4-15-2012

Problems in Extremal Combinatorics

Neal Owen Bushaw

PROBLEMS IN EXTREMAL COMBINATORICS

by

Neal Owen Bushaw

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Mathematical Sciences

The University of Memphis

May 2012

To Beh -

For teaching me that mathematics is possible.

ACKNOWLEDGMENTS

This dissertation could not exist without the help, guidance, and support of a huge number of people, both junior and senior, mathematical and non-mathematical. It is a great honor and pleasure to thank at least some of them here. Since I am given only finite space for this, I certainly cannot thank everyone here. If you are not mentioned, assume it is because you have been so crucial that I assumed you already knew you were thanked.

Without the support and guidance of my supervisor, Professor Dr. Béla Bollobás, none of this would have been possible. I would like to thank him for his steady encouragement, advice, patience, and support. Many thanks for the never ending supply of interesting problems, as well as for introducing me to many of the great mathematicians and cities of the world.

On this note, I would also like to thank Mrs. Gabriella Bollobás, for teaching me that not all mathematicians need act according to the prototype; we can be and are artistic beings, and there is more than just mathematics to be seen in the world.

I thank my examiners, Professors Paul Balister, Anna Kaminska, Jenő Lehel, and Vladimir Nikiforov for the comments and suggestions.

Prof. Paul Balister has been of immeasurable help to me throughout my time in Memphis. He is a wealth of knowledge and problems, and has been willing and eager to help me and provide advice whenever I have needed it, which has been often.

I would like to thank my Master's advisor, Professor Amites Sarkar, for showing me the beauty of extremal combinatorics, and for starting me along the path which has brought me to Memphis and to this dissertation.

Many thanks are due to my collaborators, Karen Johannson, Nathan Kettle, and Dr. Steven Kalikow, for numerous hours of interesting and fruitful discussion, as well as many more hours of fruitless discussion.

I would also like to thank all of my past teachers. From Silver Ridge Elementary, Central Kitsap Junior High, Central Kitsap High School, the University of Colorado, Waikato University, Western Washington University, and the University of Memphis, I could not have asked for a better education. In particular, I'd like to thank Mr. Mike Menefee and Mr. David Pevovar in the Central Kitsap School District for allowing me to explore my own interests during their classes, and to discover my true fields of interest at an early age.

During my University career I have innumerable outstanding teachers. I would like to thank in particular Professors Alexander Gorokhovsky and Marty Walters at the University of Colorado, Professors Richard Gardner and David Hartenstine at Western Washington University, and Professor Jim Campbell at the University of Memphis for helping me see the beauty of pure mathematics.

I am also indebted to a great many people for their non-mathematical support. Thanks to Ms. Tricia Simmons for her immense help with everything from housing, funding, and applications, to figuring out what to do when a foreign visa might not arrive in times. I would almost certainly be stranded in a foreign country without your help.

Thanks to my colleagues Ago, Andrew, Dominik, Fabricio, Karen, Ping, Richard, and Tomas for their support and emphatic discussions over the past years. Thanks also go to my friends outside of school, in particular Cathy, Chris, Emma, and Luke, for helping me not freak out during the long periods of work that have gone into this. I also must thank Marshall, the world's leading Cat Scientist, for scattering my papers around.

Without the never ending support of my parents, Bonnie and Gordon Bushaw, none of this would have been possible. From teaching me at an early age to count to ten including e and π, to supporting me throughout my university education and graduate school, I could not have asked for better; I am forever indebted to you for making me the human, teacher, and mathematician that I am. You may have been beaten out for the dedication, but this is for you.

ABSTRACT

Bushaw, Neal Owen. PhD. The University of Memphis. May 2012.
Problems in Extremal Combinatorics. Major Professor: Dr. Béla Bollobás.

This dissertation is divided into two major sections. Chapters 1 to 4 are concerned with Turán type problems for disconnected graphs and hypergraphs. In Chapter 5, we discuss an unrelated problem dealing with the equivalence of two notions of stationary processes.

The *Turán number* of a graph H, $\mathrm{ex}(n, H)$, is the maximum number of edges in any n-vertex graph which is H-free. We discuss the history and results in this area, focusing particularly on the degenerate case for bipartite graphs.

Let P_l denote a path on l vertices, and $k \cdot P_l$ denote k vertex-disjoint copies of P_l. We determine $\mathrm{ex}(n, k \cdot P_3)$ for n appropriately large, confirming a conjecture of Gorgol. Further, we determine $\mathrm{ex}(n, k \cdot P_l)$ for arbitrary l, and n appropriately large. We provide background on the famous Erdős-Sós conjecture, and conditional on its truth we determine $\mathrm{ex}(n, H)$ when H is an equibipartite forest, for appropriately large n.

In Chapter 4, we prove similar results in hypergraphs. We first discuss the related results for extremal numbers of hyperpaths, before proving the extremal numbers for multiple copies of a loose path of fixed length, and the corresponding result for linear paths. We extend this result to forests of loose hyperpaths, and linear hyperpaths. We note here that our results for loose paths, while tight, do not give the extremal numbers in their classical form; much more detail on this is given in Chapter 4.

In Chapter 5, we discuss two notions of stationary processes. Roughly, a process is a uniform martingale if it can be approximated arbitrarily well by a process in which the letter distribution depends only on a finite amount of the past. A random Markov process is a process with a coupled ‘look back’ time; that is, to determine the letter distribution, it suffices to choose a random look-back

time, and then the distribution depends only on the past up to this time. Kalikow proved that on a binary alphabet, any uniform martingale is also a random Markov process. We extend this result to any finite alphabet.

Table of Contents

Contents | Pages

List of Figures

Chapter 1. Introduction

1.1 Introduction

This dissertation is divided into two major parts. The first part, Chapters 1 to 4, are concerned with Turán theory. This chapter gives a basic overview of the structure of the paper, and also establishes the bulk of the terminology and notation used, saving those major and nonstandard definitions for their first appearances later in the work.

In Chapter 2, we provide explanation and history of the Turán problem, defining extremal numbers and graphs precisely. We give an overview of the known results, including milestones such as Turán's Theorem itself, Theorem 2.1.2 in this work, as well the Erdős-Stone Theorem, Theorem 2.1.3 in this dissertation, which determines the asymptotics of the extremal function for all graphs of chromatic number at least three. We then discuss the difficulties of finding extremal numbers for bipartite graphs, and survey the known results and major conjectures in the area. In particular, there is an extensive discussion of the extremal numbers for trees, and thus the Erdős-Sós Conjecture, as this result will be important to the results in Chapter 3.

Chapters 3 and 4 are joint work with Nathan Kettle, at the University of Cambridge. In this Chapter, we discuss the natural question of forbidding several disjoint copies of a graph. In particular, Theorem 3.1.2 determines the maximum number of edges in a graph without k vertexndisjoint copies of the path on three vertices, answering in the positive a conjecture of Gorgol. We then prove the corresponding result when forbidding k vertex disjoint copies of the path on l vertices, for $l > 3$; this is Theorem 3.1.4.

Chapter 3 continues with a proof of the extremal numbers for forests of a certain kind; those which we call equibipartite forests; this is Theorem 3.2.1. A careful definition of equibipartite forests is given in Chapter 3. It should be noted that our proof is inductive, and relies on the validity of the Erdős-Sós conjecture for those trees in the forest. In particular, for forests of paths, this is given by the Erdős-Sós, and so in this and many other cases, Theorem 3.2.1 holds without qualification.

In Chapter 4, we start out with a discussion of the hypergraph Turán problem, and again give an overview of known results. We then give definitions of several different notions of paths in hypergraphs, and state some earlier results giving extremal numbers for assorted paths. Building on this, we give the extremal numbers for k vertex disjoint linear paths in Theorem 4.5.2, and give tight bounds on the extremal number for k vertex disjoint loose paths of the same length in Theorem 4.3.1. We then extend these results to forests consisting of k paths of different length in Theorem 4.4.1 and Theorem 4.5.4. As a peculiarity of the result which we use as our base case, Theorem 4.2.4, we are only able to give tight bounds in the results for loose paths, and do not prove the exact result. For linear paths, we are able to determine these exactly.

Chapter 5 is joint work with Karen Johannson, now at the University of Cambridge, and Steven Kalikow at the University of Memphis. In this chapter, we define two notions of stationary processes: random Markov processes and uniform martingales. It was shown by Steven Kalikow in 1990 that for a binary alphabet, any uniform martingale is also a random Markov process. The reverse implication is clear, and we discuss this as well. In Section 5.2 we state Kalikow's result, and we give a reformulation of Kalikow's proof in the binary alphabet case, this allows us to introduce some notation and techniques that will be used heavily later. In Section 5.3, we state and prove the corresponding result on any finite alphabet.

1.2 Notation

Throughout, we follow the notation which is most widely accepted. For graph theoretic notation we follow primarily Bollobás [5]. We review the most commonly used terminology and notation in this section.

We use $\mathbb{N}$ to denote the natural numbers, $\{1, 2, \ldots\}$, and for $n \in \mathbb{N}$, we use $[n] = \{1, 2, \ldots, n\}$ to denote the set of the first n natural numbers. We also use $\mathbb{Z}$ for the set of all integers, and $\mathbb{Z}_-$ for the set of negative integers. Similarly, $\mathbb{Z}_{\leq 0}$ is used for the set of non-positive integers. We will not delve into such issues, but we assume the ZFC axioms of set theory and the Peano axioms for arithmetic on the natural numbers. Given a set X, we use $X^{(k)}$ to denote all the k element subsets of X; we will call a set of size k a k-set.

We use the the standard "big-oh" notations $\mathcal{O}(\cdot), \Theta(\cdot), \Omega(\cdot), o\left(\cdot\right)$ for growth of functions. Namely, $f(n) = \mathcal{O}(g(n))$ if and only if there exists constants C, n_0 such that for each $n' > n_0$, $|f(n')| \leq C\,|g(n')|$. A function $f(n) = \Omega(g(n))$ if and only if there exists C, n_0 such that for each $n' > n_0$, $|f(n')| \geq C\,|g(n')|$. A function $f(n) = \Theta(g(n))$ if and only if it is $\mathcal{O}(g(n))$ and $\Omega(g(n))$. We say that $f = o\left(g(n)\right)$ if and only if for each $\epsilon > 0$, there exists n_0 such that for every $n > n_0$, $|f(n)| \leq \epsilon\,|g(n)|$. Finally, $f(n) \sim g(n)$ if and only if $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$.

A *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 be on vertex set $V = [n]$ and edge set E.

A *hypergraph* is the natural generalization of graphs when edges are allowed more than two elements; formally, a hypergraph $H = (V, E)$ is an ordered pair, where V is again the vertex set, and $E \subseteq \mathcal{P}(V) \setminus \{\emptyset\}$. Throughout Chapter 4, we will restrict to a particular class of hypergraphs, which are defined as follows: an r*-uniform hypergraph* is an ordered pair $(V(H), E(H))$, where $V(H)$ and $E(H) \subseteq V(H)^{(r)}$ are again the *vertex* and *edge sets* of H. We will refer to these often as r-graphs, and note that a 2-graph is a graph. Where the case $r = 2$ gives the standard definition for graphs, we may give definitions only for r-hypergraphs.

A *subgraph* of an r-uniform hypergraph G is an r-uniform hypergraph G' such that $V(G') \subseteq V(G)$ and $E(G') \subseteq V(G')^{(r)} \cap E(G)$. Two hypergraphs G and H are *isomorphic* if there exists a bijection $f : V(G) \mapsto V(H)$ such that $\{v_1, \ldots, v_r\} \in E(G)$ if and only if $\{f(v_1), \ldots, f(v_r)\} \in E(H)$. When G and H are isomorphic, we write $G \cong H$. We say that G *contains* H if there exists a subgraph of G which is isomorphic to H; in this case we write $G \subseteq H$. If $H \not\subseteq G$, then G is H*-free*.

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 $N_G(V)$, is the set of all vertices which share an edge with v; that is, $N_G(V) := \{u \in V(G) : u \neq v, \{u, v\} \in E(G)\}$. We will often refer to this simply as the *neighborhood* of v, and will supress the G in the notation when it is clear from the context. The degree of a vertex $v \in G$ is the size of its open neighborhood in G; we denote this $d_G(v)$; here too, we will usually suppress the G when it is clear from the context. A graph G is called k-regular if the degree of every vertex in G is precisely k.

A set $U \subseteq V(G)$ is *independent* if it induces no edges. A bipartite graph is a graph G whose vertex set can be written as a disjoint union of two independent sets. Similarly, a graph has chromatic number k if its vertex set can be partitioned into k disjoint independent sets, but not into $k-1$ disjoint independent sets. We refer to a set of edges as independent if no two edges share an endpoint.

We use $G \cup H$ to denote the disjoint union of two r-graphs; that is, $V(G \cup H) = V(G) \cup V(G)$, and $E(G \cup H) = E(G) \cup E(H)$. Similarly, we use $k \cdot G$ to denote k vertex disjoint copies of G. We write $G + H$ for the join of G and H, the graph obtained from $G \cup H$ by adding all edges between vertices of G and vertices of H; formally, $V(G + H) = V(G) \cup V(H)$ and $E(G + H) = E(G) \cup E(H) \cup \{\{x, y\} : x \in G, y \in H\}$. K_t denotes the complete graph on t vertices, the graph with all possible edges, while E_t denotes the empty graph with t vertices with no edges. By M_t, we refer to a maximal matching on t vertices; that is, the graph on t vertices consisting of $\lfloor \frac{t}{2} \rfloor$ independent edges. We use $K_r(t)$ to denote the complete r-partite graph with t vertices in each class, and $K_{s,t}$ to denote the complete bipartite graph with s vertices in one class and t in the other.

A path in a graph G is a sequence of distinct vertices $v_1, \ldots, v_k$ such that $\{v_i, v_{i+1}\} \in E(G)$ for each $i \in [k-1]$. Such a path is of *length* $k-1$, but is *on* k *vertices*; we denote it by P_k, and note that in this we differ from many texts. An *odd path* is a path on an odd number of vertices, and an *even path* is a path on an even number of vertices.

For hypergraphs, the notion of paths is less trivial; for this reason we delay their definition until Chapter 4. A *cycle of length* k is a path on k vertices whose end vertices are adjacent; formally, $v_1, \ldots, v_k$ is a cycle if $\{v_i, v_{i+1}\} \in E(G)$ for each $i \in [k-1]$, and $\{v_1, v_k\} \in E(G)$.

Chapter 2. Turán Theory, History and Basics

2.1 History

Extremal problems are those which deal with the largest, smallest, or otherwise optimal structures with some constraint. Perhaps the earliest published example of extremal graph theory is that of Mantel's Theorem (see, e.g., [5]). Appearing originally in 1906, Mantel's Theorem says that among all triangle free graphs, the complete balanced bipartite graph contains the largest number of edges. Later, Eszter Klein demonstrated a construction giving a lower bound on the maximum number of edges in a C_4 free graph (communicated by Erdős in [9].

In 1940, Pál Turán proved that the graph on n vertices which has no K_r as a subgraph and the largest number of edges is the complete $(r-1)$-partite graph on n vertices which is 'as balanced as possible' [34, 35]. This is now known as the Turán graph $T_{r-1}(n)$, and is illustrated in Figure 2.1. Mantel's Theorem is the first, and Turán's Theorem the second, in a broad type of questions which ask, in a general sense, "Which graph on n vertices containing no copy of a fixed subgraph has the largest number of edges?". We now define Turán numbers precisely.

Definition 2.1.1. The *Turàn number*, or *extremal number*, of a graph G is defined as: $\mathrm{ex}(n, G) := \max\{|E(H)| : |V(H)| = n, G \not\subseteq H\}$. A graph H is called *extremal for* G, or simply *extremal* when G is clear from the context, if is H-free and has the maximum number of edges; that is, $H \not\subseteq G$, $|V(H)| = n$, and $|E(H)| = \mathrm{ex}(n, G)$. We use $H_{\mathrm{Ex}}(n, G)$ to denote the family of n-vertex graphs which are extremal for G.

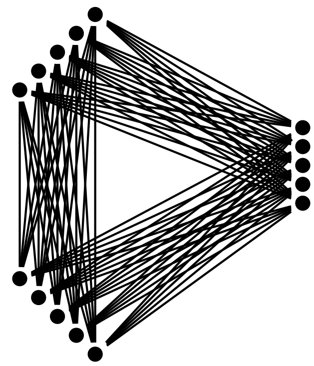

Fig. 2.1. The Turán Graph $T_3(15)$.

We now state Turán's milestone result, which in many ways was the starting point of a great amount of interest and work in the area [34, 35].

Theorem 2.1.2. *For each value of r and n, the following holds.*

$$\begin{aligned} \mathrm{ex}(n, K_r) &= |E\left(T_{r-1}(n)\right)| \\ &\leq \left(1 - \frac{1}{r-1}\right)\frac{n^2}{2} \end{aligned}$$

Further, for all n the unique extremal graph is the Turán graph $T_{r-1}(n)$, and we have equality in the second line when $r|n$.

□

While Theorem 2.1.2 determines the unique extremal graph forbidding K_r, it should be noted that in general, there is no reason to think that there is a unique extremal graph of a given size. Thus we use $\mathrm{Ex}(n, G)$ to denote the family of n-vertex extremal graphs for G, and as well use $H_{\mathrm{Ex}}(n, G)$ to denote a graph which is extremal for G and on n vertices. When this extremal graph is not unique, statements involving $H_{\mathrm{Ex}}(n, G)$ are valid regardless of the choice $H_{\mathrm{Ex}}(n, G) \in \mathrm{Ex}(n, G)$.

The next milestone in the advancement of Turán theory was a result of Paul Erdős and Arthur Stone; this theorem determines asymptotically the extremal number for the complete r-partite graph with t vertices in each class [11]. Further, as any r-chromatic graph is contained in $K_r(t)$ for large enough t, but not in $K_{r-1}(t)$, this determines the asymptotic extremal number for any r-chromatic graph. As a consequence, this theorem is often referred to as the "fundamental theorem of extremal graph theory". Somewhat later, 1966, Erdős and independently Simonovits gave a stability version of this result; that is, any graph which is $K_r(t)$-free and has nearly the maximum number of edges is almost $T_r(n)$ [31, 10]. We give the Erdős-Stone theorem here [11].

