be the size of the partition classes. By Theorem 3.5, if n is sufficiently large, there is a complete packing $\mathcal{P}$ of K_n with copies of H'. Now take two identical copies of $\mathcal{P}$, one on $\{a_1, a_2, \ldots a_n\}$ and another on $\{b_1, b_2, \ldots b_n\}$ and add a copy of H' between $a_{i1}, \ldots, a_{is}$ and $b_{j1}, \ldots, b_{js}$ and another one between $a_{j1}, \ldots, a_{js}$ and $b_{i1}, \ldots, b_{is}$ for each copy of H' in $\mathcal{P}$ between $a_{i1}, \ldots, a_{is}$ and $a_{j1}, \ldots, a_{js}$, in the obvious way. We obtain a packing on $2n$ vertices, whose union is the complement of a matching between vertices a_i and b_i.

Now let us prove the first claim. Suppose we have found an H-packing $\mathcal{P}$, whose complement is an H-free graph with about $\mathrm{ex}\,(n, H)$ edges. In order to cover each edge of it, every copy of H would use at least one out of $kn + \binom{k}{2} = (1 + o(1))kn$ new edges, where k is the number of new vertices. Since we need at least $\mathrm{ex}\,(n, H)/e(H)$ copies of H to cover all edges of the uncovered graph, we must have $k = \Omega(\mathrm{ex}\,(n, H)/n)$.

Hence, it remains to prove that such a packing $\mathcal{P}$ exists for arbitrarily large values of n. Take an (extremal) H-free graph $\overline{G}$ on n vertices with $\mathrm{ex}\,(n, H)$ edges. We would like to remove a small proportion of edges from $\overline{G}$ in order to make the complement of the remaining graph satisfy the conditions of Theorem 3.7. This would ensure the existence of the desired packing.

Let us first eliminate vertices of high degree. Suppose $\overline{G}$ has $\log n$ vertices of degree at least $\varepsilon_0 n$, where ε_0 is as in Theorem 3.7. Then by Theorem 3.3, for a sufficiently large n the bipartite graph between $m = \log n$ such vertices and the rest of $\overline{G}$ contains $K_{s,s} \supset H$, contradicting the assumption that $\overline{G}$ is H-free. It follows that $\overline{G}$ has less than $\log n$ vertices of degree at least $\varepsilon_0 n$.

Removing them, we lose at most $n \log n$ edges obtaining (unless H is a forest, in which case there is nothing to prove) a new H-free graph $\overline{G'}$ with $(1 - o(1))\mathrm{ex}\,(n, H)$ edges and no vertices of high degree.

Next we would like to remove a few more edges from $\overline{G'}$ in order to fulfil the divisibility conditions. A theorem of Pyber [28] states that a graph F that has at least $n \log n * 32r^2$ edges contains a (not necessarily spanning) r-regular subgraph. Let us set $r = 2e(H)$. Remove edge sets of r-regular subgraphs $G_1 \subset \overline{G'}$, $G_2 \subset \overline{G'} \setminus G_1$ etc. until the remaining graph $\overline{G'} \setminus (G_1 \cup G_2 \cup \cdots \cup G_k)$ has less than $n \log n * 32r^2$ edges. Then the graph $\overline{G''} = G_1 \cup G_2 \cup \cdots \cup G_k$ satisfies all conditions of Theorem 3.7 and still has about $\mathrm{ex}\,(n, H)$ edges, whence we obtain the desired packing $\mathcal{P}$.

3.12 Outlook

There is a simple sufficient condition for a graph H to be *not* matching friendly: H cannot be matching-friendly if every edge of it is contained in a 4-cycle. However, in this case, since $C_4 \subset H$ and $\mathrm{ex}\,(n, C_4) = \Theta(n^{3/2})$, we obtain $\mathrm{ex}\,(n, H)/n = \Omega(n^{1/2})$, thus being not matching-friendly does not matter, as far as Theorem 3.2 is concerned. There are examples of bipartite graphs that are not matching-friendly and C_4-free; take for instance C_8 and connect the opposite pairs of vertices by paths of length 2. Or, alternatively, take the incidence graph of the Fano plane. However, we do not know much about the extremal numbers of such graphs, so the question is: **does 'matching-friendly' ever make a difference?** In other words, is it always true that $\mathrm{ex}\,(n, H) = \Omega(n^{1/2})$ for a non-matching-friendly graph H? If this is indeed the case, then the statement of Theorem 3.2 would simplify to $f(n; H) = \Theta(\mathrm{ex}\,(n, H)/n)$ for *all* graphs H.

The constant C_H in the proof of Lemma 3.4 depends on H only when H is a disconnected forest. Is it possible to prove Lemma 3.4 with an absolute constant, perhaps even $C_H = 2 + o(1)$?

The following question was inspired by Proposition 3.9. We believe it is interesting in its own right.

Conjecture 3.11. *For every integer k and a graph G of degeneracy d there is a maximal collection $\mathcal{C}$ of edge disjoint paths of length $2k$ on G such that each vertex of G is an endvertex to at most $c_k d$ paths in $\mathcal{C}$ for some constant c_k depending only on k.*

This cannot hold for odd-length paths, as can be seen by taking, for instance paths of length 3 and $G = K_{2,m}$, where m is large. Case $k = 1$ of Conjecture 3.11 is the special case of Proposition 3.9; it was first proved by Füredi and Lehel [13]. It seems likely that using an elaboration of the method of proof of Proposition 3.9, one can also prove Conjecture 3.11 for $k = 2$ and $k = 3$. However, for $k \geq 4$ one would probably need a genuinely different approach.

More generally, can the $2k$-path in the statement of Conjecture 3.11 (or Proposition 3.9) be replaced by a tree, in which all distances between the leaves are even?

Chapter 4

Slowgrow Sequence

4.1 Introduction: the Slowgrow Sequence

Steven Kalikow introduced the following infinite sequence of positive integers for potential study. We shall denote it S and call it the *Slowgrow sequence.*

Intuitively, we build the sequence up inductively. Let $s_1 = 1$. For s_i, $i \geq 2$, write the smallest positive integer m such that the block of terms $s_{i-m+1}s_{i-m+2}\dots s_{i-1}m$ of length m ending in that m has not appeared earlier.

Set $s_1 = 1$. Having defined $s_1, s_2, \dots, s_n$, let

$$\begin{aligned} s_{n+1} = \min \{ m | (& s_{n-m+2}, s_{n-m+3}, \dots, s_n, m) \\ & \neq (s_{n-m+2-i}, s_{n-m+3-i}, \dots, s_{n-i}, s_{n+1-i}) \\ & \forall i \in [n-m+1] \}. \end{aligned}$$

Let us call the sequence $S = (s_1, s_2, \dots) = s_1 s_2 \dots$ the *Slowgrow sequence.* A *block of terms* of a sequence is a finite subsequence consisting of consecutive terms. In words, having defined the first n terms of the Slowgrow sequence, the $(n+1)$st term is the

Table 4.1: The first 1000 terms of the Slowgrow sequence, 60 per row

122323334234334434444523445334543455344544455445554545555562
345556334556434564445653455654456634455664455672345655455665
455673345665555655566645456564555667434566655656556666565665
666666753456675445674445666676345566676445667654566677344566
774455675545667754556765556766455667665566776455677555566777
454566782345667775565666776556678334567564565667674555677665
667776565675655676755566784345676665667853456776756656766755
666777565667784445676765666786345667864456777666666777854456
778554567677566666778654567777667666677873455677777676667667 6
766776667778664566678744566787545667883445677866556777875556
777884455677876456677877455667788545567788645567855556778874
556778923456777886555677893345677877556677894345677888454567
788944456778895345678655667876556778885556678855656767766777
777786745656776767677767767777787665677778776566778875566788
665667877666787775667677778876566788766667878455567866656778
785566678877566777788865656777887766778786566678886666678887
565677888766767778675567677787776677 88885

smallest positive integer m such that the suffix $(s_{n-m+2}, s_{n-m+3}, \ldots, s_n, m)$ of the Slowgrow sequence of length m ending in that m is a block of terms that has not appeared earlier in the Slowgrow sequence.

The first 1000 terms of the Slowgrow sequence that are all 1-digit numbers are given in Table 4.1. The sequence starts as 1223233342 . . . . For illustrative purposes let us verify how this was calculated. The second term of the sequence cannot be 1 since 1 has occurred in the sequence earlier. The second term of the sequence is 2 since this is the first occurrence of 12 in the sequence. The third term cannot be 1 since 1 has occurred. It will be 2 since it is the first occurrence of 22. The fourth term of the sequence cannot be 1 since 1 has occurred earlier; it also cannot be 2 since 22 has occurred earlier; it will be 3 since this is the first occurrence of 223. The fifth term of the sequence cannot be 1 since 1 has occurred earlier; it will be 2 since it is the first occurrence of 32. The sixth term cannot be 1 since 1 has occurred earlier, it cannot be 2 since 22 has occurred earlier, it will be 3 since it is the first occurrence of 323. The seventh and eighth term are found similarly to be 3s. The ninth term is not 1 since 1 has occurred earlier, not 2 since 32 has occurred earlier, not 3 since 333 has occurred earlier, so it is 4 since it is the first occurrence of 3334. The tenth term is not 1 since 1 has occurred earlier, it is 2 since it is the first occurrence of 42.

Note that the Slowgrow sequence is defined to avoid repetitions of certain kinds of blocks. Several sequences which avoid some kind of repetitions have been considered previously, most notably the Ehrenfeucht-Mycielski sequence, see [9] and the Linus sequence, see [2]. Both of the latter are sequences of binary digits 0 and 1. Both sequences are useful in ergodic theory.

The Ehrenfeucht-Mycielski sequence starts with the single bit 0; each successive digit is formed by finding the longest suffix of the sequence that also occurs earlier within the sequence, and complementing the bit following the most recent earlier occurrence of that suffix (the empty block "" is a suffix). The Ehrenfeucht-Mycielski sequence starts as $0100110\ldots$. For illustrative purposes let us verify how this was calculated. The second digit is a 1: the suffix "" of "0" occurs earlier, most-recently followed by a 0, so add the complement of 0 which is 1. The third digit is a 0: the suffix "" of "01" occurs earlier, most recently followed by a 1, so add 0. The fourth digit is a 0: the suffix "0" of "010" occurs earlier, most recently followed by a 1, so add 0. The fifth digit is a 1: the suffix "0" of "0100" occurs earlier, most recently followed by a 0, so add 1. The sixth digit is a 1: the suffix "01" of "01001" occurs earlier, most recently followed by a 0, so add 1. The seventh digit is a 0: the suffix "1" of "010011" occurs earlier, most recently followed by a 1, so add 0.

