2018

On the structure of dense graphs, and other extremal problems

Richard Snyder

ON THE STRUCTURE OF DENSE GRAPHS, AND OTHER EXTREMAL PROBLEMS

by

Richard Snyder

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Mathematical Sciences

The University of Memphis

December 2018

ACKNOWLEDGEMENTS

This dissertation is not the result of work done in pure solitude. I am indebted to my advisor, Béla Bollobás, for giving me the opportunity to learn many things about mathematics, and also many things about life.

I am more than grateful to my parents for allowing me the freedom to pursue whatever struck my imagination, no matter how far-fetched. This dissertation is their work, too.

Many thanks go to my family—my brother, Jack, and my two sisters, Dee and Donna. Their support has meant more to me than they know.

In my extended family: aunts, uncles, cousins—all have been extremely supportive of my endeavors. Thank you!

Thanks to my friends—especially to Adam, Jake, Andrew, Kenny, and Ryan—for always listening to my ravings and rantings.

Thanks to my coauthors: António, Gábor, Julian, Kamil, and Shoham. More importantly, I count them as my good friends, and I am glad to have met them.

Special thanks go to the triumvirate António, Julian, and Kamil. It has been quite a journey!

I would like to thank my fellow Memphians Amy, B, Joon, Aaron, David, Mariusz, Buddhika, Stephen, and Stephen.

I have met many other mathematical friends who have been kind and helpful (in no particular order): Ping, Seb, Tomas, István, Vytautas, Richard, and Dominik.

I have been blessed to have several wonderful teachers and lecturers. My high school calculus teacher, Mrs. Lipko, introduced me to beautiful ideas. My first analysis professor, Frank Beatrous, introduced me to even more beautiful ideas.

My undergraduate advisor, Jeff, introduced me to the world of combinatorics and gave me much needed direction. Paul Balister and Imre Leader presented mathematics in the way I hope to.

I would like to thank Mrs. Gabriella Bollobás for opening her home to me, and to many other combinatorialists in Memphis and elsewhere.

Many thanks go to Tricia, for helping me navigate the open waters.

Lastly, I would like to express my heartfelt thanks to Migle, for being there.

ABSTRACT

Snyder, Richard. Ph.D. The University of Memphis. December 2018. On the structure of dense graphs, and other extremal problems. Major Professor: Béla Bollobás, Ph.D., D.Sc., F.R.S.

Extremal combinatorics is an area of mathematics populated by problems that are easy to state, yet often difficult to resolve. The typical question in this field is the following: What is the maximum or minimum size of a collection of finite objects (e.g., graphs, finite families of sets) subject to some set of constraints? Despite its apparent simplicity, this question has led to a rather rich body of work. This dissertation consists of several new results in this field.

The first two chapters concern structural results for dense graphs, thus justifying the first part of my title. In the first chapter, we prove a stability result for edge-maximal graphs without complete subgraphs of fixed size, answering questions of Tyomkyn and Uzzell. The contents of this chapter are based on joint work with Kamil Popielarz and Julian Sahasrabudhe.

The second chapter is about the interplay between minimum degree and chromatic number in graphs which forbid a specific set of 'small' graphs as subgraphs. We determine the structure of dense graphs which forbid triangles and cycles of length five. A particular consequence of our work is that such graphs are 3-colorable. This answers questions of Messuti and Schacht, and Oberkampf and Schacht. This chapter is based on joint work with Shoham Letzter.

Chapter 3 departs from undirected graphs and enters the domain of directed graphs. Specifically, we address the connection between connectivity and linkedness in tournaments with large minimum out-degree. Making progress on a conjecture of Pokrovskiy, we show that, for any positive integer k, any $4k$-connected tournament with large enough minimum out-degree is k-linked. This chapter is based on joint work with António Girão.

The final chapter leaves the world of graphs entirely and examines a problem in finite set systems. More precisely, we examine an extremal problem on a family of finite sets involving constraints on the possible intersection sizes these sets may have. Such problems have a long history in extremal combinatorics. In this chapter, we are interested in the maximum number of disjoint pairs a family of sets can have under various restrictions on intersection sizes. We obtain several new results in this direction. The contents of this chapter are based on joint work with António Girão.

CONTENTS

List of Figures

Figure Page

CHAPTER 1

A STABILITY THEOREM FOR MAXIMAL K_{R+1}-FREE GRAPHS

1.1 Introduction

For a positive integer $r \geq 2$, a graph G is said to be $(r+1)$-*saturated* (or *maximal* K_{r+1}-*free*) if it contains no copy of K_{r+1}, but the addition of any edge from the complement $\overline{G}$ creates at least one copy of K_{r+1}. Let $T_r(n)$ denote the r-*partite Turán graph* that is, the n-vertex, complete r-partite graph for which each of the r classes is of order $\lfloor n/r \rfloor$ or $\lceil n/r \rceil$. We write $t_r(n) = e(T_r(n))$, and note that $t_r(n) = (1 - \frac{1}{r})\frac{n^2}{2} + O_r(1)$. Whenever we speak of an r-partite subgraph, we require that it is induced.

The classical theorem of Turán [67] tells us that, for an integer $r \geq 2$, the maximum number of edges in a graph not containing a K_{r+1} is $t_r(n)$, and that $T_r(n)$ is the unique K_{r+1}-free graph attaining this maximum. Erdős and Simonovits [23, 22, 60] discovered that this extremal problem exhibits a certain 'stability' phenomenon: K_{r+1}-free graphs for which $e(G)$ is close to $t_r(n)$ must resemble the Turán graph in an appropriate sense. In particular, they proved that every n-vertex, K_{r+1}-free graph with at least $t_r(n) - o(n^2)$ edges can be transformed into $T_r(n)$ by making at most $o(n^2)$ edge deletions and additions.

Beyond the seminal work of Erdős and Simonovits, we are lead to consider finer aspects of this phenomenon. More generally, it is natural to ask how the structure of a K_{r+1}-free graph G comes to resemble the Turán graph as the number of edges $e(G)$ approaches the Turán number $t_r(n)$. For instance, Nikiforov and Rousseau [56], in the context of a Ramsey-theoretic problem, showed that for $r \geq 2$ and ε sufficiently small (depending on r) the following holds: if G is an n-vertex K_{r+1}-free graph with $e(G) \geq \left(1 - \frac{1}{r} - \varepsilon\right) n^2/2$, then G contains an induced r-partite subgraph H with $|H| \geq (1 - 2\varepsilon^{1/3})n$ and $\delta(H) \geq \left(1 - \frac{1}{r} - 4\varepsilon^{1/3}\right) n$. In other words, G must contain a large r-partite subgraph with minimum degree almost as large as $\delta(T_r(n))$. The interested reader should consult the survey of Nikiforov [55] for a few other stability results in a similar vein.

Another result concerning the finer structure of stability is due to Brouwer [17], who showed that if $n \geq 2r+1$ and G is a K_{r+1}-free graph with $e(G) \geq t_r(n) - \lfloor \frac{n}{r} \rfloor + 2$, then G must be r-partite. This result has further been rediscovered by several authors [6, 35, 42], and Tyomkyn and Uzzell [68] recently gave a new proof. In this paper, we are interested in the structure of *maximal* K_{r+1}-free graphs near the Turán threshold. In this context, Brouwer's result says that if the number of edges of an $(r+1)$-saturated graph G is roughly within n/r of the Turán number $t_r(n)$, then G *is* complete r-partite. A natural question then arises, which informally is: When can one guarantee 'almost-spanning' complete r-partite subgraphs in $(r+1)$-saturated graphs?

Continuing this line of investigation, Tyomkyn and Uzzell [68] proved, among other results, that every 4-saturated graph on n vertices and with $t_3(n) - cn$ edges contains a complete 3-partite graph on $(1-o(1))n$ vertices (they also implicitly dealt with the 3-saturated case). They went on to ask if one can similarly find almost-spanning, complete r-partite subgraphs in $(r+1)$-saturated graphs with many edges, for $r \geq 4$. The main result of this paper is to resolve the question of Tyomkyn and Uzzell, in a stronger form. Not only do we show that this phenomenon persists for $(r+1)$-saturated graphs for all $r \geq 2$, but we also determine the edge threshold for which the result fails to hold. In particular, we show the following.

Theorem 1.1.1. *Let $r \geq 2$ be an integer. Every $(r+1)$-saturated graph G on n vertices with $t_r(n) - o(n^{\frac{r+1}{r}})$ edges contains a complete r-partite subgraph on $(1-o(1))n$ vertices.*

We also show that this theorem is tight in the sense that for every $\delta > 0$ there exist graphs G with $t_r(n) - \delta n^{\frac{r+1}{r}}$ edges for which the conclusion of Theorem 1.1.1 fails. Note that the fact that G is *maximal* K_{r+1}-free is important in the above theorem. Indeed, suppose that G is the graph obtained from $T_2(n)$ by removing a matching (suppose for simplicity that 2 divides n). Then G is triangle-free (but not maximal triangle-free), and has $t_2(n) - n/2$ edges. However, the largest complete bipartite subgraph is on $n/2$ vertices: very far from covering all but a vanishing fraction of the vertex set!

We actually deduce Theorem 1.1.1 from a stronger, quantitative result, which we now make precise. For a graph G and an integer $r \geq 2$, define the graph parameter

$$g_r(G) = \min\{|T| : T \subseteq V(G), G - T \text{ is complete } r\text{-partite}\}.$$

For $n, m \in \mathbb{N}$, let $\mathscr{S}_r(n,m)$ denote the set of all $(r+1)$-saturated graphs on n vertices with at least $t_r(n) - m$ edges. Then define

$$g_r(n,m) = \max\{g_r(G) : G \in \mathscr{S}_r(n,m)\}.$$

The quantitative form of our main theorem, stated below, gives an upper bound for the function $g_r(n,m)$ under some modest conditions on n.

Theorem 1.1.2. *Let r, n be integers satisfying $r \geq 2$, $n \geq 900r^6$. Every $(r+1)$-saturated graph with $t_r(n) - m$ edges contains a complete r-partite subgraph on $(1 - C_r m n^{-\frac{r+1}{r}})n$ vertices, where C_r is a constant depending only on r.*

We shall also give a construction in Section 1.3 showing that this result is tight, up to the value of C_r, in a certain range of m. More precisely, if $\varepsilon > 0$, $n \geq 2^{10r}/\varepsilon$ and $(\frac{r-1}{r} + \varepsilon)n \leq m \leq n^{\frac{r+1}{r}}$, then

$$c_{r,\varepsilon} m n^{-1/r} \leq g_r(n,m) \leq C_r m n^{-1/r},$$

where $c_{r,\varepsilon}$ is a constant depending on r and ε, and C_r is a constant depending only on r. This explicit form of our main result takes a major step towards a further question of Tyomkyn and Uzzell [68, 69], who asked for the determination of $g_3(n, cn)$. While we have determined $g_r(n,m)$ up to constants for $m \in \left[(\frac{r-1}{r}+\varepsilon)n, n^{\frac{r+1}{r}}\right]$, our construction giving the lower bound does not work for $m \in \left[\frac{n}{r}, \frac{r-1}{r}n\right]$. This is essentially due to the fact that in order to preserve maximality while avoiding creating copies of K_{r+1}, we have to remove enough edges. It seems difficult to preserve these properties (while also having small complete r-partite subgraphs) with the additional constraint of removing very few edges. We therefore leave the determination of $g_r(n,m)$ for $m \in \left[\frac{n}{r}, \frac{r-1}{r}n\right]$ as an open problem (see Section 1.5).

We also consider the situation for $(r+1)$-saturated graphs with $t_r(n) - Cn^{\frac{r+1}{r}}$ edges; that is, just beyond the edge threshold in Theorem 1.1.1. In this range it is perhaps most natural to consider "balanced" r-partite complete subgraphs or, in other words, r-partite *Turán* subgraphs. Recall that an r-partite graph with vertex classes $V_1, \ldots, V_r$ is *balanced* if $\big||V_i| - |V_j|\big| \leq 1$ for all $i, j \in [r]$. With this in mind, we set

$$g_r^*(G) = \min\{|T| : G - T \text{ is an } r\text{-partite Turán graph}\},$$

and, for $m, n \in \mathbb{N}$, define

$$g_r^*(n,m) = \max\{g_r^*(G) : G \in \mathscr{S}_r(n,m)\}.$$

Thus, $g_r^*(n,m)$ is the maximum number of vertices one is required to delete from an $(r+1)$-saturated graph on n vertices with at least $t_r(n) - m$ edges such that the remaining graph is an r-partite Turán graph.

How are the functions $g_r(n,m)$ and $g_r^*(n,m)$ related? Since $g_r^*(G) \geq g_r(G)$ for every graph G, it easily follows that $g_r^*(n,m) \geq g_r(n,m)$. Furthermore, when $m = o(n^{\frac{r+1}{r}})$ (as $n \to \infty$) we claim that $g_r^*(n,m) = g_r(n,m) = o(n)$. Essentially this is due to the fact that we can remove a relatively small fraction of the vertices to produce a Turán subgraph from a possibly unbalanced complete r-partite subgraph. More precisely, let G be an

$(r+1)$-saturated graph on n vertices and with at least $t_r(n) - o(n^{\frac{r+1}{r}})$ edges. Then Theorem 1.1.1 grants us a complete r-partite subgraph G' with vertex partition, say, $V_1 \cup \cdots \cup V_r$ on $(1-o(1))n$ vertices. Now, no two distinct sets V_i, V_j can differ by more than $o(n)$ vertices. To see this, let $|V(G')| = n'$, and suppose that, say, $|V_1| = |V_2| + \alpha n'$ for some constant $\alpha > 0$. Then clearly (by removing the $\alpha n'$ extra vertices from V_1) we have

$$e(G') \leq t_r((1-\alpha)n') + \alpha n' \left(1 - \frac{1}{r}\right)(1-\alpha)n'.$$

On the other hand, it is not too hard to check, using the basic inequalities $t_r(n) \geq (1-\frac{1}{r})\binom{n}{2}$ and $t_r(n) \leq (1-\frac{1}{r})\frac{n^2}{2}$ together with the fact that $n' = (1-o(1))n$, that

$$t_r(n) - e(G') = \Omega(n^2).$$

Since $|V(G \setminus G')| = o(n)$ we have that $e(G', G \setminus G')$ and $e(G \setminus G')$ are both $o(n^2)$. Accordingly,

$$e(G) = e(G') + e(G', G \setminus G') + e(G \setminus G') \leq t_r(n) - O(n^2),$$

as $n \to \infty$, but this contradicts the edge condition on G. Therefore, these sets can differ by at most $o(n)$ vertices. Thus, for each of the $\binom{r}{2}$ pairs V_i, V_j we remove at most $o(n)$ vertices (and hence $o(n)$ vertices in total) to create a Turán subgraph on $(1-o(1))n$ vertices. This shows that in the edge range $m = o(n^{\frac{r+1}{r}})$

$$g_r(n,m) = g_r^*(n,m) = o(n),$$

and so there is little quantitative difference between these functions.

What happens in the 'critical window', i.e., in the edge range $m = O(n^{\frac{r+1}{r}})$? In order to understand the structure of $(r+1)$-saturated graphs with $e(G) = t_r(n) - Cn^{\frac{r+1}{r}}$, we wish to rule out degenerate situations where we may find large complete r-partite subgraphs, but they are rather lopsided, with some parts much smaller or larger than others. Thus, in this edge range, we prefer to study the function $g_r^*(n,m)$. The following theorem shows that there exist $(r+1)$-saturated graphs with at least $t_r(n) - Cn^{\frac{r+1}{r}}$ edges for which the largest Turán subgraph covers a vanishing fraction of the whole vertex set. In other words, we show that $g_r^*(n, Cn^{\frac{r+1}{r}})$ increases rapidly as C increases.

Theorem 1.1.3. *Let $r \geq 2$ be an integer and let $\delta > 0$. There exists a constant $C = C(r,\delta)$ such that, for n sufficiently large, there exists an n-vertex $(r+1)$-saturated graph G that contains no copy of $T_r(\delta rn)$ and $e(G) \geq t_r(n) - Cn^{\frac{r+1}{r}}$. In terms of the function g_r^*, we show that for any sufficiently large $D > 0$ (depending on r) we have*

$$g_r^*(n, Dn^{\frac{r+1}{r}}) \geq \left(1 - \frac{c' \log(Dr)}{D}\right) n,$$

for sufficiently large n and an absolute constant c'.

Let us see how the lower bound on g_r^* follows from the first statement. We will see that in the proof of Theorem 1.1.3 we can take $C = C(r,\delta) = 2^6 r^{-1} \delta^{-1} \log(2e/\delta)$. According to the above theorem, there is an $(r+1)$-saturated graph G on n vertices (n sufficiently large) with $e(G) \geq t_r(n) - Cn^{\frac{r+1}{r}}$ and no $T_r(\delta rn)$ subgraph. Thus we have

$$g_r^*(n, Cn^{\frac{r+1}{r}}) > (1 - \delta r)n.$$

We claim that

$$\delta r \leq \frac{2^6 \log(Cr)}{C},$$

which gives the desired lower bound on g_r^*. Indeed, note that by our choice of C we have $C > 1/r$ so the right-hand side is nonnegative. Then $C\delta r = 2^6 \log(2e/\delta)$, so we need that $2^6 \log(2e/\delta) \leq 2^6 \log(2^6 \delta^{-1} \log(2e/\delta))$. But this holds, since after dividing by the factor of 2^6 and taking exponentials, we obtain the inequality $2e/\delta \leq 2^6 \log(2e/\delta)/\delta$, which is clearly true.

Unfortunately, we do not have any good upper bounds on $g_r^*(n,m)$ in this range of m, and we leave this as an open problem.

1.1.1 Organization and Notation

This chapter is organized as follows. In Section 1.2, we prove our main result, Theorem 1.1.2. Roughly speaking, we first show that any K_{r+1}-free graph with many edges has a rather substantial r-partite subgraph. We then show that one can refine this resultant r-partite graph by making each bipartite graph between partition classes complete, while removing relatively few vertices. In Section 1.3, we provide the aforementioned constructions which exhibit the tightness of Theorems 1.1.1 and 1.1.2; in Section 1.4, we prove Theorem 1.1.3. Finally, in Section 1.5 we state some further questions.

Our notation is mostly standard (see, for example, [11]). For a subset $S \subseteq V(G)$ we denote by $N_G(S) = \bigcap_{v \in S} N_G(v)$ the common (or joint) neighborhood of S in G. We shall omit the subscript 'G' if the underlying graph is understood. If $X_1, \ldots, X_r$ are disjoint subsets of $V(G)$, we denote by $G[X_1, \ldots, X_r]$ the r-partite graph induced in G with vertex classes $X_1, \ldots, X_r$. We write $f \ll g$ to mean $f(n)/g(n) \to 0$ as $n \to \infty$. All other notation we need shall be introduced as necessary.

1.2 The Proof of Theorem 1.1.2

1.2.1 Preliminary lemmas

Let us now work towards establishing Theorem 1.1.2. For that we state and prove two lemmas, the second of which is the core of the proof. For the first lemma we use the following classical theorem of Andrásfai, Erdős, and Sós [8], although the precise value of the constant $\frac{3r-4}{3r-1}$ is unimportant for us; we only need that it is strictly less than the Turán density.

Theorem 1.2.1. *For $r \geq 2$ let G be a K_{r+1}-free graph on n vertices which is not r-partite. Then there is a vertex v of G with*

$$d(v) \leq \frac{3r-4}{3r-1}n.$$

We shall also use the following result of Brouwer [17], mentioned in the introduction. We include the proof of Tyomkyn and Uzzell [68] for completeness, which uses a tool known as 'Zykov symmetrization' that we now describe. Given a graph G and nonadjacent vertices $u, v \in V(G)$ let $Z_{u,v}(G)$ be the graph obtained from G by deleting all edges incident with u and adding all edges between u and $N(v)$. Recall that $\omega(G)$ denotes the *clique number* of G, the order of the largest complete subgraph in G. Then under this operation, the clique number and chromatic number cannot increase: we have $\omega(G-u) = \omega(Z_{u,v}(G))$ and $\chi(G-u) = \chi(Z_{u,v}(G))$. Therefore, $\omega(G) - 1 \leq \omega(Z_{u,v}(G)) \leq \omega(G)$ and similarly $\chi(G) - 1 \leq \chi(Z_{u,v}(G)) \leq \chi(G)$. Note that if $d(u) < d(v)$ then the number of edges increases under the operation $Z_{u,v}$, and if $d(u) = d(v)$ we can either apply $Z_{u,v}$ or $Z_{v,u}$ to G, preserving the above properties of ω and χ, while leaving the number of edges unchanged. We call $Z_{u,v}(G)$ an *increasing Zykov symmetrization* provided $d(u) \leq d(v)$.