Theorem 2.1.3.

$$\mathrm{ex}(n, K_r(t)) = \frac{r-2}{r-1}\binom{n}{2} + o\left(n^2\right);$$

As a consequence, we have that for any graph G,

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

□

For graphs of chromatic number 3 and higher, this is somewhat the end of the road; while it is still possible to determine the extremal number of a graph exactly, the Erdős-Stone Theorem gives the asymptotic behavior. Further, determining extremal numbers exactly is quite difficult in general.

2.2 Bipartite Graphs

For a bipartite graph G, Theorem 2.1.3 gives us very little information about even the asymptotic behavior the extremal number for bipartite graphs. Indeed, for bipartite graphs this gives only that $\mathrm{ex}(n, G) = \frac{2-2}{2-1}\binom{n}{2} + o\left(n^2\right) = o\left(n^2\right)$. In general, exact extremal numbers are known for very few bipartite graphs, and even the asymptotics of $\mathrm{ex}(n, G)$ are known only for a few classes of bipartite graphs.

2.2.1 Bipartite Graphs with Cycles

Indeed for bipartite graphs which are not trees, the only results known give Turán densities, not exact results. Although the focus of Chapter 3 will be extensions of the results for trees in Section 2.2.2, we discuss these here for completeness.

Kövari, Sós and Turán proved a general upper bound on the extremal numbers for complete bipartite graphs, showing that $\mathrm{ex}(n, K_{s,t}) \leq \frac{1}{2}\sqrt[s]{t-1}n^{2-1/s} + o\left(n^{2-1/s}\right)$ [24]. This upper bound is quite simple; given a $K_{s,t}$ free graph $G = (V, E)$, this bound is implied by the inequality $\sum_{v\in V}\binom{d(v)}{s} \leq (t-1)\binom{n}{s}$. Zoltán Füredi improved this in 1996 to give a bound of $\mathrm{ex}(n, K_{s,t}) \leq \frac{1}{2}\sqrt[s]{t-s+1}n^{2-1/s} + o\left(n^{2-1/s}\right)$ [15]. This was recently improved by Nikiforov in [28] via an elegant spectral radius argument to the following bound.

$$\mathrm{ex}(n, K_{s,t}) \leq \frac{1}{2}\sqrt[s]{t-s+1}n^{2-1/s} + \frac{1}{2}(s-1)n^{2-2/s} + \frac{1}{2}(s-2)n.$$

The bound above is generally believed to give the correct term, but there exist constructions with a matching $\Omega(n^{2-1/s})$ edges in only a few cases. Paul Erdős, Alfréd Rényi, and Vera T. Sós, and independently Brown, proved the first such result, showing that $\mathrm{ex}(n, K_{2,2}) = \frac{1}{2}n^{3/2} + o\left(n^{3/2}\right)$ in 1966 [13, 7]. In the same paper, Brown further showed that $\mathrm{ex}(n, K_{3,3}) = \frac{1}{2}n^{5/3} + o\left(n^{5/3}\right)$.

Füredi also gave a bound matching construction in the case of $K_{2,t}$, showing that $\mathrm{ex}(n, K_{2,t}) = \frac{1}{2}\sqrt{t-1}n^{3/2} + o\left(n^{3/2}\right)$ [14]. The only remaining known cases are when one bipartite class is much larger than the other. For $t \geq s! + 1$, Kóllar, Rónyai, and Szabó showed that $\mathrm{ex}(n, K_{s,t}) \geq c_{s,t}n^{2-1/s}$ [22]; this construction was modified by Alon, Rónyai, and Szabo to give the same bound whenever $t \geq (s-1)! + 1$ [1].

The only other non-tree bipartite graphs for which the extremal numbers are known are cycles of length four, six, and ten. For completeness, we state these here as well. In 1974, Bondy and Simonovits proved an upper bound on the extremal numbers for all even cycles, showing that for all t, $\mathrm{ex}(n, C_{2t}) \leq 100tn^{1+1/t}$ [6]. Constructions matching the Bondy-Simonovits bound exist only for $k = 2, 3, 5$. The case $k = 2$ is the same as the case $K_{2,2}$ mentioned in the previous paragraph; the construction forbidding the hexagon was given by Benson in 1966 [4], while the result for C_{10} was proven by Wenger in 1991 [36]. It is worth noting that Wenger's algebraic construction also gave new optimal constructions for both C_4, C_6 as well.

We note that all of these constructions are quite complicated, and rely primarily on building graphs based on algebraic hypersurfaces in finite fields. As such, the constructions are valid not for any n, but typically for n which are an appropriate multiple of a prime or which are of a certain form. Regardless, these exist for infinitely n and are thus sufficient to give the asysmptotic results stated.

The careful reader may have noted that the cycle of length eight is missing from the above list of results. It is known that $c_1 n^{6/5} \leq \mathrm{ex}(n, C_8) \leq c_2 n^{5/4}$, for some constants c_1, c_2; the lower bound comes from constructions of Benson and Singleton [4, 33], while the upper bound is a recent result due to Keevash, Sudakov, and Verstraëte [21].

2.2.2 Trees

The extremal numbers for trees seem to be as hard as in the non-tree bipartite case. However, the following well known result due to Erdős and Gallai, gives a tight bound on the extremal number for paths of any length. We will use this as the base case of our induction in the proof of Theorem 3.1.4, and thus we state it carefully here [12].

Theorem 2.2.1. *For any* $n, l \in \mathbb{N}$, $\mathrm{ex}(n, P_l) \leq \frac{l-2}{2} n$.

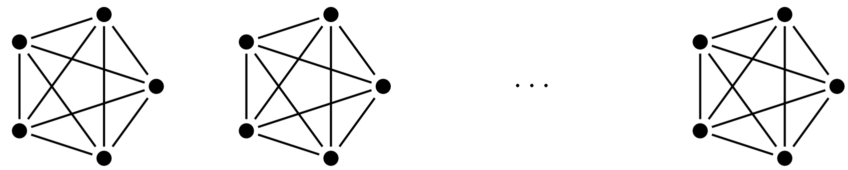

Fig. 2.2. $H_{\mathrm{Ex}}(n, P_6)$.

$\square$

The above theorem, proven in 1959, was the earliest of the bipartite extremal numbers to be determined. We note that the bound in Theorem 2.2.1 is attained by taking disjoint copies of K_{l-1} as in Figure 2.2; this extremal construction is unique as long, and gives a tight result whenever n is divisible by $l-1$.

We note that a path can be viewed as an extreme kind of tree - $l-2$ vertices have degree two, and the two leaves of course have degree one. The opposite extreme is the star - one central vertex of degree $l-1$, and the other $l-1$ vertices are leaves; we denote such a star by S_l. Forbidding the star S_l is, in fact, simply imposing a maximum degree condition, and so $\mathrm{ex}(n, S_l) \leq \frac{l-2}{2}n$. This bound is tight, with the extremal graphs being all $(l-2)$-regular graphs. Legend has it that Vera T. Sós presented the proofs of these two results to her graph theory class in Budapest in 1962, and left the following conjecture as a homework problem; by now, this is known as the notoriously difficult Erdős-Sós Conjecture.

Conjecture 2.2.2. *For any tree T on l vertices, $\mathrm{ex}(n, T) \leq \frac{l-2}{2}n$.*

In 2008, a proof of the conjecture was announced for very large trees by Ajtai, Komlós, Simonovits, and Szemerédi. For small trees, however, the conjecture is largely open. There is a sequence of results in the direction of the full theorem for smaller trees. We present a representative sample of these results here, which is certainly only the tip of the iceberg. Dobson established the conjecture for graphs of large girth; we state this theorem as an example of those partial results which exist.

Theorem 2.2.3. *If T is a tree on l vertices, and G is a graph with girth at least five and minimum degree $\delta \geq \frac{l}{2}$, then G contains T. Thus Conjecture 2.2.2 holds if the maximum number edges is taken over all T-free graphs of girth at least five.*

□

Similarly, Saclé and Woźniak [29] proved that whenever G is a graph with at least $\frac{l-2}{2}n$ edges and no C_4, G contains any tree on l vertices. In 2005, McLennan [25] proved the Erdős-Sós bound for trees of diameter at most four.

The Erdős-Sós Conjecture has also been proven for caterpillars; this result is attributed to Perles in [27]. Later, Sidorenko [30] showed that the Erdős-Sós Conjecture holds for trees of order l containing a vertex which is the parent of at least $\frac{l-1}{2}$ leaves. Many more partial results related to the Erdős-Sós Conjecture exist; see for example [2, 37].

Chapter 3. Turán Theory for Disconnected Graphs

The work in this Chapter is joint with Nathan Kettle at the University of Cambridge.

While the above theorems do not require that the forbidden graph be connected, they are typically thought of in this context. It is a natural question to consider forbidding multiple disjoint copies of a graph, or forbidding disconnected graphs. Theorem 2.1.3 still applies to these graphs, and so for graphs of chromatic number at least three the asymptotic behavior is the same. In [26] and [32] respectively, first Moon and then Simonovits showed that for large n, the extremal graph forbidding $p \cdot K_r$ is $K_{p-1} + T_{r-1}(n-p+1)$. While Moon's result gives this, it was not noted this way in his work. Simonovits independent paper expresses this result in the Turán context. We note that this is strictly larger than the number of edges in a graph forbidding only one copy of the complete graph.

Recently, Gorgol [17] proved upper and lower bounds on the extremal number for forbidding several vertex-disjoint copies of an arbitrary connected graph. We determine this number for paths of length 3 in Section 3.1.1, longer paths in Section 3.1.2, and for forests of equibipartite trees in Section 3.2.1.

3.1 Extremal Numbers for Disjoint Paths

Gorgol [17] noted that for any connected graph G on v vertices, and for any positive integers n, k such that $n \geq kv$, the graph $H_{\mathrm{Ex}}(n-kv+1, G) \cup K_{kv-1}$ does not contain $k \cdot G$. This is because K_{kv-1} simply does have enough vertices to contain k copies of G, and thus contains at most $k-1$, and by definition $H_{\mathrm{Ex}}(n-kv+1, G)$ is G-free; thus their union contains at most $k-1$ copies of G.

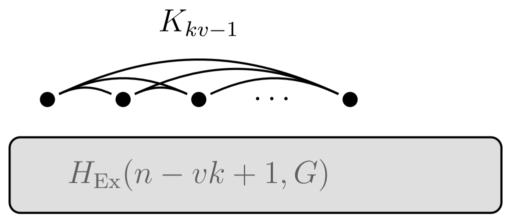

Fig. 3.1. Construction A.

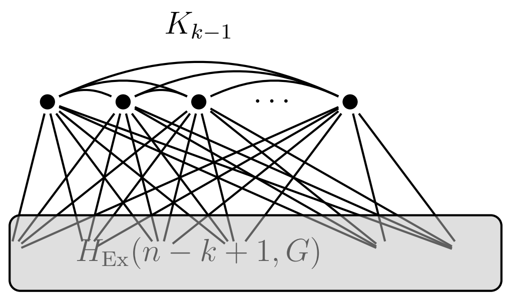

Fig. 3.2. Construction B.

Similarly, $H_{\mathrm{Ex}}(n-k+1,G)+K_{k-1}$ is $k\cdot G$-free as well. Indeed, by the definition of $H_{\mathrm{Ex}}(n-k+1,G)$, any copy of G must contain at least one vertex from the small complete graph, and since this complete graph contains only $k-1$ vertices, and so $H_{\mathrm{Ex}}(n-k+1,G)+K_{k-1}$ can contain at most $k-1$ copies of G. We shall refer to these two constructions as Construction A and Construction B, respectively.

3.1.1 Paths of Length Three

We start by looking at graphs with no disjoint paths on three vertices. The extremal case here is slightly different than for longer paths, but the proof introduces the main ideas we shall use in proving the result for all paths. Further, it serves as an introduction to the general tools needed for our results on forests, without getting extremely technical. Applying the constructions mentioned previously to the case where G is the path on three vertices, and noting that the second construction has strictly more edges for large enough n, Gorgol obtained the following bound on $\mathrm{ex}(n, k \cdot P_3)$ [17]. We again remind the reader that throughout P_l refers to the path on l vertices; in order to avoid any ambiguity, we will avoid referring to the length of a path.

$$\mathrm{ex}(n, k \cdot P_3) \geq \begin{cases} \binom{3k-1}{2} + \left\lfloor \frac{n-3k+1}{2} \right\rfloor, \text{ for } 3k \leq n < 5k-1, \\ \binom{k-1}{2} + (n-k+1)(k-1) + \left\lfloor \frac{n-k+1}{2} \right\rfloor, \text{ for } n \geq 5k-1. \end{cases} \tag{3.1}$$

Gorgol conjectured that for large enough n, Construction B is optimal for large n, and thus that the second bound in Equation (3.1) gives the correct value of $\mathrm{ex}(n, k \cdot P_3)$. In [17], she proved that this holds for $k = 2, 3$. Our first result shows that the second construction is indeed best possible for any k and large enough n. In fact, Theorem 3.1.4 and Theorem 3.2.1 show that a similar extremal structure holds for a much wider family of graphs.

Before stating the result for multiple copies of P_l, we give the following trivial lemma which will be used a the base case of our induction in Theorem 3.1.2.

Lemma 3.1.1. *If G is a graph on n vertices which contains no P_3, then G contains at most $\lfloor \frac{n}{2} \rfloor$ edges; that is, $\mathrm{ex}(n, P_3) = \lfloor \frac{n}{2} \rfloor$.*

Proof. The proof is immediate; if G contains no P_3, then no vertex can have degree ≥ 2, and so G consists of independent edges, giving $\mathrm{ex}(n, P_3) \leq \lfloor \frac{n}{2} \rfloor$. Clearly this maximum number of edges is obtained by a perfect matching when n is even and a matching leaving one vertex uncovered when n is odd, as in Figures 3.3, 3.4, respectively. $\square$

Theorem 3.1.2 (B., Kettle [8]). *For $n \geq 7k$, the following holds.*
$\mathrm{ex}(n, k \cdot P_3) = \binom{k-1}{2} + (n-k+1)(k-1) + \left\lfloor \frac{n-k+1}{2} \right\rfloor$.

Fig. 3.3. $H_{\mathrm{Ex}}(n, P_3)$, n even.

Fig. 3.4. $H_{\mathrm{Ex}}(n, P_3)$, n odd.

Further, there is a unique graph for which this bound is attained, namely $K_{k-1} + M_{n-k+1}$, as in Figures 3.5, 3.6. As mentioned in the description of Construction B in a more general context, this graph does not contain k disjoint copies of P_3, since each P_3 must contain at least one vertex from the $(k-1)$-clique.

Proof. We proceed by induction on k. The case $k = 1$ is covered by Lemma 3.1.1.

For the induction step, suppose G is a graph on n vertices, with $m > \mathrm{ex}(n, k \cdot P_3) = \binom{k-1}{2} + (n-k+1)(k-1) + \left\lfloor \frac{n-k+1}{2} \right\rfloor$ edges, $n \geq 7k$, and which contains no $k \cdot P_3$. The number of edges incident to any P_3 in G must be at least:

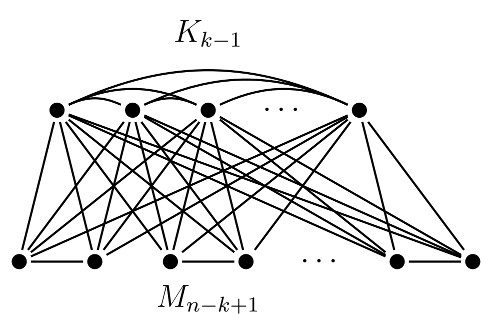

Fig. 3.5. $H_{\mathrm{Ex}}(n, k \cdot P_3)$, $n - k$ odd.

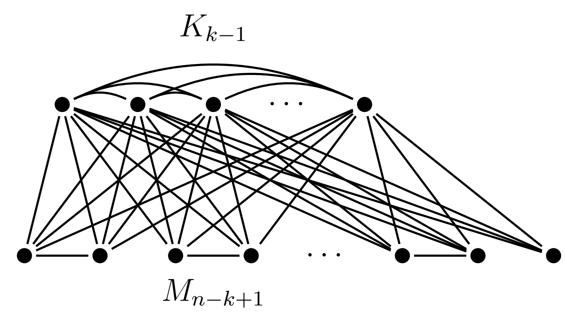

Fig. 3.6. $H_{\mathrm{Ex}}(n, k \cdot P_3)$, $n-k$ even.

$$\begin{aligned} m - &\operatorname{ex}(n-3, (k-1)\cdot P_3) && (3.2)\\ &\geq \binom{k-1}{2} + (n-k+1)(k-1) + \left\lfloor \frac{n-k+1}{2} \right\rfloor + 1 \\ &\quad - \binom{k-2}{2} - (n-k-1)(k-2) - \left\lfloor \frac{n-k-1}{2} \right\rfloor && (3.3)\\ &= n + 2k - 3. && (3.4) \end{aligned}$$

Otherwise, the graph induced by the vertices not on this P_3 contains $(k-1)\cdot P_3$ by induction, since $m - n + 2k - 3 \geq \operatorname{ex}(n-3, (k-1)\cdot P_3)$. Taking the union of this $(k-1)\cdot P_3$ and the initial P_3, which are disjoint by their construction, we are able to show that G does contain $k \cdot P_3$, a contradiction.

By the induction hypothesis, since for n in this range $m > \operatorname{ex}(n, k\cdot P_3) \geq \operatorname{ex}(n, (k-1)\cdot P_3)$, we can find $k-1$ vertex-disjoint copies of P_3 in our graph. Since each P_3 is incident to at least $n+2k-3$ edges, each must contain a vertex of degree at least $(n+2k-3)/3$. Taking such a high degree vertex from each P_3 gives us a set U of $k-1$ vertices each of degree at least $(n+2k-3)/3$.