The Linus sequence is built similarly to the Ehrenfeucht-Mycielski sequence, except here we want to minimize the longest suffix that also occurs immediately preceding the suffix. Define the Linus sequence $(L_n)_{n\geq 1}$ as a 0-1 sequence with $L_1 = 0$ and L_n chosen so as to minimize the length r of the longest immediately repeated block $L_{n-2r+1}\ldots L_{n-r} = L_{n-r+1}\ldots L_n$. The Linus sequence starts as $0100110\ldots$. For illustrative purposes let us verify how this was calculated. The second digit is a 1:

adding a 0 would produce the block "0" immediately preceded by the block "0", but adding a 1 produces the empty block "" immediately preceded by the empty block "". The third digit is a 0: adding a 1 would produce the block "1" immediately preceded by the block "1", but adding a 0 produces the empty block "" immediately preceded by the empty block "". The fourth digit is a 0: adding a 1 would produce the block "01" immediately preceded by the block "01", but adding a 0 produces the block "0" immediately preceded by the block "0". The fifth digit is a 1: adding a 0 would produce the block "0" immediately preceded by the block "0", but adding a 1 produces the block "" immediately preceded by the block "". The sixth digit is a 1: adding a 0 would produce the block "010" immediately preceded by the block "010", but adding a 1 only produces the block "1" preceded by the block "1". The seventh digit is a 0 because adding a 1 would produce the block "1" preceded by the same block but adding a 0 only produces the block "" preceded by the same block.

The first digit where the Ehrenfeucht-Mycielski and Linus sequences differ is the tenth digit.

The rest of the chapter concentrates on the Slowgrow sequence. Next we are going to present our theorems whose proofs appear in the following sections. There are three types of theorems. First, bounds on when a new number first occurs in S. Second, a result about the density of any integer in S. Third, our main result which concerns the frequency with which certain blocks of terms occur in S.

We are going to study when a new number appears in the Slowgrow sequence. Let n be a positive integer. Let us define $i_n := \min\{i : s_i = n\}$. So i_n is the index of the first appearance of n. We are going to prove the following bounds on the quantity.

Theorem 4.1. *Let i_n be the index of the first appearence of n in S. Then $i_n \geq \frac{n(n+1)(2n-5)}{6} + 3$ for every $n \geq 2$.*

Theorem 4.2. *Let i_n be the index of the first appearence of n in S. Then $i_n \leq 2 \cdot (n-1)! + 1$ for every $n \geq 1$.*

If $S = s_1, s_2, \ldots$ is the Slowgrow sequence, then the *density of k up to n in S* is $\frac{S_k(n)}{n}$ where $S_k(n) = |\{j : 1 \leq j \leq n, s_j = k\}|$. The *limiting density of k* is $\lim_{n\to\infty} \frac{S_k(n)}{n}$.

The next theorem gives a particular function as an upper bound on the density of a number k up to n in the Slowgrow sequence.

Theorem 4.3. *Let the Slowgrow sequence be $S = (s_i)_{i \geq 1}$. Let k be a positive integer. Let $S_k(n) = |\{j : 1 \leq j \leq n, s_j = k\}|$. Then the density of k up to n is*

$$\frac{S_k(n)}{n} < \max\left\{2(k+1)\left(\frac{k}{2n}\right)^{\frac{1}{k+1}}, \frac{2^k k^{k+1}}{n} + 1\right\}$$

for every positive integer n. Thus the limiting density of k is $\lim_{n\to\infty} \frac{S_k(n)}{n} = 0$.

Say a *block (of terms) $b_1 b_2 \ldots b_k$ occurs* in S if there is a positive integer i such that $s_i = b_1, s_{i+1} = b_2, \ldots, s_{i+k-1} = b_k$.

Our main result is a theorem concerning which blocks of terms occur in the Slowgrow sequence. It says that blocks which can potentially occur multiple times in S in fact occur infinitely often. However, giving any lower bounds on the number of occurrences of these blocks up to the nth term of S seems like a hopeless task. Our proof is purely existential and does not say anything about when those blocks occur. It is the most non-trivial part of this chapter.

Let θ_1 be the set of blocks of terms for which there is a $k \geq 1$ such that the block is of form $a_1 a_2 \ldots a_k$ and there is an $i \in [k-1]$ with $a_{i+1} > a_i + 1$. We shall show in Lemma 4.5 that these blocks cannot occur in S. Let θ_2 be the set of blocks of terms for which there is a $k \geq 1$ such that the block is of form $a_1 a_2 \ldots a_{k-1} k$. By the definition of S these blocks occur in S at most once. Let θ_3 be the set of blocks that do not belong to θ_1 and that have no subblocks from θ_2. Our main result says that all blocks in θ_3 occur in S infinitely often. Note that a block $a_1a_2\dots a_{k-2}a_{k-1}$ where $k \geq 2$ belongs to θ_3 if and only if $a_{i+1} \leq a_i + 1$ for every $i = 1, 2, \dots, k-2$ and $k \leq a_{k-1}$.

Theorem 4.4. *Let $k \geq 2$ and a_i, $i = 1, 2, \dots, k-1$ be positive integers such that $a_{i+1} \leq a_i + 1$ for every $i = 1, 2, \dots, k-2$ and $k \leq a_{k-1}$. Then the block of terms $a_1a_2\dots a_{k-2}a_{k-1}$ occurs in S infinitely often.*

Why do we consider the Slowgrow sequence? It has the property that it avoids a certain type of repetitions, but not only that: every new term in the sequence creates a block that has not appeared so far. It also has the following four properties which are interesting together:

1) It has a very simple natural definition

2) Every positive integer except 1 occurs infinitely often

3) Every number occurs with limiting density zero

4) A new number n seems to occur quite late, approximately around the 2^nth term.

If all one requires are properties 1), 2) and 3) then one can consider the sequence

$$1, 2, 1, 3, 2, 1, 4, 3, 2, 1, 5, 4, 3, 2, 1, \dots$$

but in that sequence the first occurrence of n is around the $\frac{n(n-1)}{2}$th index whereas in the Slowgrow sequence it is much later.

If all one requires are properties 2), 3) and 4) then one can consider a sequence something like

$$1, 2, 2, 2, 1, 1, 1, 1, 1, 1, 3, 3, 3, 3, 3, 3, 3, 3, 3, 3, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,$$

$$1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, \ldots$$

which is just the previous sequence (a_i) such that each number is repeated at most so many times that the first n occurs at around index 2^n and the limiting density of every number is 0 (for example if we arrange that the densities up to n of all numbers occurred so far are equal just before the first occurrence of a new number). However, one can argue that such a sequence is not natural.

Many of the possible blocks of terms in $\theta_2 \backslash \theta_1$ seem to occur in S. The total number of such blocks in $s_1, s_2, \ldots, s_n$ is n because, by the definition of S, every s_i defines a new block $s_{i-s_i+1}s_{i-s_i+2}\ldots s_i$ of length s_i.

The following is a property of the Slowgrow sequence that limits the growth of S from a term to the next one. However, there can be an arbitrary decrease from a term to the next one. We are going to prove in Lemma 4.6 that all positive integers appear in S. By the definition of S, every integer n is followed by a 2 at its first occurrence since this will be the first occurrence of the block $n2$. This gives an arbitrary decrease from a term to the next one.

Lemma 4.5. *In the Slowgrow sequence $S = (s_i)_{i\geq 1}$, for all positive integers i, $s_{i+1} \leq s_i + 1$.*

Proof. Let $m = s_i$. By the definition of the Slowgrow sequence, the block

$$s_{i-m+1}s_{i-m+2}\ldots s_{i-1}m$$

occurred for the first time here, so the block

$$s_{i-m+1}s_{i-m+2}\ldots s_{i-1}m(m+1)$$

has not occurred yet. Thus $s_{i+1} \leq m + 1$ by the definition of the sequence. $\square$

Lemma 4.6. *All positive integers appear in the Slowgrow sequence S.*

Proof. The first term of the Slowgrow sequence is 1 and by Lemma 4.5 the sequence does not increase by more than 1 in consecutive terms. So it is enough to prove that the sequence is not bounded. Suppose for contradiction that it is bounded and that the greatest number in the sequence is n. But each of the blocks of terms of the form $a_1a_2\dots a_{m-1}m$ where $1 \leq m \leq n$ and $1 \leq a_i \leq n$ can appear only once. The total number or terms in all these blocks is at most $1 + 2n + 3n^2 + 4n^3 + \dots + n^n$ which is finite – a contradiction with the infiniteness of the sequence. $\square$

Lemma 4.7. *If the block of terms $a_1a_2\dots a_{m-1}a_m$ occurs in S and $a_m > m$ then all blocks of terms of the form $a_1a_2\dots a_{m-1}b$ also occur in S at least once where $m \leq b \leq a_m$.*

Proof. Assume the block of terms $a_1a_2\dots a_{m-1}a_m$ occurs in S. Then by the definition of the sequence there are some positive integers $b_1, b_2, \dots, b_{a_m-m}$ preceding it, that is, $b_1, b_2, \dots, b_{a_m-m}a_1a_2\dots a_{m-1}a_m$ occurs in S. By the definition of S, the blocks

$$1, a_{m-1}2, a_{m-2}, a_{m-1}3, \dots, a_1a_2\dots a_{m-1}m,$$

$$b_{a_m-m}a_1a_2\dots a_{m-1}(m+1), ...,$$

$$b_2b_3\dots b_{a_m-m}a_1a_2\dots a_{m-1}(a_m-1)$$

have all occurred in S earlier. In particular, their subblocks of form $a_1a_2\dots a_{m-1}b$ where $m \leq b \leq a_m - 1$ have all occurred in S earlier. $\square$

These are quite general lemmas that will be useful later, in particular in proving our main result.

Table 4.2: The index of the first appearance of n in S, 2^n and their ratio for $n = 1, 2, \ldots, 19$

n	i_n	2^n	$i_n/2^n$
1	1	2	0.5
2	2	4	0.5
3	4	8	0.5
4	9	16	0.56
5	22	32	0.69
6	59	64	0.92
7	107	128	0.84
8	308	256	1.2
9	667	512	1.3
10	1680	1024	1.64
11	2945	2048	1.44
12	6255	4096	1.53
13	14023	8192	1.71
14	32735	16384	2
15	55668	32768	1.7
16	109614	65536	1.67
17	195789	131072	1.49
18	305866	262144	1.17
19	609064	524288	1.16

4.2 The first occurrence of a new number in the Slowgrow sequence

We are going to study when a new number n first occurs in the Slowgrow sequence. Recall that i_n is the index of the first appearence of n in S. The first few values of i_n, calculated by a computer, are given in Table 4.2. It is an open question if the value of i_n grows like 2^n: we observe that it is between 2^{n-1} and 2^{n+1} for all $n = 1, 2, \ldots, 19$.

Now we are going to prove a lemma that is the main part of the proof of our lower bound on i_n, given in Theorem 4.1.