The following proposition is due to Zykov [70], which he used in his proof of Turán's Theorem.

Proposition 1.2.2. *Let G be a K_{r+1}-free graph. Then there exists a sequence of increasing Zykov symmetrizations transforming G into a complete t-partite graph for some $t \leq r$.*

Proof. Call two vertices of G *twins* if they have the same neighborhood in G, and note that the application of $Z_{u,v}$ to G turns u into a twin of v. Since the relation 'is a twin of' is an equivalence relation, we can partition $V(G)$ into equivalence classes $T_1, \dots, T_q$ where $u, v \in T_i$ if and only if they are twins in G. Now, note that the bipartite graphs $G[T_i, T_j]$ for $i \neq j$ are either complete or empty. Indeed, if there is an edge xy with $x \in T_i$ and $y \in T_j$, then every vertex in T_j must be joined to x; i.e., x is joined to all of T_j, so all of T_i is joined to all of T_j. If $G[T_i, T_j]$ is empty, apply increasing Zykov symmetrizations until the classes T_i and T_j are merged into a single class. By doing this for every pair which induces an empty bipartite graph, we obtain a complete t-partite graph for some t. But since Zykov symmetrization does not increase the clique number, we must have $t \leq r$.

□

We are now in a position to give Tyomkyn and Uzzell's proof of Brouwer's theorem.

Theorem 1.2.3. *Let $r \geq 2$, $n \geq 2r+1$, and let G be an K_{r+1}-free, n-vertex graph. If $e(G) \geq t_r(n) - \lfloor \frac{n}{r} \rfloor + 2$, then G is r-partite.*

Proof. Let $G_r(n)$ denote the following graph. Take a copy of $T_r(n-1)$ with vertex partition $V_1, \dots, V_r$ and let $V(G_r(n)) = V_1 \cup \dots \cup V_r \cup \{u\}$ with $u \notin \bigcup_i V_i$. Suppose that V_1, V_2 are the smallest two classes in the partition and pick $v_1 \in V_1$ and $v_2 \in V_2$. Now join u to every vertex in $V_3 \cup \dots \cup V_r$, remove the edge v_1v_2, and join u only to v_1 and v_2 in $V_1 \cup V_2$. Note that $G_r(n)$ is K_{r+1}-free, *not* r-colorable, and has

$$e(G_r(n)) = t_r(n) - \lfloor n/r \rfloor + 1$$

edges. Thus $G_r(n)$ shows that Theorem 1.2.3 is tight. One can, in fact, classify the extremal examples for this theorem (there is more than one), and also consider the situation for $n \leq 2r$ (one must consider a different extremal example in this case). However, since we do not need these facts, we shall content ourselves with proving that if $n \geq 2r+1$ and G is an n-vertex K_{r+1}-free graph which is not r-colorable, then $e(G) \leq e(G_r(n))$.

To this end, by Proposition 1.2.2 we can apply a sequence of increasing Zykov symmetrizations so that the final graph is K_{r+1}-free and r-colorable. In particular, since $\chi(G) > r$, there exists an increasing symmetrization for which the chromatic number drops. Hence, without loss of generality, we may assume that $\chi(G) = r+1$ and for some u, v we have $\chi(Z_{u,v}(G)) = \chi(G - \{u\}) = r$. This means that G can be colored with $r+1$ colors so that u is the only vertex of its color. Hence write $V(G) = V_1 \cup \cdots \cup V_r \cup \{u\}$ where each V_i is independent and $u \notin \bigcup_i V_i$. Observe that u is joined to some vertex in each V_i (otherwise $\chi(G) = r$). Moreover, there must be at least one missing edge between some V_i and V_j (otherwise we obtain a K_{r+1} containing u).

Our graph now shares some similar structural properties with $G_r(n)$; we wish now to make some edge exchanges which make G look more like $G_r(n)$, and such that we do not decrease the total number of edges. Suppose u has neighbors $x \in V_i$, $y \in V_j$ for $i \neq j$. If $xy \notin E(G)$ and u has another neighbor, say $x' \in V_i$, then remove the edge ux and add xy. We claim that our graph remains K_{r+1}-free after this operation. Indeed, any copy of K_{r+1} must contain the vertex u, and therefore cannot contain x. Repeat this operation until there are two classes, say V_1, V_2 and two vertices $v_1 \in V_1$, $v_2 \in V_2$ such that u is joined only to v_1 and v_2, and v_1 and v_2 are nonadjacent (note that this process must terminate since the degree of u decreases at each step). Call the resulting graph G'. Now form the graph G'' from G' by adding all missing edges between each V_i and V_j (except v_1v_2), and between u and $V_3 \cup \cdots \cup V_r$. Then G'' is $(r+1)$-partite, K_{r+1}-free, and its number of edges is maximized whenever each vertex class is equal as possible. In other words,

$$e(G) = e(G') \leq e(G'') \leq e(G_r(n)),$$

as claimed.

$\square$

Here, then, is our first lemma, which grants us a sizable induced r-partite subgraph.

We remark that a lemma of this type is not new and appears in a similar form in [68].

Lemma 1.2.4. *For $r \geq 2$ there is a constant d_r, depending only on r, such that the following holds. Let $n \geq 4r$ and $0 \leq \varepsilon \leq (30r^3)^{-1}$. If G is an n-vertex K_{r+1}-free graph with $e(G) \geq t_r(n) - \varepsilon n^2$, then there is a subset $T \subseteq V(G)$ with $|T| \leq d_r \varepsilon n$ such that $G - T$ is r-partite.*

Proof. If $\varepsilon < (2rn)^{-1}$, then $e(G) > t_r(n) - \frac{n}{2r} \geq t_r(n) - \lfloor \frac{n}{r} \rfloor + 1$, where the second inequality follows by our assumption that $n \geq 4r$. Therefore by Theorem 1.2.3, G is r-partite, and there is nothing to prove. Accordingly, we may assume $\varepsilon \geq (2rn)^{-1}$.

Set $G_1 = G$. Suppose that $G_1, \ldots, G_i$ have been defined for some $i \in [n]$. If G_i is not r-partite then pick a vertex $v_i \in V(G_i)$ with $d_{G_i}(v_i) \leq \frac{3r-4}{3r-1}|G_i|$ according to Theorem 1.2.1. Set $G_{i+1} = G_i - v_i$. Suppose this process terminates at stage $t \in [n]$. Then $G_{t+1} = G - \{v_1, \ldots, v_t\}$ is r-partite. We claim that $t \leq d_r \varepsilon n$ for some constant d_r depending only on r. This follows from a simple calculation. Indeed as $e(G_{i+1}) \leq \frac{r-1}{2r}(n-i)^2$ holds for every $i \in [t]$, by Turán's theorem we have

$$\begin{aligned} e(G) &\leq \frac{3r-4}{3r-1}\Big(n + (n-1) + \cdots + (n-i+1)\Big) + \frac{r-1}{2r}(n-i)^2 \\ &= \frac{3r-4}{3r-1}\left(ni - \binom{i}{2}\right) + \frac{r-1}{2r}(n-i)^2, \end{aligned}$$

and using the lower bound on $e(G)$ we obtain

$$t_r(n) - \frac{r-1}{2r}(n-i)^2 + \frac{3r-4}{3r-1}\binom{i}{2} \leq \frac{3r-4}{3r-1}ni + \varepsilon n^2. \tag{1.1}$$

Further, using the lower bound $t_r(n) \geq (1-1/r)\binom{n}{2}$ applied to (1.1) and rearranging yields the equivalent inequality

$$i\left(1 - \frac{i}{2n} - \frac{r(3r-4)}{2n}\right) - \frac{1}{2}(r-1)(3r-1) \leq r(3r-1)\varepsilon n,$$

which is easily shown to fail if $i = 10r^2(3r-1)\varepsilon n$ when $(2rn)^{-1} \leq \varepsilon \leq (30r^3)^{-1}$. Since the resulting function in (1.1) is quadratic in i, it is indeed enough to demonstrate that it fails for one value. Accordingly, $t < 10r^2(3r-1)\varepsilon n$ as claimed. $\square$

The next lemma is the heart of the proof of our main theorem. Before stating it we introduce some notation and a bit of terminology. If G is an r-partite graph with vertex partition $V_1, \ldots, V_r$, then we denote by $\widetilde{G}[V_1, \ldots, V_r]$ the *r-partite complement of* G with respect to the partition $V_1, \ldots, V_r$. In other words $\widetilde{G}[V_1, \ldots, V_r]$ has vertex set $V_1 \cup \cdots \cup V_r$ and its edges are precisely the non-edges of G which join two vertices belonging to distinct vertex classes of $V_1, \ldots, V_r$. Often we simply speak of *the r-partite complement* in the case that the vertex partition we are using is clear from context, and we shall simply write $\widetilde{G}$. We say that a subset $S \subseteq V(G)$ of the vertices of a graph G *covers* an edge e if at least one of the endpoints of e lies in S. Further, we let $I_G(S)$ denote the collection of edges of G covered by S. An *r-saturating edge* in G is an edge of the complement $\overline{G}$ the addition of which creates a copy of K_r in G. If $X, Y \subseteq V(G)$ are subsets of vertices, then we say that a non-edge e is an *r-saturating (X,Y) edge* if it is r-saturating with one endpoint in X and the other in Y. A *K_r-matching* in a graph G is a collection of vertex disjoint copies of K_r in G. Lastly, before stating and proving the lemma, let us collect a simple observation that will be of use.

Observation 1.2.5. *Suppose that G is a bipartite graph with vertex classes V_1 and V_2 with $e(G) = \alpha|V_1||V_2|$, where $\alpha \in [0,1]$. Then for any $1 \leq t \leq |V_2|$ there is a subset $W \subseteq V_2$ of size t such that the induced graph on $V_1 \cup W$ has at least $\alpha|V_1|t$ edges.*

Proof. This assertion follows from a simple averaging argument. For $Y \subseteq V_2$ let $e(V_1, Y)$ denote the number of edges of G with an endpoint in Y. Then

$$\sum_{Y \in V_2^{(t)}} e(V_1, Y) = e(G)\binom{|V_2|-1}{t-1} = \alpha|V_1|t\binom{|V_2|}{t},$$

so there exists a subset $W \in V_2^{(t)}$ with $e(V_1, W) \geq \alpha |V_1| t$. $\square$

Lemma 1.2.6. *Let $r \geq 2$ be an integer and let G be a K_r-free, r-partite graph with vertex classes A, $B, X_1 \ldots, X_{r-2}$. Then the following statements hold.*

1. *There is a subset $R \subseteq A \cup B$ that covers all r-saturating (A, B) edges in G and*

$$|I_{\widetilde{G}}(R)| \geq c_r |R|^{\frac{r}{r-1}},$$

   for some constant $c_r > 0$ depending only on r.

2. *Suppose that $t \geq 1$ is an integer with $r - t \geq 2$, that $E \subseteq E_{\widetilde{G}}(A, B)$ is a collection of non-edges between A, B, and that there exist K_{r-t}-free subgraphs $H_1, \ldots, H_s \subseteq G$ such that every element of E is $(r-t)$-saturating in at least one of the graphs $H_1, \ldots, H_s$. Then there exists a set $R' \subseteq A \cup B$ covering every element of E with*

$$|I_{\widetilde{G}}(R')| \geq c'_{r,t} s^{-\frac{1}{r-t-1}} |R'|^{\frac{r-t}{r-t-1}},$$

   where $c'_{r,t}$ is a constant depending only on r, t.

Proof. We prove these two statements simultaneously by induction on r. The case $r = 2$ is trivial: G must be empty. The first part holds by simply choosing the smaller of the two parts of the bipartite graph G and the second part of the statement is vacuous as there is no appropriate choice for t.

So, assuming that the result holds for $r - 1 \geq 2$, we prove it for r. To this end, let G be a K_r-free, r-partite graph with vertex sets A, B, $X_1, \ldots, X_{r-2}$. We start with the proof of Part 2 as we shall need it to prove Part 1.

Proof of Part 2: Suppose we are given a collection E of non-edges between A, B and subgraphs $H_1, \ldots, H_s$ satisfying the requirements of the lemma. Start by enumerating the collection of subgraphs

$$\left\{H_i\left[A\cup B\cup X_{i_1}\cup\cdots\cup X_{i_{r-t-2}}\right] : i\in[s], 1\le i_1<\cdots<i_{r-t-2}\le r-2\right\}$$

by $H'_1,\ldots,H'_{s'}$, where $s'=\binom{r-2}{r-t-2}s$ (if $t=r-2$, then we are just listing the subgraphs $H_i[A\cup B]$ for $i=1,\ldots,s$). We now iteratively apply induction inside each of the graphs $H'_1,\ldots,H'_{s'}$: at each stage we remove a set granted by the induction hypothesis before moving to the next graph in the enumeration.

We shall define a sequence of disjoint subsets $R_1,\ldots,R_{s'}$ of $A\cup B$ and a sequence of subgraphs $G_1,\ldots,G_{s'+1}$ of G with the following properties:

1. $G_1=G$ and $G_{i+1}=G_i-R_i$ for all $i\ge 1$.
2. $|I_{\widetilde{G}_i}(R_i)|\ge c_{r,t}|R_i|^{\frac{r-t}{r-t-1}}$ for each $i\ge 1$, where $c_{r,t}$ is the constant given by the induction hypothesis of the lemma (here, the r-partite complement $\widetilde{G}_i$ is with respect to the 'obvious' r-partition of G_i).
3. Every non-edge of E is covered by $R_1\cup\cdots\cup R_{s'}$.

Suppose that, for $i\in[s']$, the graphs $G_1,\ldots,G_i$ have been defined. Apply the induction hypothesis of Lemma 1.2.6 to the $(r-t)$-partite, K_{r-t}-free graph $H'_i\cap G_i$ to find a set $R_i\subseteq V(H'_i\cap G_i)\cap(A\cup B)$ with $|I_{\widetilde{G}_i}(R_i)|\ge c_{r,t}|R_i|^{\frac{r-t}{r-t-1}}$ that covers all $(r-t)$-saturating (A,B) edges in H'_i. Finally set $G_{i+1}=G_i-R_i$. To check that every non-edge of E is covered by $R_1\cup\cdots\cup R_{s'}$, simply recall that we assumed that every non-edge of E is $(r-t)$-saturating in one of the subgraphs $H_1,\ldots,H_s$ and therefore $(r-t)$-saturating in one of the subgraphs $H'_1,\ldots,H'_{s'}$. Thus, a non-edge $e\in E$ is $(r-t)$-saturating in some H'_j for some $j\in[s']$, and so it will be covered by one of $R_1,\ldots,R_j$. That is, it will be covered in stage j, if it has not been covered already.

To finish the proof of Part 2 of the lemma, we write $R'=R_1\cup\cdots\cup R_{s'}$. Noting that the sets $R_1,\ldots,R_{s'}$ are pairwise disjoint, we apply Hölder's inequality to obtain

$$|R'| = \sum_{i=1}^{s'} |R_i| \leq s'^{\frac{1}{r-t}} \left(\sum_{i=1}^{s'} |R_i|^{\frac{r-t}{r-t-1}}\right)^{\frac{r-t-1}{r-t}},$$

and therefore

$$s'^{-\frac{1}{r-t-1}} |R'|^{\frac{r-t}{r-t-1}} \leq \sum_{i=1}^{s'} |R_i|^{\frac{r-t}{r-t-1}}.$$

Now, since the sets of edges $\{I_{\widetilde{G}_i}(R_i)\}_{i\in[s']}$ are pairwise disjoint (as the sets $R_1,\ldots,R_{s'}$ are pairwise disjoint, and we remove R_i from G_i to define G_{i+1}) we may estimate

$$\begin{aligned}
|I_{\widetilde{G}}(R')| = \sum_{i=1}^{s'} |I_{\widetilde{G}_i}(R_i)| &\geq \sum_{i=1}^{s'} c_{r,t} |R_i|^{\frac{r-t}{r-t-1}} \\
&\geq c_{r,t} s'^{-\frac{1}{r-t-1}} |R'|^{\frac{r-t}{r-t-1}} \\
&\geq c'_{r,t} s^{-\frac{1}{r-t-1}} |R'|^{\frac{r-t}{r-t-1}},
\end{aligned}$$

where $c'_{r,t}$ is a constant depending only on r,t. Note that the first equality holds since the sets $I_{\widetilde{G}_i}(R_i), i\in[s']$ are pairwise disjoint and the sum $\sum_{i=1}^{s'} |I_{\widetilde{G}_i}(R_i)|$ counts edges in $\widetilde{G}$ covered by R'. This completes the proof of Part 2 of Lemma 1.2.6.

To prove the first part we use the second part along with an extra ingredient.

Proof of Part 1 : We may assume that there is *some* saturating (A,B)-edge, otherwise we are trivially done with the choice of $R=\emptyset$. So, let $\mathscr{M}$ be a K_{r-2}-matching of maximum size in the graph $G[X_1,\ldots,X_{r-2}]$ and let Y denote the collection of vertices contained in a clique of $\mathscr{M}$. Note that $\mathscr{M}$ is nonempty as there is some saturating (A,B)-edge, and put $L=|\mathscr{M}|$ so that $|Y|=(r-2)L>0$. For each $y\in Y$, let $G(y)$ be the $(r-1)$-partite graph induced on the neighborhood of y in G with vertex classes $N(y)\cap A, N(y)\cap B$ along with $N(y)\cap X_i$ for $y\notin X_i, i\in[r-2]$. Our first claim asserts that we may assume there are many non-edges between Y and either A or B.

Claim 1.2.7. *There are either at least $\frac{1}{4(r-2)}|A||Y|$ non-edges between Y and A, or at least $\frac{1}{4(r-2)}|B||Y|$ non-edges between Y and B.*

Proof. For each $K \in \mathscr{M}$ and $S \subseteq V(G)$ we denote by $d_S(K)$ the number of vertices of S joined to every vertex of K, so that $d_S(K) = |N_G(K) \cap S|$. We may assume that, for every $K \in \mathscr{M}$, either $d_A(K) \leq \frac{1}{2}|A|$ or $d_B(K) \leq \frac{1}{2}|B|$. Indeed, suppose that there is $K \in \mathscr{M}$ with $d_A(K) > \frac{1}{2}|A|$ and $d_B(K) > \frac{1}{2}|B|$. As G is K_r-free we must then count more than $\frac{1}{4}|A||B|$ non-edges between A and B. Setting R to be the smaller of A and B, we see that trivially R covers all r-saturating (A,B) edges and

$$|I_{\widetilde{G}}(R)| > \frac{1}{4}|A||B| \geq \frac{1}{4}|R|^2,$$

so we are done (with room to spare). Therefore, we may assume that for every $K \in \mathscr{M}$ either $d_A(K) \leq \frac{1}{2}|A|$ or $d_B(K) \leq \frac{1}{2}|B|$.