Now, assume that $G\left[V \setminus U\right]$ contains P_3. Then, we can still construct another $k-1$ copies of P_3, each centered on a vertex from U, as long as each vertex in U has degree large enough to ensure it is connected to at least two vertices not contained on any of the preceding $k-1$ copies of P_3. That is, we can construct $k-1$ copies of P_3 as long as $(n+2k-3)/3 \geq 3k-1$. For $n \geq 7k$, $(n+2k-3)/3 \geq (7k+2k-3)/3 = 3k-1$, and so we are able to construct our $(k-1) \cdot P_3$. Thus it must be the case that $G[V \setminus U]$ contains no P_3, and so by Lemma 3.1.1, $|E(G[V \setminus U])| \leq \mathrm{ex}(n-k+1, P_3) = \left\lfloor \frac{n-k+1}{2} \right\rfloor$, and so G has at most $\binom{k-1}{2} + (n-k+1)(k-1) + \left\lfloor \frac{n-k+1}{2} \right\rfloor$ edges, a contradiction. $\square$

The above proof shows that Construction B is extremal for $n \geq 7k$. No construction is known giving a better bound for $n \geq 5k-1$, where Construction B first has more edges than Construction A, and we conjecture that the above example is optimal in this range. We note that the bound on n in Theorem 3.1.2 comes from guaranteeing that our collection of large degree vertices has large enough degree that we can construct $k-1$ non-intersecting copies of P_3 from this set, and that the bound on n is tight at this step.

3.1.2 Longer Paths

In the proof of Theorem 3.1.2, in order to find a P_3 it was enough to find a vertex of degree two. To find subsequent copies of P_3, it sufficed to find vertices of large degree; this is what is happening in the part of the proof surrounding (3.2), (3.3) . In a slightly more general sense, we will use collections of vertices with large shared neighborhood in order to 'build' copies of a fixed graph. This idea is a crucial tool in the proofs of both Theorem 3.1.4 and Theorem 3.2.1, and in fact will even be a primary tool in the hypergraph Turán problems discussed in Chapter 4. We formalize this notion for graphs in Lemma 3.1.3. The proof is straightforward, but having this Lemma as a standalone tool will be extremely useful.

Lemma 3.1.3. *Let G be a graph on n vertices with m edges, $t \in \mathbb{N}$, and let F_1, F_2 be arbitrary graphs. Letting $r = |V(F_1)|$ and $m' = m - \mathrm{ex}(n-r, F_2) - \binom{r}{2}$, if $F_1 \cup F_2 \not\subseteq G$, any F_1 in G contains t vertices with shared neighborhood of size at least*

$$n' \geq \frac{m' - (n-r)(t-1)}{(r-t+1)\binom{r}{t}}$$

Proof. Assume that there is some copy of F_1 inside G, say on vertex set U. Since G doesn't contain $F_1 \cup F_2$, clearly $G[V \setminus U]$ contains no F_2. Thus $G[V \setminus U]$ contains at most $\mathrm{ex}(n-r, F_2)$ edges. Since there are at most $\binom{|V(F_1)|}{2} = \binom{r}{2}$ edges with both endpoints in U, it must be that U has at least $m - \mathrm{ex}(n-r, F_2) - \binom{r}{2} = m'$ edges to $V \setminus U$.

Define n_0 to be the number of vertices in $V \setminus U$ with neighborhood of size at least t in U; that is, $n_0 = |\{v \in V \setminus U : |N_U(v)| \geq t\}|$. Since U has at most $n_0 r + (n - r - n_0)(t-1)$ edges to $V \setminus U$, $n_0 r + (n - r - n_0)(t-1) \geq m'$, and so $n_0 \geq \frac{m'-(n-r)(t-1)}{r-t+1}$. Trivially, there are only $\binom{r}{t}$ subsets of size t in F_1. Putting these together, we see that some subset has shared neighborhood of size $n' \geq \frac{m'-(n-r)(t-1)}{r-t+1} / \binom{r}{t}$, as claimed. □

The proof of Lemma 3.1.2 also required the value of $\mathrm{ex}(n, P_3)$ as the base case of the induction; for longer paths, Theorem 2.2.1 plays this role, giving that $\mathrm{ex}(n, P_l) \leq \frac{l-2}{2} n$. In the next theorem, we show that a construction similar to Construction B, namely a split graph, is optimal when forbidding k disjoint copies of a longer path. This Theorem is joint work with Nathan Kettle, and appears in [8].

Theorem 3.1.4 (B., Kettle [8])**.** *For $k \geq 2$, $l \geq 4$, and $n \geq 2l + 2kl\left(\lceil \frac{l}{2} \rceil + 1\right)\binom{l}{\lfloor \frac{l}{2} \rfloor}$,*

$$\mathrm{ex}(n, k \cdot P_l) = \binom{k\lfloor \frac{l}{2} \rfloor - 1}{2} + \left(k\left\lfloor \frac{l}{2} \right\rfloor - 1\right)\left(n - k\left\lfloor \frac{l}{2} \right\rfloor + 1\right) + c_l,$$

where $c_l = 1$ if l is odd, and $c_l = 0$ if l is even.

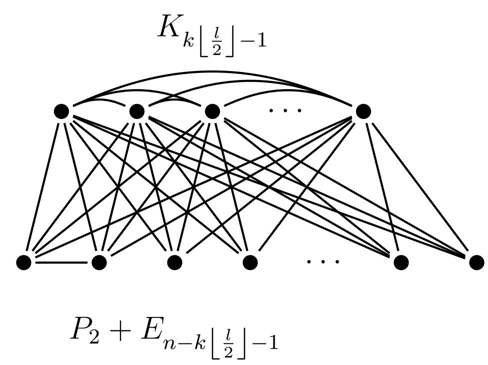

Fig. 3.7. $H_{\mathrm{Ex}}(n, k \cdot P_l) = G(n, k, l)$, l odd.

The extremal graph here is $G(n, k, l) := K_t + E_{n-t}$, with a single edge added to the empty class when l is odd, and $t = k \left\lfloor \frac{l}{2} \right\rfloor - 1$, as seen in Figures 3.7, 3.8 respectively. We note that the result above for $k \cdot P_l$ for $l \geq 4$ does not match the earlier result for $k \cdot P_3$ in Theorem 3.1.2; this is primarily due to the difference in the structure of the extremal graphs for the base cases.

Remark: It should be noted that for paths of even lengths, the above bound can be proven, and the extremal structure determined, via a paper of Balister, Győri, Lehel, and Schelp as a consequence of a theorem regarding the maximal number of edges in a connected graph containing no path of some fixed length [3]. One can, of course, divide a long path into many short even paths, and this allows one to deduce our Theorem 3.1.4 from their Theorem 1.3; for odd length paths this result gives a nonoptimal number of edges due to parity issues. In both the even and odd cases, deducing this precisely from the Balister, Győri, Lehel, Schelp result is nontrivial; in particular, determining the extremal structure is quite difficult. This extremal number within connected graphs was also determined earlier by Kopylov in 1977, but the approach in the proof given there did not give the extremal structure [23]. Regardless, this proof uses very different methods, and was discovered independently.

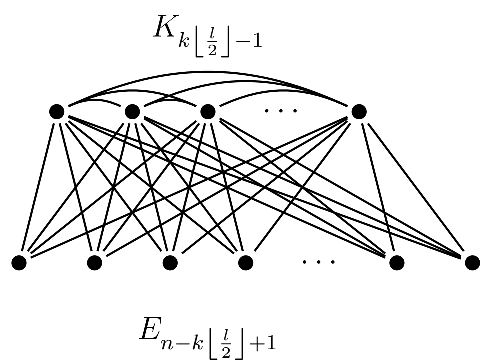

Fig. 3.8. $H_{\mathrm{Ex}}(n, k\cdot P_l) = G(n,k,l)$, l even.

Proof. We proceed by induction on k, starting with the base case, $k=2$.

Let G be a graph with $|V(G)| = n \geq 2l + 4l\left(\left\lceil \frac{l}{2}\right\rceil + 1\right)\binom{l}{\left\lfloor \frac{l}{2}\right\rfloor}$, $\left|\left[\begin{smallmatrix}E\\G\end{smallmatrix}\right]\right| \geq \binom{2\left\lfloor \frac{l}{2}\right\rfloor -1}{2} + \left(2\left\lfloor \frac{l}{2}\right\rfloor - 1\right)\left(n - 2\left\lfloor \frac{l}{2}\right\rfloor + 1\right) \; + \; c_l$, and which contains no $2\cdot P_l$. For $n \geq l^2$, $|E(G)|$ is then greater than $\mathrm{ex}(n, P_l)$, and so G contains a P_l on vertex set U, say.

Using Lemma 3.1.3 with $F_1 = P_l$, $F_2 = P_l$, and $m = \binom{2\left\lfloor \frac{l}{2}\right\rfloor -1}{2} + \left(2\left\lfloor \frac{l}{2}\right\rfloor - 1\right)\left(n - 2\left\lfloor \frac{l}{2}\right\rfloor + 1\right) \; + \; c_l$, some elementary simplifications show that any P_l contained in G must have at least $\left\lfloor \frac{l}{2}\right\rfloor$ vertices sharing a neighborhood of size at least

$$\begin{aligned}
n' =& \frac{m - \mathrm{ex}(n-l, P_l) - \binom{l}{2} - (n-l)\left(\left\lfloor \frac{l}{2}\right\rfloor - 1\right)}{\left(\left\lceil \frac{l}{2}\right\rceil + 1\right)\binom{l}{\left\lfloor \frac{l}{2}\right\rfloor}} \\
\geq& \frac{\binom{2\left\lfloor \frac{l}{2}\right\rfloor -1}{2} + \left(2\left\lfloor \frac{l}{2}\right\rfloor - 1\right)\left(n - 2\left\lfloor \frac{l}{2}\right\rfloor + 1\right)}{\left(\left\lceil \frac{l}{2}\right\rceil + 1\right)\binom{l}{\left\lfloor \frac{l}{2}\right\rfloor}} \\
&+ \frac{c_l - (n-l)\left(\frac{l}{2} - 1\right) - \binom{l}{2} - (n-l)\left(\left\lfloor \frac{l}{2}\right\rfloor - 1\right)}{\left(\left\lceil \frac{l}{2}\right\rceil + 1\right)\binom{l}{\left\lfloor \frac{l}{2}\right\rfloor}} \\
\geq& \frac{\left(1 - \frac{c_l}{2}\right)(n-l)}{\left(\left\lceil \frac{l}{2}\right\rceil + 1\right)\binom{l}{\left\lfloor \frac{l}{2}\right\rfloor}} \qquad (3.5)
\end{aligned}$$

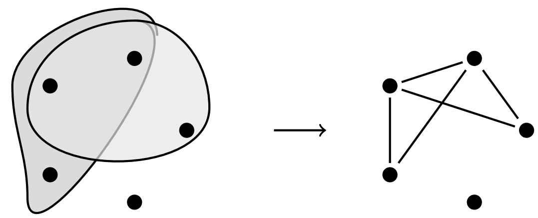

Fig. 3.9. Flattening a Hypergraph.

For n in our range, (3.5) is at least $2l$. Thus we have that for any $P_l \subset G$, we can find $\lfloor \frac{l}{2} \rfloor$ vertices with shared neighborhood of size at least $2l$.

We now create an $\lfloor \frac{l}{2} \rfloor$-uniform hypergraph $\mathcal{H}$ with $V(\mathcal{H}) = V(G)$ as follows: for any $P_l \subseteq G$, we find a subset U' of $\lfloor \frac{l}{2} \rfloor$ vertices with a large common neighborhood, as guaranteed above, and add U' as an edge in $\mathcal{H}$. We now flatten this hypergraph to form a simple graph G' on the same vertex set, with $uv \in E(G')$ whenever u and v are contained in the same hyperedge.

Since vertices adjacent in G' were in a common edge of $\mathcal{H}$, this means they have large common neighborhood in G. Therefore, a path of length $\lfloor \frac{l}{2} \rfloor$ in G' lets us find a path of length l in G, simply by following edges from each vertex on the G' path, to the common neighborhood with the next vertex on the G' path, and from this common neighborhood to the next vertex on the G' path; following this pattern, we obtain a path in G which is twice as long as the one in G'. More formally, as $n' \geq 2l$, if G' contains $2 \cdot P_{\lfloor \frac{l}{2} \rfloor}$, we can choose distinct common neighbors for each pair of consecutive vertices in these paths, and distinct neighbors for the end vertices, giving us $2 \cdot P_l$ in G. Thus since $2 \cdot P_l \nsubseteq G$, G' cannot contain $2 \cdot P_{\lfloor \frac{l}{2} \rfloor}$; this is seen in Figure 3.10.

We further note that certainly two disjoint hyperedges in $\mathcal{H}$ give rise to two such disjoint paths. Thus every pair of edges in $\mathcal{H}$ intersect; such a hypergraph is

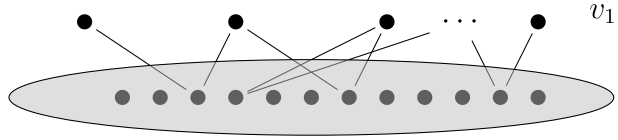

Fig. 3.10. Building a path from vertices with large shared neighborhood.

called *intersecting*. We will further call a hypergraph k-*intersecting* if every pair of edges intersect in at least k vertices.

Claim A If there exists $X \subseteq V(\mathcal{H})$, with $|X| = t < \lfloor \frac{l}{2} \rfloor$, and such that X contains some vertex from each edge in $\mathcal{H}$, then $|E(G)| < |E(G(n, 2, l))|$.

Indeed, assume X is such a set. By the construction of $\mathcal{H}$, since $\mathcal{H}[V(\mathcal{H}) \setminus X]$ contains no hyperedges, $G[V(G) \setminus X]$ contains no P_l, and so Theorem 2.2.1 tells us that

$$|E(G)| \leq \binom{t}{2} + t(n-t) + \frac{l-2}{2}(n-t) \leq \left(2\left\lfloor\frac{l}{2}\right\rfloor - \frac{3}{2}\right)n.$$

Recall that

$$\begin{aligned}|E\left(G\left(n,2,l\right)\right)| &= \binom{2\left\lfloor\frac{l}{2}\right\rfloor - 1}{2} + \left(2\left\lfloor\frac{l}{2}\right\rfloor - 1\right)\left(n - 2\left\lfloor\frac{l}{2}\right\rfloor + 1\right) + c_l \\ &\geq \left(2\left\lfloor\frac{l}{2}\right\rfloor - 1\right)n - l^2,\end{aligned}$$

and so as $n > 2l^2$, $|E(G)| < |E\left(G\left(n,2,l\right)\right)|$. Thus Claim A holds.

Now, assume we have at least $2\left\lfloor\frac{l}{2}\right\rfloor$ vertices contained in edges of $\mathcal{H}$, but without $2 \cdot P_{\lfloor \frac{l}{2} \rfloor}$ in G'. We will now show that no two hyperedges can intersect in only a single vertex.

If $E_1, E_2 \in E(\mathcal{H})$ with $E_1 \cap E_2 = \{x\}$, then $|E_1 \cup E_2| = 2\left\lfloor\frac{l}{2}\right\rfloor - 1$ vertices, and so $\mathcal{H}$ contains an edge E_3 not contained in their union. We may assume that

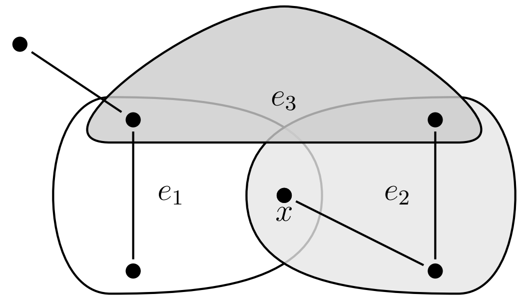

Fig. 3.11. Case 1, in the proof of Theorem 3.1.4.

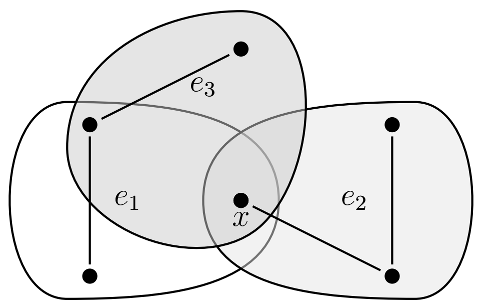

Fig. 3.12. Case 2, in the proof of Theorem 3.1.4.

this edge intersects $E_1 \cup E_2$ outside $\{x\}$; if no such edge exists, we are done by Claim A applied to the set $\{x\}$. Without loss of generality, $E_3 \cap E_1 \not\subseteq E_2$.

Let us consider two cases.

<u>Case 1:</u> $E_3 \cap (E_2 \setminus E_1) \neq \emptyset$. Then we can find a cycle in G' through all the vertices in $E_1 \cup E_2$. Since we have at least $2\left\lfloor \frac{l}{2} \right\rfloor$ vertices in edges of G', there is at least one other vertex adjacent to this cycle. This gives us a path of length $2\left\lfloor \frac{l}{2} \right\rfloor$, and so G' contains $2 \cdot P_{\left\lfloor \frac{l}{2} \right\rfloor}$.

<u>Case 2:</u> $E_3 \cap (E_2 \setminus E_1) = \emptyset$. Then there at least one vertex $y \in E_3 \setminus (E_1 \cup E_2)$, and so we can form one $P_{\left\lfloor \frac{l}{2} \right\rfloor}$ in $(E_1 \setminus \{x\}) \cup \{y\}$ and a disjoint $P_{\left\lfloor \frac{l}{2} \right\rfloor}$ entirely inside E_2; again, in this case G' contains $2 \cdot P_{\left\lfloor \frac{l}{2} \right\rfloor}$.

We now have that $\mathcal{H}$ is an intersecting hypergraph, with at least $2\left\lfloor \frac{l}{2} \right\rfloor$ vertices contained in its edges, and with no two edges intersecting in a unique vertex; $\mathcal{H}$ is then 2-intersecting. The edge set of $\mathcal{H}$ is nonempty, so pick an edge E, and any vertex in $x \in E$. Since each edge in $\mathcal{H}$ intersects E in at least two vertices, any edge in $\mathcal{H}$ intersects $E \setminus \{x\}$, a set of size $\left\lfloor \frac{l}{2} \right\rfloor - 1$. No such set of vertices exists by Claim A.