Lemma 4.8. *$i_{n+1} \geq i_n + n(n-2)$ for every positive integer n.*

Proof. By the definition of i_n, the first time n occurs in the Slowgrow sequence is at index i_n. By the definition of the sequence $s_{i_n+1} = 2$. By Lemma 4.5, $s_{i_n+2} \leq 3$ and by induction $s_{i_n+m-1} \leq m$ for every positive m, so $s_{i_n+n-1} \leq n$. So the second time n occurs is at least at index $i_n + n - 1$. Let the index where n occurs the ith time be j_i (suppressing the dependence on n in the notation). So $j_2 \geq i_n + n - 1$. We know $s_{j_2+1} = 3$. Since $n33$ has not occurred before, $s_{j_2+2} = 3$. Now by Lemma 4.5, $s_{j_2+3} \leq 4$ and by induction $s_{j_2+m-1} \leq m$ for every m, so $s_{j_2+n-1} \leq n$. So the third time n occurs is at index $j_3 \geq j_2 + n - 1$. Now by Lemma 4.5, $s_{j_3+1} \leq 4$. Let k_i be the first index such that $s_{k_i} = n$ and $s_{k_i+1} = i$ (again suppressing the dependence on n in the notation), then $k_4 \geq j_3$. Since $n43$ has not occurred before, $s_{k_4+2} = 3$. Now by Lemma 4.5, $s_{k_4+3} \leq 4$ and by induction $s_{k_4+n-1} \leq n$. Analogously $k_5 \geq k_4 + n - 1$ and by induction $k_m \geq k_{m-1} + n - 1$ for every m such that $5 \leq m \leq n$. Now $s_{k_n+1} = n$, $s_{k_n+2} = 3$, and by Lemma 4.5, $s_{k_n+3} \leq 4$ and by induction $s_{k_n+m} \leq m + 1$ for every $m = 3, 4, \ldots, n$, so $i_{n+1} \geq k_n + n$. Hence

$$i_{n+1} \geq k_n + n \geq k_{n-1} + 2n - 1 \geq \ldots \geq k_4 + (n-3)n - (n-2)$$

$$\geq j_3 + (n-3)n - (n-2) \geq j_2 + (n-2)n - (n-1)$$

$$\geq i_n + (n-1)n - n = i_n + n(n-2).$$

□

We are going to repeat Theorem 4.1 here for ease of reference.

Theorem. *Let i_n be the index of the first appearence of n in S. Then $i_n \geq \frac{n(n+1)(2n-5)}{6} + 3$ for every $n \geq 2$.*

Theorem 4.1 follows from induction on n using Lemma 4.8 and $i_2 = 2$.

Next we are going to give an upper bound on the time that a new number n first appears in S. We are going to prove Theorem 4.2, repeated here for ease of reference.

Theorem. *Let i_n be the index of the first appearence of n in S. Then $i_n \leq 2 \cdot (n-1)! + 1$ for every $n \geq 1$.*

Proof. We shall work in the part $s_1, \ldots s_{i_n - 1}$ of S. By the definition of i_n this only contains the numbers $1, 2, \ldots, n-1$. Consider the types of blocks of terms 1, _ 2, _ _ 3, _ _ _ 4, …, _ _ … _ $(n-1)$, i.e. blocks in θ_2 containing only numbers up to $n-1$. The first of them appears once. By Lemma 4.5, the second appears at most $n-1$ times (12, 22, 32, …, $(n-1)2$). By Lemma 4.5, the third appears at most $(n-2)(n-1)$ times

$$(123,\ 223,\ 323, \ldots, (n-1)23,\ 233,\ 333,\ \ldots, (n-1)33, \ldots, (n-1)(n-1)3),$$

since the second term of the block can be $2, 3, \ldots, n-1$ and the first term of the block can be $1, 2, \ldots, n-1$. The fourth type of block of terms _ _ _ 4 appears at most $(n-3)(n-2)(n-1)$ times. In general, the kth type of block appears at most $(n-k+1)(n-k+2)\ldots(n-1)$ times since by Lemma 4.5 the number immediately before k can be $k-1, k, \ldots, n-1$, by Lemma 4.5 the number before that can be $k-2, k-1, \ldots, n-1$ and inductively, by Lemma 4.5, the ith number from the end of the block can be $k-i, k-i+1, \ldots, n-1$. Thus, by induction on n, the total number of the blocks appearing in $s_1, \ldots s_{i_n - 1}$ is at most

$$1 + (n-1) + (n-2)(n-1) + (n-3)(n-2)(n-1) + \ldots + 2 \cdot 3 \ldots (n-1)$$

$$\leq 2 \cdot (n-1)!.$$

Since, by the definition of S, every number s_i determines exactly one block from θ_2, the number of blocks is equal to the number of terms. So in $(s_1, \ldots, s_{2\cdot(n-1)!+1})$ we must have a number at least n. Since by Lemma 4.5 the increase in consecutive terms is not more than 1 we must have seen the number n. $\square$

These theorems give basic bounds on the index i_n of the first appearence of n in the Slowgrow sequence. We hope that better bounds can be proved.

4.3 Limiting density of a number in the Slowgrow sequence

Recall that the density up to n in S of a positive integer k is the number of occurrences of k in $s_1, \ldots, s_n$ divided by n. We are going to prove a relatively easy upper bound on the density of k up to n in S which grows like $O(\frac{1}{\sqrt[k+1]{n}})$ as $n \to \infty$. By computer experiment, the density of 2 up to n decreases as $\theta(\frac{n}{2^n})$. We conjecture that the true density of k up to n decreases like $O(\frac{n^{k-1}}{2^n})$. It would be interesting to bridge the gap between these rates. We are going to give the proof of Theorem 4.3 which is repeated here for ease of reference.

Theorem. *Let the Slowgrow sequence be $S = (s_i)_{i \geq 1}$. Let k be a positive integer. Let $S_k(n) = |\{j : 1 \leq j \leq n, s_j = k\}|$. Then the density of k up to n is*

$$\frac{S_k(n)}{n} < \max\left\{2(k+1)\left(\frac{k}{2n}\right)^{\frac{1}{k+1}}, \frac{2^k k^{k+1}}{n} + 1\right\}$$

for every positive integer n. Thus the limiting density of k is $\lim_{n \to \infty} \frac{S_k(n)}{n} = 0$.

Proof. Fix an integer $b \geq 2k$. Let us call a block of terms in S given as $k_1 k_2 \ldots k_{k-1} k$ *bad* if $k_i \leq b$ for all $i = 1, 2, \ldots, k-1$. Crudely, there are at most b^k bad strings in total and hence the total number of occurrences of k in them is at most kb^k. Thus the density up to n in S of occurrences of k in bad strings is at most $\frac{kb^k}{n}$. Let us call a block $k_1 k_2 \ldots k_{k-1} k$ *good* if it is not bad, i.e. there is an i, $1 \leq i \leq k-1$ such that $k_i > b$.

The density up to n in S of occurrences of k in good blocks is at most $\frac{2k}{b}$: This is because a good block has a number at least b, so as we go towards lower indices from it in S, we arrive at an k only after at least $b-k$ steps (since, by Lemma 4.5, $s_{i-1} \geq s_i - 1$ for every $i \geq 2$). So, the number of k's in the good block and $b-k$ steps toward lower indices from it is at most k. So the proportion of k's is at most $\frac{k}{b-k} \leq \frac{2k}{b}$.

Thus a function that bounds the density of k up to n from above is $f_{k,b}(n) = \frac{kb^k}{n} + \frac{2k}{b}$. Now let us vary b as a function of n. We minimize $f_{k,b}$ with respect to b to get the function $f_k(n) = \max\left\{2(k+1)\left(\frac{k}{2n}\right)^{\frac{1}{k+1}}, \frac{2^k k^{k+1}}{n} + 1\right\}$, given by $b = \max\left\{\left\lceil\left(\frac{2n}{k}\right)^{\frac{1}{k+1}}\right\rceil, 2k\right\}$. The function $f_k(n)$ bounds the density of k up to n from above. In particular, the limiting density of k is 0 since $f_k(n)$ approaches 0 as n approaches infinity. $\square$

We hope that a better upper bound on the density of a number k up to n in the Slowgrow sequence can be proved. Indeed, we conjecture that a bound like $O(c^{-n})$ for a suitable constant $c > 1$ would be closer to the actual value of the density.

4.4 Infinite occurrence of blocks

In this section we are going to prove our main result. It says that any block that can potentially occur in S multiple times in fact occurs infinitely often. This is interesting because, even though the Slowgrow sequence is chosen specifically so as to avoid repetition of certain kinds of blocks, all the blocks except those that appear at most once by the definition of the sequence and those that do not appear by Lemma 4.5 actually appear infinitely often.

Lemma 4.9. *Let $k \geq 2$ be an integer and $b_1, b_2, \ldots, b_{k-2}$ be positive integers such that the block $ab_1b_2\ldots b_{k-2}$ occurs in S for all sufficiently large a, i.e. for all $a > A_1$*

where A_1 is a positive integer. Then $ab_1b_2\dots b_{k-2}k$ occurs in S for all sufficiently large a, i.e. for all $a > A_2$ where A_2 is a positive integer.

Proof. We are given that the block $ab_1b_2\dots b_{k-2}$ occurs in S for all $a > A_1$ where A_1 is an integer. By the definition of the sequence, each of the blocks 1, $b_{k-2}2$, $b_{k-3}b_{k-2}3$, ..., $b_1b_2\dots b_{k-2}(k-1)$ can occur in S at most once. Thus there are finitely many a such that $ab_1b_2\dots b_{k-2}b$ occurs for some $b \leq k-1$, let them be $a_1, a_2, \dots a_m$ where $m \leq k-1$. So by the definition of S, the block $ab_1b_2\dots b_{k-2}k$ occurs in S for all $a > A_2 := \max\{A_1, a_1, a_2, \dots, a_m\}$. □

The next proposition is the major step in proving our main theorem. Roughly, it says that all blocks in which numbers decrease sufficiently fast occur in S infinitely often (provided the last number is greater than the length of the block). It will be proved by induction where the induction step is divided into two lemmas.

Let us define the following statements.

P_d is "Let $p_1, \dots, p_d$ be positive integers such that $d+1 \leq p_d$ and $p_{y+1} \leq \frac{p_y}{2} - 1$ for all $y = 1, 2, \dots, d-1$. Any block of terms $ap_1p_2\dots p_d$ occurs in S for all sufficiently large a." for any $d \geq 0$.

$B_{k,d}$ is "Any block of terms $ab_1b_2\dots b_{k-1}$ where $b_1 = d+1$, $b_{k-1} = k$ and $b_{j+1} \in \{b_j, b_j+1\}$ for every $j = 1, 2, \dots, k-2$ occurs in S for all sufficiently large a." for any $d \geq 0$ and any $k \geq d+1$.

Proposition 4.10. P_d holds for every $d \geq 0$.

For this result we shall require the following lemmas.

Lemma 4.11. *Let k and d be positive integers such that $k \geq d+1$. If P_{d-1} holds then $B_{k,d}$ holds.*

Proof. Let $p_0 = a$. By Lemma 4.9, since $p_0p_1p_2\dots p_{d-1}$ occurs in S for all sufficiently large p_0 we have that the block $p_0p_1p_2\dots p_{d-1}b_1$ occurs in S for all sufficiently large p_0 (since $b_1 = d+1$).

We are going to use induction on k. Our induction hypothesis is that there is an integer-valued function f_i (which may also depend on $b_1,\dots,b_i$) such that the block

$$p_{d+i-b_i}p_{d+i-b_i+1}\dots p_{d-1}b_1b_2\dots b_i$$

occurs in S for all tuples $(p_{d+i-b_i}, p_{d+i-b_i+1},\dots,p_{d-1})$ such that $p_{d-1} > f_i(0)$ and $p_{j-1} > f_i(p_j)$ for every $j = d+i-b_i+1,\dots,d-1$. It is true for $i=1$ as can be seen from the statement of the lemma and the first paragraph of this proof. We shall prove the claim for $i+1$, assuming the claim for i, splitting the proof into two cases.