Write $\mathscr{M} = \mathscr{M}_A \cup \mathscr{M}_B$, where $\mathscr{M}_A$ are those $K \in \mathscr{M}$ which satisfy $d_A(K) \leq \frac{1}{2}|A|$ and $\mathscr{M}_B$ are those that satisfy $d_B(K) \leq \frac{1}{2}|B|$. Then, without loss of generality, we have $|\mathscr{M}_A| \geq \frac{1}{2}|\mathscr{M}|$. Now since each $K \in \mathscr{M}_A$ sends at least $\frac{1}{2}|A|$ non-edges to A and since each clique in $\mathscr{M}$ is vertex-disjoint, we have that there are at least $\frac{1}{4}|A||\mathscr{M}| = \frac{1}{4(r-2)}|A||Y|$ non-edges between Y and A. $\square$

Now, observe that, by the maximality of $\mathscr{M}$, every r-saturating (A,B) edge is $(r-1)$-saturating in one of the graphs $\{G(y)\}_{y \in Y}$. Hence we may apply the bound in Part 2 of the lemma to obtain a set R_0 which covers every r-saturating (A,B) edge and

$$|I_{\widetilde{G}}(R_0)| \geq c'_{r,1}(r-2)^{-\frac{1}{r-2}} L^{-\frac{1}{r-2}} |R_0|^{\frac{r-1}{r-2}}. \tag{1.2}$$

However, this bound is not useful if L is too large. In order to deal with this issue we shall randomly augment R_0 with a set R'_0 of $|R_0|$ vertices. The resulting set $R = R_0 \cup R'_0$

will only be a factor of two larger than R_0 but will cover 'many' edges of $\widetilde{G}$ — enough to achieve a better lower bound on $|I_{\widetilde{G}}(R)|$.

To this end, note that by Claim 1.2.7 we may assume that, without loss of generality, there are at least $\frac{1}{4(r-2)}|A||Y|$ non-edges between Y and A. Further, we may assume that $|R_0| \le |A|$. Indeed, suppose otherwise that $|R_0| > |A|$. If $|A||\mathscr{M}| \ge |A|^{\frac{r}{r-1}}$, we are done by choosing $R = A$, since then $|I_{\widetilde{G}}(A)| \ge \frac{1}{4}|A||\mathscr{M}| \ge \frac{1}{4}|A|^{\frac{r}{r-1}}$. Otherwise, $L = |\mathscr{M}| < |A|^{\frac{1}{r-1}} < |R_0|^{\frac{1}{r-1}}$, and using (1.2) yields $|I_{\widetilde{G}}(R_0)| \ge c'_r|R_0|^{\frac{r}{r-1}}$, so we are done with the choice $R = R_0$.

Hence, assuming that $|R_0| \le |A|$, by Observation 1.2.5, one can find a subset $R'_0 \subseteq A$ of size $|R_0|$ such that the number of non-edges between R'_0 and Y is at least $\frac{1}{4(r-2)}|R_0||Y| = \frac{1}{4}|R_0|L$.

We now set $R = R_0 \cup R'_0$ and claim that R is our desired set. First note that R covers all r-saturating (A,B) edges in G, as R_0 already does. To count the total number of non-edges covered by R, we note that $|R| \le 2|R_0|$, and so we have (using (1.2))

$$\begin{aligned}
2|I_{\widetilde{G}}(R)| &\ge |I_{\widetilde{G}}(R_0)| + |I_{\widetilde{G}}(R'_0)| \\
&\ge c'_{r,1}(r-2)^{-\frac{1}{r-2}}L^{-\frac{1}{r-2}}|R_0|^{\frac{r-1}{r-2}} + \frac{1}{4}|R_0|L \\
&\ge c'L^{-\frac{1}{r-2}}|R|^{\frac{r-1}{r-2}} + \frac{1}{8}|R|L, \qquad (1.3)
\end{aligned}$$

where $c' = c'_{r,1}2^{-\frac{r-1}{r-2}}(r-2)^{-\frac{1}{r-2}}$. A simple analysis reveals that the quantity on the right-hand side of (1.3) is minimized in L if $L = (8c'/(r-2))^{\frac{r-2}{r-1}}|R|^{\frac{1}{r-1}}$. Substituting this value of L back into (1.3) yields

$$|I_{\widetilde{G}}(R)| \ge c_r|R|^{\frac{r}{r-1}},$$

where c_r is a constant depending only on r. $\square$

We remark that a more careful reading of the proof shows that we may take $c_r \ge \frac{1}{2^{r+2}}$.

Therefore, we have $c_{r,t} \geq \frac{1}{2^{r+2-t}}$ and $c'_{r,t} = c_{r,t} \binom{r-2}{r-t-2}^{-\frac{1}{r-t-1}} \geq \frac{1}{(r-2)2^{r+2-t}}$.

1.2.2 Finishing the proof

We can now proceed to finish the proof of Theorem 1.1.2.

Proof (of Theorem 1.1.2). Let r, n be integers with $r \geq 2$ and $n \geq 900r^6$, and suppose that G is an n-vertex $(r+1)$-saturated graph with $e(G) \geq t_r(n) - m$. For notational convenience we shall write $m = \varepsilon n^2$. Thus we must find a complete r-partite subgraph of G on at least $(1 - C_r \varepsilon n^{\frac{r-1}{r}})n$ vertices, for some constant C_r depending only on r. We shall additionally insist that $C_r \geq 1$. The result is then trivial if $\varepsilon > n^{-\frac{r-1}{r}}$ and so we may assume that $\varepsilon \leq n^{-\frac{r-1}{r}}$. Since $n \geq (30r^3)^2$ we have that $\varepsilon \leq (30r^3)^{-1}$, so we may apply Lemma 1.2.4 to obtain a subset $T \subseteq V(G)$ such that $|T| \leq d_r \varepsilon n$ and $G - T$ is r-partite. Let the vertex classes of $G - T$ be $V_1, \ldots, V_r$. We now simply apply Part 2 of Lemma 1.2.6 to common neighborhoods of appropriate subsets of T. But before we do this we need a bound on $e(\widetilde{G}[V_1, \ldots, V_r])$, the number of non-edges between the parts $V_1, \ldots, V_r$, which is the content of the following claim.

Claim 1.2.8. $e(\widetilde{G}[V_1, \ldots, V_r]) \leq (d_r + 1)\varepsilon n^2$.

Proof. First note that if $|T| = 0$, then G is r-partite and $e(\widetilde{G}[V_1, \ldots, V_r]) = 0$ since G is $(r+1)$-saturated. So, we may assume that $|T| \geq 1$. In this case, the number of non-edges $e\left(\overline{G}\right)$ satisfies $e\left(\overline{G}\right) \leq \binom{n}{2} - t_r(n) + \varepsilon n^2$, and also

$$e\left(\overline{G}\right) \geq \sum_{i=1}^{r} \binom{|V_i|}{2} + e(\widetilde{G}[V_1, \ldots, V_r]) \geq r \binom{\frac{n-|T|}{r}}{2} + e(\widetilde{G}[V_1, \ldots, V_r]),$$

by convexity of the function $x \mapsto \binom{x}{2}$. By using the estimate $t_r(n) \geq \left(1 - \frac{1}{r}\right)\binom{n}{2}$, combining the lower and upper bounds on $e\left(\overline{G}\right)$, and rearranging, we get

$$e(\widetilde{G}[V_1,\dots,V_r]) \le \varepsilon n^2 + \frac{1}{r}\binom{n}{2} - r\binom{\frac{n-|T|}{r}}{2} < \varepsilon n^2 + \frac{r-1}{2r}n + \frac{n|T|}{r} \tag{1.4}$$

$$= \varepsilon n^2 + 2n|T|\left(\frac{r-1}{4r|T|} + \frac{1}{2r}\right). \tag{1.5}$$

Now, if $|T| \ge r/2$, then (1.5) is at most $\varepsilon n^2 + \frac{2n|T|}{r}$, and we are done. If $|T| < r/2$, then by (1.4) we have $e(\widetilde{G}[V_1,\dots,V_r]) < \varepsilon n^2 + n = \left(1+\frac{1}{\varepsilon n}\right)\varepsilon n^2$. But clearly $\frac{1}{\varepsilon n} \le d_r$, as otherwise $|T| < 1$. Hence, the desired bound on $e(\widetilde{G}[V_1,\dots,V_r])$ holds. $\square$

For $t \in [r-1]$ let $\mathscr{C}_t$ denote the collection of copies of K_t contained in $G[T]$, the graph induced on T. We say a non-edge e is of *type t* if it lies between two of the classes $V_1,\dots,V_r$, and the addition of e to G creates a K_{r+1} with exactly t vertices in T. Since G is $(r+1)$-saturated and $G[V_1,\dots,V_r]$ is a K_{r+1}-free graph, every non-edge between two of the classes $V_1,\dots,V_r$ is of type t for some $t \in [r-1]$. For $t \in [r-1]$ we let E_t denote the collection of type t non-edges.

Set $V = V_1 \cup \cdots \cup V_r$ and define $\mathscr{G}_t = \{G[N(K)\cap V] : K \in \mathscr{C}_t\}$ for $t \in [r-1]$. For each $i \ne j \in [r]$, we show that one can make the induced bipartite graph $G[V_i,V_j]$ complete by removing a relatively small number of vertices. Doing this in succession for each of the $\binom{r}{2}$ pairs V_i, V_j with $i \ne j$ then yields a complete r-partite subgraph.

So fix $i \ne j \in [r]$ and note that for each $t \in [r-1]$, each graph in the collection $\mathscr{G}_t$ is K_{r+1-t}-free and every (V_i,V_j) non-edge of E_t is $(r+1-t)$-saturating in one of the graphs of $\mathscr{G}_t$. So for each $t \in [r-1]$ we may invoke Part 2 of Lemma 1.2.6 to obtain a set $S_t(i,j) \subseteq V_i \cup V_j$ that covers every $(r+1)$-saturating (V_i,V_j) edge of type t and

$$\left|I_{\widetilde{G}[V_1,\dots,V_r]}(S_t(i,j))\right| \ge c'_{r+1,t}|\mathscr{C}_t|^{-\frac{1}{r-t}}|S_t(i,j)|^{\frac{r+1-t}{r-t}}.$$

Moreover, by Claim 1.2.8 we have $|I_{\widetilde{G}[V_1,\dots,V_r]}(S_t(i,j))| \le e(\widetilde{G}[V_1,\dots,V_r]) \le (d_r+1)\varepsilon n^2$, and using the bound $|\mathscr{C}_t| \le |T|^t \le (d_r \varepsilon n)^t$, we obtain

$$
\begin{aligned}
|S_t(i,j)|^{\frac{r+1-t}{r-t}} &\le (c'_{r+1,t})^{-1}(d_r+1)\varepsilon n^2 |\mathscr{C}_t|^{\frac{1}{r-t}} \\
&\le (c'_{r+1,t})^{-1}(d_r+1) d_r^{\frac{t}{r-t}} \varepsilon^{\frac{r}{r-t}} n^{\frac{2r-t}{r-t}}.
\end{aligned}
$$

It follows that

$$
|S_t(i,j)| \le C_{r,t}\left(\varepsilon^{\frac{r}{r+1-t}} n^{\frac{r-1}{r+1-t}}\right) n,
$$

where $C_{r,t}$ is a constant depending only on r,t, for each $t \in [r-1]$, and $i \ne j \in [r]$.

As every edge between the parts $V_1, \dots, V_r$ is of type t for some $t \in [r-1]$, we conclude that the set $S = \bigcup_{t=1}^{r-1} \bigcup_{i \ne j \in [r]} S_t(i,j)$ covers every non-edge between the parts $V_1, \dots, V_r$. It follows that $G - S - T$ is a complete r-partite graph. To bound $|S|$ recall that $\varepsilon \le n^{-\frac{r-1}{r}}$. Then we have

$$
\begin{aligned}
|S| &\le \sum_{t=1}^{r-1} \sum_{i \ne j \in [r]} C_{r,t}\left(\varepsilon^{\frac{r}{r+1-t}} n^{\frac{r-1}{r+1-t}}\right) n \\
&\le (r-1)\binom{r}{2} \max_{t \in [r-1]} \{C_{r,t}\} \left(\varepsilon n^{\frac{r-1}{r}}\right) n \\
&\le C'_r \left(\varepsilon n^{\frac{r-1}{r}}\right) n,
\end{aligned}
$$

where the constant C'_r depends only on r. It is here that we have used the condition $\varepsilon \le n^{-\frac{r-1}{r}}$, since this implies that the dominating term in the sum above is the one with $t = 1$. Hence we have found a complete r-partite subgraph on

$$
n - |S| - |T| \ge n - C'_r \left(\varepsilon n^{\frac{r-1}{r}}\right) n - d_r \varepsilon n \ge \left(1 - C_r \varepsilon n^{\frac{r-1}{r}}\right) n
$$

vertices, for some constant C_r. This completes the proof. $\square$

1.3 Constructions

The aim of this section is to describe a family of constructions that demonstrate the optimality of Theorem 1.1.1. Indeed, the construction described in the next two subsections will show that in the large edge range $(\frac{r-1}{r}+\varepsilon) \leq m \leq n^{\frac{r+1}{r}}$ we have

$$g_r(n,m) \geq c_{r,\varepsilon} m n^{-1/r},$$

for some constant $c_{r,\varepsilon}$ depending on r, ε.

1.3.1 Removed edges

We begin by inductively constructing a family of auxiliary graphs $G_{r,s}$, for each $r,s \in \mathbb{N}$, $r,s \geq 2$. It is useful to keep in mind that the edges of the r-partite graph $G_{r,s}$ record edges to be removed from a later graph. First let us introduce a family of r-partite graphs $G_{r,s_1,s_2,\ldots,s_{r-1}}$ for which $G_{r,s}$ will be a special case.

Construction of $G_{r,s_1,\ldots,s_{r-1}}$ *:* Let $s_1,\ldots,s_{r-1} \geq 2$ be integers. We define a sequence of graphs $G_{2,s_1}, G_{3,s_1,s_2}, \ldots, G_{r,s_1,\ldots,s_{r-1}}$ inductively, where $G_{i,s_1,\ldots,s_{i-1}}$ will be an i-partite graph. First, we define G_{2,s_1} to be the complete bipartite graph K_{s_1,s_1}. Now let $2 \leq t \leq r-1$ and assume that we have defined the t-partite graph $G_{t,s_1,\ldots,s_{t-1}}$. We define $G_{t+1,s_1,\ldots,s_t}$ as follows. Let $H_1,\ldots,H_{s_t}$ be vertex disjoint copies of $G_{t,s_1,\ldots,s_{t-1}}$ and suppose H_p has vertex classes $A_1^p,\ldots,A_t^p$, for each $p \in [s_t]$. Define $G_{t+1,s_1,\ldots,s_t}$ to be the $(t+1)$-partite graph with the first t vertex classes defined as $A_i := A_i^1 \cup \cdots \cup A_i^{s_t}$, for $i \in [t]$, and with the $(t+1)$st vertex class defined as a collection of new vertices $A_{t+1} = \{x_1,\ldots,x_{s_t}\}$. We define the edge set $E(G_{t+1,s_1,\ldots,s_t}) = \bigcup_{p=1}^{s_t} E(H_p) \cup \big\{x_p y : y \in H_q, p,q \in [s_t], p \neq q\big\}$.

Now for simplicity we let $G_{r,s} := G_{r,2s,s,\ldots,s}$, for $s \geq 2$. The following proposition records several useful properties of our family of graphs $G_{r,s}$.

Proposition 1.3.1. *The graph $G_{r,s}$ has the following properties.*

1. *$G_{r,s}$ is r-partite with vertex partition $A_1 \cup \cdots \cup A_r$ (and hence it makes sense to consider the r-partite complement of $G_{r,s}$ with respect to this partition).*

2. *The r-partite complement $\widetilde{G}_{r,s}$ is K_r-free.*

3. *For each $i \in [r]$, there is a copy of K_{r-1} in $\widetilde{G}_{r,s} \setminus A_i$.*

4. *Every edge between two different vertex classes of $G_{r,s}$ is r-saturating in $\widetilde{G}_{r,s}$.*

5. $|G_{r,s}| = \sum_{i=1}^{r-2} s^i + 4s^{r-1} = \frac{s}{s-1}(4s^{r-1} - 3s^{r-2} - 1) \leq 4\frac{s^r}{s-1}$.

6. $e(G_{r,s}) \leq 4(r-1)s^r$.

7. *The size of the largest two vertex classes is $2s^{r-1}$.*

8. *All other vertex classes have size at most s^{r-2}.*

9. *There is a matching between the largest two vertex classes of $G_{r,s}$.*

10. *Any independent set in $G_{r,s}$ has at most $|G_{r,s}| - 2s^{r-1}$ vertices.*

Proof. We shall use induction on r. The base case $r = 2$ is trivial. Suppose the assertions hold for $r \geq 2$. Clearly $G_{r+1,s}$ is $(r+1)$-partite and $\widetilde{G}_{r+1,s}$ is K_{r+1}-free. To show Part 3, suppose first that $i = r+1$. By induction hypothesis there is a copy of K_{r-1} in $\widetilde{H}_1 \setminus A_1 = \widetilde{H}_1 \setminus A_1^1$ which together with any $x \in \widetilde{H}_2 \cap A_1$ form a copy of K_r in $\widetilde{G}_{r+1,s} \setminus A_{r+1}$. The argument is very similar for the case when $i \in [r]$. To show Part 4, notice that the only edges between vertex classes in $G_{r+1,s}$ are either inside H_p or between x_p and H_q, for some $p, q \in [s]$, $p \neq q$. If we add an edge to $\widetilde{G}_{r+1,s}$ (which corresponds to removing that edge from $G_{r+1,s}$) of the former type, the assertion holds simply by induction. If we add an edge $x_p y$ with $y \in A_i$, $i \in [r]$, of the latter type, first observe that it follows from Part 3 of the induction hypothesis that $\widetilde{H}_p \setminus A_i$ contains a copy of K_{r-1}, say K. Hence, both x_p

and y are joined to every vertex in K, thus forming a K_{r+1} in $\widetilde{G}_{r+1,s}$. The number of vertices satisfies the relation $|G_{r+1,s}| = s + s|G_{r,s}|$ while $|G_{2,s}| = 4s$, and thus the claim follows. The number of edges satisfies the recurrence
$e(G_{r+1,s}) = s \cdot e(G_{r,s}) + s(s-1)|G_{r,s}| \le s \cdot e(G_{r,s}) + s(s-1)4\frac{s^r}{s-1} = s \cdot e(G_{r,s}) + 4s^{r+1}$ so, by induction, $e(G_{r+1,s}) \le 4(r-1)s^{r+1} + 4s^{r+1} = 4rs^{r+1}$. Parts 7,8,9 follow immediately by induction. Finally, to argue Part 10, simply notice that for each $p \in [s]$, by induction, there is no independent set in H_p with more than $|H_p| - 2s^{r-1}$ vertices. Therefore, from disjointness of the H_p's, any independent set in $G_{r+1,s}$ has at most $|G_{r+1,s}| - 2s^r$ vertices. $\square$

1.3.2 The final construction

We can now proceed to construct a family of graphs $H_{r,s,t}(n)$ that will demonstrate the tightness of Theorem 1.1.2. We let $H_1, \dots, H_t$ be vertex disjoint copies of $G_{r,s}$ with vertex partitions $H_p = A_1^p \cup \dots \cup A_r^p$ for each $p \in [t]$. We now augment the vertex set of the H_p's to be the vertex set for our G. First note that since $n \ge 4s^{r-1}tr + t$, we can find $\ell_1, \dots, \ell_r \in \mathbb{N}$, so that for each $i \in [r]$ we have $\sum_{p=1}^t |A_i^p| + \ell_i \in \left\{ \lfloor \frac{n-t}{r} \rfloor, \lceil \frac{n-t}{r} \rceil \right\}$ and $\sum_{i=1}^r \left(\sum_{p=1}^t |A_i^p| + \ell_i \right) = n - t$. Note that as n is large enough, we may assume that $\ell_1, \dots, \ell_r > 0$. We now define the sets $A_1, \dots, A_r$ as

$$A_i = A_i^1 \cup \dots \cup A_i^t \cup Y_i,$$

for $i \in [r]$, where Y_i is a collection of ℓ_i new vertices. We additionally define $A_{r+1} = \{x_1, \dots, x_t\}$ as a collection of t new vertices and finally set $V(G) = \bigcup_{i=1}^{r+1} A_i$.