Thus we now know that all edges of $\mathcal{H}$ are contained in a set A of vertices with $|A| \leq 2\left\lfloor \frac{l}{2} \right\rfloor - 1$, and hence by construction any P_l in G contains at least $\left\lfloor \frac{l}{2} \right\rfloor$ vertices from A. We define three more sets of vertices as follows:

$$
\begin{aligned}
B &= \left\{ x \in G \setminus A \,|\, d_A(x) \geq \left\lfloor \frac{l}{2} \right\rfloor \right\}, \\
C &= \left\{ x \in G \setminus A \,|\, \left\lfloor \frac{l}{2} \right\rfloor > d_A(x) > 0 \right\}, \\
D &= \left\{ x \in G \setminus A \,|\, d_A(x) = 0 \right\}.
\end{aligned}
$$

Certainly D can contain no P_l, since every P_l meets A. Thus the number of edges entirely within D is at most $\frac{l-2}{2}|D|$ by Theorem 2.2.1.

We now claim that every vertex $x \in B \cup C$ is the end vertex of a P_l in G, with alternate vertices in A, which also misses any given $y_1, y_2 \in B \cup C$. Since x is adjacent to some $y \in A$, and y is contained in some hyperedge E, as long as $n' > |A| + \left\lfloor \frac{l}{2} \right\rfloor + 2$, we can find $\left\lfloor \frac{l}{2} \right\rfloor$ vertices in $(B \cup C) \setminus \{x, y_1, y_2\}$ adjacent to all vertices in E, allowing us to find such a P_l. This is the "path building" discussed early in a vague form and illustrated in Figure 3.10.

Further, no vertex in D can have degree more than 1 to $B \cup C$. Indeed, assume uv, uw are both edges with $u \in D$, and $v, w \in B \cup C$. We can find a P_l leaving v, that misses w, with alternate vertices in A. This gives a P_l starting at w with only $\left\lfloor \frac{l}{2} \right\rfloor - 1$ vertices from A, as in Figure 3.13. A vertex in $B \cup C$ with degree 2 to $B \cup C$ allows us to create a path in the same way, so our graph contains none

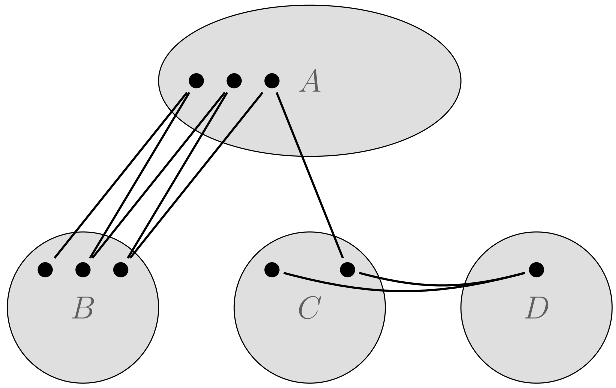

Fig. 3.13. Building a path off a degree two vertex in D.

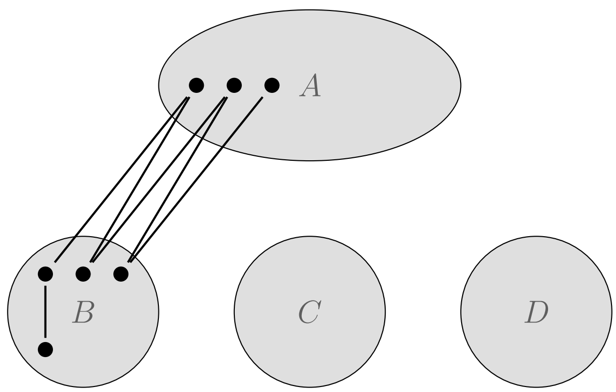

Fig. 3.14. Building a path with an edge inside B.

of these; in either case, this is a contradiction since every P_l contains at least $\left\lfloor \frac{l}{2} \right\rfloor$ vertices from A by construction.

Similarly, if l is even, an edge inside B allows us to create a P_l using only $\left\lfloor \frac{l}{2} \right\rfloor - 1$ vertices from A, as in Figure 3.14, so in this case B must be empty. If l is odd, since every vertex in B is adjacent to vertices in every edge of $\mathcal{H}$, then the existence of two disjoint edges in B allows us to create a P_l with only $\left\lfloor \frac{l}{2} \right\rfloor - 1$ vertices from A. A single edge does not create this problem, however; this is where odd/even distinction and the c_l in the theorem arises.

We've now counted edges between B and C and between $B \cup C$ and D respectively, and counted the edges inside each of B, C, and D. We can use the degree conditions in their definitions to bound the number of edges from A to B, C, and D. Putting all of these together, we see that

$$
\begin{aligned}
|E(G)| \leq& \binom{|A|}{2} + (n - |A| - |C| - |D|)\,|A| \qquad (3.6)\\
&+ \left(1 + \left\lfloor \frac{l}{2} \right\rfloor - 1\right)|C| + \left(1 + \frac{l-2}{2}\right)|D| + c_l,\\
=& \binom{|A|}{2} + (n - |A|)\,|A| + \left(\left\lfloor \frac{l}{2} \right\rfloor - |A|\right)|C| + \left(\frac{l}{2} - |A|\right)|D| \qquad (3.7)
\end{aligned}
$$

As any P_l in G contains at least $\lfloor \frac{l}{2} \rfloor$ vertices of A, and G contains some P_l by Theorem 2.2.1, $|A| \geq \lfloor \frac{l}{2} \rfloor$. If $|A| = \lfloor \frac{l}{2} \rfloor$, then $|D| \leq n - |A|$, and so whenever $n > |A| + 2$,

$$
\begin{aligned}
|E(G)| &\leq \binom{|A|}{2} + (n - |A|)|A| + \frac{c_l}{2}(n - |A|) + c_l\\
&= \binom{|A|+1}{2} + (n - |A| - 1)(|A| + 1) + (\frac{c_l}{2} - 1)(n - |A| - 1) + \frac{3}{2}c_l
\end{aligned}
$$

In fact, $|A| + 1 \leq 2\lfloor \frac{l}{2} \rfloor - 1$, and so $|E(G)| \leq |G(n, 2, k)|$.

If $|A| > \lfloor \frac{l}{2} \rfloor$, the coefficients of $|C|$ and $|D|$ in (3.6) are negative, and so $|E(G)|$ is maximized when both C and D are empty. This gives the bound on $|E(G)|$ as claimed. Further, since C and D must be empty to attain this bound, it also shows that the extremal graph is in fact $G(n, 2, l) = K_{2\lfloor \frac{l}{2} \rfloor - 1} + E_{n-2\lfloor \frac{l}{2} \rfloor + 1}$ with an extra edge in the empty class for odd l, as in the statement of the theorem.

We have now established a second base case, $k = 2$. Somewhat surprisingly, the inductive step is easy to show. Establishing the case $k = 2$ from Theorem 2.2.1 is difficult largely because of the significant differences between the extremal graphs for $k = 2$, $k = 1$. For $k \geq 2$, establishing the case $k + 1$ from the case k is relatively easy, since the extremal graph for $k + 1$ contains the extremal graph for k; thus the use of Lemma 3.1.3 gives us precisely the number of common neighbors which is most useful, and establishes precisely the structure of the rest of the graph.

We now prove the induction step; thus let $k \geq 3$, and assume that the theorem holds for $k' < k$. Let G be a graph on n vertices with

$$m \geq \binom{k\left\lfloor \frac{l}{2} \right\rfloor - 1}{2} + \left(k\left\lfloor \frac{l}{2} \right\rfloor - 1\right)\left(n - k\left\lfloor \frac{l}{2} \right\rfloor + 1\right) \ + \ c_l$$

edges, which does not contain $k \cdot P_l$. This graph does contain some P_l, by Theorem 2.2.1, and again by Lemma 3.1.3 we can find $\left\lfloor \frac{l}{2} \right\rfloor$ vertices with shared neighborhood of size at least

$$\begin{aligned} n' =& \frac{\binom{k\lfloor \frac{l}{2}\rfloor - 1}{2} + \left(k\left\lfloor \frac{l}{2}\right\rfloor - 1\right)\left(n - k\left\lfloor \frac{l}{2}\right\rfloor + 1\right)}{\left(\left\lceil \frac{l}{2}\right\rceil + 1\right)\binom{l}{\lfloor \frac{l}{2}\rfloor}} \\ &+ \frac{c_l - \mathrm{ex}(n-l, (k-1)\cdot P_l) - \binom{l}{2} - (n-l)\left(\left\lfloor \frac{l}{2}\right\rfloor - 1\right)}{\left(\left\lceil \frac{l}{2}\right\rceil + 1\right)\binom{l}{\lfloor \frac{l}{2}\rfloor}} \\ =& \frac{\binom{k\lfloor \frac{l}{2}\rfloor - 1}{2} + \left(k\left\lfloor \frac{l}{2}\right\rfloor - 1\right)\left(n - k\left\lfloor \frac{l}{2}\right\rfloor + 1\right)}{\left(\left\lceil \frac{l}{2}\right\rceil + 1\right)\binom{l}{\lfloor \frac{l}{2}\rfloor}} \\ &+ \frac{c_l - \binom{(k-1)\lfloor \frac{l}{2}\rfloor - 1}{2} - \left((k-1)\left\lfloor \frac{l}{2}\right\rfloor - 1\right)\left(n - l - (k-1)\left\lfloor \frac{l}{2}\right\rfloor + 1\right)}{\left(\left\lceil \frac{l}{2}\right\rceil + 1\right)\binom{l}{\lfloor \frac{l}{2}\rfloor}} \\ &+ \frac{-c_l - \binom{l}{2} - (n-l)\left(\left\lfloor \frac{l}{2}\right\rfloor - 1\right)}{\left(\left\lceil \frac{l}{2}\right\rceil + 1\right)\binom{l}{\lfloor \frac{l}{2}\rfloor}}. \end{aligned}$$

The equality here is valid since for n in our range, $n - l \geq 2l + 2l(k-1)\left(\left\lceil \frac{l}{2}\right\rceil + 1\right)\binom{l}{\lfloor \frac{l}{2}\rfloor}$. Simplifying, we see that this is equal to

$$
\begin{aligned}
&\frac{n+k\left\lfloor\frac{l}{2}\right\rfloor^2-\frac{3}{2}\left\lfloor\frac{l}{2}\right\rfloor^2+c_lk\left\lfloor\frac{l}{2}\right\rfloor-\frac{5+4c_l}{2}\left\lfloor\frac{l}{2}\right\rfloor-2c_l}{\left(\left\lceil\frac{l}{2}\right\rceil+1\right)\binom{l}{\left\lfloor\frac{l}{2}\right\rfloor}}\\
\geq&\frac{n-l}{\left(\left\lceil\frac{l}{2}\right\rceil+1\right)\binom{l}{\left\lfloor\frac{l}{2}\right\rfloor}}.
\end{aligned}
$$

Further, for the values of n under consideration, we have $n' \geq 2kl$. Write U for the set of vertices given by Lemma 3.1.3.Then $G\left[V \setminus U\right]$ is a graph on $n - \left\lfloor\frac{l}{2}\right\rfloor$ vertices, and at least $\mathrm{ex}(n - \left\lfloor\frac{l}{2}\right\rfloor, (k-1)\cdot P_l)$ edges. If we can find $(k-1)\cdot P_l$, then since $n' \geq kl$, we can find another P_l in G disjoint from these $k-1$, as before. Thus we do not have $k-1$ disjoint copies of P_l in $G[V \setminus U]$, so by the inductive hypothesis, $|E\left(G\left[V \setminus U\right]\right)| = \mathrm{ex}(n - \left\lfloor\frac{l}{2}\right\rfloor, (k-1)\cdot P_l)$. Thus $G\left[V \setminus U\right] = G\left(n - \left\lfloor\frac{l}{2}\right\rfloor, k-1, l\right)$, and so $G = G(n,k,l)$. □

The above proof shows that our construction is optimal for $n = \Omega(kl^{\frac{3}{2}}2^l)$. We conjecture that this construction is optimal for $n = \Omega(kl)$. We also note a comparison between Theorem 3.1.4 for even paths and Theorem 2.2.1: certainly if one forbids $k \cdot P_{2l}$, then one is also forbidding P_{2kl}. Thus an easy upper bound on $\mathrm{ex}(n, k\cdot P_{2l})$ is $\mathrm{ex}(n, P_{2kl})$. The difference between this bound and the precise result established above is relatively small, $(kl-1)(\frac{kl}{2})$. In particular, it is not dependent on n for fixed k and l, despite the enormous difference between the extremal graphs.

3.2 Trees

Throughout the following section, we need an analogue of Lemma 3.1.1 as a starting point. For longer paths, we used the Erdős-Gallai result, Lemma 2.2.1. The analogous result for trees is the Erdős-Sós Conjecture discussed in Section 2.2.2.

3.2.1 Forests of Equibipartite Trees

Our proof of Theorem 3.1.4 can be adapted to work on a significantly larger class of graphs. A key element of our proof was finding a set of vertices which intersected every long path in at least half its vertices. This continues to be an essential idea, and thus we restrict ourselves to trees which have the same number of vertices in each vertex class, when viewed as a bipartite graph. This restriction also provides a great deal of information on how such trees can embed into our claimed extremal graphs. We call such trees *equibipartite*, and a forest in which each component is an equibipartite tree is called an equibipartite forest. Clearly any equibipartite tree or equibipartite forest has an even number of vertices.

If we allow ourselves the considerable benefit of assuming that Erdős-Sós holds for all equibipartite trees, we can determine the extremal number for any equibipartite forest, for large n. Somewhat suprisingly, there is a not insignificant difference in the structure of the extremal graph and thus in the extremal number depending on whether the forest admits a perfect matching. We note that the restriction to equibipartite trees is not artificial; as evidenced by both the result for odd paths and for stars of any size, the result does not hold for trees which have different sized partitions in their bipartite classes. This theorem is also joint work with Nathan Kettle.

Theorem 3.2.1 (B., Kettle [8])**.** *Let H be an equibipartite forest on $2l$ vertices which is comprised of at least two trees. If the Erdős-Sós Conjecture holds for each component tree in H, then for $n \geq 3l^2 + 32l^5\binom{2l}{l}$,*

$$\mathrm{ex}(n, H) = \begin{cases} \binom{l-1}{2} + (l-1)(n-l+1), \; \textit{if } H \textit{ admits a perfect matching} \\ (l-1)(n-l+1) \; \textit{otherwise.} \end{cases}$$

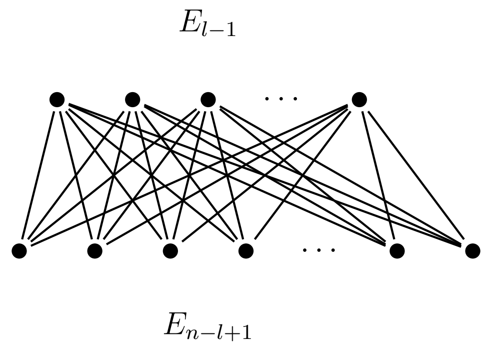

Fig. 3.15. Extremal graph for an equibipartite forest with no perfect matching.

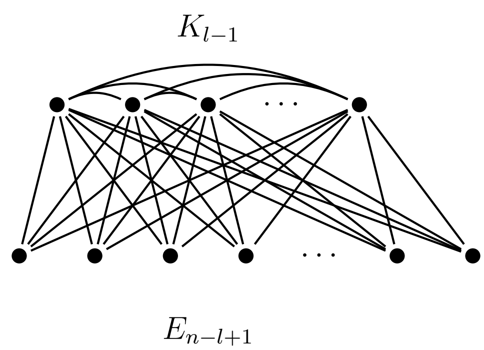

Fig. 3.16. Extremal graph for an equibipartite forest with a perfect matching.

The extremal graphs here are $K_{l-1} + E_{n-l+1}$ for any forest with a perfect matching, and $E_{l-1} + E_{n-l+1}$ for any forest with no perfect matching, as in Figures 3.16, 3.15. To prove the eventual extremal number for equibipartite trees as in Theorem 3.2.1, we do not need the full strength of the Erdős-Sós Conjecture; in fact, it suffices to know that $\mathrm{ex}(n, T) \leq \frac{|T|-2}{2}n + o(n)$ for any of the equibipartite trees $T \subseteq H$. In this case, however, the bound on n for which the result holds is worse. We also note that in order to avoid the complication of many lower order terms, the bound on n in the statement of the theorem has not been optimized.

Before proving Theorem 3.2.1, we will need the following structure lemmas for equibipartite trees, Lemma 3.2.2. These lemmas also give some intuition into

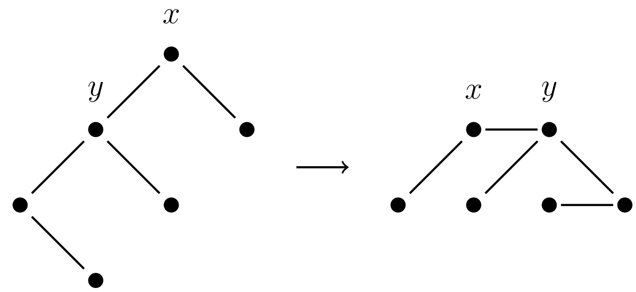

Fig. 3.17. Lemma 3.2.2, for a tree with a perfect matching.

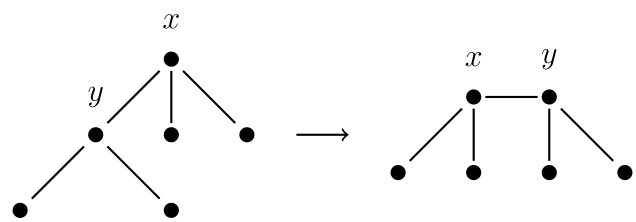

Fig. 3.18. Lemma 3.2.3, for a tree with no perfect matching.

why the distinction between forests with and without perfect matchings occurs. The partitions discussed in the lemmas are illustrated in Figures 3.17, 3.18.