Firstly, if $b_{i+1} = b_i + 1$ then, by Lemma 4.9,

$$p_{d+i-b_i}p_{d+i-b_i+1}\dots p_{d-1}b_1b_2\dots b_ib_{i+1}$$

occurs in S for all sufficiently large p_{d+i-b_i}. Rewriting,

$$p_{d+(i+1)-b_{i+1}}p_{d+(i+1)-b_{i+1}+1}\dots p_{d-1}b_1b_2\dots b_ib_{i+1}$$

occurs in S for all sufficiently large $p_{d+(i+1)-b_{i+1}}$. We can define $f_{i+1}(j) = (f_i(j)$ plus the increase on the lower bound on the possible $p_{d+i-b_i})$ for every j.

Secondly, if $b_{i+1} = b_i$, then first of all we notice that

$$p_{d+i-b_i+1}p_{d+i-b_i+2}\dots p_{d-1}b_1b_2\dots b_i$$

occurs in S for all sufficiently large p_{d+i-b_i+1} by the existence of f_i. Now, by Lemma 4.9,

$$p_{d+i-b_i+1}p_{d+i-b_i+2}\dots p_{d-1}b_1b_2\dots b_ib_{i+1}$$

occurs in S for all sufficiently large p_{d+i-b_i+1}. Rewriting,

$$p_{d+(i+1)-b_{i+1}}p_{d+(i+1)-b_{i+1}+1}\dots p_{d-1}b_1b_2\dots b_ib_{i+1}$$

occurs in S for all sufficiently large $p_{d+(i+1)-b_{i+1}}$. We can define $f_{i+1}(j) = (f_i(j)$ plus the increase on the lower bound on the possible $p_{d+i-b_i+1})$ for every j.

To finish the proof, notice that $p_{d+k-1-b_{k-1}} = p_{d-1}$ and define $a := p_{d-1}$. $\square$

Recall that $B_{k,d}$ says that any block of terms $ab_1b_2\dots b_{k-1}$ where $b_1 = d+1$, $b_{k-1} = k$ and $b_{j+1} \in \{b_j, b_j+1\}$ for every $j = 1, 2, \dots, k-2$ occurs in S for all sufficiently large a. To prove the next lemma, we use that $B_{k,d}$ means that the block $ab_1\dots b_{k-1}$ occurs in S for all possible choices of the set of j such that $b_{j+1} = b_j$. Note that this set has $d-1$ elements, denote them $j_0 < \dots < j_{d-2}$. We are going to do induction on d within the proof of the next lemma. On the ith induction step we assume that $j_0, \dots, j_{i-1}$ have been fixed previously, we are going to vary j_i from roughly $k - \frac{k}{2^i}$ to $k - \frac{k}{2^{i+1}}$ and assume that we can make any choices for $j_{i+1}, \dots, j_{d-2}$ such that roughly $j_{i+1} > k - \frac{k}{2^{i+1}}$. The idea is that we vary j_i in the same kind of space that the $j_{i+1}, \dots, j_{d-2}$ are restricted to. Recall that P_d says that if we let $p_1, \dots, p_d$ be positive integers such that $d+1 \leq p_d$ and $p_{y+1} \leq \frac{p_y}{2} - 1$ for all $y = 1, 2, \dots, d-1$ then any block of terms $ap_1p_2\dots p_d$ occurs in S for all sufficiently large a. By varying j_i we obtain roughly $\frac{k}{2^{i+1}}$ occurrences of the block $b_{k+i-p_i}b_{k+1+i-p_i}\dots b_{k-1}p_1p_2\dots p_i$. So this gives us that there is an occurrence in S of $b_{k+i+1-p_{i+1}}b_{k+i+2-p_{i+1}}\dots b_{k-1}p_1p_2\dots p_ip_{i+1}$ (which is a subblock of the previous followed by p_{i+1}) roughly for every $p_{i+1} = i+1, i+2, \dots, \frac{k}{2^{i+1}}$ and an occurrence in S of $p_{i+2-p_{i+1}}\dots p_{i+1}$ for every $p_{i+1} = 1, 2, \dots, i$.

Lemma 4.12. *Let d and k be positive integers such that $k \geq d+1$. If $B_{k,d}$ holds then P_d holds.*

Proof. Assume $b_{k-1} > b_{k-2}$. Let $a := b_{k-2}$ such that $p_1 \leq a$. Then $b_{k-p_1} b_{k-p_1+1} \ldots b_{k-3} a p_1$ occurs in S by Lemma 4.7.

Let us define $j_0 = \min\{j : b_j = b_{j+1}\}$ and let $j_i = \min\{j > j_{i-1} : b_j = b_{j+1}\}$. Now we are going to vary the b_i (within the allowed constraints) so that j_0 varies from $k - p_1$ to $k - \lceil \frac{p_1}{2} \rceil$ and the other j_i stay fixed so that $j_1 \geq k - \lceil \frac{p_1}{2} \rceil + 1$. We examine what term comes after $b_{k-p_1} b_{k-p_1+1} \ldots b_{k-3} a p_1$. If it was a 2 then it was the first occurrence of $p_1 2$; varying the position of j_0 we may assume not. If it was a 3 then it was the first occurrence of $a p_1 3$; varying the position of j_0 we may assume not. Continuing, we may assume it was any p_2, $d + 1 \leq p_2 \leq \frac{p_1}{2} - 1$. Now we are going to treat j_0 as fixed. Letting the b_i vary so that j_1 varies from $k - p_2 + 1$ to $k - \lceil \frac{p_2}{2} \rceil + 1$ and fixing $j_2 \geq k - \lceil \frac{p_2}{2} \rceil + 2$ and fixing all $j_i, i \geq 3$, we prove analogously that the p_2 can be followed by any p_3, $d + 1 \leq p_3 \leq \frac{p_2}{2} - 1$. Continuing inductively, treating $j_0, \ldots, j_{i-2}$ as fixed and varying j_{i-1} from $k - p_i + (i-1)$ to $k - \lceil \frac{p_i}{2} \rceil + (i-1)$ and fixing $j_i \geq k - \lceil \frac{p_i}{2} \rceil + i$ and fixing all $j_{i+1}, \ldots, j_{d-2}$, we prove analogously that p_i can be followed by any p_{i+1}, $d + 1 \leq p_{i+1} \leq \frac{p_i}{2} - 1$. Finally, varying the position of j_{d-2}, we have proved that p_{d-1} can be followed by any p_d, $d + 1 \leq p_d \leq \frac{p_{d-1}}{2} - 1$. □

Now we are ready to prove Proposition 4.10.

Proof. By Lemma 4.6, every positive integer a occurs in S. That is, we have P_0. We proceed by induction on d, using Lemmas 4.11 and 4.12. Their statements together give us: if P_{d-1} then P_d for any $d \geq 0$. So P_d for any $d \geq 0$. □

Now we are ready to prove our main result Theorem 4.4. It says that all possible blocks of terms that can potentially occur in S multiple times actually occur infinitely often. We repeat Theorem 4.4 here for ease of reference.

Theorem. *Let $k \geq 2$ and a_i, $i = 1, 2, \ldots, k-1$ be positive integers such that $a_{i+1} \leq a_i + 1$ for every $i = 1, 2, \ldots, k-2$ and $k \leq a_{k-1}$. Then the block of terms $a_1 a_2 \ldots a_{k-2} a_{k-1}$ occurs in S infinitely often.*

Proof. Let $d = a_1 - 2$. By Proposition 4.10 we know that there is an integer-valued function f such that whenever p_i are such that $p_{i-1} \geq f(p_i)$ for every $i = 1, 2, \ldots, d$ and $p_d \geq f(0)$ then $p_0 p_1 \ldots p_d$ occurs in S. By Lemma 4.9, $p_0 p_1 \ldots p_d$ is followed by a_1 for all sufficiently large p_0. So, by increasing the values $f(p_1)$ by the bound on p_0 given by Lemma 4.9, we can obtain a new function f_1 such that $p_0 p_1 \ldots p_d a_1$ occurs for all sufficiently large p_d and for all p_i such that $p_i \geq f_1(p_{i+1})$, $i = 0, 1, \ldots, d-1$.

Now our induction hypothesis is that there is an integer-valued function f_j such that

$$p_{d-a_j+j+1} p_{d-a_j+j+2} \ldots p_d a_1 a_2 \ldots a_j$$

occurs for all sufficiently large p_d and for all p_i such that $p_i \geq f_j(p_{i+1})$, $i = d - a_j + j + 1, d - a_j + j + 2, \ldots, d-1$. Now, by Lemma 4.9,

$$p_{d-a_j+j+1} p_{d-a_j+j+2} \ldots p_d a_1 a_2 \ldots a_j$$

is followed by a_{j+1} for all sufficiently large $p_{d-a_{j+1}+(j+1)+1}$. We can define a new function f_{j+1} by increasing all $f_j(p_{d-a_{j+1}+(j+1)+2})$ by the bound given on $p_{d-a_{j+1}+(j+1)+1}$ by Lemma 4.9. This gives an integer-valued function f_{j+1} such that

$$p_{d-a_{j+1}+(j+1)+1} p_{d-a_{j+1}+(j+1)+2} \ldots p_d a_1 a_2 \ldots a_{j+1}$$

occurs for all sufficiently large p_d and for all p_i such that $p_i \geq f_{j+1}(p_{i+1})$, $i = d - a_{j+1} + (j+1) + 1, d - a_{j+1} + (j+1) + 2, \ldots, d-1$.

Note that $d - a_{k-1} + (k-1) + 1 \leq d$. Thus by induction we get that

$$p_{d-a_{k-1}+(k-1)+1}p_{d-a_{k-1}+(k-1)+2} \dots p_d a_1 a_2 \dots a_{k-2} a_{k-1}$$

occurs for all sufficiently large $p_{d-a_{k-1}+(k-1)+1}$. Hence $a_1 a_2 \dots a_{k-2} a_{k-1}$ occurs in S infinitely often. $\square$

Corollary 4.13. *Let a_i, $k \geq 2$ be positive integers such that $a_{i+1} \leq a_i + 1$ for every $i = 1, 2, \dots, k-3$, $k \leq a_{k-2} + 1$. Then the block of terms $a_1 a_2 \dots a_{k-2}(k-1)$ occurs in S once.*

Proof. It follows from Theorem 4.4 and Lemma 4.7. $\square$

Corollary 4.14. *The block of terms $nn \dots n$ where n appears n times occurs in S once.*

Proof. The occurrence follows from the previous corollary. It can only occur once by the definition of S. $\square$

Our next proposition says that some blocks which can potentially occur in S, i.e. that are in $\theta_2 \backslash \theta_1$, actually never occur.

Proposition 4.15. Let $n \geq 2$. Assume that the block of terms $a_1 a_2 \dots a_{k-1} k$ occurs in

$$(s_1, s_2, \dots, s_{i_n - 1}).$$

Then the block of terms $a_1 a_2 \dots a_{k-1} k(k+1) \dots (n+1)$ never occurs in S.

Proof. Since the block $a_1 a_2 \dots a_{k-1} k$ occurs in S at most once, it would have to be followed by $(k+1) \dots (n+1)$ at its only occurrence. However, assuming that it is indeed followed by $(k+1) \dots n$ this is in turn followed by a 2 since it will be the first occurrence of $n2$. $\square$

We have established that blocks which can potentially occur several times in the Slowgrow sequence actually occur infinitely often. A far-reaching goal would be to give bounds on the density of such blocks up to n.