We define the edge set as follows: the vertex x_p is joined to $V(H_p)$, for each $p \in [t]$, and for $i, j \in [r]$, $x \in A_i, y \in A_j$, xy is an edge if and only if $i \neq j$ and the edge xy is *not* in any of the graphs $H_1, \dots, H_t$. We then add a maximal set of edges among A_{r+1} that leaves the graph K_{r+1}-free. That is, we first define a graph G' by $V(G') = V(G)$ and

$$E(G') = \{x_p y : y \in V(H_p), p \in [t]\} \cup \left\{xy : x \in A_i, y \in A_j, 1 \le i < j \le r\right\} \setminus \bigcup_{p=1}^{t} E(H_p),$$

and then augment the edge set to form $E(G)$:

$$E(G) = E(G') \cup X,$$

where $X \subseteq A_{r+1}^{(2)}$ is maximal in the sense that adding any further edge of $A_{r+1}^{(2)}$ will yield a K_{r+1} in G. Call this final graph $H_{r,s,t}(n)$.

The following Proposition shows that $H_{r,s,t}(n)$ has all of the properties that are of interest to us. Before proceeding, let us note the following easy observation.

Observation 1.3.2. *For integers $r, t \le n$ with $r \ge 2$ we have*

$$t_r(n-t) \ge t_r(n) - (1 - 1/r)tn.$$

Proof. If x is a vertex of minimum degree in $T_r(n)$, then $T_r(n) - x = T_r(n-1)$ and so $t_r(n) = t_r(n-1) + \delta(T_r(n))$. Iterating this fact yields

$$t_r(n) = t_r(n-t) + \sum_{j=0}^{t-1} \delta(T_r(n-j)) \le t_r(n-t) + t \cdot \delta(T_r(n)) \le t_r(n-t) + (1-1/r)tn,$$

as claimed. □

Proposition 1.3.3. *Suppose that $n, r, s, t \in \mathbb{N}$ with $r, s \ge 2$ satisfy $n \ge 4s^{r-1}tr + t$. Then there exists an $(r+1)$-saturated graph G on n vertices with $e(G) \ge t_r(n) - \frac{r-1}{r}tn - 4(r-1)ts^r$ such that any complete r-partite subgraph has at most $n - 2ts^{r-1}$ vertices.*

Proof. Let $G = H_{r,s,t}(n)$. We see that G satisfies

$$e(G) \geq t_r(n-t) - t \cdot e(G_{r,s}) \geq t_r(n) - \frac{r-1}{r}tn - t \cdot e(G_{r,s})$$
$$\geq t_r(n) - \frac{r-1}{r}tn - 4(r-1)ts^r,$$

where in the second inequality we have used Observation 1.3.2. We first note that any complete r-partite subgraph is of order at most $n - 2ts^{r-1}$, as for each $p \in [t]$, at most $|H_p| - 2s^{r-1}$ vertices from $V(H_p)$ can be included in a complete r-partite subgraph of G, by Part 10 of Proposition 1.3.1.

To see that G is $(r+1)$-saturated we may argue as we did in the proof of Proposition 1.3.1. First, notice that G is K_{r+1}-free. Indeed, if there were a copy of K_{r+1} in G then, by construction, it would contain exactly one vertex from A_{r+1}, say x_p for some $p \in [t]$. Since the neighborhood of x_p outside A_{r+1} is exactly H_p, which is K_r-free, it follows that x_p cannot be contained in any copy of K_{r+1}, which yields a contradiction. There are only three types of edges that one could add to G: edges from $E(H_p)$, for some $p \in [t]$, edges between A_{r+1} and one of the A_i, $i \in [r]$, and edges within a vertex class. Note that the first option must create a K_r by Proposition 1.3.1, which then extends to a K_{r+1} when we include x_p. If we add an edge $x_p y$, for some $y \in A_i$, $p \in [t]$, $i \in [r]$, first notice that by Part 3 of Proposition 1.3.1 we may choose a copy of K_{r-1}, say K, in the graph induced on $V(H_p) \setminus A_i$. We then form a K_{r+1} by observing that x_p and y are joined to all of K. If we add an edge within one of the classes $A_1, \ldots, A_r$, then we find a K_{r-1} among $Y_1, \ldots, Y_r$ that does not intersect the class that contains the added edge. Clearly this K_{r-1} is in the common neighborhood of both points of the added edge and hence we extend to a K_{r+1}. Adding an edge within A_{r+1} guarantees a K_{r+1} by the construction of G. $\square$

By choosing s and t appropriately, we arrive at the following.

Theorem 1.3.4. *Let $r \geq 2$ be an integer and let $\varepsilon > 0$. Then there exist $n_0, b_0, c_0 > 0$*

which are constants depending on r and ε such that the following holds. Let $n \in \mathbb{N}$ and $m > 0$ be such that $n \geq n_0$ and $(\frac{r-1}{r} + \varepsilon)n \leq m \leq b_0 n^{\frac{r+1}{r}}$. Then there exists an $(r+1)$-saturated graph G on n vertices and $e(G) \geq t_r(n) - m$, with no complete r-partite subgraph on more than $\left(1 - c_0 m n^{-\frac{r+1}{r}}\right) n$ vertices.

Proof. Fix any $\varepsilon > 0$ and let $c' = \left(\frac{r-1}{r} + \varepsilon\right)^{-1}$, $c = \varepsilon/4(r-1)$. We set $s = \left\lfloor (cn)^{\frac{1}{r}} \right\rfloor$ and $t = \left\lfloor c' m n^{-1} \right\rfloor$ in Proposition 1.3.3. Observe that $t \geq 1$, and as long as $n \geq n_0 \geq \frac{2^r}{c}$, then $s \geq 2$. It is easy to check that for $b_0 \leq \left(8c' c^{\frac{r-1}{r}} r\right)^{-1}$ the condition $n \geq 4s^{r-1}tr + t$ holds for this choice of s and t. Indeed, we have:

$$\begin{aligned} 4s^{r-1}tr + t \leq 8s^{r-1}tr \leq 8\,(cn)^{\frac{r-1}{r}}\, rc' m n^{-1} &\leq 8(cn)^{\frac{r-1}{r}} rc' b_0 n^{\frac{r+1}{r}} n^{-1} \\ &= 8c' c^{\frac{r-1}{r}} r b_0 n \leq n, \end{aligned}$$

where the penultimate inequality follows from the assumption that $m \leq b_0 n^{\frac{r+1}{r}}$.

Let G be as in the conclusion of Proposition 1.3.3. It follows that

$$\begin{aligned} e(G) &\geq t_r(n) - \frac{r-1}{r} tn - 4(r-1)ts^r \geq t_r(n) - \frac{r-1}{r} tn - 4(r-1)ctn \\ &\geq t_r(n) - tn\left(\frac{r-1}{r} + 4(r-1)c\right) \geq t_r(n) - m. \end{aligned}$$

To finish the proof, notice that $t \geq \frac{c' m n^{-1}}{2}$ and $s \geq \frac{(cn)^{\frac{1}{r}}}{2}$. Therefore we have

$$g_r(G) \geq 2ts^{r-1} \geq 2\frac{c' m n^{-1}}{2} \frac{(cn)^{\frac{r-1}{r}}}{2^{r-1}} \geq \frac{c' c^{\frac{r-1}{r}}}{2^{r-1}} m n^{-\frac{1}{r}}.$$

Hence we have that there is no complete r-partite subgraph on more than $(1 - c_0 m n^{-\frac{r+1}{r}})n$ vertices, where $c_0 = \frac{c' c^{\frac{r-1}{r}}}{2^{r-1}}$.

$\square$

1.4 Beyond the threshold: $(r+1)$-saturated graphs on $t_r(n) - O(n^{\frac{r+1}{r}})$ edges

If G is an $(r+1)$-saturated graph with $t_r(n) - o(n^{\frac{r+1}{r}})$ edges, then Theorem 1.1.1 tells us that G has a complete r-partite subgraph $G' = V_1 \cup \cdots \cup V_r$ on $(1-o(1))n$ vertices. It is easy to see that no two classes V_i, V_j can differ by more than $o(n)$ vertices (otherwise, there would be too few edges in G), and so we may remove at most $o(n)$ vertices to make G' a r-partite Turán graph. In other words, as we showed in the introduction to this chapter, there is little quantitative difference between the maximum sized (in terms of number of vertices) r-partite Turán subgraph and the maximum sized complete r-partite subgraph in the edge regime $t_r(n) - o(n^{\frac{r+1}{r}})$. However, if $e(G) = t_r(n) - O(n^{\frac{r+1}{r}})$ we find it most natural to restrict our attention to *balanced* complete r-partite subgraphs or, equivalently, r-partite Turán subgraphs.

Recall that for $n,m \in \mathbb{N}$, the quantity $g_r^*(n,m)$ is the maximum number vertices that one must remove from an $(r+1)$-saturated graph on $t_r(n) - m$ edges so that the remaining graph is an r-partite Turán graph. In this section, we show that, for C sufficiently large compared to r, we have

$$g_r^*(n, Cn^{\frac{r+1}{r}}) \geq \left(1 - \frac{c' \log(Cr)}{C}\right)n,$$

for an absolute constant c' and sufficiently large n. In other words, the vertex set of the largest r-partite Turán subgraph can cover an arbitrarily small fraction of the vertices in the edge range $e(G) = t_r(n) - O(n^{\frac{r+1}{r}})$. We remind the reader of the statement of Theorem 1.1.3 for convenience.

Theorem 1.4.1. *Let $r \geq 2$ be an integer and let $\delta > 0$. There exists a constant $C = C(r,\delta)$ such that, for n sufficiently large, there exists an n-vertex $(r+1)$-saturated graph G that contains no $T_r(\delta rn)$ and $e(G) \geq t_r(n) - Cn^{\frac{r+1}{r}}$. In terms of the function g_r^*, we show that,* *for sufficiently large $D > 0$, we have*

$$g_r^*(n, Dn^{\frac{r+1}{r}}) \geq \left(1 - \frac{c' \log(Dr)}{D}\right) n,$$

for sufficiently large n.

Proof. Fix $\delta \in (0,1)$ and choose $C(\delta) = 2^6 r^{-1} \delta^{-1} \log(2e/\delta) = 4rB(\delta)$, where we have set $B(\delta) = 16r^{-2}\delta^{-1}\log(2e/\delta)$. With foresight, we select $s = n^{\frac{1}{r}}$, $t = B(\delta)n^{\frac{1}{r}}$ and note that for large enough n we have $\frac{\delta}{4}n > 2s^{r-1}$.

We build our desired graph G in three stages. We start by defining our first stage graph G_I. Let $T_r(n-t)$ be the Turán graph on $n-t$ vertices with vertex classes $V_1, \ldots, V_r$ and let $V_{r+1} = \{x_1, \ldots, x_t\}$ be a set of vertices disjoint from $V(T_r(n-t))$. Define $V(G_I) = V(T_r(n-t)) \cup V_{r+1}$ and $E(G_I) = E(T_r(n-t))$. In the second stage, we use a probabilistic construction to form the graph G_{II} by removing edges between the classes V_i, V_j, where $i, j \in [r]$, and adding edges between the classes V_{r+1}, V_i, $i \in [r]$. After this second stage we will almost be finished: G_{II} will be a K_{r+1}-free graph with many edges; G_{II} will not contain a $T_r(\delta rn)$; and adding non-saturating edges to G_{II} will not ruin these properties. In the final stage we augment G_{II} by choosing an arbitrary maximal K_{r+1}-free graph which contains G_{II}. This will serve as our final graph G.

We now prepare for the second stage. For each $i \in [r]$, fix a vertex $v_i \in V_i$ and then define $V_i' = V_i \setminus \{v_i\}$. The edges incident to the vertices $v_1, \ldots, v_r$ will go unaltered throughout this construction. This is to ensure that the addition of any edge within any of the classes $V_1, \ldots, V_r$ creates a K_{r+1} with $v_1, \ldots, v_r$, even after the edge deletions in stage II. We now define an auxiliary graph H on $V_1', \ldots, V_r'$ which records edges that we shall delete from $T_r(n-t)$ to form G_{II}.

For $p \in [t]$, let H^p be a copy of the r-partite graph $G_{r,s}$, as defined in Section 1.3, where we think of the vertex sets of the H^p as being disjoint and H_1^p, H_2^p as being the two largest vertex classes (each of order $2s^{r-1}$) in the vertex partition $H_1^p, \ldots, H_r^p$.

We shall randomly embed each H^p into $V_1' \cup \cdots \cup V_r'$ in a manner that respects the partition $V_1', \ldots, V_r'$. To this end, we define a probability space on tuples of injections $(f_1, \ldots, f_t)$ with $f_p : V(H^p) \to \bigcup_{i=1}^r V_i'$. We choose each f_p so that $\{f_p(H_i^p)\}_p$ are (fixed) vertex disjoint sets for each $3 \le i \le r$, while for $i \in \{1,2\}$, $f_p(H_i^p)$ is a uniform random subset of V_i' of size $|H_i^p|$ and each $\{f_p(H_i^p) : i \in \{1,2\}, p \in [t]\}$ is chosen independently. Note that since $|H_3^p|, \ldots, |H_r^p| \le s^{r-2}$ (by Part 8 of Proposition 1.3.1) it is indeed possible to request that $|H_i^1|, \ldots, |H_i^t|$ are disjoint subsets of V_i', as
$s^{r-2}t = B(\delta)n^{1-1/r} < (n-1-t)/r$, for large enough n.

Define the graph $H(f_1, \ldots, f_t)$ to have vertex set $V_1' \cup \cdots \cup V_r'$ and edge set

$$E(H(f_1, \ldots, f_t)) = \bigcup_{p \in [t]} \{xy : f_p^{-1}(x) f_p^{-1}(y) \in E(H^p)\}.$$

We define $G(f_1, \ldots, f_t)$ to be a graph on the same vertex set as G_I and with edge set

$$E(G(f_1, \ldots, f_t)) = \{x_p y : y \in f_p(H^p),\ p \in [t]\} \cup E(T_r(n-t)) \setminus E(H(f_1, \ldots, f_t)).$$

In what follows, we show that the probability of making a "good" choice for $G(f_1, \ldots, f_t)$ is non-zero.

Claim 1.4.2. *Let $f_1, \ldots, f_t$ be any functions as described above. The graph $G(f_1, \ldots, f_t)$ is K_{r+1}-free.*

Proof of Claim 1.4.2: If a copy of K_{r+1} is contained in $G(f_1, \ldots, f_t)$, it must have exactly one vertex in each class $V_1, \ldots, V_{r+1}$. Hence there must exist $p \in [t]$ so that $G(f_1, \ldots, f_t)$ induced on $f_p(V(H^p))$ contains a copy of K_r. This induced graph is contained in a copy of $\widetilde{G}_{r,s}$ (as in Proposition 1.3.1), which is K_r-free, a contradiction. $\square$

We now show that every "missing" edge between V_1, V_2 are saturating edges. This is important as we need to ensure that the edges we remove in stage II are not just added back in, in the final stage.

Claim 1.4.3. *Let $f_1,\dots,f_t$ be functions as described above. Adding any edge, which is not already present, between the classes V_1,V_2 in $G(f_1,\dots,f_t)$ creates a K_{r+1}.*

Proof of Claim 1.4.3: Suppose that $e \notin E_{G(f_1,\dots,f_t)}(V_1,V_2)$. This means that $e \in E_{H(f_1,\dots,f_t)}(V_1,V_2)$ and thus $e \in E_{f_p(H^p)}(V_1,V_2)$, for some $p \in [t]$. Every such edge in $f_p(H^p)$, if deleted from H^p, is contained in an independent set I with exactly one vertex in each part $V_1,\dots,V_r$; this holds by Part 4 in Proposition 1.3.1. Since each of the $H^1,\dots,H^t$ are disjoint on $V_3,\dots,V_r$, I is a set containing only e, in H. This is the same as saying that e is a r-saturating edge in $G(f_1,\dots,f_t)$ in the graph induced on $f_p(V(H^p))$. Since the vertex $x_p \in V_{r+1}$ joins to all of $f_p(V(H^p))$, e is $(r+1)$-saturating in $G(f_1,\dots,f_t)$. $\square$

The following claim will help us show that we cannot find a large r-partite Turán graph in our final graph.

Claim 1.4.4. *The probability that $G(f_1,\dots,f_t)$ contains a complete bipartite graph $K_{\delta n/2,\delta n/2}$ between V_1,V_2 is less than $1/2$.*

Proof of Claim 1.4.4: Let $E(A,B)$ be the "bad" event that the pair $A \subset V_1'$, $B \subset V_2'$ have no edge of $H(f_1,\dots,f_t)$ between them. We define the random variable X to be the number of pairs of subsets $A \subset V_1'$, $B \subset V_2'$ of size $\delta n/2$ each, that have no edge of $H(f_1,\dots,f_t)$ between them.

To estimate the expectation of X we fix two sets $A \subseteq V_1'$, $B \subseteq V_2'$ of size $\delta n/2$, and let $E_p = E_p(A,B)$, for $p \in [t]$, denote the event that $f_p(H^p)$ has no edge between A,B. By independence, $\mathbb{P}(E(A,B)) = \prod_p \mathbb{P}(E_p)$. We fix $p \in [t]$ and look to bound $\mathbb{P}(E_p)$. We explicitly express the two largest vertex classes of H^p, $H_1^p = \{y_1,\dots,y_{2s^{r-1}}\}$, and $H_2^p = \{z_1,\dots,z_{2s^{r-1}}\}$, where y_iz_i, $i \in [2s^{r-1}]$, are the edges of a perfect matching in H^p between the two largest classes (which is guaranteed by Proposition 1.3.1). For ease of notation, let $f = f_p$ and let us say that a pair $f(y_i), f(z_i)$ *hits* A, B if $f(y_i) \in A$ and $f(z_i) \in B$. We will say that $f(y_i), f(z_i)$ *misses* the pair, otherwise. We define $E_p(i)$ to be the event that $f(y_i), f(z_i)$ misses A, B.

Note that $\mathbb{P}(E_p)$ is at most

$$\mathbb{P}\left(\bigcap_{i=1}^{2s^{r-1}} E_p(i)\right) = \prod_{i=1}^{2s^{r-1}} \mathbb{P}\left(E_p(i) | E_p(i-1), \ldots, E_p(1)\right). \tag{1.6}$$

So to bound $\mathbb{P}(E_p)$, we need only to bound the terms in the above product. This is easily done as the conditional probabilities $\mathbb{P}\left(E_p(i) | E_p(i-1), \ldots, E_p(1)\right)$ do not differ too much from the unconditioned probabilities $\mathbb{P}(E_p(i))$. To this end, note that $E_p(1), \ldots, E_p(i-1)$ depend only on the choices of $Y_{i-1} = \{f(y_1), \ldots, f(y_{i-1})\}, Z_{i-1} = \{f(z_1), \ldots, f(z_{i-1})\}$. Thus, we have

$$\begin{aligned}
\mathbb{P}\left(E_p(i) | E_p(i-1), \ldots, E_p(1)\right) &\leq \max_{Y_{i-1}, Z_{i-1}} \mathbb{P}\left(E_p(i) | Y_{i-1}, Z_{i-1}\right) \\
&= 1 - \min_{Y_{i-1}, Z_{i-1}} \mathbb{P}\left(f(y_i), f(z_i) \text{ hits } A, B | Y_{i-1}, Z_{i-1}\right) \\
&\leq 1 - \min_{Y_{i-1}, Z_{i-1}} \frac{|A \setminus Y_{i-1}||B \setminus Z_{i-1}|}{(|V_1'| - (i-1))^2} \\
&\leq 1 - \left(\frac{r(\delta n/2 - 2s^{r-1})}{n}\right)^2 \\
&\leq \exp(-\delta^2 r^2/16),
\end{aligned}$$

where the third inequality follows by recalling that $|V_1'|, |V_2'| \leq n/r$ and the last inequality follows by recalling that $\frac{\delta}{4} n > 2s^{r-1}$. So, from (1.6), we have

$$\mathbb{P}(E_p) \leq \exp(-r^2 \delta^2 s^{r-1}/8),$$

for each $p \in [t]$, and therefore

$$\mathbb{P}(E(A,B)) = \prod_{p=1}^{t} \mathbb{P}(E_p) \leq \exp\left(-\frac{r^2\delta^2 s^{r-1}t}{8}\right).$$

So, by linearity of expectation, we have

$$\mathbb{E}X \leq \binom{n}{\delta n/2}^2 \exp\left(-\frac{r^2\delta^2 s^{r-1}t}{8}\right).$$

Using the standard inequality $\binom{n}{k} \leq \left(\frac{ne}{k}\right)^k$, we have

$$\mathbb{E}X \leq (2e/\delta)^{\delta n} \exp\left(-\frac{r^2\delta^2 s^{r-1}t}{8}\right) = \exp\left(\delta n \log(2e/\delta) - \frac{r^2\delta^2 s^{r-1}t}{8}\right).$$

Recalling our choices of $s = n^{1/r}$ and $t = B(\delta)n^{1/r} = 16r^{-2}\delta^{-1}\log(2e/\delta)n^{1/r}$, we have $\mathbb{E}X < 1/2$ for sufficiently large n. This completes the proof of Claim 1.4.4. $\square$

We now define G_{II} to be a graph of the form $G(f_1,\ldots,f_t)$ for which there are no copies of $K_{\delta n/2,\delta n/2}$ between vertex classes V_1', V_2'. Such a graph $G(f_1,\ldots,f_t)$ exists with non-zero probability, by Claim 1.4.4.