Lemma 3.2.2. *Let H be a equibipartite tree on $2l$ vertices. If H contains a perfect matching, then every partition of $V(H)$ into two classes of different sizes is such that the larger class induces at least one edge.*

Proof. If H contains a perfect matching, $M \subseteq E(H)$, then for any partition of $V(H)$ into nonequal classes, $|V_1| < |V_2|$, the number of edges in M which meet V_1 is at most $|V_1| < l$, and so some edge lies inside V_2. $\square$

Lemma 3.2.3. *Let H be a equibipartite tree on $2l$ vertices. If H does not contain a perfect matching, then there exists a partition of $V(H)$ into two classes of different sizes such that the larger class induces no edges and the smaller class induces exactly one edge.*

Proof. Consider H as a bipartite graph with bipartition $V(H) = (A, B)$. Since H contains no perfect matching, there is a set $S \subseteq A$ for which Hall's condition (see, e.g., [5]) fails; namely $|N(S)| < |S|$. If we take S minimal, then $H[S \cup N(S)]$ is connected, as otherwise one of its components would fail Hall's condition.

Consider $H[(A \setminus S) \cup (B \setminus N(S))]$. Each component of this graph is joined to $N(S)$ by a single edge. Since the union of these components has larger intersection with B than with A, at least one of the components does. Let C be such a component, and let xy be the unique edge between C and $N(S)$, with $x \in C$ and $y \in N(S)$.

Now consider the partition $(C, V(H) \setminus C)$. Then taking the set of vertices $V_{x,y}$ which are in the same bipartite class as x in C or in the same bipartite class as y in $V(H) \setminus C$ as one class of our new partition, and $V(H) \setminus V_{x,y}$ as the other forms a partition of $V(H)$ with exactly one edge in $V_{x,y}$, and none in $V(H) \setminus V_{x,y}$.

Since our tree is equibipartite, $|V_{x,y} \cap (V(H) \setminus C)| + |\,(V(H) \setminus V_{x,y}) \cap C| = l$. By our definition of C, $|V_{x,y} \cap C| < |\,(V(H) \setminus V_{x,y}) \cap C|$. By construction, each of the sets $V_{x,y} \cap (V(H) \setminus C)$, $(V(H) \setminus V_{x,y}) \cap C$, and $V_{x,y} \cap C$ are nonempty. Then $|V_{x,y}| = |V_{x,y} \cap C| + |V_{x,y} \cap (V(H) \setminus C)| < |(V(H) \setminus V_{x,y}) \cap C| + |V_{x,y} \cap (V(H) \setminus C)| = l$, and so our partition is an unbalanced partition with no edges in the larger class and exactly one edge in the smaller class, as claimed.

□

Proof of Theorem 3.2.1. We again proceed by induction, using the Erdős-Sós Conjecture as our base case. Here we avoid the difficulty in going from the case of a forest with only a single tree to a forest with two trees that was present for paths since we are restricting ourselves to equibipartite trees.

Let H have components $H_1, H_2, \ldots, H_k$, each on $2l_1, 2l_2, \ldots, 2l_k$ vertices respectively, and G be a graph on n vertices with m edges which does not contain

H, and with $m \geq (l-1)(n-l+1)$. Without loss of generality, $l_1 \leq l_i$, for each i. For notational ease, we also define $H' = H_2 \cup \ldots \cup H_k$ and $l' = \frac{|H'|}{2} = l - l_1$.

As $n \geq l^2$, $m \geq \mathrm{ex}(n, H')$ by induction (or Erdős-Sós, if H' is a tree), and so we can find a copy of $H' \subseteq G$. As in the proof of Lemma 3.1.3, for any copy of H' we can bound from below the size of the set E' of edges between H' and $G \setminus H'$ by $m - \binom{2l'}{2} - \mathrm{ex}(n-2l', H_1)$. By the Erdős-Sós Conjecture, this is at least $(l-1)(n-l+1) - \binom{2l'}{2} - (n-2l')(l_1-1) \geq l'n - 3l^2$.

Consider the set of vertices $X = \{v \in G \setminus H' : |N(v) \cap H'| \geq l'\}$. Then

$$2l'|X| + (l'-1)(n-2l'-|X|) \geq |E'| \geq l'n - 3l^2.$$

Thus $|X| \geq \frac{n-3l^2}{l'+1}$. As there are only $\binom{2l'}{l'}$ sets of l' vertices in H', we can find a set A of l' vertices in H' with at least $n' = \frac{n-3l^2}{(l'+1)\binom{2l'}{l'}}$ common neighbors. By our assumption on n, $n' \geq 32l^3$.

Interchangine the roles of H_1 and H', for any H_1 we similarly bound from below the size of the set E_1 of edges between H_1 and $G \setminus H_1$ by $m - \binom{2l_1}{2} - \mathrm{ex}(n-2l_1, H')$. Note that $n - 2l_1$ is much larger than needed in the condition of the inductive hypothesis, and so

$$\begin{aligned} |E_1| &\geq (l-1)(n-l+1) - \binom{2l_1}{2} - (n-2l_1-l'+1)(l'-1) - \binom{l'-1}{2} \\ &\geq l_1 n - 3l^2. \end{aligned} \tag{3.8}$$

With this in mind, we define the following set of vertices which are not in A, but which are still of large degree:

$$B = \left\{ w \in G \mid w \notin A \text{ and } d_G(w) \geq \frac{n - 3l^2}{l_1 + 1} \right\}$$

Now, any copy of H_1 in G must contain at least l_1 vertices from $A \cup B$, as otherwise the sum of the degrees of vertices in H_1 is less than $(l_1 + 1) \frac{n-3l^2}{l_1+1} + (l_1 - 1) n$, contradicting (3.8) above.

As a rough bound on the number of edges in G, we note that if G contained more than $2ln$ edges, we can find a copy of H' by induction (or by the Erdős-Sós Conjecture if H' is a single tree). Removing this copy of H' leaves a graph on $n - 2l'$ vertices with more than $2l_1 n \geq 2l_1 (n - 2l')$ edges, since each vertex is of course adjacent to at most n edges. Again by Conjecture 2.2.2, we can find a copy of H_1. Thus our graph can have at most $2ln$ edges.

This means that for any $c > 0$, there are at most $\frac{4ln}{c}$ vertices of degree at least c. Choosing $c = \frac{8ln}{n'} = \frac{8l(l'+1)\binom{2l'}{l'}n}{n-3l^2}$, there are at least $\frac{n'}{2}$ common neighbors of A with degree at most c. Since $n \geq 6l^2$, $c \leq 16l\,(l' + 1) \binom{2l'}{l'}$. Then since $\frac{n'}{2} \geq l'$, we can find a copy of H' with l' vertices in A and the other l' vertices having degree at most c.

Since this copy of H' is incident to at least $l'n - 3l^2$ edges, any vertex in A has degree at least

$$\begin{aligned} & l'n - 3l^2 - l'c - (l' - 1)(n - 1) \\ & \geq n - 3l^2 - l'c \\ & = n - c'. \end{aligned} \tag{3.9}$$

There are at most $\frac{4ln}{c} = \frac{n'}{2}$ vertices of degree at least c, and at most $l'c'$ vertices not adjacent to all of A. Since

$$\frac{n-3l^2}{l_1+1} - \frac{n'}{2} \geq \frac{n-3l^2}{2\,(l_1+1)} \geq 16l^4\binom{2l}{l} > l'c',$$

by the definition of B each vertex $x \in B$ is adjacent to a vertex y which is adjacent to all of A and such that $d_G(y) \leq c$.

This condition on the vertices in B enables us to find, for each $x \in B$, a copy of H' from which $l'-1$ of the vertices have small degree, and whose intersection with B contains x as a leaf. Further, we can find a set U of $l'-1$ vertices of degree at most c which are each adjacent to all of A, so for any $z \in A$, $G\left[(U \cup \{x\} \cup \{y\} \cup (A \setminus \{z\}))\right]$ is a graph on $2l'$ vertices which contains a copy of $K_{l',l'-1}$ with an extra vertex x adjacent to some vertex in the larger set.

We can find a copy of H' in this by letting a leaf of H' correspond to x, and so as in (3.9), every vertex in B must have degree at least $n-c'$. If B contained at least l_1 vertices, they would have common neighborhood of size at least $n - l_1c' \geq l$, allowing us to find H_1 in $G[V(G) \setminus A]$, and again as the common neighborhood of A is of size at least $2l$, we can find a disjoint copy of H', giving a copy of H in G. Thus $|B| \leq l_1 - 1$, and so $|A \cup B| \leq l' + l_1 - 1 = l - 1$.

We now define two more sets of vertices as follows:

$$D = \{x \in G \setminus (A \cup B) \,|\, d_{A\cup B}(x) \geq l_1\},$$
$$E = \{x \in G \setminus (A \cup B) \,|\, d_{A\cup B}(x) < l_1\}.$$

We note that any vertex not in $A \cup B$ which is adjacent to all of A is in D, and thus $|E| \leq l'c'$. There can be no H_1 in E, so the number of edges in E is at most $(l_1-1)\,|E|$ by Erdős-Sós. We now claim that no vertex $v \in D$ can have a neighbor $y \in D \cup E$. Indeed, we can find a set U of $l_1 - 1$ vertices in $A \cup B$

adjacent to v since each vertex in $A \cup B$ has degree at least $n - c'$. Further, we can find $W \subseteq (D \cup E) \setminus \{v, y\}$ consisting of $l_1 - 1$ vertices adjacent to all of U. As before we can find a copy of H_1 on $U \cup W \cup \{v, y\}$ with only $l_1 - 1$ vertices from $A \cup B$; a contradiction. Thus all edges in $G[D \cup E]$ are in E.

Letting $|A \cup B| = t$, we bound the number of edges in G by

$$\begin{aligned} &\binom{t}{2} + t\,(n - t - |E|) + (l_1 - 1)\,|E| + (l_1 - 1)\,|E| \\ &= \binom{t}{2} + t\,(n - t) + (2l_1 - 2 - t)\,|E|\,. \end{aligned} \tag{3.10}$$

If $t < l - 1$, then since $|E| \leq l'c'$ the number of edges in G is at most $\binom{t}{2} + t\,(n - t) + 2l_1 l'c' < (l - 1)\,(n - l + 1)$, for $n \geq 2l^2c' + l^2$, and hence $|A \cup B| = l - 1$.

The common neighborhood of $A \cup B$ has size at least $n - (l - 1) - (l - 1)c'$, as each vertex in $A \cup B$ is adjacent to all but c' vertices in G. Thus we can find a copy of $K_{l-1,n-(l-1)(c'+1)} \subseteq G$, where the smaller class is $A \cup B$. If H does not contain a perfect matching, then by Lemma 3.2.3 we can partition the vertices into unequal sets X,Y, the larger of which is empty, and the smaller of which contains exactly one edge. This is clearly present in G if $A \cup B$ contains an internal edge.

Counting all edges in G, we see that by (3.10),

$$|E(G)| \leq (l - 1)(n - l - 1) - (l - 2l_1 + 1)|E| + C_H,$$

where $C_H = \binom{l-1}{2}$ if H admits a perfect matching, and $C_H = 0$ otherwise. As l_1 is minimal, $(l - 2l_1 + 1) > 0$, and so the number of edges is maximized when $|E| = 0$. □

It is unlikely that the bound on n in Theorem 3.2.1 is optimal. Determining the minimal value of n for which this construction is optimal remains an open question. We have not carefully analyzed the best possible bound on n from our proof, but it is not significantly better than as stated. We conjecture that the given construction is for even fairly small values of n, but we are unaware of better constructions for anything but extremely small (i.e. linear in l) n.

Chapter 4. Turán Theory for Hypergraphs

4.1 Background

We can extend the notion of Turán numbers in a very natural way to the case of hypergraphs; in fact, there are several logical ways to do this. As before, we will want the extremal number for a hypergraph $\mathcal{H}$ to be the largest number of edges in an n vertex hypergraph. However, we will restrict ourselves to the case where $\mathcal{H}$ is an r-uniform hypergraph, as allowing the extremal number to count edges of different sizes somehow obscures the true extremal structure. With this in mind, we define the Turán numbers for hypergraphs as follows.

Definition 4.1.1. The *r-uniform hypergraph Turán Number* of an r-uniform hypergraph $\mathcal{H}'$ is defined as the following.

$$\mathrm{ex}_r(n, \mathcal{H}') = \max\{|E(\mathcal{H})| : |V(\mathcal{H})| = n, \mathcal{H}' \not\subseteq \mathcal{H}\}.$$

Similarly, when $\mathcal{H}'$ is a family of hypergraphs, this maximum is taken over all graphs not containing any member of $\mathcal{H}'$ as a subgraph.

In general, Turán theory for r-uniform hypergraphs with $r \geq 3$ is much less developed than the theory outlined in Chapter 2 for 2-graphs. In the same paper in which Turán proved Theorem 2.1.2, he posed the natural question of determining $\mathrm{ex}_r(n, K_t^r)$, where K_t^r denotes the complete r uniform graph on t vertices [34]. This problem remains open in all cases for $r > 2$, even up to asymptotics.

Unlike in the case for $r = 2$, graphs, it is not immediately clear that $\frac{\mathrm{ex}_r(n,\cdot)}{\binom{n}{r}}$ is monotone. Thus at the outset, it is unclear even whether the asymptotic behavior of $\mathrm{ex}_r(n, \mathcal{H}')$ is well defined. However, Katona, Nemetz and Simonovits in 1964 [19] showed via a simple and elegant averaging argument that, in fact, $\lim_{n\to\infty} \frac{\mathrm{ex}_r(n,\mathcal{H}')}{\binom{n}{r}}$, the *Turán density of* $\mathcal{H}'$, does exist in general for fixed $\mathcal{H}'$. We present their result and short proof here.

Theorem 4.1.2. *For any r-uniform hypergraph $\mathcal{H}$, $\frac{\mathrm{ex}_r(n,\mathcal{H})}{\binom{n}{r}}$ is monotone decreasing, and so the Turán density $\pi_r(\mathcal{H}) := \lim_{n\to\infty} \frac{\mathrm{ex}_r(n,\mathcal{H})}{\binom{n}{r}}$ exists.*

Proof. Let $1 \le h < n$ and consider an $\mathcal{H}$-free hypergraph $\mathcal{F}$ on n vertices which is extremal for $\mathcal{H}$. Choose $H \subseteq [n]^{(h)}$ uniformly at random. Then $\left|H^{(r)} \cap \mathcal{F}\right| \le \mathrm{ex}_r(h, \mathcal{H})$ for all such H. Further, $\mathbb{E}\left(\left|H^{(r)} \cap \mathcal{F}\right|\right)$ is $|\mathcal{F}|$ times the probability that a fixed $F \in [n]^{(r)}$ is in $H^{(r)}$; that is, $|\mathcal{F}| \binom{h}{r}/\binom{n}{r}$. Thus $|\mathcal{F}| \binom{h}{r}/\binom{n}{r} = \mathrm{ex}_r(n,\mathcal{H})\binom{h}{r}/\binom{n}{r} \le \mathrm{ex}_r(h,\mathcal{H})$. Dividing by $\binom{h}{r}$ gives the result as desired. □

As one might guess from the subtlety in proving even the existence of hypergraph Turán densities, determining extremal numbers precisely for hypergraphs is extremely difficult indeed. Part of the difficulty lies in the fact that for many classes of hypergraphs, the conjectured families of extremal hypergraphs are very large; thus proving extremal results involves a great many cases. For paths, however, we are lucky enough to have some results. We collect them in the following section, and build on them in the following sections.

4.2 Extremal numbers for Hyperpaths

We note first that unlike in the case of graphs, there is no obvious definition of a path. In a graph, knowing that an edge intersects two vertices gives full information about the edge; this is of course not true in r-graphs, for $r \ge 3$. Thus we give three different definitions of paths; we present these from most general to most specific. The first of these is due to Berge.

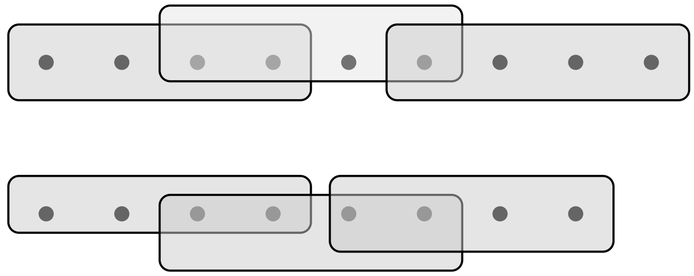

Fig. 4.1. Two 4-uniform loose paths, each on 3 edges.

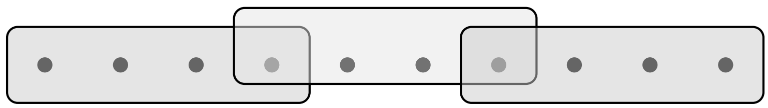

Fig. 4.2. A 4-uniform linear path on 3 edges.

Definition 4.2.1. A *Berge path* of length l in an r-uniform hypergraph is a family of edges $\{F_1, \ldots, F_l\}$ along with a family of vertices $\{v_1, \ldots, v_{l+1}\}$ such that for each $i \in \{1, \ldots, l\}$, $v_i, v_{i+1} \in F_i$.

Definition 4.2.2. A *loose path* of length l in an r-uniform hypergraph is a family of edges $\{F_1, \ldots, F_l\}$ such that $F_i \cap F_j \neq \emptyset$ iff $|i - j| = 1$, as seen in Figure 4.1. We use $\mathcal{P}_l^{(r)}$ to denote the family of r-uniform loose paths on l edges.

Definition 4.2.3. A *linear path* of length l in an r-uniform hypergraph is a family of edges $\{F_1, \ldots, F_l\}$ such that $|F_i \cap F_j| = 1$ if $|i - j| = 1$, and $|F_i \cap F_j| = \emptyset$ otherwise, as seen in Figure 4.2. We use $\mathfrak{P}_l^{(r)}$ to denote an r-uniform linear path on l edges.

The following results were proven by Füredi, Jiang, and Seiver [16].