4.4.1 Summary

Let us call a block of terms $a_1 a_2 \ldots a_k$ *good* if it does not belong to θ_1, i.e. $a_{i+1} \le a_i + 1$ for every $i = 1, 2, \ldots, k-1$. Since, by Lemma 4.5, $s_i \ge s_{i+1} - 1$ for every positive integer i, only good blocks can occur in S.

As a summary, the following blocks of terms occur in S infinitely often: $a_1 a_2 \ldots a_k$ where the block is good and $a_k > k$.

The following blocks of terms occur in S once: $a_1 a_2 \ldots a_{k-1} k$ where the block is good and $a_{k-1} > k - 1$.

Some good blocks never occur in S as seen in Proposition 4.15.

4.5 Open questions

We state a few open questions, asked by Steven Kalikow and the author.

When does a new number n first appear in the Slowgrow sequence? Recall that i_n is the index of the first appearance of n in S. It would be interesting to give better bounds on i_n. By computer experiments, it seems that, for every positive integer n, i_n is about 2^n, see Table 4.2. It would be interesting to know any good intuitive argument why it may be so. On some level, the Slowgrow sequence may be treated as (pseudo)random, so perhaps some probabilistic argument may give an intuitive explanation.

For every postitive integer i, it takes some time before the first i appears in the Slowgrow sequence. Then i becomes more "frequent" until it eventually appears less and less often – since its density up to n approaches zero as n approaches infinity. It would be interesting to know for what n the density of i up to n is the largest. Computer experiments suggest that there are large subblocks in the Slowgrow sequence where i occurs often. One could also somehow define a 'local density' of i in the Slowgrow sequence and study when it will be the largest.

Chapter 5

co-Sidon: Additive Properties of a Pair of Sequences

5.1 Introduction

For a given set $A \subset \mathbb{N}_0$ of non-negative integers, here and throughout the chapter, the *counting function* $A(n)$ is defined as the number of elements of A not exceeding n, i.e., $A(n) = |A \cap \{0, 1, 2, \ldots, n\}|$. Consider the following functions

$$\begin{aligned} r(A, n) =& |\{(a_1, a_2) \in A \times A : a_1 + a_2 = n\}|, \\ r_1(A, n) =& |\{(a_1, a_2) \in A \times A : a_1 + a_2 = n \quad \text{and} \quad a_1 \leq a_2\}|, \\ r_2(A, n) =& |\{(a_1, a_2) \in A \times A : a_1 + a_2 = n \quad \text{and} \quad a_1 < a_2\}|. \end{aligned}$$

A well-studied problem concerning these functions is to determine necessary and sufficient conditions on A for their (eventual) monotonicity. Here and throughout the chapter, monotonicity refers to monotonicity in n. In other words, for what sets A we can find an n_0 such that $r(A, n+1) \geq r(A, n)$ for all $n > n_0$? Although the three functions look similar, and in fact $|r(A, n) - 2r_2(A, n)| \leq 1$ and

$|r_1(A,n) - r_2(A,n)| \leq 1$, the (partial) answers to these questions may be quite different.

Erdős, Sárközy and Sós [10] proved that $r(A,n)$ is eventually monotone increasing if and only if A contains all the positive integers from a certain point on. On the other hand, they obtained only a partial answer for r_1 and r_2. In particular, they proved that if

$$\lim_{n \to +\infty} \frac{n - A(n)}{\log n} = +\infty$$

then $r_1(A,n)$ is not eventually monotone increasing. (This result was also obtained independently by Balasubramanian [1].)

Also, for $r_2(A,n)$ they proved that if

$$A(n) = o\left(\frac{n}{\log n}\right)$$

then $r_2(A,n)$ cannot be monotone increasing from a certain point on.

Motivated by these results, Sárközy asked the following question in his valuable paper on unsolved problems in number theory [30] (see Problem 4 in [30]).

Problem 5.1. *If A, B are given infinite sets of non-negative integers, what can one say about the monotonicity of the number of solutions of the equation*

$$a + b = n,\ a \in A,\ b \in B?$$

We can naturally rephrase this question by defining the following function.

Definition 5.2. The *representation function* for two sets $A, B \subset \mathbb{N}_0$ is

$$r(A,B,n) = |\{(a,b) \in A \times B : a + b = n\}|.$$

The main goal of the present chapter is to give some sufficient conditions on A, B for the monotonicity of this function. This new representation function acts in a surprisingly different way from the prequel. Our main result is as follows.

Theorem 5.3. *For all* $0 \leq \alpha, \beta < 1$, $1/2 < c_1, c_2 \leq 1$, *there exist sets* $A, B \subset \mathbb{N}_0$ *such that* $r(A, B, n)$ *is monotone increasing in* n*;*

$$\limsup_{n\to\infty} \frac{A(n)}{n^{c_1}} = \alpha; \quad \limsup_{n\to\infty} \frac{B(n)}{n^{c_2}} = \beta.$$

In the next sections we develop tools to approach Theorem 5.3 and prove some related results. Then we will return to the proof of Theorem 5.3.

5.2 co-Sidon Sets

Before proving Theorem 5.3, we introduce a generalized notion of Sidon sets and study some of its properties. Recall that a set $A \subset \mathbb{N}_0$ is called *Sidon* if $r_1(A, n) \leq 1$ for all $n \in \mathbb{N}$, i.e., the sums of unordered pairs of elements of A are all distinct. We remark that it is possible to extend the notion of a Sidon set to a pair of sets in different ways. In this chapter, we consider the following generalization.

Definition 5.4. Two sets $A, B \subset \mathbb{N}_0$ are called *co-Sidon* if $r(A, B, n) \leq 1$ for all $n \in \mathbb{N}_0$, i.e., the sums $a + b$ are distinct for all $(a, b) \in A \times B$.

Note that if A, B are co-Sidon then $|A \cap B| \leq 1$.

For sets A and B of integers we denote their *sum set* by $A + B = \{a + b : a \in A, b \in B\}$. For simplicity if the set B is a single element b we denote their sum set by $A + b = A + B$.

When A, B are finite sets, we prove a simple but sharp result about $|A|, |B|$.

Proposition 5.5. If $A, B \subset \{0, 1, 2, \ldots, n\}$ are co-Sidon, then

$$\min\{|A|, |B|\} \leq \lfloor \sqrt{2n} \rfloor.$$

Furthermore, equality can be obtained for infinitely many values of n.

Proof. Since A and B are finite (and co-Sidon) we have $|A + B| = |A||B|$. Without loss of generality assume $|A| \leq |B|$. Then, $|A|^2 \leq |A + B|$.

Clearly for an element $c \in A + B$ we have $0 \leq c \leq 2n$. However, either 0 or $2n$ is not an element of $A + B$, otherwise we would have $0, n \in A \cap B$ and there would be two distinct solutions to $a + b = n$ with $a \in A$ and $b \in B$. Thus, $|A + B| \leq 2n$ which yields $|A| \leq \lfloor \sqrt{2n} \rfloor$ and the upper-bound is established.

To see that the upper bound is best possible for infinitely many n, consider the following construction for A and B. Let $m \in \mathbb{N}$ be fixed and define

$$A := \{0, m, 2m, \ldots, (2m - 1)m\}$$

and

$$B := \{0, 1, 2, \ldots, m - 1, 2m^2, 2m^2 + 1, 2m^2 + 2, \ldots, 2m^2 + m - 1\}.$$

Note that $|A| = |B| = 2m$ and $A + B = \{0, 1, \ldots, 4m^2 - 1\}$. Therefore A and B are co-Sidon. As $A, B \subseteq \{0, 1, 2, \ldots, 2m^2 + m - 1\}$, we can take $n = 2m^2 + m - 1$. This gives

$$2m = \sqrt{4m^2} \leq \sqrt{4m^2 + 2m - 2} = \sqrt{2n} < \sqrt{4m^2 + 4m + 1} = 2m + 1.$$

Hence $\min\{|A|, |B|\} = 2m = \lfloor \sqrt{2n} \rfloor$. As the choice of m was arbitrary, there are infinitely many n for which we can reach the upper bound in the statement of the theorem. $\square$

It is worth to compare the above result to the following theorem of Erdős and Turán [11] on finite Sidon sets.

Theorem 5.6. *There is an absolute positive constant c such that if $n \in \mathbb{N}$ and $A \subset [n]$ is a Sidon set, then $|A| < n^{1/2} + cn^{1/4}$.*

On the other hand, the best known constructions give Sidon sets of size $n^{1/2}$ for infinitely many n (see e.g. [18, 29] for details). The reduction of this gap is a well-known hard problem.

We consider now the case where A, B are infinite co-Sidon. Defining $A_n = A \cap \{0, 1, \ldots, n\}$ and $B_n = B \cap \{0, 1, \ldots, n\}$, we have that A_n,B_n are co-Sidon. So, by Theorem 5.5, for any n we have

$$\min\{A(n), B(n)\}/\sqrt{n} = \min\{|A_n|, |B_n|\}/\sqrt{n} \leq \lfloor\sqrt{2n}\rfloor/\sqrt{n} \leq \sqrt{2}.$$

A simple example shows that we can come close to achieving this bound.

Construction 5.7. *Let A be the set of integers which can be written in the form $\sum_{i=0}^{k} \alpha_i 2^{2i}$ where $\alpha_i \in \{0, 1\}$ and $k \in \mathbb{N}$. Let B be the set of integers which can be written in the form $\sum_{i=0}^{k} \alpha_i 2^{2i+1}$ where $\alpha_i \in \{0, 1\}$ and $k \in \mathbb{N}$. It is clear that A and B are co-Sidon and $A + B = \mathbb{N}_0$. It can easily be verified that*

$$\liminf_{n\to\infty} \frac{A(n)}{\sqrt{n}} = 1$$

$$\liminf_{n\to\infty} \frac{B(n)}{\sqrt{n}} = \frac{\sqrt{2}}{2}$$

$$\limsup_{n\to\infty} \frac{A(n)}{\sqrt{n}} = \sqrt{3}$$

$$\limsup_{n\to\infty} \frac{B(n)}{\sqrt{n}} = \frac{\sqrt{6}}{2}$$

Thus,

$$\liminf_{n\to\infty} \frac{\min\{A(n), B(n)\}}{\sqrt{n}} = \sqrt{2}/2.$$

Comparing this with the following result of Erdős (see [31, 18]), we conclude that infinite Sidon sets and infinite co-Sidon sets also behave differently. In general, we have more freedom when working with co-Sidon sets.

Theorem 5.8. *There is an absolute, positive constant c such that for any infinite Sidon set $A \subset \mathbb{N}$ we have*

$$\liminf_{n\to\infty} \frac{A(n)}{\sqrt{n/\log n}} < c.$$

It is also worth mentioning the following theorem of Krückeberg [22] for infinite Sidon sets.

Theorem 5.9. *There is a Sidon set $A \subset \mathbb{N}$ such that*

$$\limsup_{n\to\infty} \frac{A(n)}{\sqrt{n}} \geq \sqrt{2}/2.$$

The following definition will be useful for us.

Definition 5.10. We call sets $A, B \subset \mathbb{N}_0$ *perfect* if the sum set $A + B$ is an interval (possibly unbounded) of consecutive integers.