To define our final graph G, we choose a maximal K_{r+1}-free graph which contains G_{II}. Since G_{II} is K_{r+1}-free, G is also K_{r+1}-free and, trivially, G is $(r+1)$-saturated. Using inequalities $t_r(n-t) \geq t_r(n) - tn$ and $e(G_{r,s}) \leq 4(r-1)s^r$, we have that

$$\begin{aligned} e(G) &\geq e(G_{II}) \\ &\geq t_r(n) - tn - te(G_{r,s}) \\ &\geq t_r(n) - B(\delta)n^{\frac{r+1}{r}} - 4(r-1)s^r t \\ &\geq t_r(n) - 4rB(\delta)n^{\frac{r+1}{r}} = t_r(n) - C(r,\delta)n^{\frac{r+1}{r}}. \end{aligned}$$

We now observe that G cannot contain a copy T of $T_r(\delta rn)$. Suppose, towards a contradiction, that G contains T. First note that $G - V_{r+1}$ is r-partite with vertex partition $V_1, \ldots, V_r$. This is because the addition of any pair $e = uv$ to G_{II}, *within* some V_i, would form a copy of K_{r+1} on vertex set $\{u,v\} \cup (\{v_1, \ldots, v_r\} - \{v_i\})$. Therefore $G - V_{r+1}$ is r-partite. This means that G must contain a copy K of $K_{\delta n/2, \delta n/2}$ between V_1', V_2', as $|T \cap V_{r+1}| \leq |V_{r+1}| = t < \delta n/4$ and therefore $|T \cap (V_1 \cup \cdots \cup V_r)| \geq \delta rn/2$. Now since all non-edges between V_1, V_2 are $(r+1)$-saturating in G_{II} (Claim 1.4.3), we have that no edges were added between V_1, V_2 in forming G. In other words, $G_{II}[V_1, V_2] = G[V_1, V_2]$. This implies that $K \cong K_{\delta n/2, \delta n/2}$ is also a subgraph of G_{II}, which contradicts Claim 1.4.4. This completes the proof of Theorem 1.1.3. $\square$

1.5 Final remarks and open problems

Recall that $g_r(n,m)$ is defined to be the maximum number of vertices that one is required to remove from an n-vertex, $(r+1)$-saturated graph with at least $t_r(n) - m$ edges, so that the remaining graph is complete r-partite. Combining Theorems 1.1.2 and 1.3.4 we see that for any $\varepsilon > 0$, if $n \geq n_0(r, \varepsilon)$ and $(\frac{r-1}{r} + \varepsilon)n \leq m \leq n^{\frac{r+1}{r}}$ one has

$$c_{r,\varepsilon} mn^{-1/r} \leq g_r(n,m) \leq C_r mn^{-1/r},$$

where $c_{r,\varepsilon}$ depends on r, ε, and C_r depends only on r. However, our construction does not work if $n/r \leq m \leq \frac{r-1}{r}n$ (note that when $m < n/r$ the problem becomes trivial, since Brouwer's theorem (Theorem 1.2.3) implies that the graph must be r-partite, in which case, by maximality, must be complete r-partite). We leave the determination of $g_r(n,m)$ in this range of m as an open problem.

Problem 1.5.1. *Determine $g_r(n,m)$ for $n/r \leq m \leq \frac{r-1}{r}n$.*

We remark that when $r = 2$ these ranges coincide; the lower bound coming from our construction in Theorem 1.3.4 by plugging in $r = 2$ yields that for any $\varepsilon > 0$

$$g_2(n,(1/2+\varepsilon)n) \geq \frac{1}{4}(\varepsilon n)^{1/2}.$$

However, we can do slightly better with the following modified construction, G_0. Consider a copy of the Turán graph $T_2(n-1)$ with vertex partition $V_1 \cup V_2$ and let $V(G_0) = V_1 \cup V_2 \cup \{u\}$ for a vertex $u \notin V_1 \cup V_2$. Let $X \subset V_1, Y \subset V_2$ with $|X| = (\varepsilon n)^{1/2} = |Y|$ and remove from $T_2(n-1)$ all edges between X and Y. Finally, join u to every vertex in $X \cup Y$, but to no other vertices. It is easy to see that G_0 is triangle-saturated and has at least $t_2(n) - (1/2+\varepsilon)n$ edges. On the other hand, the largest complete bipartite subgraph is on $n - (\varepsilon n)^{1/2}$ vertices, showing that

$$g_2(n,(1/2+\varepsilon)n) \geq (\varepsilon n)^{1/2}.$$

Unfortunately, we do not see how to extend this construction to larger values of r.

It is also natural to consider the largest k such that the Turán subgraph $T_r(k)$ must appear in every $(r+1)$-saturated graph G, with $e(G) \geq t_r(n) - m$ edges, where $m \sim Cn^{\frac{r+1}{r}}$. This amounts to the following problem regarding the function $g_r^*(n,m)$, the "balanced" analogue of $g_r(n,m)$.

Problem 1.5.2. *Determine $g_r^*(n, Cn^{\frac{r+1}{r}})$, for each $C \in \mathbb{R}^+$ and sufficiently large n.*

Recall that Theorem 1.1.3 shows that $g_r^*(n,Cn^{\frac{r+1}{r}}) \geq \left(1 - \frac{c' \log(Cr)}{C}\right) n$, for C large (depending on r) and fixed, and $n \to \infty$, but we have no non-trivial upper bounds for $g_r^*(n,Cn^{\frac{r+1}{r}})$, when C is large.

CHAPTER 2

THE HOMOMORPHISM THRESHOLD OF $\{C_3, C_5\}$-FREE GRAPHS

2.1 Introduction

In this chapter, we are interested in the structure of graphs of high minimum degree which forbid specific subgraphs. For a fixed graph H, a graph is said to be *H-free* if it does not contain H as a subgraph. Let $\mathrm{Forb}(H)$ denote the class of H-free graphs, and let $\mathrm{Forb}_n(H)$ denote the class of n-vertex graphs in $\mathrm{Forb}(H)$. Furthermore, let $\mathrm{Forb}(H,d)$ denote the class of H-free graphs G with minimum degree at least $d|V(G)|$. Analogous definitions hold if we replace H by some family $\mathscr{H}$ of graphs. Finally, we say that a graph G is homomorphic to a graph H if there exists a map $f : V(G) \to V(H)$ such that $f(u)f(v) \in E(H)$ whenever $uv \in E(G)$. For example, G is homomorphic to K_r if and only if $\chi(G) \leq r$

A classical result of Andrásfai, Erdős and Sós [9] (mentioned in the previous chapter) states that if G is a K_{r+1}-free graph on n vertices with minimum degree $\delta(G) > \frac{3r-4}{3r-1}n$, then G is r-colorable. This result can be viewed as a significant strengthening of the following fact, which is a consequence of Turán's theorem: the minimum degree of a K_{r+1}-free graph on n vertices is at most $(1-1/r)n$. Note also here that the chromatic number $\chi(G)$ of G is bounded by a constant independent of n. In general, one may ask whether or not this behavior persists when the minimum degree condition is weakened. Along these lines, Häggkvist [33] showed that any n-vertex triangle-free graph of minimum degree greater than $3n/8$ is homomorphic to a 5-cycle, and accordingly has chromatic number at most 3. Note that this is indeed an extension of the Andrásfai-Erdős-Sós result when the minimum degree condition is weakened, since a balanced blow-up of a 5-cycle exhibits the tightness of that result. Jin [38] took up the investigation and significantly extended the work of Häggkvist: he proved that for all

$1 \leq d \leq 9$, any n-vertex triangle-free graph with minimum degree larger than $\frac{d+1}{3d+2}n$ is homomorphic to the graph F_d^2, which is obtained by adding all chords joining vertices at distances $1 \pmod 3$ along a cycle of length $3d-1$. Observe that F_d^2 is triangle-free and 3-colorable for every d. The graphs F_d^2 are a special case of a larger family of graphs, F_d^k, which we shall discuss shortly. We note that Jin's result [38] is best possible, in the sense that such a statement does not hold for $d = 10$. Indeed, by taking a suitably chosen unbalanced blow-up of the Grötzsch graph (see Figure 2.1) one can obtain a triangle-free graph on n vertices and minimum degree $\lfloor 10n/29 \rfloor$ which is not 3-colorable, so in particular it is not homomorphic to F_d^2 for any d. This blow-up is defined as follows. Give weight $3t$ to the vertices along the outer 5-cycle of the Grötzsch graph, weight $4t$ to the central vertex, and weight $2t$ to the remaining 5 vertices. Blow-up each vertex proportional to its weight: we obtain a graph on $29t$ vertices which is $10t$-regular, as required.

Building on this work, Chen, Jin, and Koh [18] showed, in particular, that any n-vertex 3-colorable triangle-free graph G with $\delta(G) > n/3$ is homomorphic to F_d^2, for some d. Again, the Grötzsch graph shows that the assumption that the graph is 3-colorable is necessary.

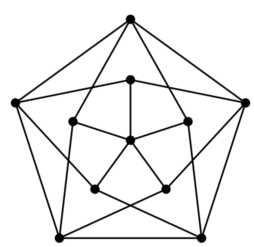

Figure 2.1: The Grötzsch graph

In general, one may ask for the *smallest* minimum degree condition we may impose on an H-free graph which guarantees that it has bounded chromatic number. To be precise, this prompts us to define the *chromatic threshold* $\delta_\chi(H)$ of a graph H:

$$\delta_\chi(H) = \inf\{d : \text{there exists } C = C(H,d) \text{ such that if } G \in \text{Forb}(H,d), \text{ then } \chi(G) \leq C\}.$$

In other words, $\delta_\chi(H)$ is the infimum over all $d \in [0,1]$ such that every H-free graph on n vertices and with minimum degree at least dn has bounded chromatic number (independent of n). This definition was implicit in the works of Andrásfai [7] and Erdős and Simonovits [25], and was first explicitly formulated by Łuczak and Thomassé [49].

For every $\varepsilon > 0$, Hajnal (appearing in [25]) constructed graphs in $\mathrm{Forb}(K_3, 1/3-\varepsilon)$ with arbitrarily large chromatic number, thereby proving the bound $\delta_\chi(K_3) \geq 1/3$. Thomassen [65] thereafter established the matching upper bound, showing that $\delta_\chi(K_3) = 1/3$. In fact, Brandt and Thomassé [16] strengthened this by showing that triangle-free graphs of minimum degree larger than $n/3$ have chromatic number at most four, answering a question of Erdős and Simonovits [25]. Extensions of these results were obtained by Goddard and Lyle [32] and Nikiforov [54], who showed that $\delta_\chi(K_r) = \frac{2r-5}{2r-3}$. More precisely, they showed that any K_r-free graph with minimum degree larger than $\frac{2r-5}{2r-3}|V(G)|$ has $\chi(G) \leq r+1$. Finally, building off of ideas of Łuczak and Thomassé [49] and Lyle[50], Allen, Böttcher, Griffiths, Kohayakawa and Morris [4] determined the value of $\delta_\chi(H)$ for every graph H with $\chi(H) > 2$.

Note that the results of Häggkvist [33], Jin [38], and Chen, Jin, and Koh [18] mentioned earlier not only show that triangle-free graphs of large enough minimum degree have bounded chromatic number, but that they are actually homomorphic to some specific 3-colorable triangle-free graph. One may ask then, with respect to the above discussion, whether we can replace the property of having bounded chromatic number with the property of admitting a homomorphism to a graph of bounded order with additional properties. This question was posed by Thomassen [65] in the specific case of triangle-free graphs, and motivated Oberkampf and Schacht [57] to introduce the *homomorphism threshold* $\delta_{\mathrm{hom}}(H)$ of a graph H:

$$\delta_{\mathrm{hom}}(H) = \inf\{d : \exists C = C(H,d) \text{ s.t. } \forall G \in \mathrm{Forb}(H,d) \\ \exists G' \in \mathrm{Forb}_C(H) \text{ s.t. } G \text{ is homomorphic to } G'\}.$$

In words, $\delta_{\text{hom}}(H)$ is the infimum over all $d \in [0,1]$ such that every H-free graph with n vertices and minimum degree at least dn is homomorphic to an H-free graph of bounded order (independent of n). Note that the definition of $\delta_{\text{hom}}(H)$ extends naturally if we replace H by a family $\mathscr{H}$ of graphs.

Łuczak [48] proved that $\delta_{\text{hom}}(K_3) \leq 1/3$. Note that if G is homomorphic to G', then $\chi(G) \leq |V(G')|$. Accordingly, we always have $\delta_{\text{hom}}(H) \geq \delta_\chi(H)$, and so, since $\delta_\chi(K_3) = 1/3$, it follows that $\delta_{\text{hom}}(K_3) = 1/3$. The earlier mentioned papers of Goddard and Lyle [32] and Nikiforov [54] actually provide a structural characterization of K_r-free graphs with $\delta(G) > \frac{2r-5}{2r-3}|V(G)|$ (for $r \geq 4$). In particular, we know that $\delta_{\text{hom}}(K_r) = \delta_\chi(K_r) = \frac{2r-5}{2r-3}$.

Aside from these results, not much is known in general about the homomorphism threshold. Note that if $\mathscr{H}$ contains a bipartite graph, then $\delta_{\text{hom}}(\mathscr{H}) = 0$. Thus, it suffices to determine the homomorphism threshold for families which do not contain any bipartite graphs. Let $\text{ex}(n,\mathscr{H})$ denote the *Turán function* of the family $\mathscr{H}$, i.e., $\text{ex}(n,\mathscr{H})$ is the maximum number of edges in an $\mathscr{H}$-free graph on n vertices. Further, let $\pi(\mathscr{H}) = \lim_{n\to\infty} \frac{\text{ex}(n,\mathscr{H})}{\binom{n}{2}}$ denote the *Turán density* of $\mathscr{H}$. Then it is easy to see that $\delta_{\text{hom}}(\mathscr{H}) \leq \pi(\mathscr{H})$. Summarizing, we have the general bounds

$$\delta_\chi(\mathscr{H}) \leq \delta_{\text{hom}}(\mathscr{H}) \leq \pi(\mathscr{H}).$$

In studying the homomorphism threshold, we are looking at the global structural properties of graphs in $\text{Forb}(\mathscr{H},d)$ as d ranges from 0 to $\pi(\mathscr{H})$. Moreover, as d ranges from 0 to $\pi(\mathscr{H})$ we expect this to place more restrictions on the possible homomorphism types, as this restricts membership in $\text{Forb}(\mathscr{H},d)$ (*all* $\mathscr{H}$-free graphs belong to $\text{Forb}(\mathscr{H},0)$).

Oberkampf and Schacht [57] gave a new proof of the fact that $\delta_{\text{hom}}(K_r) = \frac{2r-5}{2r-3}$ avoiding the Regularity Lemma (which was used in Łuczak's proof), and asked for the determination of the homomorphism threshold of the two simplest yet unknown families of graphs: the odd cycle, $\delta_{\hom}(C_{2k-1})$, and graphs of odd-girth $2k+1$, $\delta_{\hom}(\{C_3,\ldots,C_{2k-1}\})$ for $k \geq 3$. As our first main result, we determine the value of the second of these two parameters in the case $k=3$.

Theorem 2.1.1. *The homomorphism threshold of* $\{C_3,C_5\}$ *is* $1/5$.

In other words, Theorem 2.1.1 states that, for every $\varepsilon > 0$, if G is a $\{C_3,C_5\}$-free graph on n vertices and minimum degree at least $(1/5+\varepsilon)n$, then G is homomorphic to a $\{C_3,C_5\}$-free graph of order at most C, where C depends on ε but not on n. Moreover, there is a sequence of graphs $(G_n)_{n\in\mathbb{N}}$ where G_n has order n, is $\{C_3,C_5\}$-free, has minimum degree approximately $n/5$, but does not admit a homomorphism to any $\{C_3,C_5\}$-free graph of bounded order. We also establish an upper bound on $\delta_{\hom}(C_5)$. This is a consequence of Theorem 2.1.1, since C_5-free graphs of large minimum degree end up being triangle-free as well. In particular, we have the following.

Corollary 2.1.2. *The homomorphism threshold of* C_5 *is at most* $1/5$.

By a result of Thomassen [66], we know that $\delta_{\chi}(\{C_3,C_5\}) = 0$. Therefore, Theorem 2.1.1 gives the first example of a family of graphs for which the chromatic threshold and the homomorphism threshold differ. We do not know if there is a single graph for which these two parameters differ.

We are able to say much more about the structure of $\{C_3,C_5\}$-free graphs with n vertices and minimum degree larger than $n/5$. First we need to define a family of graphs, sometimes known as *generalized Andrásfai graphs*. For $k \geq 2$, let $\mathscr{C}_{2k-1}$ denote the family of odd-cycles $\{C_3,\ldots,C_{2k-1}\}$. For integers $d \geq 1$ and $k \geq 2$, denote by F_d^k the graph obtained from a $((2k-1)(d-1)+2)$-cycle (an edge, when $d=1$) by adding all chords joining vertices at distances clockwise along the cycle of the form $j(2k-1)+1$ for $j=1,\ldots,d-2$ (with the convention of adding no chords when $d=2$). Equivalently, F_d^k is the Cayley graph on $\mathbb{Z}_{(2k-1)(d-1)+2}$ with generating set C consisting of all elements in $\mathbb{Z}_{(2k-1)(d-1)+2}$ congruent to 1 modulo $2k-1$. In other words, we have

$$C = \{j(2k-1)+1 : j = 0,1,\ldots,d-1\},$$

and it is easy to see that C is an inverse closed subset of $\mathbb{Z}_{(2k-1)(d-1)+2}$ so that this Cayley graph is indeed undirected.

We can also give an alternative description of these graphs as the complement of the $(k-1)(d-1)$-th power of a cycle of length $(2k-1)(d-1)+2$ (recall that for an integer $m \geq 1$, the m-th power G^m of G is the graph on vertex set $V(G)$ where we join all pairs of vertices which are distance at most m in G). However, we shall not utilize this description in the sequel.

It is not difficult to check that F_d^k is d-regular, maximal $\mathscr{C}_{2k-1}$-free (i.e., has *odd-girth* $2k+1$) and 3-colorable. Let us summarize some facts about these graphs in the following proposition.