Theorem 4.2.4. *Let $r \geq 3$, $l \geq 1$. Then for n sufficiently large, we have*

$$\mathrm{ex}_r(n, \mathcal{P}_l^{(r)}) = \binom{n-1}{r-1} + \binom{n-2}{r-1} + \ldots + \binom{n-t}{r-1} + c_l,$$

where $t = \left\lfloor \frac{l+1}{2} \right\rfloor - 1$, and $c_l = \mathbb{1}_{\{l \text{ is even}\}}$.

The unique extremal family consists of all the r-sets in $[n]$ which meet some fixed set S of t vertices, plus one additional r-set disjoint from S when l is even.

Theorem 4.2.5. *Let $r \geq 4$, $l \geq 1$. Then for n sufficiently large, we have*

$$\mathrm{ex}_r(n, \mathfrak{P}_l^{(r)}) = \binom{n-1}{r-1} + \binom{n-2}{r-1} + \ldots + \binom{n-t}{r-1} + d_l,$$

where $t = \left\lfloor \frac{l+1}{2} \right\rfloor - 1$, and $d_l = \begin{cases} 0, \text{ for } l \text{ odd}, \\ \binom{n-t-2}{r-2}, \text{ for } l \text{ even}. \end{cases}$

For l odd, the unique extremal family consists of all the r-sets in $[n]$ which meet some fixed set S of t vertices. For l even, we have these edges plus all the r-sets in $[n] \setminus S$ containing some two fixed elements not in S.

There is a significant difference between these theorems that may be overlooked by the casual reader: Theorem 4.2.4 is valid for 3-uniform hypergraphs, while Theorem 4.2.5 only applies to 4-uniform and higher hypergraphs. Because of this difference, and since we use these theorems as the base case of our induction in both Theorem 4.3.1 and Theorem 4.5.2, this distinction will appear in our theorems as well. Füredi, Jiang, and Seiver conjecture [16] that Theorem 4.2.5 holds for $r = 3$, and if this is true then our methods will immediately give Theorem 4.5.2 for $r = 3$ as well.

4.3 Multiple Loose Paths of Fixed Length

We abuse notation slightly, and let $k \cdot \mathcal{P}_l^{(r)} \subseteq \mathcal{H}$ denote the event that $\mathcal{H}$ contains k vertex disjoint loose paths. We note that this does not require the paths to be isomorphic.

Theorem 4.3.1 (B., Kettle). *Let $r \geq 3$, $l \geq 3$, $k \geq 1$, and n sufficiently large. Then letting $t = k \left\lfloor \frac{l+1}{2} \right\rfloor - 1$, $c_l = \mathbb{1}_{\{l \text{ is even}\}}$, any graph on at least $\binom{n-1}{r-1} + \ldots + \binom{n-t}{r-1} + c_l$ edges contains k vertex disjoint loose paths each of length l. For ease of notation, we define $f(n, r, k, l) = \binom{n-1}{r-1} + \ldots + \binom{n-t}{r-1} + c_l$.*

We note that the graph on n vertices in which each edge is incident to at least one of a specified set S of t vertices, along with a single edge disjoint from S

when l is even, gives a graph with exactly $f(n,r,k,l)$ and without k vertex disjoint paths of length l.

Proof. The case $k=1$ is provided by 4.2.4. We proceed by induction on k; thus assume that $k \geq 2$, and that $\mathcal{H}$ is a hypergraph on n vertices and with $|E(\mathcal{H})| = m > f(n,r,k,l)$. Since $f(n,r,k,l) > f(n,r,1,l)$, for n large enough, we can find at least one loose path inside $\mathcal{H}$.

Consider one of these loose paths, say on vertex set P. Certainly $|E(V(\mathcal{H}) \setminus P)| \leq f(n-|P|, r, k-1, l)$, or else by induction, the graph on $V(H) \setminus P$ contains $(k-1) \cdot \mathcal{P}_l^{(r)}$; these along side the loose path on P form $k \cdot \mathcal{P}_l^{(r)}$.

Letting N_P denote the number of edges of $\mathcal{H}$ incident to vertices in P, we have that

$$\begin{aligned} N_P &\geq m - f(n-|P|, r, k-1, l) && (4.1) \\ &> f(n,r,k,l) - f(n-(l+1), r, k-1, l) && (4.2) \\ &= \frac{\left\lfloor \frac{l+1}{2} \right\rfloor n^{r-1}}{(r-1)!} + O(n^{r-2}) && (4.3) \end{aligned}$$

We now focus on counting sets of vertices which can be used to easily 'finish' edges started by vertices in P. With this in mind, for every set R of $r-1$ vertices from $V(\mathcal{H}) \setminus P$, we define

$$A_R = \{e \in E(\mathcal{H}) : R \subseteq e \text{ and } e \setminus R \in P\}.$$

We now break the $(r-1)$ subsets of $V(\mathcal{H}) \setminus P$ into two sets, dependent on the size of their respective A_R:

$$A = \left\{ R \in (V(\mathcal{H}) \setminus P)^{(r-1)} : |A_R| \leq \left\lfloor \frac{l+1}{2} \right\rfloor - 1 \right\}$$

$$B = \left\{ R \in (V(\mathcal{H}) \setminus P)^{(r-1)} : |A_R| \geq \left\lfloor \frac{l+1}{2} \right\rfloor \right\}$$

Counting edges entirely contained in P and the edges incident to the sets A, B defined above, we have that

$$N_P \leq \binom{|V(P)|}{2}\binom{n}{r-2} + \left(\left\lfloor \frac{l+1}{2} \right\rfloor - 1\right)|A| + |V(P)|\,|B|\,, \tag{4.4}$$

$$\leq \binom{|V(P)|}{2}\binom{n}{r-2} + \left(\left\lfloor \frac{l+1}{2} \right\rfloor - 1\right)\binom{n}{r-1} + rl|B|. \tag{4.5}$$

By comparison of the upper and lower bounds on N_P, (4.3) and (4.5), we have that

$$|B| \geq \frac{\frac{n^{r-1}}{(r-1)!} + O(n^{r-2})}{rl}. \tag{4.6}$$

To each set $R \in B$ we associate a set of $\left\lfloor \frac{l+1}{2} \right\rfloor$ vertices from A_R arbitrarily. From (4.6), we see that some set of $\left\lfloor \frac{l+1}{2} \right\rfloor$ vertices is chosen many times; here 'many' is at least:

$$\frac{n^{r-1}}{(r-1)!rl\binom{|V(P)|}{\lfloor \frac{l+1}{2} \rfloor}} + O(n^{r-2}) \tag{4.7}$$

$$\geq \frac{n^{r-1}}{(r-1)!rl\binom{rl}{\lfloor \frac{l+1}{2} \rfloor}} + O(n^{r-2}). \tag{4.8}$$

Thus each loose path found in $\mathcal{H}$ has a set of $\left\lfloor \frac{l+1}{2} \right\rfloor$ vertices which have many common edge finishing $(r-1)$ sets in the rest of the graph.

Let U be such a set of $\left\lfloor \frac{l+1}{2} \right\rfloor$ vertices. Since $|E\left(V\left(\mathcal{H}\right) \setminus U\right)| > f(n - \left\lfloor \frac{l+1}{2} \right\rfloor, r, k-1, l)$, we can find $(k-1) \cdot \mathcal{P}_l^{(r)}$ on vertices inside $V(\mathcal{H}) \setminus U$, say on vertex set W. By (4.8), we can pick one of U's edge finishing sets which is disjoint from both W and the rest of U. Repeating this

$l - \left\lfloor \frac{l+1}{2} \right\rfloor$ times gives us a loose path which is disjoint from W, and thus we have constructed $k \cdot \mathcal{P}_l^{(r)}$.

This gives us the desired upper bound on $\mathrm{ex}_r(n, k \cdot \mathcal{P}_l^{(r)})$, and we note that this bound is attained by the graph formed by taking all edges incident to a set of $k \left\lfloor \frac{l+1}{2} \right\rfloor - 1$ vertices, and an extra disjoint edge when l is even. $\square$

4.4 Multiple Loose Paths of Different Lengths

Building on the results above for loose paths, we are able to apply methods similar to those used in Chapter 3 to give bounds on extremal numbers for multiple loose paths of different lengths. We note some ambiguity in the base case: instead of forbidding a particular loose path of length l the family of loose paths of length l is forbidden; thus, as in the previous theorem, we are not able to guarantee the existence of a particular loose path of a given length, only that some l-edge loose path exists.

Theorem 4.4.1 (B., Kettle). *Let $r \geq 3$, $k \geq 2$, and n sufficiently large. For any $l_1, \ldots, l_k \geq 3$, letting $t = \sum_{i \in [k]} \left\lfloor \frac{l_i+1}{2} \right\rfloor - 1$, $c_l = \mathbb{1}_{\{any\ l_k\ is\ even\}}$, any graph on at least $\binom{n-1}{r-1} + \ldots + \binom{n-t}{r-1} + c_l$ edges contains k vertex disjoint loose paths of lengths $l_1, \ldots, l_k$, respectively. For ease of notation, we define $h(n, r, \{l_1, \ldots, l_k\}) = \binom{n-1}{r-1} + \ldots + \binom{n-t}{r-1} + c_l$.*

We again note that the hypergraph on n vertices in which each edge is incident to a specified set S of t vertices, along with a single edge disjoint from S when one of the paths is even, gives a graph with exactly $h(n, r, \{l_1, \ldots, l_k\})$ and without k vertex disjoint paths of the specified lengths.

Proof. The case $k = 1$ is provided by 4.2.4. We proceed by induction on k; thus assume that $k \geq 2$, and that $\mathcal{H}$ is a hypergraph on n vertices and with $|E(\mathcal{H})| = m > h(n, r, \{l_1, \ldots, l_k\})$. If any of the l_i is even, we rearrange the list so that l_1 is even.

Since $h(n,r,\{l_1,\ldots,l_k\}) > h(n,r,l_1)$, for n large enough, we can find at least one loose path on l_1 vertices inside $\mathcal{H}$. Consider one of these l_1-paths, say on vertex set P. Certainly $|E(V(\mathcal{H}) \setminus P)| \leq h(n-|P|, r, \{l_2,\ldots,l_k\})$, or else by induction, the graph on $V(H) \setminus P$ contains $l_2 \cdot \ldots \cdot l_k$; these along side the loose path on P form $k \cdot \mathcal{P}_l^{(r)}$.

Letting N_P denote the number of edges of $\mathcal{H}$ incident to vertices in P, we have that

$$\begin{aligned} N_P &\geq m - h(n-|P|, r, \{l_2,\ldots,l_k\}) && (4.9)\\ &\geq h(n,r,\{l_2,\ldots,l_k\}) - h(n-(l+1), r, \{l_2,\ldots,l_k\}) && (4.10)\\ &= \frac{\left\lfloor \frac{l_1+1}{2} \right\rfloor n^{r-1}}{(r-1)!} + O(n^{r-2}) && (4.11) \end{aligned}$$

We now focus on counting sets of vertices which can be used to easily 'finish' edges started by vertices in P. With this in mind, for every set R of $r-1$ vertices from $V(\mathcal{H}) \setminus P$, we define

$$A_R = \{e \in E(\mathcal{H}) : R \subseteq e \text{ and } e \setminus R \in P\}.$$

We now break the $(r-1)$ subsets of $V(\mathcal{H}) \setminus P$ into two sets, dependent on the size of their respective A_R:

$$A = \left\{ R \in (V(\mathcal{H}) \setminus P)^{(r-1)} : |A_R| \leq \left\lfloor \frac{l_1+1}{2} \right\rfloor - 1 \right\}$$

$$B = \left\{ R \in (V(\mathcal{H}) \setminus P)^{(r-1)} : |A_R| \geq \left\lfloor \frac{l_1+1}{2} \right\rfloor \right\}$$

Counting edges entirely contained in P and the edges incident to the sets A, B defined above, we have that

$$N_P \leq \binom{|V(P)|}{2}\binom{n}{r-2} + \left(\left\lfloor \frac{l_1+1}{2} \right\rfloor - 1\right)|A| + |V(P)|\,|B|\,, \tag{4.12}$$

$$\leq \binom{|V(P)|}{2}\binom{n}{r-2} + \left(\left\lfloor \frac{l_1+1}{2} \right\rfloor - 1\right)\binom{n}{r-1} + rl|B|. \tag{4.13}$$

By comparison of the upper and lower bounds on N_P, (4.11)) and (4.13), we have that

$$|B| \geq \frac{\frac{n^{r-1}}{(r-1)!} + O(n^{r-2})}{rl}. \tag{4.14}$$

To each set $R \in B$ we associate a set of $\lfloor \frac{l_1+1}{2} \rfloor$ vertices from A_R arbitrarily. From (4.14), we see that some set of $\lfloor \frac{l_1+1}{2} \rfloor$ vertices is chosen many times; here 'many' is at least:

$$\frac{n^{r-1}}{(r-1)!rl\binom{|V(P)|}{\lfloor \frac{l+1}{2} \rfloor}} + O(n^{r-2}) \tag{4.15}$$

$$\geq \frac{n^{r-1}}{(r-1)!rl\binom{rl}{\lfloor \frac{l+1}{2} \rfloor}} + O(n^{r-2}). \tag{4.16}$$

Thus each loose path found in $\mathcal{H}$ has a set of $\lfloor \frac{l_1+1}{2} \rfloor$ vertices which have many common edge finishing $(r-1)$ sets in the rest of the graph.

Let U be such a set of $\lfloor \frac{l+1}{2} \rfloor$ vertices. Since $|E\left(V\left(\mathcal{H}\right) \setminus U\right)| > h(n - \lfloor \frac{l+1}{2} \rfloor, r, \{l_2, \ldots, l_k\})$, we can find k vertex disjoint loose paths of appropriate lengths on vertices inside $V(\mathcal{H}) \setminus U$, say on vertex set W. By (4.16), we can pick one of U's edge finishing set which is disjoint from both W and the rest of U. Repeating this $l_1 - \lfloor \frac{l_1+1}{2} \rfloor$ times gives us a loose path which is disjoint from W, and thus we have constructed our k disjoint loose paths.. $\square$

4.5 Multiple Linear Paths

We shall need the following result due to Keevash, Mubayi, and Wilson (Theorem 1.3 in [20]), which gives an upper bound on the number of edges in a hypergraph where no two edges intersect at *exactly* one vertex.

Theorem 4.5.1. *Let $\mathcal{H}$ be an r-uniform hypergraph on n vertices with no singleton intersection, where $r \geq 3$. Then*

$$|E(\mathcal{H})| \leq \binom{n}{r-2}.$$

We now state our result for the case of linear paths.

Theorem 4.5.2 (B., Kettle). *Let $r \geq 4$, $l \geq 3$, $k \geq 2$, and n sufficiently large. Then letting $t = k\left\lfloor \frac{l+1}{2} \right\rfloor - 1$, $d_l = \begin{cases} 0, \text{ for } l \text{ odd}, \\ \binom{n-t-2}{r-2}, \text{ for } l \text{ even}. \end{cases}$, any graph on n vertices with at least $\binom{n-1}{r-1} + \ldots + \binom{n-t}{r-1} + d_l$ edges contains k vertex disjoint linear paths each of length l.*

For ease of notation, we define $a(n, r, k, l) = \binom{n-1}{r-1} + \ldots + \binom{n-t}{r-1} + d_l$.

We note that the hypergraph on n vertices in which each edge is incident to a specified set S of t vertices, along with all edges disjoint from S containing some two fixed elements not in S when k is even, gives a graph with exactly $a(n, r, k, l)$ and without k vertex disjoint paths of length l. We note that for approximately the first half of the proof, this is identical to the proof for loose paths. The difference arises in the last steps, where we are building a linear path out of common neighborhoods. Instead of simply taking any two intersecting edges, as in the case for loose paths, we need to find edges which intersect appropriately for building linear paths.

Proof. The case $k = 1$ is provided by 4.2.4. We proceed by induction on k. Let $k \geq 2$, and let $\mathcal{H}$ be a hypergraph on n vertices and with $|E(\mathcal{H})| = m > a(n, r, k, l)$.

Since $a(n, r, k, l) > a(n, r, 1, l)$, for n large enough, we can find at least one linear path inside $\mathcal{H}$.

Consider one of these linear paths, say on vertex set P. Certainly $|E(V(\mathcal{H}) \setminus P)| \leq a(n - |P|, r, k - 1, l)$, or else by induction, the graph on $V(H) \setminus P$ contains $(k - 1) \cdot \mathfrak{P}_l^{(r)}$; these along side the linear path on P form $k \cdot \mathfrak{P}_l^{(r)}$.

Letting N_P denote the number of edges of $\mathcal{H}$ incident to vertices in P, we have that

$$N_P \geq m - a(n - |P|, r, k - 1, l) \tag{4.17}$$

$$\geq a(n, r, k, l) - a(n - (r - 1)l - 1, r, k - 1, l) \tag{4.18}$$

$$= \frac{\left\lfloor \frac{l+1}{2} \right\rfloor n^{r-1}}{(r - 1)!} + O(n^{r-2}) \tag{4.19}$$

Again we focus on counting sets of vertices which can be used to build edges started by vertices in P. For every set R of $r - 1$ vertices from $V(\mathcal{H}) \setminus P$, we define

$$A_R = \{e \in E(\mathcal{H}) : R \subseteq e \text{ and } e \setminus R \in P\}.$$

We now classify the $(r - 1)$ subsets of $V(\mathcal{H}) \setminus P$ into two sets based on the size of their respective A_R:

$$A = \left\{ R \in (V(\mathcal{H}) \setminus P)^{(r-1)} : |A_R| \leq \left\lfloor \frac{l+1}{2} \right\rfloor - 1 \right\}$$

$$B = \left\{ R \in (V(\mathcal{H}) \setminus P)^{(r-1)} : |A_R| \geq \left\lfloor \frac{l+1}{2} \right\rfloor \right\}$$

Counting the edges lying entirely on vertices in P along with edges incident to A, B defined above, we have that

$$N_P \leq \binom{|V(P)|}{2}\binom{n}{r-2} + \left(\left\lfloor \frac{l+1}{2} \right\rfloor - 1\right)|A| + |V(P)|\,|B|\,, \tag{4.20}$$

$$\leq \binom{|V(P)|}{2}\binom{n}{r-2} + \left(\left\lfloor \frac{l+1}{2} \right\rfloor - 1\right)\binom{n}{r-1} + rl|B|. \tag{4.21}$$

By comparison of the upper and lower bounds on N_P, (4.19) and (4.21), we have that

$$|B| \geq \frac{\frac{n^{r-1}}{(r-1)!} + O(n^{r-2})}{rl}. \tag{4.22}$$

To each set $R \in B$ we now associate a set of $\lfloor \frac{l+1}{2} \rfloor$ vertices from A_R arbitrarily. From (4.22), we see that some set of $\lfloor \frac{l+1}{2} \rfloor$ vertices is chosen many times; here 'many' is again at least:

$$\frac{n^{r-1}}{(r-1)!rl\binom{|V(P)|}{\lfloor \frac{l+1}{2} \rfloor}} + O(n^{r-2}) \tag{4.23}$$

$$\geq \frac{n^{r-1}}{(r-1)!rl\binom{rl}{\lfloor \frac{l+1}{2} \rfloor}} + O(n^{r-2}). \tag{4.24}$$

Thus each linear path found in $\mathcal{H}$ has a set of $\lfloor \frac{l+1}{2} \rfloor$ vertices which have many common edge finishing $(r-1)$ sets in the rest of the graph.