The next proposition will be helpful in building new perfect co-Sidon sets from other co-Sidon sets.

Proposition 5.11. Let $A, B \subset \mathbb{N}_0$ be finite perfect co-Sidon sets. Let $c = \max(A) + \max(B) - \min(A) - \min(B) + 1$. Then for any $k \in \mathbb{N}_0$, the sets A and $C = \bigcup_{i=0}^{k} (B + ic)$ are perfect co-Sidon.

Proof. Let $r = \min(A) + \min(B)$. By assumption, $A + B = \{r, r+1, \ldots, c+r-1\}$. For each i, the sets A and $B + ic$ are co-Sidon. Furthermore, the sets

$$
\begin{aligned}
A + (B + c) &= \{c + r, c + r + 1, \ldots, 2c + r - 1\} \\
A + (B + 2c) &= \{2c + r, 2c + r + 2, \ldots, 3c + r - 1\} \\
&\vdots \\
A + (B + kc) &= \{kc + r, kc + r + 1, \ldots, (k+1)c + r - 1\}
\end{aligned}
$$

are all pairwise disjoint consecutive intervals. Therefore A and $\bigcup_{i=0}^{k}(B + ic)$ are perfect co-Sidon with sum set $\{r, r+1, \ldots, (k+1)c + r - 1\}$. □

Clearly the proposition also holds for $C = \bigcup_{i=0}^{\infty}(B + ic)$.

Next we characterize all infinite perfect co-Sidon sets $A, B \subset \mathbb{N}_0$ using the mixed radix representation. Note that both the co-Sidon and perfect properties are invariant under translation of each of the sets (i.e. addition or subtraction by a constant), so without loss of generality we may assume $0 \in A \cap B$.

Theorem 5.12. *Let $A, B \subset \mathbb{N}_0$ be infinite, such that $0 \in A \cap B$. Then A, B are perfect co-Sidon if and only if there exists an infinite sequence of integers $(k_i)_{i=1}^{\infty}$ such that $\forall i, k_i \geq 2$ and (up to an exchange of A and B),*

$$A = \left\{ \sum_{i=1}^{\infty} k_1 k_2 \ldots k_{2i-2} a_{2i-1} : \forall j, 0 \leq a_{2j-1} < k_{2j-1}, \quad \text{finitely many } a_{2i-1} \text{ non-zero} \right\}$$

and

$$B = \left\{ \sum_{i=1}^{\infty} k_1 k_2 \ldots k_{2i-1} a_{2i} : \forall j, 0 \leq a_{2j} < k_{2j}, \quad \text{finitely many } a_{2i} \text{ non-zero} \right\}.$$

Proof. A sum of the form $\sum_{i=1}^{\infty} k_1 k_2 \ldots k_{i-1} a_i$ where $0 \leq a_j < k_j$, and only finitely many a_i are non-zero, is precisely the so-called *mixed-radix representation* with bases

$(k_1, k_2, \ldots, k_i, \ldots)$. Thus the base r representation is the special case where $k_i = r$ for all i. For any sequence $(k_i)_{i=1}^{\infty}$ of integers with $k_i \geq 2$, every non-negative integer is uniquely representable with bases (k_i).

Let $(k_i)_{i=1}^{\infty}$ be a sequence of integers such that $\forall i, k_i \geq 2$. Suppose A and B are of the form determined by the bases k_i as above. As every non-negative integer is uniquely representable by with bases (k_i), A and B are co-Sidon. Also observe that

$$A + B = \left\{ \sum_{i=1}^{\infty} k_1 k_2 \ldots k_{i-1} a_i : \forall j, 0 \leq a_j < k_j, \quad \text{finitely many } a_i \text{ non-zero} \right\}.$$

Thus $A + B = \mathbb{N}_0$ and therefore A and B are perfect.

Now assume that A, B are perfect co-Sidon. Unless $A = B = \{0\}$, we can assume without loss of generality that $1 \in A$. To show that A, B are of the required form, we need to construct a sequence of base elements $(k_i)_{i \in \mathbb{N}}$ that represents A and B as in the statement of the theorem.

Our construction of the integers k_i is recursive. Let $k_0 = 1$. For $t \geq 1$ define $c_t = k_{t-1} k_{t-2} \cdots k_0$ and let

$$k_t = \begin{cases} \max\{a : \{c_t, 2c_t, \ldots, (a-1)c_t\} \subset A\}, & \text{if } t \text{ is odd} \\ \max\{b : \{c_t, 2c_t, \ldots, (b-1)c_t\} \subset B\}, & \text{if } t \text{ is even} \end{cases}$$

Note that $\forall t > 0, k_t < \infty$. Otherwise, one of A or B contains an infinite arithmetic progression, whose consecutive terms differ by c_t. But as they are co-Sidon, this implies that the other set is finite (in fact of cardinality at most c_t), a contradiction.

Now define two families of sets. Let $A_0 = B_0 = \{0\}$ and for each $t \geq 1$,

$$A_t = \left\{ \sum_{i=1}^{t} k_1 k_2 \ldots k_{i-1} a_i : \forall j, 0 \leq a_j < k_j \quad \text{and} \quad a_{2j} = 0 \right\}$$

and

$$B_t = \left\{ \sum_{i=1}^{t} k_1 k_2 \dots k_{i-1} b_i : \forall j, 0 \le b_j < k_j \quad \text{and} \quad b_{2j-1} = 0 \right\}.$$

Note that for all j, $A_{2j} = A_{2j-1}$ and $B_{2j-1} = B_{2j-2}$. Let $A^* = \bigcup_{i=0}^{\infty} A_t$ and $B^* = \bigcup_{i=0}^{\infty} B_t$. It only remains to prove that $A = A^*$ and $B = B^*$. We will use the following claim.

Claim 5.13. *For all $t \ge 0$*

$$\begin{aligned} A \cap \{0, 1, \dots, k_1 \cdots k_t - 1\} &= A_t \\ B \cap \{0, 1, \dots, k_1 \cdots k_t - 1\} &= B_t. \end{aligned}$$

Proof. Suppose not and let t be minimal such that the claim does not hold. Thus there must exist an $x \in \mathbb{N}$ such that either $x \in (A \cap \{0, 1, \dots, k_1 k_2 \cdots k_t - 1\}) \Delta A_t$ or $x \in (B \cap \{0, 1, \dots, k_1 k_2 \cdots k_t - 1\}) \Delta B_t$, where Δ denotes the symmetric difference of sets. Pick a minimal such x. Let us assume that t is odd and $t \ge 3$; the proof is trivial for $t = 0$ or $t = 1$ and similar when $t \ge 2$ is even. As t is odd (and minimal) $B_t = B_{t-1} = B \cap \{0, 1, \dots, k_1 \cdots k_{t-1} - 1\} \subset B \cap \{0, 1, \dots, k_1 \cdots k_t - 1\}$, thus $B_t \backslash (B \cap \{0, 1, \dots, k_1 \cdots k_t - 1\})$ is empty.

Now write

$$x = \sum_{i=1}^{t} k_1 k_2 \dots k_{i-1} a_i$$

in the mixed-radix representation with bases $(k_i)_{i=1}^{\infty}$. Set

$$z = \sum_{i=0}^{\lfloor \frac{t}{2} \rfloor} k_1 \cdots k_{2i} a_{2i+1}$$

and

$$w = \sum_{i=1}^{\lfloor \frac{t}{2} \rfloor} k_1 \cdots k_{2i-1} a_{2i}.$$

By definition, $z \in A_t$, $w \in B_t = B_{t-1}$ and $x = z + w$. By the minimality of t, $B_{t-1} \subset B$, thus $w \in B$. We now distinguish the remaining three cases.

(i) Suppose $x \in (A \cap \{0, 1, \ldots k_1 \cdots k_t - 1\}) \backslash A_t$. Since $x \notin A_t$, we have $x \neq z$, thus $z \in A$ by minimality of x. Now we have that $x, z \in A$ and $0, w \in B$. But $x + 0 = z + w$, contradicting the fact that A and B are co-Sidon.

(ii) Suppose $x \in A_t \backslash (A \cap \{0, 1, \ldots, k_1 \cdots k_t - 1\})$. As $A + B = \mathbb{N}_0$, we can write $x = a + b$ with $a \in A$, $b \in B$. Note that $x \leq k_1 k_2 \cdots k_t - 1$ and this implies $x \notin A$. In particular, $x \neq a$. We claim that $x = b$. If not, then $0 < a, b < x$ and the minimality of x implies that $a \in A_t$ and $b \in B_t$. But $a + b = x \in A_t$, which contradicts the definition of A_t and B_t. Thus we may suppose $x = b$, i.e., $x \in A_t \cap B$.

For $0 \leq i \leq \left\lfloor \frac{t}{2} \right\rfloor - 1$, define

$$\alpha_{2i+1} = \begin{cases} k_{2i+1} - a_{2i+1} & \text{if} \quad a_{2i+1} > 0 \\ 0 & \text{if} \quad a_{2i+1} = 0 \end{cases}$$

and

$$\beta_{2i+2} = \begin{cases} 0 & \text{if} \quad \alpha_{2i+1} = 0 \\ 1 & \text{if} \quad \alpha_{2i+1} > 0. \end{cases}$$

Let

$$u = (\alpha_{t-1} 0 \alpha_{t-4} \ldots \alpha_3 - \alpha_1)_{(k_i)} = \sum_{i=0}^{\left\lfloor \frac{t}{2} \right\rfloor - 1} k_1 \cdots k_{2i} \alpha_{2i+1} \in A_{t-2},$$

$$v = (\beta_{t-1} 0 \beta_{t-3} 0 \ldots \beta_2 0)_{(k_i)} = \sum_{i=1}^{\left\lfloor \frac{t}{2} \right\rfloor} k_1 \cdots k_{2i-1} \beta_{2i}.$$

By definition of k_t, $a_t \prod_{i=0}^{t-1} k_i \in A$ and by minimality of t, we have $u \in A$ and $v \in B$. Clearly, $u \neq a_t \prod_{i=0}^{t-1} k_i$. But $u + x = a_t \prod_{i=0}^{t-1} k_i + v$, contradicting the fact that A and B are co-Sidon.

(iii) Suppose $x \in (B \cap \{0, 1, \ldots, k_1 \cdots k_t - 1\}) \backslash B_t$. Clearly $x \notin A$, otherwise $0, x \in A \cap B$ which contradicts A, B being co-Sidon. Also $x \notin A_t$, otherwise $x \in A_t \cap B$ and we can continue as at the end of case (ii). Thus $x \neq z$, this implies $z \in A$ by the minimality of x. Also $w \in B_t$ implies $x \neq w$. Now $0 + x = z + w$, with $0, z \in A$ and $x, w \in B$ contradicting the fact that A and B are co-Sidon. □

To complete the proof of the theorem, we must show $\forall t > 0, k_t \geq 2$. Suppose that $k_{t_0} = 1$. That is, $c_{t_0} = k_1 k_2 \cdots k_{t_0 - 1}$ is in neither A nor B. But then as A and B are perfect co-Sidon, there exist $a \in A$ and $b \in B$ such that $a + b = c_{t_0}$. By assumption, $a, b < c_{t_0}$. But clearly $(a, b) \notin A_{t_0} \times B_{t_0}$ as $A_{t_0} + B_{t_0} \subset \{0, 1, \ldots, c_{t_0} - 1\}$ contradicting Claim 5.13. □

Theorem 5.12 allows us to make a useful observation about the structure of perfect co-Sidon sets.