Proposition 2.1.3. *Let $d \geq 1, k \geq 2$ be integers. The following properties of F_d^k hold:*

1. *F_d^k is d-regular.*
2. *F_d^k is a subgraph of F_{d+1}^k.*
3. *$\chi(F_d^k) = 3$ for all $d \geq 2$.*
4. *F_d^k is $\mathscr{C}_{2k-1}$-free.*
5. *Every two distinct vertices in F_d^k ($d \geq 2$) are contained in a $(2k+1)$-cycle. Therefore, for any two distinct vertices there is a path of length $1,3,\ldots,$ or $2k-1$ between them.*

Proof. The first item above is immediate from the definition, so let us turn to the second item. Clearly it is true for $d=1$, and if $d \geq 2$, then removing any consecutive $2k-1$

vertices from the cycle of F^k_{d+1} results in a copy of F^k_d. For the third item, we can write $V(F^k_d) = A \cup B \cup C$, where A consists of all residues $0, 3, 6, \ldots$ modulo $(2k-1)$, B consists of all residues $1, 4, 7, \ldots$ modulo $(2k-1)$, and C consists of all residues $2, 5, 8, \ldots$ modulo $(2k-1)$. If $x, y \in A$, then $x - y$ is a multiple of 3 and at most $2k-1$, hence is not congruent to 1 modulo $(2k-1)$. The same holds for B and C. So each of A, B, and C are independent sets giving us our desired 3-coloring. Of course, we cannot do better when $d \geq 2$, since F^k_d contains an odd-cycle.

The fourth item is clearly true for $d = 1, 2$, so we may assume that $d \geq 3$. Let us view F^k_d as a Cayley graph with vertex set $\mathbb{Z}_{(2k-1)(d-1)+2}$, and suppose that there is an odd cycle $(a_0 \ldots a_{l-1})$ of length $l = 2j-1$ for $2 \leq j \leq k$. Note that the sum (indices modulo l) $\sum_{i=0}^{l-1}(a_{i+1} - a_i)$ is 0 in $\mathbb{Z}_{(2k-1)(d-1)+2}$; thus there is an integer m such that

$$\sum_{i=0}^{l-1}(a_{i+1} - a_i) = ((2k-1)(d-1)+2)m. \tag{2.1}$$

On the other hand, each term $c_i := a_{i+1} - a_i$ in that sum is congruent to 1 mod $2k-1$, and so reducing the above equation mod $2k-1$ yields $2m = l \pmod{2k-1}$. This implies that $m \geq k + j - 1 \geq l$. Using this bound, Equation (2.1), and the fact that each c_i has the form $1 + (2k-1)b_i$ for some $b_i \in \{0, \ldots, d-1\}$, we get

$$l + (2k-1)\sum_{i=0}^{l-1} b_i = \sum_{i=0}^{l-1} c_i \geq (2k-1)(d-1)l + 2l.$$

Since $b_i \leq d-1$ for each i we have $(2k-1)(d-1)l \geq (2k-1)(d-1)l + l$, a contradiction.

Finally, let us prove the last item in the proposition. We prove this by induction on d. Note that this easily holds for $d = 2$, since F^k_2 by definition is a $(2k+1)$-cycle. Now, given $d \geq 3$, let us suppose the result holds for smaller values of d. Let x, y be two distinct vertices. If these vertices are at distance at most $2k$ clockwise along the cycle, then they are contained in a C_{2k+1}, so we may suppose otherwise. Thus there exists an interval I of length $2k-1$ separating x and y. Removing this interval creates a copy of F_{d-1}^{k} containing x and y, so by induction we obtain that x and y are contained in a cycle of length $2k+1$.

□

In the odd-girth 7 case we have $k=3$ and we shall write F_d instead of F_d^3 for simplicity. In particular, F_1 is an edge, F_2 is a C_7 (a cycle of length 7) and F_3 is the graph obtained by adding all diagonals to a C_{12} (by a *diagonal* in an even cycle C_{2r}, $r \geq 2$, we mean an edge joining vertices at distance r along the cycle). This graph is also known as the Möbius ladder on 12 vertices (see Figure 2.2).

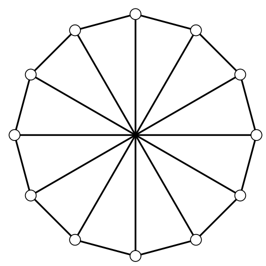

Figure 2.2: F_3, the Möbius ladder

As our second main result, we determine the structure of $\{C_3, C_5\}$-free graphs on n vertices with minimum degree larger than $n/5$, thus answering a question of Messuti and Schacht [53].

Theorem 2.1.4. *Let G be a $\{C_3, C_5\}$-free graph on n vertices with $\delta(G) > n/5$. Then G is homomorphic to F_d, for some d.*

As a consequence of Theorem 2.1.4, we are able to obtain the following.

Corollary 2.1.5. *Let G be a $\{C_3, C_5\}$-free graph on n vertices with $\delta(G) > \frac{d}{5d-3}n$. Then G is homomorphic to F_{d-1}.*

As F_d is not homomorphic to F_{d-1}, a suitable blow-up of F_d shows that this result is tight. For $\{C_3, C_5\}$-free graphs and graphs of higher odd-girth, similar results have been obtained before. Häggkvist and Jin [34] proved that any n-vertex $\{C_3, C_5\}$-free graph with minimum degree larger than $n/4$ is homomorphic to C_7, which is a special case of Corollary 2.1.5 when $d = 3$.

2.1.1 Organization and Notation

The remainder of this chapter is organized as follows. In Section 2.2 we shall provide an outline of the technical results needed to prove our main theorem. Many of these state that certain subgraphs cannot appear in maximal $\{C_3, C_5\}$-free graphs of minimum degree larger than $n/5$. In the next three sections (Section 2.3 to Section 2.5) we shall prove each of these technical results. In Section 2.6, we deduce our main theorem, Theorem 2.1.4. Finally, Section 2.7 includes our results concerning homomorphism thresholds, Theorem 2.1.1 and Corollary 2.1.2.

Our notation is standard. In particular, for a graph G, we use $|V(G)|$ to denote the number of vertices of G, $V(G)$ denotes the vertex set, $E(G)$ the edge set, and $\delta(G)$ denotes the minimum degree. For a vertex v, $N_G(v)$ denotes the neighborhood of v, and for a subset $X \subseteq V(G)$, $N_G(v,X)$ denotes the neighborhood of v in X, i.e. $N_G(v,X) = N_G(v) \cap X$. We shall often omit the use of the subscript 'G'. If $X, Y \subseteq V(G)$, then we say an edge e is an $X - Y$ *edge* if one endpoint of e is in X, the other in Y. If X=$\{x\}$, then we simply say e is an $x - Y$ edge. We denote by $(v_1 \dots v_\ell)$ the cycle on vertices $v_1, \dots v_\ell$ taken in this order. Similarly, we denote by $v_0 \dots v_\ell$ the path on vertices $v_0, \dots, v_\ell$ taken in this order. A cycle (path) with ℓ edges is an *ℓ-cycle* (*ℓ-path*). For any path P we shall let $l(P)$ denote the length of P (similarly, for any cycle); moreover, if $u, v \in V(P)$, then uPv denotes the subpath of P with endpoints u and v. If C is a cycle in G and $u, v \in V(C)$, then uCv denotes the *shortest* path along C joining u, v. For a path $P = ax_1 \dots x_t b$ between a and b, the *interior* of the path, denoted int(P), is the subpath $x_1 \dots x_t$.

2.2 Overview

In this section we provide a tour through the technical results needed to establish our main theorem. Note that in proving Theorem 2.1.4 we may assume our graph is maximal $\{C_3, C_5\}$-free. Accordingly, the following results concern maximal $\{C_3, C_5\}$-free graphs. The main tool needed for the proof of Theorem 2.1.4 is the following result.

Theorem 2.2.1. *Let G be a maximal $\{C_3, C_5\}$-free graph on n vertices with $\delta(G) > n/5$. Then every vertex in G has a neighbor in every 7-cycle in G.*

We remark that Jin [39] proved the analogous theorem for 5-cycles in triangle-free graphs of large enough minimum degree. In order to establish Theorem 2.2.1 we shall need a sequence of lemmas which show that certain subgraphs cannot appear in maximal $\{C_3, C_5\}$-free graphs of large minimum degree. The first of these lemmas, which shows that $\{C_3, C_5\}$-free graphs with large minimum degree do not have induced 6-cycles, proves very useful, and we shall use it throughout the paper. Brandt and Ribe-Baumann [15] mention it without proof. In fact, we are able to prove that induced 6-cycles do not appear in dense graphs of given odd-girth, which are edge-maximal with respect to this property.

Lemma 2.2.2. *Let $k \geq 3$ be an integer and suppose G is an n-vertex maximal odd-girth $2k+1$ graph with $\delta(G) > \frac{n}{2k-1}$. Then G contains no induced 6-cycle.*

We believe that this result may of independent interest, as it may be useful for tackling further structural problems for dense graphs of given odd-girth. In particular, the triangle-free case admits 4-colorable examples, which contain induced 6-cycles. We believe that Lemma 2.2.2 is evidence that all odd-girth $2k+1$ ($k \geq 3$) graphs with minimum degree larger than $\frac{n}{2k-1}$ are, in fact, 3-colorable.

We shall also need the fact that a 'partial' Möbius ladder cannot appear as a subgraph in the graphs we consider. More precisely, we need the following lemma.

Lemma 2.2.3. *Let G be a maximal $\{C_3, C_5\}$-free graph on n vertices with $\delta(G) > n/5$. If $(x_1 \dots x_{12})$ is a 12-cycle with two consecutive diagonals x_1x_7 and x_2x_8 present. Then either $(x_1 \dots x_{12})$ or $(x_2x_3 \dots x_7x_1x_{12} \dots x_8)$ induces a Möbius ladder.*

We note, and prove, the following useful corollary of Lemma 2.2.3.

Corollary 2.2.4. *Let G be a maximal $\{C_3, C_5\}$-free graph on n vertices with $\delta(G) > n/5$. If u is a vertex with no neighbors in a 7-cycle C, then u has no neighbor with two neighbors in C.*

Proof. Suppose that $C = (x_1 \dots x_7)$ and u has no neighbors in C, but a neighbor v of u has two neighbors in C. Say, v is adjacent to x_2 and x_7 (see Proposition 2.2.6 below).

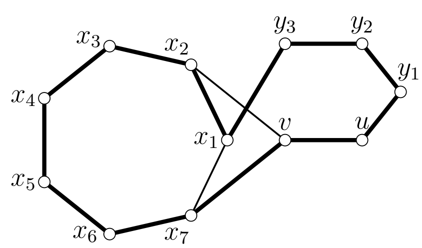

Figure 2.3: Creating a Möbius Ladder

Since u is not adjacent to x_1 and G is maximal $\{C_3, C_5\}$-free, there must be a path of length 2 or 4 between them; but a path of length 2 is impossible (it would complete the path uvx_2x_1 to a 5-cycle), so there is a 4-path $uy_1y_2y_3x_1$. One may check that none of y_1, y_2, y_3 is equal to one of the vertices of C or to v (see Figure 2.3). But then $(x_1 \dots x_7vuy_1y_2y_3)$ is a 12-cycle with two consecutive diagonals x_1x_7 and x_2v. It follows from Lemma 2.2.3 that all diagonals in the cycle must be present (or we need to consider the 12-cycle $(x_2 \dots x_7x_1y_3y_2y_1uv)$ with diagonals x_1x_2 and x_7v). In particular, u has a neighbor in C, a contradiction. $\square$

Finally, in order to prove Theorem 2.2.1, we establish the following, which is the last of our results regarding forbidden subgraphs in maximal $\{C_3, C_5\}$-free graphs of large minimum degree.

Lemma 2.2.5. *Let G be a maximal $\{C_3, C_5\}$-free graph on n vertices with $\delta(G) > n/5$. Then G does not contain, as an induced graph, the graph obtained by two 7-cycles whose intersection is a path of length 3 (see Figure 2.5).*

Before proceeding to the proofs of the above forbidden subgraph lemmas, we shall show how to prove Theorem 2.2.1 using Lemma 2.2.2, Lemma 2.2.3, and Lemma 2.2.5. The proofs of these lemmas shall be deferred to Sections 2.3, 2.4, and 2.5, respectively. In order to aid in their proofs, we introduce the following definition.

Definition. A subgraph H of a graph G is called *well-behaved* (in G) if for every vertex u in G, there is a vertex v in H, such that $N_G(u,H) \subseteq N_H(v)$.

In particular, this implies that $G[H \cup \{u\}]$ is homomorphic to H for every $u \in V(G)$. Many of the subgraphs we consider are actually well-behaved (in their respective host graphs). For example, we note the following useful fact.

Proposition 2.2.6. *Let $k \geq 2$ be an integer and let G be an odd-girth $2k+1$ graph. Then C_{2k+1} is well-behaved in G.*

Proof. Let $k \geq 2$ and let $C = (x_1 \ldots x_{2k+1})$ be a $(2k+1)$-cycle in G, labelled counterclockwise. Suppose without loss of generality that $w \in V(G) \setminus V(C)$ is joined to x_1. We claim that either $N(w,C) \subseteq N_C(x_2)$ or $N(w,C) \subseteq N_C(x_{2k+1})$. Let w' be another neighbor of w in C and suppose to the contrary that $w' \neq x_3, x_{2k}$. Let P denote the path $x_1x_2x_3 \ldots w'$ and P' denote the path $x_1x_{2k+1}x_{2k} \ldots w'$. Now, note that $l(P) \leq (2k+1) - 3 = 2k-2$ and similarly $l(P') \leq 2k-2$. Moreover, one of P, P' must be odd, say P. But then the cycle (wx_1Pw') is odd and has length at most $2k-1$, a contradiction. □

We need the following observation before proving Theorem 2.2.1.

Observation 2.2.7. *Let G be a maximal $\{C_3, C_5\}$-free graph on n vertices with $\delta(G) > n/5$. Suppose that u has no neighbors in a 7-cycle C. Then u has a common neighbor with at most one of the vertices in C.*

Proof. Suppose that u has no neighbor in the cycle $C = (x_1 \dots x_7)$. Furthermore, suppose that u has a common neighbor v with x_1. By symmetry, it suffices to show that u has no common neighbors with x_2, x_3 or x_4. It easily follows that u and x_2 have no common neighbors (otherwise, a cycle of length 3 or 5 is formed). Suppose that u and x_3 have a common neighbor w. Consider the 6-cycle $(vuwx_3x_2x_1)$. Recall that G has no induced 6-cycles; thus one of the pairs ux_2, vx_3, wx_1 is an edge in G. But ux_2 is not an edge, by the assumption that u has no neighbor in C, and if one of vx_3 and wx_1 is an edge, a contradiction to Corollary 2.2.4 is reached. Finally, if u and x_4 have a common neighbor w, then the set $\{u, v, w, x_1, \dots, x_7\}$ induces a graph that consists of two 7-cycles whose intersection is a path of length 3, contradicting Lemma 2.2.5. □

Proof of Theorem 2.2.1. Suppose that the theorem is false and choose a vertex u and a 7-cycle C which minimise the distance between u and C such that u has no neighbor in C. Since G must be connected, it easily follows that there is a path of length two between u and C. Therefore, we may assume without loss of generality that u has no neighbors in the 7-cycle $C = (x_1 \dots x_7)$ and v is a common neighbor of u and x_1. Since u is not joined to x_2 and G is maximal $\{C_3, C_5\}$-free, there is a 4-path $uy_1y_2y_3x_2$ between u and x_2 (a 2-path would create a 5-cycle). We note that y_1 cannot be joined to x_1, otherwise a 5-cycle is formed, so in particular $y_1 \neq v$. Thus, by Observation 2.2.7, y_1 has no neighbors in C. We note that no two of the four vertices $\{u, x_2, x_3, x_6\}$ have a common neighbor; this follows from Observation 2.2.7 and the assumption that G is $\{C_3, C_5\}$-free. It follows from the minimum degree condition that y_1 has a common neighbor with one of u, x_2, x_3, x_6. But y_1 does not have a common neighbor with either u or x_2 (otherwise, a cycle of length 3 or 5 if formed). Thus y_1 has a common neighbor with either x_3 or x_6. Assume that y_1 has a common neighbor with x_3 (with x_6). Then, by Observation 2.2.7, y_1 has no common neighbors with any other vertex in C. It follows that no two of the vertices in $\{u, y_1, x_2, x_5, x_6\}$ (in $\{u, y_1, x_3, x_4, x_7\}$) have a common neighbor, a contradiction to the minimum degree condition. $\square$

In the next three sections we shall prove Lemma 2.2.2, Lemma 2.2.3, and Lemma 2.2.5. The general strategy is the following. We want to show that some graph F cannot appear in a maximal $\{C_3, C_5\}$-free graph G of large minimum degree. If F is a subgraph of G, and if every vertex has a 'small' number of neighbors in F, then double counting the edges between $V(F)$ and $V(G) \setminus V(F)$ will produce a contradiction with the minimum degree condition. Often the original target graph F does not satisfy this goal, and we shall need to pass to some suitable subgraph of F which meets our needs. This requires detailed analysis of the possible neighborhoods of vertices in F (or some subgraph of F).

2.3 No induced 6-cycles

Brandt and Ribe-Baumann [15] stated that maximal $\{C_3, C_5\}$-free graphs of high minimum degree forbid induced 6-cycles. However, they did not provide a proof. In this section, we prove that induced 6-cycles do not appear in dense graphs of any given odd-girth which is at least 7. The proof of Lemma 2.2.2 is broken up into the next three subsections. A brief sketch of the proof goes as follows. Assuming there is an induced C_6 in G, we use the edge-maximality of G to conclude that there must exist three paths joining the diametrically opposite pairs of vertices in the 6-cycle. We shall argue that each of these paths must have length precisely $2k-2$, and are pairwise vertex disjoint. Call this resulting graph H. The idea is then to show that each vertex of G must send 'few' neighbors into H, or to some small subgraph of H. In order to execute this part of the proof, we need to show that some auxiliary graphs do not appear in dense odd-girth $2k+1$

graphs. These auxiliary graphs can be formed from suitable subdivisions of K_4, which we describe next.

2.3.1 No graphs $\Theta_{k,\ell}$ and $\Theta^*_{k,\ell}$

For an integer $k \geq 3$ and $\ell \in [k-1]$ we define the graphs $\Theta_{k,\ell}$ and $\Theta^*_{k,\ell}$ from a particular subdivision of K_4 as follows. Label the branch vertices of the subdivision as a,b,c,d (written counterclockwise). Subdivide the edges ab,cd with an additional $2k-3$ vertices each; call the resulting paths $P_{ab} = ax_1 \ldots x_{2k-3}b$ and $P_{cd} = cy_1 \ldots y_{2k-3}d$, respectively. Subdivide the edge ac with a new vertex e, and subdivide bd with a new vertex f. Call this graph Θ_k. Note that each of the paths P_{ab} and P_{cd} creates two $(2k+1)$-cycles. For example, P_{ab} creates the $(2k+1)$-cycles $C_{abce} := (aP_{ab}bcea)$ and $C_{abfd} := (aP_{ab}bfda)$. We shall label the other two cycles analogously. Construct a 'diagonal' path $D = tuv$ such that $t = x_\ell \in P_{ab}$ and $v = y_\ell \in P_{cd}$. We shall call this graph $\Theta_{k,\ell}$. Finally, $\Theta^*_{k,\ell}$ shall denote the graph obtained from $\Theta_{k,\ell}$ by adding a 'reflected' diagonal path $D' = t'u'v'$ such that $t' = y_{2k-2-\ell} \in P_{cd}$ and $v' = x_{2k-2-\ell}$. Note that $l(t'P_{cd}d) = l(v'P_{ab}b) = \ell$. A priori, u could be any common neighbor of t,v, and u' could be any common neighbor of t',v'. However, if $\Theta_{k,\ell}$ or $\Theta^*_{k,\ell}$ is a subgraph of an odd-girth $2k+1$ graph, then we have the following.

Observation 2.3.1. *Suppose that G is a graph of odd-girth $2k+1$ and suppose that $\Theta_{k,\ell} \subset G$ and $\Theta^*_{k,\ell} \subset G$ are labelled as above. Then $u \notin V(\Theta_k)$ and $u' \notin V(\Theta_{k,\ell})$.*

Proof. Every 'face' of Θ_k and $\Theta_{k,\ell}$ is a $(2k+1)$-cycle. □

Before proving the main lemma in this section we need one small result concerning neighborhoods in the graphs Θ_k.