Remark: It is at this point that the proof differs from the proofs for loose paths.

Let U be such a set of $\lfloor \frac{l+1}{2} \rfloor$ vertices. Since $|E\left(V\left(\mathcal{H}\right) \setminus U\right)| > a(n - \lfloor \frac{l+1}{2} \rfloor, r, k-1, l)$, we can find $(k-1) \cdot \mathfrak{P}_l^{(r)}$ on vertices inside $V(\mathcal{H}) \setminus U$, say on vertex set W.

By (4.8), since

$$\frac{n^{r-1}}{(r-1)!rl\binom{rl}{\lfloor\frac{l+1}{2}\rfloor}} + O(n^{r-2}) \geq \binom{n}{r-2}, \tag{4.25}$$

we can find two $r-1$ sets from those edge finishing sets which are disjoint from both W and the rest of U. For the case of loose paths, this was enough to construct our paths. For the case of linear paths, however, we need edges which intersect in precisely one vertex. However, by the inequality in (4.25), we have enough of these $r-1$ sets that by Theorem 4.5.1, we can find among these edge finishing sets two which intersect in exactly one vertex. This is precisely the structure we need to build linear paths, and for n large enough removing these sets from our pool still leaves us enough to apply the Keevash, Mubayi, Wilson result again, and so repeating this $l - \lfloor\frac{l+1}{2}\rfloor$ times gives us a linear path which is disjoint from W; thus we have constructed $k \cdot \mathfrak{P}_l^{(r)}$. □

We further note that as all r-uniform linear paths of length l are isomorphic, we have the following immediate corollary.

Corollary 4.5.3 (B., Kettle)**.** *Let* $r \geq 4$*,* $l \geq 3$*,* $k \geq 1$*, and* n *sufficiently large. The following holds.*

$$\mathrm{ex}_r(n, k \cdot \mathfrak{P}_l^{(r)}) = a(n, r, k, l).$$

Further, by modifiying the above proof in exactly the same manner as was used to prove Theorem 4.4.1 based on the proof of Theorem 4.3.1, we obtain the following result for multiple linear paths of varying lengths. The modifications necessary are identical to those differences found between the two theorems for loose paths; thus we omit it here.

Theorem 4.5.4 (B., Kettle). *Let $r \geq 4$, $k \geq 2$, $l_1, \ldots, l_k \geq 3$, and n sufficiently large. Setting $t = \sum_{i \in [k]} \left\lfloor \frac{l_i+1}{2} \right\rfloor - 1$ and*

$$d_l = \begin{cases} 0, \ for\ l\ \ odd, \\ \binom{n-t-2}{r-2}, \ for\ l\ \ even \end{cases},$$

the following holds.

$$\mathrm{ex}_r(n, \mathfrak{P}^(r)_{l_1} \cdot \ldots \cdot \mathfrak{P}^(r)_{l_k}) = \binom{n-1}{r-1} + \ldots + \binom{n-t}{r-1} + d_l.$$

The upper bound is given by the discussed modification above, the lower bound is given by the hypergraph on n vertices in which each edge is incident to at least one of a specified set S of t vertices, along with all edges disjoint from S containing some two fixed elements not in S when one of the l_k is even.

Chapter 5. Doubly Infinite Words

This chapter is joint work with Steven Kalikow and Karen Johannson.

In this chapter, we discuss two notions of doubly stationary processes on doubly infinite words. In particular, we show that any *uniform martingale* is a *random Markov process*. These two notions were shown to be equivalent for a two letter alphabet by Steven Kalikow in 1990 [18]. In Section 5.1, we discuss first a little bit of background and the relevant definitions are given. Then in Section 5.2, we give a slightly reformulated version of Kalikow's proof, before extending this to finite alphabets in Section 5.3.

5.1 Definitions

We let A be any set, and let $\Omega = A^{\mathbb{Z}}$. We call A the *alphabet*, and Ω the set of doubly infinite words on alphabet A. For each $\omega = (\omega_j)_{j\in\mathbb{Z}} \in \Omega$, and each $i \in \mathbb{Z}$, we define $X_i(\omega) = \omega_i$.

Given an alphabet A, set of doubly infinite words Ω, a sigma algebra $\mathcal{A}$, and a probability measure $\mathbb{P}$ on $\mathcal{A}$, we will investigate the triple $(\Omega, \mathcal{A}, \mathbb{P})$. Such a triple is called a *stationary process* if for $i, j \in \mathbb{Z}$, $r \in \mathbb{N}$, and $a_0, \ldots, a_r \in A$,

$$\mathbb{P}(X_i = a_0 \wedge X_{i+1} = a_1 \wedge \cdots \wedge X_{i+r} = a_r)$$
$$= \mathbb{P}(X_j = a_0 \wedge X_{j+1} = a_1 \wedge \cdots \wedge X_{j+r} = a_r).$$

That is, a triple is a stationary process if the law of cylinder sets in the word is independent of its location within the word. In general, we use $\{X_i\}_{i\in\mathbb{Z}}$ to represent an element of Ω chosen at random according to the probability measure

$\mathbb{P}$. The stationary property of the process allows us to focus simply on the law of ω_0; since the process is invariant under shifts, the distribution here defines the whole process.

For a given word $\omega \in \Omega$, we call $(w_i)_{i\in\mathbb{Z}_-} \in A^{\mathbb{Z}_-}$ the *past*, or *whole past*, of w_0. Similarly, we call $(w_{-i})_{i=-m}^{-1}$ the m*-past* of X_0. To ease notation, we will write, for example, $\mathbb{P}\left(X_0 = \omega_0 \mid (X_i)_{-m}^{-1} = (\omega_i)_{-m}^{-1}\right)$ as shorthand for $\mathbb{P}\left(X_0 = \omega_0 \mid X_{-1} = \omega_{-1}, \ldots, X_{-m} = \omega_{-m}\right)$. We now define two particular families of stationary processes.

Definition 5.1.1. A stationary process $(\Omega, \mathcal{A}, \mathbb{P})$ on alphabet A is called a *uniform martingale* if for each $\epsilon > 0$, there exists an $N \in \mathbb{N}$ such that for all $m \geq N$ and each sequence $(w_i)_{i\in\mathbb{Z}_-} \in A^{\mathbb{Z}_-}$, the following holds almost surely.

$$\left|\mathbb{P}\left(X_0 = \omega_0 \mid (X_i)_{-m}^{-1} = (\omega_i)_{-m}^{-1}\right) - \mathbb{P}\left(X_0 = \omega_0 \mid (X_i)_{i<0} = (\omega_i)_{i<0}\right)\right| < \epsilon.$$

Thus if a process is a uniform martingale, the probability measure on ω_0 given the whole past can be approximated arbitrarily close by a measure which only looks at the m-past, for m large enough.

Definition 5.1.2. A stationary process $(\Omega, \mathcal{A}, \mathbb{P})$ is called a *random Markov process* if there exists an independent stationary process $(\Omega', \mathcal{A}', \mathbb{P}')$ on alphabet $\mathbb{Z}^+$ and a coupling $\hat{\mathbb{P}}$ of Ω and Ω', $\left\{((X_i, L_i))_{i\in\mathbb{Z}} : X_i \in A, L_i \in \mathbb{Z}^+\right\}$, such that L_0 is independent of $\{X_i : i < 0\}$ and so that the following holds.

$$\hat{\mathbb{P}}\left(X_0 = \omega_0 \mid (X_i)_{-n}^{-1} = (\omega_i)_{-n}^{-1} \wedge L_0 = n\right) = \hat{\mathbb{P}}\left(X_0 = \omega_0 \mid (X_i)_{i<0} = (\omega_i)_{i<0} \wedge L_0 = n\right).$$

We call the coupled L_i the *look back time for* X_i, for each $i \in \mathbb{Z}$. This is since if the process is a random Markov process, one need know only L_0 and then look at the L_0-past to determine the law of X_0 exactly.

5.2 Two Letter Alphabets

The following theorem is due to Steven Kalikow, and states the equivalence of random Markov processes and Uniform martingales, for two letter alphabets. We present a reformulation of Kalikow's proof here; the ideas are similar to those used in the proof of Theorem 5.3.1 for finite alphabets, and this reformulation uses similar notation to that used later in order to give a sketch of the proof for larger alphabets.

Theorem 5.2.1. *Any stationary process,* $\left(\{0,1\}^{\mathbb{Z}}, \mathcal{A}, \mathbb{P}\right)$*, which is a uniform martingale is also a random Markov process.*

Proof. Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a uniform martingale. We will in fact show that for any sequence $(p_i)_{i\geq 0} \subseteq (0,1)$ with $\sum_{i\geq 0} p_i = 1$, we can choose a sequence of rapidly increasing look back times $(n_i)_{i\geq 0}$ so that $\mathbb{P}'\left(L_0 = n_i\right) = p_i$, and for all other $k \in \mathbb{N} \setminus \{n_i\}_{i\geq 0}$, $\mathbb{P}'\left(L_0 = k\right) = 0$. Having done so, will have shown that Ω is a random Markov process.

Since $(\Omega, \mathcal{A}, \mathbb{P})$ is a uniform martingale, we can choose n_0 large enough that for any $m \geq n_0$ and for each $\omega \in \Omega$,

$$\left|\mathbb{P}\left(X_0 = \omega_0 \mid (X_i)_{-m}^{-1} = (\omega_i)_{-m}^{-1}\right) - \mathbb{P}\left(X_0 = \omega_0 \mid (X_i)_{i<0} = (\omega_i)_{i<0}\right)\right| < \frac{1}{3}.$$

We now will define what amounts to a table of what the associated probabilities ought to be for a given word and amount of past; we define this naïvely, and then show that it has the desired properties.

For each $n \geq n_0$ and each $\omega \in \Omega$, if $\mathbb{P}\left(X_0 = 1 \mid (X_i)_{-n}^{-1} = (\omega_i)_{-n}^{-1}\right) \geq \frac{1}{2}$, we define

$$P(\omega, n) = \sup_{\omega' : (\omega'_i)_{-n}^{-1} = (\omega_i)_{-n}^{-1}} \mathbb{P}\left(X_0 = 1 \mid (X_i)_{i<0} = (\omega'_i)_{i<0}\right).$$

If $\mathbb{P}\left(X_0 = 1 \mid (X_i)_{-n}^{-1} = (\omega_i)_{-n}^{-1}\right) < \frac{1}{2}$, we define

$$P(\omega, n) = \inf_{\omega' : (\omega'_i)_{-n}^{-1} = (\omega_i)_{-n}^{-1}} \mathbb{P}\left(X_0 = 1 \mid (X_i)_{i<0} = (\omega'_i)_{i<0}\right).$$

We take a moment to explain the above definitions, as the notation somewhat obscures the simple idea at play. Both the infimum and supremum are taken over all words which agree with ω for the n-past. We define the sequence $P(\omega, n)$ in this way so that if $\mathbb{P}\left(X_0 = 1 \mid (X_i)_{-n}^{-1} = (\omega_i)_{-n}^{-1}\right) \geq \frac{1}{2}$, the values of $P(\omega, n)$ remain bounded away from 0; if $\mathbb{P}\left(X_0 = 1 \mid (X_i)_{-n}^{-1} = (\omega_i)_{-n}^{-1}\right) < \frac{1}{2}$, the values of $P(\omega, n)$ remain bounded away from 1. We note to the reader that whenever one encounters $\mathbb{P}\left(\cdot, (\omega_i)_{-n}^{-1}\right)$, it is best thought of as '$\cdot$ given the n_0-past'. This view makes the intuition as to what is happening in this proof, and many in the area, much clearer.

Remark: The above definition of $P(\omega, n)$ will need some modification in the proof for finite alphabets.. This will be the primary difficulty in the finite alphabet case, and is where much of the complication in the proof occurs. The modification of the $P(\cdot, n)$ function will ensure that it still remains bounded away from both 0 and 1.

By the choice of n_0, in the first case $P(\omega, n) \geq \frac{1}{6}$ for all $n \geq n_0$, and by definition $P(\omega, n)$ is decreasing for $n \geq n_0$. That is,

$$\frac{1}{6} \leq \cdots \leq P(\omega, n_0 + 2) \leq P(\omega, n_0 + 1) \leq P(\omega, n_0).$$

In the second case, $P(\omega, n)$ is bounded above by $\frac{5}{6}$ for all n, and is increasing for n_0. That is,

$$P(\omega, n_0) \leq P(\omega, n_0 + 1) \leq P(\omega, n_0 + 2) \leq \cdots \leq \frac{5}{6}.$$

We now define $(n_i)_{i>0}$ recursively. Since $(\Omega, \mathcal{A}, \mathbb{P})$ is a uniform martingale, we can find n_i large enough that for every ω,

$$|P(\omega, n_i) - P(\omega, n_{i+1})| < \frac{p_{i+1}}{6 \sum_{j \leq i} p_j}.$$

With these definitions in place, we will now show that we can find a coupling $\hat{\mathbb{P}}$ of Ω and $((n_i)_{i \geq 0})^{\mathbb{Z}}$ which satisfies the following property.

$$P(\omega, n) = \hat{\mathbb{P}} \left(X_0 = 1 \mid \omega_{-1}, \ldots, \omega_{-n_i} \text{ and } L_0 \leq n_i \right). \tag{5.1}$$

Since we wish L_0 to be independent of $(X_i)_{i<0}$, (5.1) is the same as

$$P(\omega, n) = \frac{\hat{\mathbb{P}} \left(X_0 = 1 \text{ and } L_0 \mid \omega_{-1}, \ldots, \omega_{-n_i} \right)}{\hat{\mathbb{P}} \left(L_0 \leq n_i \right)}.$$

To define this coupling, we first define another function $T(\cdot, \cdot)$ based on $P(\cdot, \cdot)$, so that

$$T(w, n_i) = \hat{\mathbb{P}}(X_0 = 1 \mid \omega_{-1}, \ldots, \omega_{-n_i} \text{ and } L_0 = n_i). \tag{5.2}$$

Indeed, define $T(\omega, n_0) = P(\omega, n_0)$, and for $i \geq 1$, define $T(\omega, n_i)$ so that

$$p_i T(w, n_i) = \left(\sum_{j \leq i-1} p_j \right) P(\omega, n_{i-1}) - \left(\sum_{j \leq i} p_j \right) P(\omega, n_i)$$

We note that

$$
\begin{aligned}
\hat{\mathbb{P}}\left(X_0 = 1 \mid \omega\right) &= \sum_{i\leq 0} \hat{\mathbb{P}}\left(X_0 = 1 \text{ and } L_0 = n_i \mid \omega\right) \\
&= \sum_{i\geq 0} p_i T(\omega, n_i) \\
&= p_0 P(\omega, n_0) + \sum_{i\geq 1}\left(\left(\sum_{j\leq i} p_j\right) P(\omega, n_i) - \left(\sum_{j\leq i-1} p_j\right) P(\omega, n_{i-1})\right) \\
&= \lim_{i\to\infty}\left(\sum_{j\leq i} p_j\right) P(\omega, n_i) \\
&= \mathbb{P}(X_0 = 1 \mid \omega).
\end{aligned}
$$

Thus $\mathbb{P}$ and $\hat{\mathbb{P}}$ agree on all cylinder sets in Ω, and thus 5.2 defines a propability measure that is a couple of Ω and $(\{n_i\}_{i\geq 0})^{\mathbb{Z}}$, as long as $0 \leq T(\omega, n_i) \leq 1$.

To show that $T(\omega, n_i) \in [0, 1]$, we consider two cases.

Case 1: Suppose that $\mathbb{P}\left(X_0 = 1 \mid (\omega_i)_{-n_0}^{-1}\right) \geq \frac{1}{2}$. We have already that $P(\omega, n_{i-1}) \geq P(\omega, n_i) \geq \frac{1}{6}$. Since $P(\omega, n_i)$ is a convex combination of $T(\omega, n_i)$ and $P(\omega, n_{i-1})$, certainly $T(\omega, n_i) \leq P(\omega, n_i) \leq 1$. For the lower bound, we note that

$$
\begin{aligned}
p_i T(\omega, n_i) &= \left(\sum_{j\leq i} p_j\right) P(\omega, n_i) - \left(\sum_{j\leq i-1} p_j\right) P(\omega, n_{i-1}) \\
&= p_i P(\omega, n_i) - \left(\sum_{j\leq i-1} p_j\right)\left(P(\omega, n_{i-1}) - P(\omega, n_i)\right) \\
&\geq p_i \frac{1}{6} - \left(\sum_{j\leq i-1} p_j\right) \frac{p_i}{6\sum_{j\leq i-1} p_j} = 0.
\end{aligned}
$$

Thus in this case, $T(\omega, n_i) \in [0, 1]$ as claimed.