Corollary 5.14. *If A and B are infinite perfect co-Sidon sets then for all $m \in \mathbb{N}$ there are infinitely many $n \in \mathbb{N}$ such that $\{n, n+1, \ldots, 2n+m\} \cap A = \emptyset$.*

Proof. As the statement remains true when we translate A or B, it suffices to prove it for A and B with $0 \in A \cap B$. There exists an infinite sequence of integers (k_i) $\forall i, k_i \geq 2$ such that A and B are represented by the bases k_i as in Theorem 5.12. Fix $m \in \mathbb{N}$ and let t be such that $2 \prod_{i=0}^{t-1} k_i - 3 \geq m$ and $(k_t - 1) \prod_{i=0}^{t-1} k_i \in A$. Then by Theorem 5.12 the next element in A is exactly $\prod_{i=0}^{t+1} k_i$. Let $n = (k_t - 1) \prod_{i=0}^{t-1} k_i + 1$. Now

$$
\begin{aligned}
\prod_{i=0}^{t+1} k_i &= k_{t+1} \{(k_t - 1) + 1\} \prod_{i=0}^{t-1} k_i \\
&\geq 2 \left\{ n - 1 + \prod_{i=0}^{t-1} k_i \right\} \\
&\geq 2n - 2 + m + 3 = 2n + m + 1.
\end{aligned}
$$

Thus $\{n, n+1, \ldots, 2n+m\} \cap A = \emptyset$. Since A is infinite, it follows that for every m there are infinitely many such n. $\square$

Of course, the claim also holds for B.

It is natural to ask whether all co-Sidon sets A, B are subsets of perfect co-Sidon sets A^*, B^*. The answer turns out to be no as the following proposition shows.

Proposition 5.15. The sets $A = \{2^k : k \in \mathbb{N},\ k \geq 9\}$ and $B = \{3^l : l \in \mathbb{N},\ l \geq 9\}$ are co-Sidon and there are no perfect co-Sidon sets A^*, B^* such that $A \subseteq A^*$ and $B \subseteq B^*$.

Proof. The Diophantine equation $2^k + 3^l = 2^m + 3^n$ with $k < m$ and $l > n$ has only five solutions (see [32]); all have exponents less than 9. This implies that A and B are co-Sidon.

Note that, for all $n \geq 2^9$, A contains numbers between n and $2n$. That is, for all n, $A \cap \{n, n+1, \ldots, 2n\} \neq \emptyset$. However, if A^* and B^* are perfect co-Sidon sets such that $A \subset A^*$ and $B \subset B^*$, then according to Corollary 5.14 there is an n with $A^* \cap \{n, n+1, \ldots, 2n+m\} = \emptyset$. $\square$

The discussion of perfect and co-Sidon pairs of sets in this section prepares us for the study of the representation function in the next section.

5.3 Representation Function

We seek to provide sufficient conditions on A and B so that the representation function $r(A, B, n) = |\{(a, b) \in A \times B : a + b = n\}|$ is (eventually) monotone increasing. For $C \subset \mathbb{N}_0$ let us denote its complement $\overline{C} = \mathbb{N}_0 \backslash C$.

It is easy to see that if either A or $\overline{A}$ is finite and either B or $\overline{B}$ is finite then $r(A, B, n)$ is eventually monotone. To see this, if $\overline{A}$ and B are finite, then for all

$n > \max(\overline{A}) + \max(B)$ we have that $b \in B$ implies $n - b \in A$ and thus $r(A, B, n) = |B|$. Also, if $\overline{A}$ and $\overline{B}$ are finite, then for all $n > \max(\overline{A}) + \max(\overline{B})$ we have $r(A, B, n) = n + 1 - |\overline{A}| - |\overline{B}|$. Finally, if A and B are both finite then it is obvious that $r(A, B, n)$ is eventually monotone. So the study is non-trivial only in the case when A and $\overline{A}$ are both infinite.

Proposition 5.16. Let $A, B \subset \mathbb{N}_0$ be infinite perfect co-Sidon sets such that $A + B = \mathbb{N}_0$. Then, for any $A' \subset A$ and $B' \subset B$, the representation function $r(A + B', B + A', n)$ is monotone increasing.

Proof. Note that

$$\begin{aligned} r(A + B', B + A', n) &= r\left(\bigcup_{b \in B'} A + b, \bigcup_{a \in A'} B + a, n\right) \\ &= \sum_{a \in A', b \in B'} r(A + b, B + a, n) \end{aligned}$$

The second equality holds because the unions are disjoint.

From $A + B = \mathbb{N}_0$ it follows that $(A + b) + (B + a) = \mathbb{N}_0 + a + b$ and thus each summand is

$$r(A + b, B + a, n) = \begin{cases} 0 & \text{if} \quad n < a + b, \\ 1 & \text{if} \quad n \geq a + b. \end{cases}$$

Therefore, the representation function $r(A + B', B + A', n)$ is monotone increasing. □

It follows from Theorem 5.12 that sets A and B which are infinite perfect co-Sidon exist. Since the subsets in Proposition 5.16 are arbitrary, we can construct many sets A and B such that $r(A, B, n)$ is monotone increasing. The next theorem allows us to choose sets A and B whose representation function is monotone and increasing and whose counting functions $A(n)$ and $B(n)$ grow at a controlled rate.

Theorem 5.17. *Let $A, B \subset \mathbb{N}_0$ be infinite perfect co-Sidon such that $A + B = \mathbb{N}_0$. Let $f : \mathbb{N}_0 \to \mathbb{R}$ be such that $A(n) \leq f(n)$ and for every $M > 0$ there exists n_0 such that for $n > n_0$ we have $f(n) < n + 1 - MA(n)$. Then there exists a $B' \subseteq B$ such that*

$$(A + B')(n) \leq f(n) \quad \text{for all} \quad n \in \mathbb{N}_0$$

and

$$(A + B')(n) \geq f(n) - A(n) \quad \text{for infinitely many} \quad n \in \mathbb{N}_0.$$

Proof. Let A and B be as in the statement and write $B = \{b_0 < b_1 < \ldots\}$. By assumption, $b_0 = 0$. Let us construct $B' \subseteq B$ greedily as follows: set $B'_0 = \{0\}$ and for $i > 0$ let

$$B'_{i+1} = \begin{cases} B'_i \cup \{b_{i+1}\} & \text{if} \quad \big(A + (B'_i \cup \{b_{i+1}\})\big)(n) \leq f_A(n) \quad \text{for all} \quad n \in \mathbb{N}_0, \\ B'_i & \text{otherwise.} \end{cases}$$

Then let $B' = \bigcup_{i=0}^{\infty} B'_i$. We claim that this B' satisfies the conditions of the theorem. By the construction,

$$(A + B')(n) \leq f(n) \quad \text{for all } n \in \mathbb{N}_0.$$

To prove that the other inequality holds for infinitely many values of n, we first need to show that $B \setminus B'$ is infinite. Suppose that $B \setminus B'$ is finite, and let $M = |B \backslash B'|$. Since $A + B \backslash B' = \cup_{b \in B \backslash B'} (A + b)$ we have $(A + B \backslash B')(n) \leq MA(n)$ for every n. Now, clearly

$$\bigcup_{b \in B'} (A + b) = \mathbb{N}_0 \backslash \left(\bigcup_{b \in B \backslash B'} (A + b) \right).$$

It follows that $\big(A + B'\big)(n) = n + 1 - \big(A + (B \backslash B')\big)(n) \geq n + 1 - MA(n)$ for all n.

But, for large enough n, we have $n+1-MA(n) > f(n)$. Then, for large enough n we would have $\big(A+B'\big)(n) > f(n)$, which contradicts the construction of B'. Hence $B \setminus B'$ is infinite.

Therefore, for infinitely many i, we have $b_{i+1} \notin B'$. For such an i we have $B'_{i+1} = B'_i$. Therefore, by definition of B'_{i+1}, there exists n_{i+1} such that $(A + B'_i \cup \{b_{i+1}\})(n_{i+1}) > f(n_{i+1})$. Note that $n_{i+1} \geq b_{i+1}$, because for all $n < b_{i+1}$,

$$\big(A + B'_i \cup \{b_{i+1}\}\big)(n) = \big(A + B'_i\big)(n) \leq f_A(n).$$

Therefore there are infinitely many n such that,

$$(A + B')(n) \geq (A + B'_i)(n) \geq f(n) - A(n).$$

□

Our main theorem follows as a corollary of Theorem 5.17. We restate our main theorem here for ease of reference:

Theorem. *For all $0 \leq \alpha, \beta < 1$, $1/2 < c_1, c_2 \leq 1$, there exist sets $A, B \subset \mathbb{N}_0$ such that $r(A, B, n)$ is monotone increasing in n;*

$$\limsup_{n\to\infty} \frac{A(n)}{n^{c_1}} = \alpha; \qquad \limsup_{n\to\infty} \frac{B(n)}{n^{c_2}} = \beta.$$

Proof. Suppose we are given constants $0 \leq \alpha < 1$ and $1/2 < c_1 \leq 1$ Let A_0, B_0 be perfect co-Sidon sets such that $A_0(n) = \Theta(n^{1/2})$, $B_0(n) = \Theta(n^{1/2})$ (e.g. Construction 5.7.) Let $f(n) = \alpha n^{c_1} + d$ where d is a constant large enough such that $f(n) \geq A_0(n)$ for all n. Clearly for all $m > 0$ there exists an n_0 such that for $n > n_0$, $f(n) < n + 1 - mA_0(n)$. By Theorem 5.17, there is a $B' \subset B_0$ such that $(A_0 + B')(n) \leq f(n)$ for all n and $(A_0 + B')(n) \geq f(n) - A_0(n)$ for infinitely many n.

Set $A = A_0 + B'$. Then

$$\alpha = \lim_{n\to\infty} \frac{f(n)}{n^{c_1}} \geq \limsup_{n\to\infty} \frac{A(n)}{n^{c_1}} \geq \lim_{n\to\infty} \frac{f(n) - A_0(n)}{n^{c_1}} = \alpha.$$

We can construct B in the same manner. By Proposition 5.16, the representation function $r(A, B, n)$ is monotone increasing. □

By modifying the previous two proofs, we can restate Theorem 5.3 with either (or both) limit superiors replaced with limit inferiors. The details are left to the interested reader. Theorem 5.3 gives a strong answer about the densities of sets A and B with monotone representation function $r(A, B, n)$.

When $c_1 = c_2 = 1$ and $\alpha, \beta \in \mathbb{Q}$ we can restate Theorem 5.3 by replacing the limit superiors with standard limits.

Theorem 5.18. *For all rational* $0 \leq \alpha, \beta \leq 1$, *there exist sets* $A, B \subset \mathbb{N}_0$ *such that* A *has density* α, B *has density* β *and* $r(A, B, n)$ *is monotone increasing in* n.