Proposition 2.3.2. *Let $k \geq 3$, G be a graph of odd-girth $2k+1$, and suppose $H \subset G$ is a copy of Θ_k labelled as above. Then the only possible 3-vertex neighborhoods in $int(P_{ab}) \cup int(P_{cd})$ are (up to relabelling) of the form $\{x_i, x_{i+2}, y_{2k-3-i}\}$ for $i \in [2k-5]$.*

Proof. Let w be a vertex of G and suppose it has 3 neighbors $w_1, w_2, w_3 \in \mathrm{int}(P_{ab}) \cup \mathrm{int}(P_{cd})$. We may assume $w_1, w_2 \in \mathrm{int}(P_{ab})$ and $w_3 \in \mathrm{int}(P_{cd})$. Consider the paths $P_{ab}^+ = a \ldots w_1$, $P_{ab}^- = w_2 \ldots b$, $P_{cd}^- = c \ldots w_3$, and $P_{cd}^+ = w_3 \ldots d$, with lengths l_{ab}^+, l_{ab}^-, l_{cd}^-, and l_{cd}^+, respectively. We have $w_3 = y_j$ for some $j \in [2k-3]$ and $w_1 = x_i$ for some $1 \leq i \leq 2k-5$, and therefore by Proposition 2.2.6 (using the $(2k+1)$-cycle C_{abfd}), $w_2 = x_{i+2}$. Then

- $l_{ab}^+ = i$ and $l_{ab}^- = 2k-4-i$.
- $l_{cd}^- = j$ and $l_{cd}^+ = 2k-2-j$.

Now, each pairing of an $a-b$-path with a $c-d$-path yields a cycle. For instance, pairing P_{ab}^+ with P_{cd}^- yields the cycle $C_{+,-} := (ww_1 P_{ab}^+ aec P_{cd}^- w_3 w)$ which has length $l_{ab}^+ + l_{cd}^- + 4 = i + j + 4$. Similarly, we have

- $C_{+,+} = (ww_1 P_{ab}^+ ad P_{cd}^+ w_3 w)$ with length $l_{ab}^+ + l_{cd}^+ + 3 = 2k+1+i-j$.
- $C_{-,+} = (ww_2 P_{ab}^- bfd P_{cd}^+ w_3 w)$ with length $l_{ab}^- + l_{cd}^+ + 4 = 4k-2-(i+j)$.
- $C_{-,-} = (ww_2 P_{ab}^- bc P_{cd}^- w_3 w)$ with length $l_{ab}^- + l_{cd}^- + 3 = 2k-1+j-i$.

Now, since $l(P_{ab}) = l(P_{cd}) = 2k-2$ is even, we have that P_{ab}^+, P_{ab}^- have the same parity, and P_{cd}^+, P_{cd}^- have the same parity. Suppose first that P_{cd}^- is even. If P_{ab}^+ is odd, then the cycles $C_{+,-}$ and $C_{-,+}$ are odd. So if $i+j \leq 2k-4$, then $l(C_{+,-}) \leq 2k$, a contradiction with the odd-girth assumption. Otherwise, $l(C_{-,+}) \leq 4k-2-(2k-3) \leq 2k+1$, which is a contradiction unless $i+j = 2k-3$, i.e., $j = 2k-3-i$. If P_{ab}^+ is even then the cycles $C_{+,+}$ and $C_{-,-}$ are odd. If $j > i$, then since i and j are even, we have $j - i \geq 2$ and so $l(C_{+,+}) \leq 2k-1$, a contradiction. Therefore, $i \geq j \geq 2$. But if $i \geq j$, then $l(C_{-,-}) \leq 2k-1$, again a contradiction.

Thus, we may suppose that P_{cd}^- is odd. We shall proceed along the lines of the argument in the previous paragraph. Suppose that P_{ab}^+ is even. Then $C_{+,-}$ and $C_{-,+}$ are odd cycles, and the same argument as in the previous paragraph applies: we get a contradiction unless $j = 2k-3-i$. We may therefore assume that P_{ab}^{+} is an odd path. It follows that $C_{+,+}$ and $C_{-,-}$ are odd cycles. Similarly as before, if $j \geq i$ then since they are both odd, we have $j \geq i+2$, so $l(C_{+,+}) \leq 2k-1$; and if $i \geq j$, then $l(C_{-,-}) \leq 2k-1$. It follows that if w has 3 neighbors in $\mathrm{int}(P_{ab}) \cup \mathrm{int}(P_{cd})$, then they are precisely x_i, x_{i+2}, y_{2k-3-i} for some $i \in [2k-5]$, as claimed. □

In what follows we shall show that $\Theta_{k,\ell}$ for $\ell \in [2, k-1]$ and $\Theta_{k,1}^{*}$ cannot appear as a subgraph of a graph with odd-girth $2k+1$ and minimum degree larger than $\frac{n}{2k-1}$. This is the key tool to show that induced 6-cycles cannot appear as subgraphs either.

Lemma 2.3.3. *Let $k \geq 3$ be an integer and let G be an odd-girth $2k+1$ graph on n vertices with $\delta(G) > \frac{n}{2k-1}$. Then the following hold:*

1. *$\Theta_{k,\ell}$ is not a subgraph of G, for any $\ell \in [2, k-1]$.*
2. *$\Theta_{k,1}^{*}$ is not a subgraph of G.*

Proof. Let G be as in the statement of the Lemma, and suppose that there is a copy Θ of $\Theta_{k,\ell}$ in G, for some $1 \leq \ell \leq k-1$. Moreover, if $\ell = 1$, then we shall assume Θ is actually a copy of $\Theta_{k,1}^{*}$. Label this copy as usual, and note that by Observation 2.3.1, $u \notin P_{ab} \cup P_{cd} \cup \{e, f\}$ (and $u' \notin P_{ab} \cup P_{cd} \cup \{e, f, u\}$, if $\ell = 1$). Let $H = G[V(\Theta) \setminus \{e, f, u\}] = G[V(P_{ab}) \cup V(P_{cd})]$ if $\ell \geq 2$; if $\ell = 1$, H shall denote the graph induced on $V(\Theta) \setminus \{e, f, u, u'\}$. Observe that $|H| = 2(2k-1)$. There are a number of paths and cycles that will be of interest to us during the course of the proof. In addition to the four $(2k+1)$-cycles $C_{abce}, C_{abfd}, C_{cdae}$, and C_{cdfb} defined earlier, we shall consider the following:

- $C_{+}^{\ell} = (tuvP_{cd}daP_{ab}t)$, $C_{-}^{\ell} = (tuvP_{cd}cbP_{ab}t)$ — the two $(2k+1)$-cycles created by the addition of the diagonal path $D = tuv$.

- $P_{ab}^{+} = aP_{ab}t$, $P_{ab}^{-} = tP_{ab}b$.
- $P_{cd}^{-} = cP_{cd}v$, $P_{cd}^{+} = vP_{cd}d$.

With notation out of the way, the following claim asserts that every vertex of G has few neighbors in H.

Claim 2.3.4. *Every vertex of G has at most 2 neighbors in H.*

Proof. Let $w \in V(G)$. We shall break the proof up according to the number of neighbors of w in the set of branch vertices $X = \{a,b,c,d\}$. Note that w can have at most 2 neighbors in X.

So let us suppose first that $|N(w,X)| = 2$; by symmetry, we may assume $N(w,X) = \{a,c\}$. Then w can have no other neighbors in $P_{ab} \cup P_{cd}$, since otherwise w would have 3 neighbors in one of the $(2k+1)$-cycles C_{cdae}, C_{abce}, contradicting Proposition 2.2.6.

Now assume $|N(w,X)| = 1$ and, by symmetry, let $N(w,X) = \{a\}$. Suppose that w has 2 additional neighbors $w_1, w_2 \in \mathrm{int}(P_{ab}) \cup \mathrm{int}(P_{cd})$, and note that by Proposition 2.2.6 we may assume $w_1 \in P_{ab}$ and $w_2 \in P_{cd}$. Suppose first that $w_1 \in P_{ab}^{+}$. Then, since C_{+}^{ℓ} is a $(2k+1)$-cycle and by Proposition 2.2.6, it follows that $w_2 \in P_{cd}^{-}$. However, this contradicts the fact that C_{cdae} is well-behaved. Therefore, we may assume that $w_1 \in P_{ab}^{-}$ so that $\ell = 1, w_1 = x_2$. In particular, since $\ell = 1$, we are assuming that Θ is a copy of $\Theta_{k,1}^{*}$ with 'reflected diagonal' $D' = t'u'v'$ where $t' = y_{2k-3}$ and $v' = x_{2k-3}$. Now, we cannot have $w_1 = v$; otherwise the 5-cycle $(wvutw_1w)$ is formed (recall that we are assuming $k \geq 3$, so in particular G has odd-girth at least 7). It follows that $w_2 \in P_{cd}^{+}$, and so w_2 belongs to the $(2k+1)$-cycle C_{+}^{ℓ}. By Proposition 2.2.6 we must have that $w_2 = y_{2k-3} = t'$. However, this creates the cycle $(wt'u'v'P_{ab}w_1w)$ which has length exactly $2k-1$, a contradiction. Hence, we have shown that w can have at most 2 neighbors in H whenever $|N(w,H)| = 1$.

Finally, assume that $N(w,X) = \emptyset$ so that $N(w,H) \subset \mathrm{int}(P_{ab}) \cup \mathrm{int}(P_{cd})$. Since $V(H) \cup \{e,f\}$ induces a copy of Θ_k in G, we have that if w has 3 neighbors in H, then they are x_i, x_{i+2}, and y_{2k-3-i} for some $i \in [2k-5]$ by Proposition 2.3.2. Now, we cannot have $i \leq \ell$ as then $x_i, y_{2k-3-i} \in C_+^\ell$; but since they are both in the interiors of their respective paths, they cannot satisfy Proposition 2.2.6 applied to C_+^ℓ. It follows that we have $i > \ell$ and $x_i, x_{i+2} \in C_-^\ell$, $y_{2k-3-i} \in C_+^\ell$. Consider the cycle $C = (wy_{2k-3-i}P_{cd}vutP_{ab}x_iw)$, which has length precisely $((2k-2-\ell)-(i+1))+(i-\ell)+4 = 2k-3-2\ell+4$, which is odd and has length at most $2k-1$, a contradiction. Accordingly, w can have at most 2 neighbors whenever $N(w,X) = \emptyset$. This completes the proof of Claim 2.3.4.

□

Using Claim 2.3.4, we can complete the proof of Lemma 2.3.3. Indeed, let us double count the edges between $V(H)$ and $V(G) \setminus V(H)$. Recall that $|H| = 2(2k-1)$. By the minimum degree condition,

$$e(H, G \setminus H) > 2(2k-1)\left(\frac{n}{2k-1} - 2\right) = 2n - 4(2k-1).$$

On the other hand, by Claim 2.3.4, $e(H, G \setminus H) \leq 2(n - 2(2k-1)) = 2n - 4(2k-1)$, a contradiction. This completes the proof of Lemma 2.3.3. □

2.3.2 Neighborhoods in Φ_6^k

Let Φ_6^k be the following graph. Let $C = (aecbfda)$ be a 6-cycle (labelled counterclockwise). Then Φ_6^k is the graph obtained from C by adding pairwise vertex disjoint $(2k-2)$-paths $P_{ab} = ax_1 \dots x_{2k-3}b$, $P_{cd} = cy_1 \dots y_{2k-3}d$, and $P_{ef} = ez_1 \dots z_{2k-3}f$, whose interiors are vertex disjoint from C. We shall call C the *outer cycle* of Φ_6^k. Later (see Section 2.3.3) we shall show that if an induced C_6 appears in a maximal odd-girth $(2k+1)$ graph G, then Φ_6^k must be a subgraph of G. The aim of this section is to characterize the possible neighborhoods in Φ_6^k whenever it appears as a subgraph of a graph with odd-girth $2k+1$.

Suppose that G has odd-girth $2k+1$ and minimum degree $\delta(G) > \frac{n}{2k-1}$, and let H be a copy of Φ_6^k in G. If every vertex of G has at most 3 neighbors in H, then we arrive at the following contradiction:

$$3(2k-1)\left(\frac{n}{2k-1}-3\right) < e(H, G\setminus H) \leq 3(n-3(2k-1)),$$

where we have used the minimum degree condition and the fact that $|H| = 3(2k-1)$. Unfortunately, there are possible 4-vertex neighborhoods in H. Our aim in this subsection is to classify such neighborhoods. We shall first show that if a vertex has either 0 or at least 2 neighbors in the outer cycle C of H, then it must have at most 3 neighbors in H. To state this result, if H is a copy of Φ_6^k with outer cycle C, then we shall denote by $\mathrm{int}(H)$ the graph $G[V(H)\setminus V(C)]$.

Lemma 2.3.5. *Let $k \geq 3$ be an integer, and let G be an n-vertex graph with odd-girth $2k+1$ and $\delta(G) > \frac{n}{2k-1}$. Suppose that $H \subset G$ is a copy of Φ_6^k with outer cycle C. If w is a vertex of G such that either $|N(w,C)| \geq 2$ or $N(w,H) \subset int(H)$, then w has at most 3 neighbors in H.*

Proof. Let $H \subset G$ be labelled as in the definition of Φ_6^k, and suppose first that $w \in V(G)$ such that $|N(w,C)| \geq 2$; in particular, we may assume that w is joined to a and c. Recall that each path P_{ab}, P_{cd}, P_{ef} creates two $(2k+1)$-cycles. For example, P_{ab} creates the $(2k+1)$-cycles $C_{abce} = (aP_{ab}bcea)$ and $C_{abfd} = (aP_{ab}bfda)$. We label the other cycles analogously. Since w is joined to both a and c, it follows by Proposition 2.2.6 that w has no more neighbors in P_{cd} nor in P_{ab} (simply consider the $(2k+1)$-cycles C_{cdae} and C_{abce}). Therefore, w can only have additional neighbors in P_{ef}. However, if w has 2 neighbors in P_{ef}, then it has 3 neighbors in the $(2k+1)$-cycle C_{efda}, a contradiction with Proposition 2.2.6. So, w has at most one additional neighbor in P_{ef}.

Let us assume now that $N(w,H) \subset \text{int}(H)$. We shall show that $|N(w,H)| \leq 3$. Indeed, if w has 2 neighbors in the interior of one of the paths P_{ab}, P_{cd}, P_{ef}, then it must have no other neighbors in $\text{int}(H)$. Otherwise, if w has 2 neighbors in, say $\text{int}(P_{ab})$, then look at the subdivision Θ_k with branch vertices $\{a,b,c,d\}$, and note that by Proposition 2.3.2, its neighbors are from the set $\{x_i, x_{i+2}, y_{2k-3-i}\}$, for some $i \geq 1$. On the other hand, looking at the the copy of Θ_k with branch vertices $\{a,b,f,e\}$, we see that w's neighbors belong to the set $\{x_i, x_{i+2}, z_{i+1}\}$. However, we claim that w cannot be joined to both y_{2k-3-i} and z_{i+1}. Indeed, look at the copy of Θ_k with branch vertices $\{c,d,f,e\}$. If w is joined to y_{2k-3-i} and z_{i+1}, then we obtain a copy of $\Theta_{k,i+1}$ (with diagonal path $z_{i+1}wy_{2k-3-i}$), contradicting the first part of Lemma 2.3.3. Therefore, w has at most one neighbor in the interior of each of the paths P_{ab}, P_{cd} and P_{ef}, or has at most 2 neighbors in $\text{int}(H)$. This completes the proof of Lemma 2.3.5. $\square$

By Lemma 2.3.5 we have that no vertex has more than 4 neighbors in H, and moreover, we may assume that if w has 4 neighbors in H, then $|N(w,C)| = 1$. The following asserts that there is essentially only one possible 4-vertex neighborhood in H.

Lemma 2.3.6. *The only* 4*-vertex neighborhood in H (up to isomorphism) is* $\{a, x_2, y_{2k-3}, z_1\}$.

Proof. By symmetry we may suppose that $N(w,C) = \{a\}$. Note that w cannot have 2 neighbors in the interior P_{ab}. Indeed, otherwise w has 3 neighbors in the $(2k+1)$-cycle C_{abce}, contradicting Proposition 2.2.6. Similarly, w cannot have 2 neighbors in the interior of P_{cd} or P_{ef}; so it has precisely one neighbor in each of P_{ab}, P_{cd}, and P_{ef}. Using the cycle C_{abce} and Proposition 2.2.6 we must have $N(w, \text{int}(P_{ab})) = \{x_2\}$. Similarly, using the appropriate $(2k+1)$-cycle we have that $N(w, \text{int}(P_{cd})) = \{y_{2k-3}\}$ and $N(w, \text{int}(P_{ef})) = \{z_1\}$, and the proof is complete. $\square$

2.3.3 Finishing the proof of Lemma 2.2.2

In this subsection, we complete the proof of Lemma 2.2.2.

Proof. Suppose that $C = (aecbfda)$ is an induced 6-cycle in G, where the vertices are labelled counterclockwise. Owing to the missing edge ab and the edge-maximality of G, there must exist an even path P_{ab} between a and b of length at most $2k-2$. Similarly, there exist paths P_{cd} and P_{ef} between c,d and e,f, respectively. Let $P_{ab} = ax_1 \dots x_r b$, $P_{cd} = cy_1 \dots y_s d$, and $P_{ef} = ez_1 \dots z_t f$, where $r,s,t \leq 2k-3$ are odd. Let $H = G[V(P_{ab}) \cup V(P_{cd}) \cup V(P_{ef})]$. Our first claim asserts that H is, in fact, a copy of the graph Φ_6^k.

Claim 2.3.7. *H is a copy of Φ_6^k.*

Proof. We first note that $V(P_{ab}) \cap V(C) = \{a,b\}$. Indeed, if otherwise, let i be maximal such that $x_i \in V(C)$. By symmetry, we may assume that $x_i \in \{e,c\}$. Suppose $x_i = e$. It follows that the path $x_i P_{ab} b$ is even and so the subpath $ax_1 \dots x_i$ must be even. But then $ax_1 \dots x_i a$ is an odd-cycle of length at most $2k-1$, contradicting the odd-girth assumption. If $x_i = c$, then $x_i P_{ab} b$ must be odd, so that $ax_1 \dots x_i$ is also odd. But then $ax_1 \dots x_i Ca$ is an odd closed walk of length at most $2k-1$, so contains an odd cycle of length at most $2k-1$. The same argument shows that $V(P_{cd}) \cap V(C) = \{c,d\}$ and $V(P_{ef}) \cap V(C) = \{e,f\}$. Observe that this implies $l(P_{ab}) = l(P_{cd}) = l(P_{ef}) = 2k-2$.

The only remaining task is to show that the paths P_{ab}, P_{cd}, P_{ef} are pairwise vertex disjoint. To the contrary, suppose that $V(P_{ab}) \cap V(P_{cd}) \neq \emptyset$, and let u be the first vertex in which they intersect. Let $i = l(aP_{ab}u), i' = l(uP_{ab}b)$ and $j = l(cP_{cd}u), j' = l(uP_{cd}d)$, so that $(i+j)+(i'+j') = 2(2k-2)$. Then without loss of generality, assume $i+j \leq 2k-2$. It follows that the cycle $(aP_{ab}uP_{cd}cea)$ has length at most $2k$, and so i and j must be of the same parity. On the other hand, since $l(P_{ab}) = l(P_{cd}) = 2k-2$ is even, it follows that i' and j' also have the same parity. Now, if $i+j' \leq 2k-2$, then we obtain a contradiction with the odd-girth assumption considering the cycle $(aP_{ab}uP_{cd}da)$. It follows that $i' + j \leq 2k-3$. Then $(uP_{cd}cbP_{ab}u)$ is an odd closed walk (which contains an odd cycle) of length at most $2k-2$, a contradiction. Therefore, $P_{ab} \cap P_{cd} = \varnothing$, and the other cases follow symmetrically.