Case 2: Suppose $\mathbb{P}\left(X_0 = 1 \mid (\omega_i)_{-n_0}^{-1}\right) < \frac{1}{2}$. By definition, $P(\omega, n_{i-1}) \leq P(\omega, n_i) \leq \frac{5}{6}$, and so by convexity we get the upper bound. As in the first case, we have that

$$\begin{aligned}
p_i T(\omega, n_i) &= p_i P(\omega, n_i) + \left(\sum_{j \leq i-1} p_j\right)(P(\omega, n_i) - P(\omega, n_{i-1}) \\
&\leq p_i \frac{5}{6} + \left(\sum_{j \leq i-1} p_j\right) \frac{p_i}{6 \sum_{j \leq i-1} p_j} \\
&\leq p_i.
\end{aligned}$$

Therefore we have that in all cases, $T(\omega, n_i) \in [0, 1]$, and so the measure given by 5.2 is indeed a coupling. Thus we have shown $(\Omega, \mathcal{A}, \mathbb{P})$ is a random Markov process, as claimed. □

5.3 Finite Alphabets

With Steven Kalikow and Karen Johannson, we proved the natural extension of Theorem 5.2.1. We state and prove this result here. While the proof is similar in overall structure to the proof in the two letter case, there are a number of subtleties which do not arise in the binary alphabet case.

Theorem 5.3.1. *Let $k \geq 2$, $A = [k]$, and $\Omega = A^{\mathbb{Z}}$. If $(\Omega, A, \mathbb{P})$ is a uniform martingale, then Ω is a random Markov process.*

Proof. As before, let $(p_i)_{i \geq 0}$ be a sequence of nonnegative real numbers, with $\sum_{i \geq 0} p_i = 1$. We will construct a sequence of lookback times $(n_i)_i \geq 0$ such that Ω is a random Markov process with $\mathbb{P}(L_0 = n_i) = p_i$.

Ω is a uniform martingale, and so we can choose n_0 large enough so that for all $m \geq n_0$ and every $\omega \in \Omega$, the following holds.

$$\left|\mathbb{P}\left(X_0 = \omega_0 \mid (X_i)_{-m}^{-1} = (\omega_i)_{-m}^{-1}\right) - \mathbb{P}\left(X_0 = \omega_0 \mid (X_i)_{i<0} = (\omega_i)_{i<0}\right)\right| < \frac{1}{k^2}.$$

Now, fix ω and consider for each $a \in A$ the quantity $\mathbb{P}\left(X_0 = a \mid (X_i)_{-n_0}^{-1} = (\omega_i)_{-n_0}^{-1}\right)$. By reordering the alphabet, we may assume without loss of generality the following.

$$\begin{aligned}\mathbb{P}\left(X_0 = 1 \mid (X_i)_{-n_0}^{-1} = (\omega_i)_{-n_0}^{-1}\right) &\leq \mathbb{P}\left(X_0 = 2 \mid (X_i)_{-n_0}^{-1} = (\omega_i)_{-n_0}^{-1}\right) \\ &\leq \cdots \leq \mathbb{P}\left(X_0 = k \mid (X_i)_{-n_0}^{-1} = (\omega_i)_{-n_0}^{-1}\right) \qquad (5.3)\end{aligned}$$

We note here that since we have only a k letter alphabet, and we have arranged the letters so that $\mathbb{P}\left(X_0 = k \mid (X_i)_{-n_0}^{-1} = (\omega_i)_{-n_0}^{-1}\right)$ is the largest probability, we have that $\mathbb{P}\left(X_0 = k \mid (X_i)_{-n_0}^{-1} = (\omega_i)_{-n_0}^{-1}\right) \geq \frac{1}{k} \geq \frac{2}{k^2}$. Now, let $l \in [0, \ldots, k-1]$ be such that the following holds.

$$\mathbb{P}\left(X_0 = l \mid (X_i)_{-n_0}^{-1} = (\omega_i)_{-n_0}^{-1}\right) < \frac{2}{k^2} \leq \mathbb{P}\left(X_0 = l+1 \mid (X_i)_{-n_0}^{-1} = (\omega_i)_{-n_0}^{-1}\right)$$

We note here that if $l = 0$, then we have that for all $a \in A$, $\mathbb{P}\left(X_0 = a \mid (X_i)_{-n_0}^{-1} = (\omega)_{-n_0}^{-1}\right) \geq \frac{2}{k^2}$; that is, all letters occur with relatively high probability. This will cause no difficulties.

Now, for every $a \in A \setminus \{k\}$ and each $n \geq n_0$, we will define a function $P_a(\omega, n)$ as in the proof of Theorem 5.2.1; the primary difference in the proof will come in examining the properties of $P_a(\omega, n)$, and showing that it gives rise to the proper coupling. We also note that we won't define $P_k(\omega, n)$; this lines up with the two letter case, where we defined only one P function. With this in mind, we define P in two parts as follows. For $a \in [1, l]$, we define

$$P_a(\omega, n) = \sup_{\omega' : (\omega'_i)_{-n}^{-1} = (\omega_i)_{-n}^{-1}} \mathbb{P}\left(X_0 = a \mid (X_i)_{i<0} = (\omega'_i)_{i<0}\right).$$

For $a \in (l, k-1]$, we define

$$P_a(\omega, n) = \inf_{\omega' : (\omega'_i)_{-n}^{-1} = (\omega_i)_{-n}^{-1}} \mathbb{P}\left(X_0 = a \mid (X_i)_{i<0} = (\omega'_i)_{i<0}\right).$$

We define the sequence $P_a(\omega, n)$ in this way so that if in the first case the values of $P(\omega, n)$ remain bounded away from 0, while in the second case the values of $P_a(\omega, n)$ remain bounded away from 1.

We note that by the choice of n_0, if $a \in [1, l]$, then $P_a(\omega, n)$ is increasing in n for $n \geq n_0$, and bounded above by $\frac{3}{k^2}$; that is,

$$P_a(\omega, n) \leq P_a(\omega, n+1) \leq \cdots \leq \frac{3}{k^2} = \frac{2}{k^2} + \frac{1}{k^2}.$$

Similarly, if $a \in (l, k-1]$, then $P_a(\omega, n)$ is decreasing in n for $n \geq n_0$ and bounded below by $\frac{1}{k^2}$; that is,

$$P_a(\omega, n) \geq P_a(\omega, n+1) \geq \cdots \geq \frac{1}{k^2} = \frac{2}{k^2} - \frac{1}{k^2}.$$

Now, we are prepared to define the remaining lookback times. Indeed, since $(\Omega, \mathcal{A}, \mathbb{P})$ is a uniform martingale we can define $(n_i)_{i \geq 1}$ recursively so that for all ω and all $a \in A$,

$$|P_a(\omega, n_{i+1}) - P_a(\omega, n_i)| < \frac{p_i}{k^3 \sum_{j \leq i-1} p_j}.$$

We now let $(\Omega', \mathbb{P}')$ be an independent stationary distribution on alphabet $\{n_i\}_{i \geq 0}$ with $\mathbb{P}'(L_0 = n_i) = p_i$.

The goal now is to show that there is a coupling $\hat{\mathbb{P}}$ of $(\Omega, \mathcal{A}, \mathbb{P})$ and $(\Omega', \mathbb{P}')$ with the desired property that for each $a \in [1, k-1]$, each $\omega \in \Omega$, and each $i \in \mathbb{N}$,

$$P_a(\omega, n_i) = \hat{\mathbb{P}}\left(X_0 = a \mid \omega_{-1}, \omega_{-2}, \ldots, \omega_{-n_i} \text{ and } L_0 \leq n_i\right).$$

Note that by the choice of n_0,

$$\begin{aligned}\sum_{a\in[k-1]} P_a(\omega,n_i) &\leq \sum_{a\in[k-1]}\left(\mathbb{P}\left(X_0=a \mid \omega_{-1},\ldots,\omega_{-n_0}\right)+\frac{1}{k^2}\right)\\ &= 1-\mathbb{P}\left(X_0=k \mid \omega_{-1},\ldots,\omega_{-n_0}\right)+(k-1)\frac{1}{k^2}\\ &\leq 1-\frac{1}{k}+\frac{k-1}{k^2}\\ &= 1-\frac{1}{k^2}. \end{aligned} \tag{5.4}$$

We now define, as before, the function $T_a(\omega,n)$. Indeed, define for each $a\in[k-1]$, $T_a(\omega,n_0)=P_a(\omega,n_0)$, and for $i\geq 1$, define $T_a(\omega,n_i)$ such that the following holds.

$$p_iT_a(\omega,n_i)+\left(\sum_{j<i}p_j\right)P_a(\omega,n_{i-1})=\left(\sum_{j\leq i}p_j\right)P_a(\omega,n_i).$$

This defines our table function every except at k; thus we define $T_k(\omega,n_i)$ to be the 'left over' probability. That is,

$$T_k(\omega,n_i)=1-\sum_{a\in[k-1]}T_a(\omega,n_i).$$

Ideally, we would like to define the coupling $\hat{\mathbb{P}}$ so that $\hat{\mathbb{P}}\left(X_0=1 \mid \omega_{-1},\ldots,\omega_{-n_i} \text{ and } L_0=n_i\right)=T_a(\omega,n_i)$. As before, $T_a(\omega,n_i)$ is defined in such a way that this will be true given that $T_a(\omega,n_i)\in[0,1]$. It suffices to show that for each $a\in[k-1]$, $T_a(\omega,n_I)\geq 0$, and that $\sum_{a\in[k-1]}T_a(\omega,n_i)\leq 1$. For the first part, we consider two cases.

Case 1: For $a<l$, $P_a(\omega,n_{i-1})$ is a convex combination of $T_a(\omega,n_i)$ and $P_a(\omega,n_i)$. Since $P_a(\omega,n_i)\leq P_a(\omega,n_{i-1})$, then $T_a(\omega,n_i)\geq P_a(\omega,n_{i-1}\geq 0$ by its definition.

Case 2: For $a \geq l$, $P_a(\omega, n_{i-1}) \geq P_a(\omega, n_i)$. Thus by the definition of the n_i's and the choice of l, we have that

$$\begin{aligned}
p_i T_a(\omega, n_i) &= \left(\sum_{j \leq i} p_j\right) P_a(\omega, n_i) - \left(\sum_{j<i} p_j\right) P_a(\omega, n_{i-1}) \\
&= p_i P_a(\omega, n_i) - \left(\sum_{j<i} p_j\right) (P_a(\omega, n_{i-1}) - P_a(\omega, n_i)) \\
&\geq p_i \frac{1}{k^2} - \left(\sum_{j<i} p_j\right) \frac{p_i}{k^3 \sum_{j \leq i-1} p_j} \\
&= p_i \left(\frac{1}{k^2} - \frac{1}{k^3}\right) = p_i \frac{k-1}{k^3} > 0.
\end{aligned}$$

Now consider $T_k(\omega, n_i)$. By definition, $T_k(\omega, n_i) \leq 1$; it remains to show that this is nonnegative. Note first that for $a > l$, $P_a(\omega, n_i) - P_a(\omega, n_{i-1}) \leq 0$ since $P_a(\omega, \cdot)$ is decreasing here. Then we have that for each $i \in \mathbb{N}$,

$$\begin{aligned}
p_i \sum_{a \in [k-1]} T_a(\omega, n_i) &= p_i \sum_{a \in [k-1]} P_a(\omega, n_i) + \left(\sum_{j<i} p_j\right) \sum_{a \in [k-1]} (P_a(\omega, n_i) - P_a(\omega, n_{i-1})) \\
&\leq p_i \sum_{a \in [k-1]} P_a(\omega, n_i) + \left(\sum_{j<i} p_j\right) \sum_{a \in [l]} (P_a(\omega, n_i) - P_a(\omega, n_{i-1})) \\
&\leq p_i \left(1 - \frac{1}{k^2}\right) + \left(\sum_{j<i} p_j\right) \sum_{a \in l} \frac{p_i}{k^3 \sum_{j \in [i-1]} p_j} \\
&\leq p_i \left(1 - \frac{1}{k^2}\right) + k \frac{1}{k^3} = p_i.
\end{aligned}$$

Since $p_i > 0$, $\sum_{a\in[k-1]} T_a(\omega, n_i) \leq 1$. Thus defining the probability on $\Omega \times \Omega'$ by

$$\hat{\mathbb{P}}\left(X_0 = a \mid \omega_{-1}, \omega_{-2}, \ldots, \omega_{-n_i} \text{ and } L_0 = n_i\right) = T_a(\omega, n_i)$$

indeed gives a coupling, as it agrees with $\mathbb{P}$ on all cylinder sets of Ω, and hence on all measurable sets of Ω. $\square$

References

[1] N. Alon, L. Rónyai, and T. Szabó, *Norm-graphs: variations and applications*, J. Combin. Theory Ser. B **76** (1999), 280–290.

[2] S. Balasubramanian and E. Dobson, *Constructing trees in graphs with no k2,s*, Journal of Graph Theory **56** (2007), 301–310.

[3] P. N. Balister, E. Győri, J. Lehel, and R. H. Schelp, *Connected graphs without long paths*, Discrete Math. **308** (2008), 4487–4494.

[4] C.T. Benson, *Minimal regular graphs of girths eight and twelve*, Canad. J. Math. **18** (1966), 1091–1094.

[5] B. Bollobás, *Modern graph theory*, Graduate Texts in Mathematics, vol. 184, Springer-Verlag, New York, 1998.

[6] J. A. Bondy and M. Simonovits, *Cycles of even length in graphs*, J. Combinatorial Theory Ser. B **16** (1974), 97–105.

[7] W. G. Brown, *On graphs that do not contain a Thomsen graph*, Canad. Math. Bull. **9** (1966), 281–285.

[8] N. Bushaw and N. Kettle, *Turán numbers of multiple paths and equibipartite forests*, Combin. Probab. Comput. **20** (2011), 837–853.

[9] P. Erdős, *On sequences of integers no one of which divides the product of two others and on some related problems.*, Mitt. Forsch.-Inst. Math. Mech. Univ. Tomsk **2** (1938), 74–82 (English).

[10] P. Erdős and M. Simonovits, *A limit theorem in graph theory.*, Stud. Sci. Math. Hung. **1** (1966), 51–57 (English).

[11] P. Erdős and A. H. Stone, *On the structure of linear graphs*, Bull. Amer. Math. Soc. **52** (1946), 1087–1091.

[12] P. Erdős and T. Gallai, *On maximal paths and circuits of graphs*, Acta Math. Acad. Sci. Hungar **10** (1959), 337–356 (unbound insert).

[13] P. Erdős, A. Rényi, and V. T. Sós, *On a problem of graph theory*, Studia Sci. Math. Hungar. **1** (1966), 215–235.

[14] Z. Füredi, *New asymptotics for bipartite Turán numbers*, J. Combin. Theory Ser. A **75** (1996), 141–144.

[15] ______, *An upper bound on Zarankiewicz' problem*, Combin. Probab. Comput. **5** (1996), 29–33.

[16] Z. Füredi, T. Jiang, and R. Seiver, *Exact solution of the hypergraph Turán problem for k-uniform linear paths*, ArXiv e-prints (2011).

[17] I. Gorgol, *Turán numbers for disjoint copies of graphs*, Graphs Combin. **27** (2011), 661–667.

[18] S. Kalikow, *Random Markov processes and uniform martingales*, Israel J. Math. **71** (1990), 33–54.

[19] Gy. Katona, T. Nemetz, and M. Simonovits, *On a problem of Turán in the theory of graphs*, Mat. Lapok **15** (1964), 228–238.

[20] P. Keevash, D. Mubayi, and R.M. Wilson, *Set systems with no singleton intersection*, SIAM J. Discrete Math. **20** (2006), 1031–1041 (electronic).

[21] P. Keevash, B. Sudakov, and J. Verstraete, *On a conjecture of Erdos and Simonovits: Even Cycles*, ArXiv e-prints (2011).

[22] J. Kollár, L. Rónyai, and T. Szabó, *Norm-graphs and bipartite Turán numbers*, Combinatorica **16** (1996), 399–406.

[23] G. N. Kopylov, *Maximal paths and cycles in a graph*, Dokl. Akad. Nauk SSSR **234** (1977), 19–21.

[24] T. Kövari, V. T. Sós, and P. Turán, *On a problem of K. Zarankiewicz*, Colloquium Math. **3** (1954), 50–57.

[25] A. McLennan, *The Erdős-Sós conjecture for trees of diameter four*, J. Graph Theory **49** (2005), 291–301.

[26] J. W. Moon, *On the number of complete subgraphs of a graph*, Canad. Math. Bull. **8** (1965), 831–834.

[27] W. Moser and J. Pach, *Recent developments in combinatorial geometry*, New trends in discrete and computational geometry, Algorithms Combin., vol. 10, Springer, Berlin, 1993, pp. 281–302.

[28] V. Nikiforov, *A contribution to the Zarankiewicz problem*, Linear Algebra Appl. **432** (2010), 1405–1411.

[29] J.F. Saclé and M. Woźniak, *The Erdős-Sós conjecture for graphs without* C_4, J. Combin. Theory Ser. B **70** (1997), 367–372.

[30] A. F. Sidorenko, *Asymptotic solution for a new class of forbidden* r*-graphs*, Combinatorica **9** (1989), 207–215.

[31] M. Simonovits, *A method for solving extremal problems in graph theory, stability problems*, Theory of Graphs (Proc. Colloq., Tihany, 1966), Academic Press, New York, 1968, pp. 279–319.

[32] ______, *A method for solving extremal problems in graph theory, stability problems*, Theory of Graphs (Proc. Colloq., Tihany, 1966), Academic Press, New York, 1968, pp. 279–319.

[33] Robert Singleton, *On minimal graphs of maximum even girth*, J. Combinatorial Theory **1** (1966), 306–332.

[34] P. Turán, *Eine Extremalaufgabe aus der Graphentheorie*, Mat. Fiz. Lapok **48** (1941), 436–452.

[35] ______, *On the theory of graphs*, Colloquium Math. **3** (1954), 19–30.

[36] R. Wenger, *Extremal graphs with no C^4's, C^6's, or C^{10}'s*, J. Combin. Theory Ser. B **52** (1991), 113–116.

[37] M. Woźniak, *On the Erdős-Sós conjecture*, J. Graph Theory **21** (1996), 229–234.