Proof. We construct A and B using mixed radix representation to describe its elements. Write $\alpha = p_1/q_1$ and $\beta = p_2/q_2$ where $p_i, q_i \in \mathbb{N}$. Set $k_1 = q_1$, $k_2 = q_2$ and $k_i = 2$ for all $i > 2$. Let A_0 be the set of all integers that can be written in the form

$$\sum_{i=0}^{k} k_1 k_2 \cdots k_{2i} a_{2i+1}$$

where for each i, $0 \leq a_{2i+1} < k_{2i+1}$. Similarly let B_0 be the set of all integers that can be written in the form

$$\sum_{i=1}^{k} k_1 k_2 \cdots k_{2i-1} b_{2i}$$

where for each i, $0 \leq b_{2i} < k_{2i}$. Note that A_0 and B_0 are perfect co-Sidon.

Let A' be the subset of A_0 consisting of all those integers whose k_1-digit (in the mixed radix representation) lies in the set $\{0, 1, \ldots, p_1 - 1\}$. As $p_1 \leq q_1$ we must have

$p_1 - 1 \leq k_1 - 1$. Thus A' is well-defined. Then $B = A' + B_0$ is the set of all numbers whose k_1-digit lies in $\{0, \ldots, p_1 - 1\}$ that is, B consists of the numbers congruent to $0, 1, \ldots, p_1 - 1 \pmod{q_1}$. The density of this set is clearly p_1/q_1.

Similarly, let B' be the subset of B_0 consisting of all those integers whose k_2-digit (in the mixed radix representation) lies in the set $\{0, 1, \ldots, p_2 - 1\}$. Again as $p_2 \leq q_2$ we have $p_2 - 1 \leq k_2 - 1$ so B' is also well-defined. A similar argument holds when we are considering $A = A_0 + B'$. Here, A is the set of numbers whose k_2-digit is in $\{0, 1, \ldots, p_2 - 1\}$. Thus A consists exactly of the numbers less than or equal to $(p_2 - 1)q_1 \pmod{q_1 q_2}$. This follows as the base of the first digit is q_1. Again it is clear that A has density $(p_2 q_1)/(q_1 q_2) = p_2/q_2$.

By Proposition 5.16, $r(A, B, n)$ is monotone increasing. $\square$

Finally, we determine for which sets A, B the representation function $r(A, B, n)$ is eventually *strictly* increasing. The corresponding question for a single set has been considered by Chen and Tang [8] who discuss when the functions r, r_1, r_2 are strictly increasing. When considering two sets and the function r, the problem turns out to be easy.

Proposition 5.19. Let $A, B \subset \mathbb{N}_0$, then the representation function $r(A, B, n)$ is eventually strictly monotone increasing if and only if $\overline{A}$ and $\overline{B}$ are finite.

Proof. First, let us assume that $r(A, B, n)$ is eventually strictly increasing. We will use the trivial identity that

$$n + 1 = r(\mathbb{N}_0, \mathbb{N}_0, n) = r(A, B, n) + r(\overline{A}, B, n) + r(A, \overline{B}, n) + r(\overline{A}, \overline{B}, n),$$

which is equivalent to

$$n + 1 - r(A, B, n) = r(\overline{A}, B, n) + r(A, \overline{B}, n) + r(\overline{A}, \overline{B}, n).$$

In the last identity the left hand side is bounded, since we assumed that $r(A, B, n)$ is eventually strictly increasing. Thus so is the right hand side. Hence $r(\overline{A}, B, n)$, $r(A, \overline{B}, n)$ and $r(\overline{A}, \overline{B}, n)$ are all bounded. From this it follows that $r(\overline{A}, \mathbb{N}_0, n) = r(\overline{A}, B, n) + r(\overline{A}, \overline{B}, n)$ and $r(\mathbb{N}_0, \overline{B}, n) = r(A, \overline{B}, n) + r(\overline{A}, \overline{B}, n)$ are bounded. Thus $\overline{A}$ and $\overline{B}$ must be finite.

Now we assume that $\overline{A}$ and $\overline{B}$ are finite. For any $n > \max(\overline{A}) + \max(\overline{B})$ we know that $a \in \overline{A}$ implies $n - a \notin \overline{B}$ and vice versa, so we can write

$$\begin{aligned} r(A, B, n) &= n + 1 - |\overline{A}| - |\overline{B}| \\ &< n + 2 - |\overline{A}| - |\overline{B}| = r(A, B, n+1) \end{aligned}$$

Thus for $n > \max(\overline{A}) + \max(\overline{B})$ the representation function is strictly increasing. $\square$

This concludes the discussion of our generous sufficient conditions on sets A and B for the representation function to be eventually monotone increasing. It would also be interesting to give some necessary conditions.

5.4 Open Problems

A far-reaching goal would be to completely characterize co-Sidon sets. Which co-Sidon sets are subsets of some perfect co-Sidon sets? Are two random sets likely to be co-Sidon?

Can we completely characterize sets A, B whose representation function is monotone increasing? Are there constructions that do not come from perfect co-Sidon sets?

Bibliography

Note: Numbers following each bibliography item indicate the pages they are cited from in the dissertation.

[1] Balasubramanian, R., *A note on a result of Erdős, Sárközy and Sós*, Acta Arith. 49 (1987), 45–53. 80

[2] Balister, P., Kalikow, S. and Sarkar, A., The linus sequence, *Comb. Probab. Comput.* **19, 1 (January 2010)**, (2010), 21–46. 60

[3] Benevides, F., Hulgan, J., Lemons, N., Palmer, C., Riet, A. and Wheeler, J., Additive Properties of a Pair of Sequences, *Acta Arith.* **139** (2009), no. 2, 185–197. 3

[4] Biró, Cs., Horn, P. and Wildstrom, D.J., On Hajnal's triangle free game, *slides for the conference Infinite and finite sets, June 13-17, 2011, Budapest.* 2, 8

[5] Bollobás, B., *Extremal Graph Theory*, 1978, Academic Press. 7, 39

[6] Bollobás, B., *Modern Graph Theory*, 1998, Springer. 3, 41

[7] Bryant, D. E., Khodkar, A. and El-Zanati, S. I., Small embeddings for partial G-designs when G is bipartite, *Bull. Inst. Combin. Appl.* **26** (1999), 86–90. 37

[8] Chen, Y. and Tang, M., *On the monotonicity properties of additive representation functions, II*, preprint. 95

[9] Ehrenfeucht, A. and Mycielski, J., A pseudorandom sequence: how random is it?, in Guy, R. K., Unsolved problems, *American Mathematical Monthly* **99** (1992), 373–375. 60

[10] Erdős, P., Sárközy, A. and Sós, V. T., *Problems and results on additive properties of general sequences, III,* Studia Sci. Math. Hungar. 22 (1987), 53–63; *IV*, in: Number Theory, Proceedings, Ootacamund, India, 1984, ed. K. Alladi, Lecture Notes in Mathematics 1122, Spring-Verlag, Berlin-Heidelberg-New York, 1985, 85–104; and *V*, Monatshefte Math. 102 (1986), 183–197. 80

[11] Erdős, P. and Turán, P. *On a problem of Sidon in additive number theory, and on some related problems*, J. London Math. Soc. 16 (1941), 212–215. 83

[12] Ferrara, M., Harris, A. and Jacobson, M., The game of $\mathcal{F}$-saturator, *Discrete Appl. Math.* **158** (2010), no. 3, 189–197. 7

[13] Füredi, Z. and Lehel, J., Tight embeddings of partial quadrilateral packings, *Journal of Combinatorial Theory, Series A* **117(4)** (2010), 466–474. 2, 37, 46, 57

[14] Füredi, Z., Reimer, D. and Seress, A., Triangle-Free Game and Extremal Graph Problems, *Congr. Numer.* **82** (1991), 123–128. 1, 2, 8

[15] Füredi, Z., Riet, A. and Tyomkyn, M., Completing Partial Packings of Bipartite Graphs, *J. Comb. Th. Ser. A* **118** (2011), no. 8, 2463–2473. 2

[16] Gustavsson, T., Decompositions of large graphs and digraphs with high minimum degree, *Doctoral Dissertation*, Dept. of Mathematics, Univ. of Stockholm, 1991. 42

[17] Häggkvist, R., Decompositions of complete bipartite graphs. Surveys in combinatorics, 1989 (Norwich, 1989), 115–147, London Math. Soc. Lecture Note Ser., 141, Cambridge Univ. Press, Cambridge, 1989. 43

[18] Halberstam, H. and Roth, K. F. *Sequences* (Vol 1), Oxford University Press, 1966, 89–90. 83, 84

[19] Hilton, A. J. W. and Lindner, C. C., Embedding partial 4-cycle systems, *manuscript*. 37

[20] Hoffman, D. G., Lindner, C. C. and Rodger, C. A., A partial $2k$-cycle system of order n can be embedded in a $2k$-cycle system of order $kn + c(k); k \geq 3$, where $c(k)$ is a quadratic function of k. *Discrete Math.* **261** (2003), 325-336. 37

[21] Jenkins, P., Embedding partial G-designs where G is a 4-cycle with a pendant edge, *Discrete Math.* **292** (2005), 83-93. 37

[22] Krückeberg, F. *B_2-Folgen und verwandte Zahlenfolgen*, J. Reine Angew. Math. 206 (1961), 53–60. 84

[23] Küçükçifçi, S., Lindner, C. C. and Rodger, C. A., A partial kite system of order n can be embedded in a kite system of order $8n + 9$, *Ars Combinatoria*, **79** 2006, 257-268. 37

[24] Lindner, C. C., A partial 4-cycle system of order n can be embedded in a 4-cycle system of order at most $2n + 15$, *Bull. Inst. Combin. Appl.* **37** (2003), 88-93. 37

[25] Lindner, C. C., A small embedding for partial 4-cycle systems when the leave is small, *J. Autom. Lang. Comb.* **8** (2003), 659-662. 37

[26] Lindner, C. C. and Rodger, C. A., Decomposition into cycles II: Cycle systems, in *Contemporary Design Theory: A collection of surveys* (J. H. Dinitz and D. R. Stinson, Eds.), John Wiley and Sons, 1992, pp. 325-369. 37

[27] Lindner, C.C., Rodger, C. A. and Stinson, D. R., Embedding cycle systems of even length. *J. Combin. Math. Combin. Comput.* **3** (1988), 65-69. 37

[28] Pyber, L., Regular subgraphs of dense graphs. *Combinatorica.* **5** (1985), no. 4, 347-349. 56

[29] Ruzsa, I. Z. *Solving a linear equation in a set of integers I*, Acta Arith. 65 (1993), 259–282. 83

[30] Sárközy, A. *Unsolved problems in number theory*, Period. Math. Hungar. 42 (2001), 17–35. 80

[31] Stöhr, A. *Gelöste und ungelöste Fragen über Basen der natürlichen Zahlenreihe II*, J. Reine Angew. Math. 194 (1955), 111–140. 84

[32] Tijdeman. R. and Wang, L. *Sums of products of powers of given prime numbers*, Pacific J. Math. 132 (1988), 177-193.; Corr: Pacific J. Math. 135 (1988), 396-398. 90

[33] West, D., The F-Saturation Game (2009) and Game Saturation Number (2011), *http://www.math.uiuc.edu/~west/regs/fsatgame.html* (last visited 11/25/2012). 7

[34] Wilson, R. M., Decomposition of complete graphs into subgraphs isomorphic to a given graph, *Congr. Numer.* **15** (1975), 647-659. 36, 42