$\square$

Now, as mentioned earlier, if every vertex has at most 3 neighbors in H, then we reach a contradiction by double counting the edges between H and $G-H$. Therefore, we may assume there is at least one vertex of degree 4 in H. In particular, by Lemma 2.3.6, we may assume that x_1y_{2k-3} and x_1z_1 are edges. We may therefore obtain the following subdivision of K_4: the branch vertices are $\{z_1, f, c, y_{2k-3}\}$; z_1f is subdivided with vertices $z_2, \ldots, z_{2k-3}$; cy_{2k-3} is subdivided with vertices $y_1, \ldots, y_{2k-4}$; z_1y_{2k-3} with x_1; z_1c with e; fy_{2k-3} with d; and fc with b. We shall denote by H' the graph induced in G on the vertex set of this subdivided K_4. The following claim asserts that certain pairs of vertices of H' cannot have a common neighbor in G.

Claim 2.3.8. *Let H' be as above. Then the pairs $\{y_i, z_{2k-i-2}\}$ for $i \in [2k-4]$ have disjoint neighborhoods in G.*

Proof. Consider the copy Θ of Θ_k induced on $V(P_{cd}) \cup V(P_{ef}) \cup \{a,b\}$. If y_1, z_{2k-3} have a common neighbor we construct a copy of $\Theta^*_{k,1}$ from Θ using the fact that z_1, y_{2k-3} have common neighbor x_1. If y_i and z_{2k-i-2} have a common neighbor for some $2 \leq i \leq 2k-4$ then we may construct a copy of $\Theta_{k,i}$ from Θ. This contradicts Lemma 2.3.3. $\square$

For ease of notation, let $x := x_1, z := z_1$, and $y := y_{2k-3}$. Furthermore, let $P_{zf} = zz_2 \ldots z_{2k-3}f$, $P_{cy} = cy_1 \ldots y_{2k-4}y$, and $P_{xb} = xx_2 \ldots x_{2k-3}b$. Let H'' be the graph induced on $V(P_{zf}) \cup V(P_{cy}) \cup \{x, b\}$ and note that $|H''| = 2(2k-2)+2 = 2(2k-1)$. Our final claim asserts that every vertex has at most 2 neighbors in H'', which leads to the usual contradiction via double counting.

Claim 2.3.9. *Every vertex of G has at most 2 neighbors in H''.*

Proof. Let w be a vertex of G and suppose that w has 3 neighbors w_1, w_2, w_3 in H''. First assume that $\{w_1, w_2, w_3\} \cap \{x, b\} \neq \emptyset$; without loss of generality suppose $w_1 = x$. Note that w cannot be joined to b since $l(P_{xb}) = 2k - 3$. By considering the appropriate $(2k+1)$-cycles and using Proposition 2.2.6, we have that the only possibility for w_2 and w_3 is $w_2 = z_2$ and $w_3 = y_{2k-4}$. However, this is impossible, since by Claim 2.3.8, z_2 and y_{2k-4} have no common neighbor. Accordingly, w can have at most 2 neighbors in H'' whenever $\{w_1, w_2, w_3\} \cap \{x, b\} \neq \emptyset$.

Hence, we may suppose that $\{w_1, w_2, w_3\} \subset V(P_{zf}) \cup V(P_{cy})$, and that $w_1, w_2 \in V(P_{zf})$, $w_3 \in V(P_{cy})$. We first claim that we may assume $\{w_1, w_2, w_3\} \subset \text{int}(P_{zf}) \cup \text{int}(P_{cy})$. Indeed, suppose otherwise that $w_1 = z$. Then we must have $w_2 = z_3$ and, by considering the $(2k+1)$-cycle $(fdyxzP_{zf}f)$, we also must have that $w_3 = y$. Then w has 3 neighbors in this $(2k+1)$-cycle, a contradiction.

Accordingly, we shall assume that $w_1, w_2 \in \text{int}(P_{zf})$ and $w_3 \in \text{int}(P_{cy})$. Suppose $w_1 = z_{2k-i-2}$ and $w_2 = z_{2k-i} = z_{2k-(i-2)-2}$ for some $i \in [2k-4]$, and let $w_3 = y_j$ for some j. By Claim 2.3.8, we have that $j \neq i, i-2$; moreover, a simple parity argument shows that $j \neq i-1$. So we have $j \notin \{i-2, i-1, i\}$. Hence we have that either $1 \leq j \leq i-3$ or $i+1 \leq j \leq 2k-4$. By symmetry of the graph H', we may assume the former holds. We shall consider the following paths:

- $P_{zf}^{+} = zP_{zf}w_1$; $P_{zf}^{-} = w_2P_{zf}f$.
- $P_{cy}^{+} = yP_{cy}y_j$; $P_{cy}^{-} = y_jP_{cy}c$.

Now, since $l(P_{zf}) = 2k-3$ is odd, one of P_{zf}^{+} and P_{zf}^{-} has different parity from P_{cy}^{-}; assume first this is P_{zf}^{+} and consider the cycle $C_0 := (wy_jP_{cy}cezP_{zf}w_1w)$. It has length precisely

$$
\begin{aligned}
l(C_0) = l(P_{cy}^-) + l(P_{zf}^+) + 4 &= j + (2k - i - 3) + 4 \\
&\leq (i-3) + (2k - i - 3) + 4 = 2k - 2,
\end{aligned}
$$

which is a contradiction, since C_0 is odd. Now, suppose that P_{zf}^- and P_{cy}^- have different parity and consider the cycle $C_0' := (wy_jP_{cy}cbfP_{zf}w_2w)$. Then $l(C_0') = l(P_{cy}^-) + l(P_{zf}^-) + 4 = j + (i-2) + 4 = i + j + 2$. Since C_0' is an odd-cycle, we reach a contradiction provided $i + j \leq 2k - 3$. Therefore, we may assume that $i + j \geq 2k - 2$. In that case, P_{cy}^+ and P_{zf}^+ have different parity and so the cycle $(wy_jP_{cy}yxzP_{zf}w_1w)$ is odd and has length

$$
\begin{aligned}
l(P_{cy}^+) + l(P_{zf}^+) + 4 &= (2k - i - 3) + (2k - 3 - j) + 4 \\
&= 4k - (i + j) - 2 \leq 4k - (2k-2) - 2 = 2k,
\end{aligned}
$$

a contradiction.

In all possibilities for $w_3 = y_j$ we have reached a contradiction with the odd-girth assumption. It follows that every vertex must have at most 2 neighbors in H'', completing the proof of Claim 2.3.9.

□

The proof of Lemma 2.2.2 is now complete using Claim 2.3.9: double counting the edges between H'' and $G - H''$ produces the usual contradiction.

□

2.4 12-cycles with few diagonals

Our aim in this section is to prove Lemma 2.2.3. We divide the proof into steps, according to the number of diagonals. Note that the case of having precisely five diagonals is immediate from Lemma 2.2.2 that forbids induced 6-cycles. The subsequent subsections deal with the remaining cases.

Proposition 2.4.1. *Let G be a maximal $\{C_3, C_5\}$-free graph on n vertices with $\delta(G) > n/5$. Then G has no 12-cycle with exactly four diagonals.*

Proof. Suppose that $(x_1 \dots x_{12})$ is a 12-cycle with exactly four diagonals. Let H be the graph induced by $\{x_1, \dots, x_{12}\}$. In light of Lemma 2.2.2, G has no induced 6-cycle, so we may assume that the edges $x_1x_7, x_2x_8, x_3x_9, x_4x_{10}$ are present in the graph and that x_5x_{11}, x_6x_{12} are non-edges. In fact, it is easy to verify that the only edges in H are the edges of the 12-cycle and the four aforementioned diagonals.

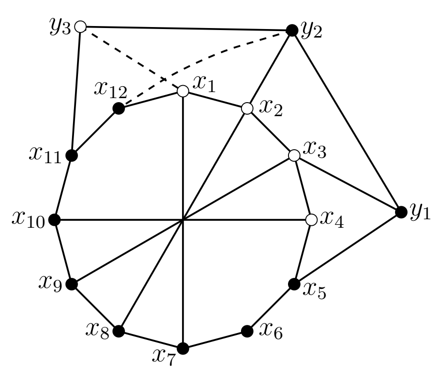

Figure 2.4: Constructing the graph H' in the case of four diagonals

Recall that a subgraph F of a graph G is called well-behaved if for every vertex u in G there is a vertex v in F such that $N_G(u, F) \subseteq N_F(v)$. Before returning to the proof, we make the following observation.

Claim 2.4.2. *H is well-behaved as a subgraph of G.*

Proof. We first point out that no vertex u in G can be adjacent to all of $\{x_4,x_6,x_{11}\}$. Indeed, otherwise, $(ux_{11}x_{12}x_1x_7x_6)$ is an induced C_6 (the addition of any chord to this cycle creates a triangle or a pentagon), contradicting Lemma 2.2.2. By symmetry, no vertex can be adjacent to all vertices in one of the following sets: $\{x_5,x_7,x_{12}\}$, $\{x_1,x_6,x_{11}\}$, $\{x_5,x_{10},x_{12}\}$. We conclude that no vertex can be adjacent to both x_6 and x_{11}. Indeed, by considering the 6-cycle $(x_1x_7x_6ux_{11}x_{12})$, since there is no induced C_6, u must be adjacent to x_1, contradicting the above. Similarly, no vertex is adjacent to both x_5 and x_{12}. One may check that any other possible neighborhood of a vertex of G in H is contained in the neighborhood of a vertex in H. □

The pair $\{x_5,x_{11}\}$ is a non-edge in G, and so, by the assumption that G is maximal $\{C_3,C_5\}$-free, there is a path of length two or four between x_5 and x_{11}. In fact, the length must be four because, otherwise, a cycle of length 3 or 5 will be created. Let $x_5y_1y_2y_3x_{11}$ be this 4-path. One may verify that $y_2 \notin V(H)$, and possibly $y_3 = x_{12}$ or $y_1 = x_6$, but not both. We shall assume, without loss of generality, that $y_1 \neq x_6$.

Claim 2.4.3. *We may assume that y_1x_3 is an edge in G.*

Proof. No two of the following vertices have a common neighbor: x_3,x_6,x_9,x_{12} (they are at distance one or three from each other). In other words, their neighborhoods are pairwise disjoint, and so, by the minimum degree condition, every vertex in G has a common neighbor with at least one of these four vertices. Note that y_2 does not have a common neighbor with either x_6 or x_{12} (this will create a C_5). By symmetry, we may assume that y_2 and x_3 have a common neighbor u. If $u = y_1$, Claim 2.4.3 follows. Thus, we suppose otherwise. Consider the 6-cycle $(uy_2y_1x_5x_4x_3)$. Since there are no induced 6-cycles, one of the following is an edge: y_1x_3, y_2x_4, ux_5. If y_1x_3 is an edge, the claim follows; y_2x_4 cannot be an edge (because of the 5-cycle $(y_2x_4x_{10}x_{11}y_3)$); if ux_5 is an edge, we replace y_1 by u to obtain the required property. □

Claim 2.4.4. *We may assume that y_2x_2 is an edge.*

Proof. As before, by considering the neighbors of x_2, x_5, x_8, x_{11}, we have that y_3 has a common neighbor with x_2 or x_8. If u is a common neighbor of y_3 and x_2, we may assume that $u \neq y_2$ (otherwise, we are done). By considering the 6-cycle $(ux_2x_3y_1y_2y_3)$, either y_2x_2 or uy_1 is an edge. We may assume that uy_1 is an edge. Then, by replacing y_2 by u we obtain the required property. Now suppose that y_3 and x_8 have a common neighbor u. By considering $(ux_8x_9x_{10}x_{11}y_3)$, u is adjacent to x_{10}. This, in turn, implies that u is adjacent to x_3 (see $(ux_8x_2x_3x_4x_{10})$), a contradiction: the 5-cycle $(ux_3y_1y_2y_3)$ is formed. □

Denote by H' the graph induced by $\{x_5, \ldots, x_{12}, y_1, y_2\}$ (see the black vertices in Figure 2.4). We shall show that every vertex of G has few neighbors in H', yielding a contradiction to the minimum degree condition on G. More precisely, we have the following:

Claim 2.4.5. *No vertex in G has more than two neighbors in H'.*

Proof. We note that by Claim 2.4.2, no vertex in G has more than two neighbors in $V(H') \cap V(H)$. Thus, if a vertex u has three neighbors in H', at least one of them is either y_1 or y_2. If u is adjacent to y_1, then the only other neighbors u can have in H' are x_6, x_9, x_{12}, but no two of these vertices may have a common neighbor. Similarly, if u is adjacent to y_2, its other possible neighbors in H' are x_5, x_8, x_{11}, no two of which have a common neighbor. The claim follows. □

Using Claim 2.4.5, we may now finish the proof of Proposition 2.4.1 by double counting the number of edges between H' and $V(G) \setminus V(H')$, as usual. □

Now we deal with the remaining case, of a 12 cycle with two or three diagonals, and thereby complete the proof of Lemma 2.2.3.

Proposition 2.4.6. *Let G be a maximal $\{C_3, C_5\}$-free graph on n vertices with $\delta(G) > n/5$. Then G induces no 12-cycle with two diagonals and at most one additional chord.*

Proof. Suppose that $C = (x_1 \dots x_{12})$ is a 12-cycle with two consecutive diagonals x_1x_7 and x_2x_8, and at most one additional chord. We note that any additional chord is a diagonal in one of the following 12-cycles $(x_1 \dots x_{12})$ or $(x_2 \dots x_7x_1x_{12} \dots x_8)$, both of which have two consecutive diagonals. Hence, and by symmetry, we may assume that the additional chord is either x_6x_{12} or x_5x_{11}. However, if x_5x_{11} is the additional chord, then $(x_1x_7x_6x_5x_{11}x_{12})$ is an induced 6-cycle, contradicting Lemma 2.2.2. Thus we assume that, if there is an additional chord, it is x_6x_{12}. Furthermore, if x_6x_{12} is not an edge, we assume that G contains no 12-cycles with two consecutive diagonals and exactly one extra chord. Let H be the graph induced by $\{x_1, \dots, x_{12}\}$ and denote $H' = H \setminus \{x_1, x_7\}$.

Claim 2.4.7. *Every vertex of G has at most two neighbors in H'.*

Proof. Observe that for any 7-cycle C and any vertex u in G, u has at most two neighbors in C, and if it does have two neighbors, they are at distance 2 in C. Suppose that u has three neighbors in H'. It follows by symmetry that u has two neighbors in $\{x_2, \dots, x_6\}$, which we can denote by x_{i-1} and x_{i+1} for some $i \in \{3,4,5\}$, and another neighbor x_j for some $j \in \{8, \dots, 12\}$. But then, by replacing x_i by u, we may assume that x_i is joined to x_j. This is a contradiction: either to Proposition 2.4.1 (if C had three chords, i.e. if x_6x_{12} is an edge, then now it has four chords); or, if x_6x_{12} is not an edge, to the assumption that there is no 12-cycle with two consecutive diagonals and an additional chord. □

Proposition 2.4.6 follows from Claim 2.4.7 by double counting the number of edges between H' and $V(G) \setminus V(H')$. The proof of Lemma 2.2.3 is therefore complete. □

2.5 Two 7-cycles intersecting in a 3-path

In this section we prove Lemma 2.2.5; that is, the graph in Figure 2.5 cannot appear as an induced subgraph of a maximal $\{C_3, C_5\}$-free graph on n vertices and minimum degree

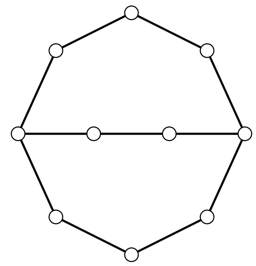

Figure 2.5: Two 7-cycles intersecting in a 3-path

larger than $n/5$. The proof follows the same strategy as the proofs in the previous two sections, though it requires more effort.

Proof of Lemma 2.2.5. Suppose that H is an induced subgraph of G which is the union of two 7-cycles intersecting in a path of length 3. Denote the two 7-cycles by $(x_1x_2x_3x_4x_5x_6x_7)$ and $(x_1x_2x_3x_4x_8x_9x_{10})$ (see Figure 2.5). We start by showing that H is a well-behaved subgraph of G.

Claim 2.5.1. *The graph H is well-behaved.*

Proof. Suppose that H is not well-behaved. Then, up to relabelling, one of the two following pairs has a common neighbor in G: $\{x_6, x_9\}$ or $\{x_5, x_{10}\}$. If u is a neighbor of x_6 and x_9 then, by Lemma 2.2.2, u is also a neighbor of x_1 (consider the 6-cycle $(ux_6x_7x_1x_{10}x_9)$). But then, x_9 has no neighbors in the 7-cycle $C = (x_1 \dots x_7)$ and its neighbor u has two neighbors in C (x_1 and x_6), contradicting Corollary 2.2.4. Now suppose that u is a neighbor of both x_5 and x_{10}. Consider the 6-cycle $(ux_5x_4x_8x_9x_{10})$. Since G contains no induced 6-cycle, u must be adjacent to x_8. Now consider the 7-cycle $(ux_{10}x_1x_2x_3x_4x_8)$. The vertex x_6 has no neighbors in C (x_6 cannot be adjacent to u), but x_5 has two neighbors in C (x_4 and u). This is a contradiction to Corollary 2.2.4. $\square$

Arguments as in Claim 2.5.1, using Corollary 2.2.4 and Lemma 2.2.2 will appear frequently in the proof of Lemma 2.2.5.

Since x_6 and x_8 are nonadjacent, there is a 4-path with ends x_6 and x_8 (a 2-path would create a C_5). Up to relabelling, three cases arise:

1. There is a 3-path $x_6y_1y_2x_9$ between x_6 and x_9. The vertices y_1 and y_2 are not in H.
2. There is a 3-path $x_7y_1y_2x_8$ between x_7 and x_8. The vertices y_1 and y_2 are not in H.
3. There is a 4-path $x_6y_1y_2y_3x_8$ between x_6 and x_8. The vertices y_1, y_2, y_3 are not in H.

In the rest of the proof, we show that each of the three cases is impossible, thus completing the proof of Lemma 2.2.5. Case 2 will be the most difficult to resolve. Case 1 follows since we have already shown, during the course of proving that induced 6-cycles do not exist, that the graph induced on $\{x_1, \ldots, x_{10}, y_1, y_2\}$ cannot exist as a subgraph (see Claim 2.3.8 and Claim 2.3.9).

2.5.1 **Case 2:** a 3-path between x_7 and x_8

Denote by H' the graph induced by $\{x_1, \ldots, x_{10}, y_1, y_2\}$ (see Figure 2.6).

Claim 2.5.2. *The graph H' is well-behaved.*

Proof. If H' is not well-behaved, then up to relabelling, y_1 and x_3 have a common neighbor u (recall that H is well-behaved by Claim 2.5.1). Consider the 6-cycle $(uy_1x_7x_1x_2x_3)$. Since there is no induced 6-cycle (Lemma 2.2.2), either y_1 is adjacent to x_2, or u is adjacent to x_1. The former case leads to a contradiction: then x_1 has two neighbors in the 7-cycle $(x_2x_3x_4x_5x_6x_7y_1)$ whereas its neighbor x_{10} has no neighbors there, contradicting Corollary 2.2.4. So, suppose the latter case holds, i.e. u is adjacent to x_1. But then u has two neighbors in the 7-cycle $(x_1x_2x_3x_4x_8x_9x_{10})$ whereas y_1 has none, a contradiction. □

As before, in light of the missing edge x_6x_{10}, one of the following three cases holds.

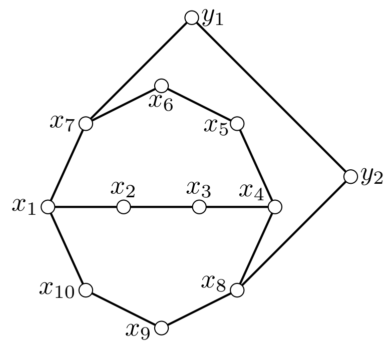

Figure 2.6: Case 2: a path of length 3 between x_7 and x_8

(a) There is a 3-path $x_6z_1z_2x_9$ between x_6 and x_9.

(b) There is a 3-path $x_5z_1z_2x_{10}$ between x_5 and x_{10}.

(c) There is a 4-path $x_6z_1z_2z_3x_{10}$ between x_6 and x_{10}.

However, (a) does not hold, as we have seen in the previous subsection. So it remains to consider (b) and (c).

Case 2b: 3-paths between x_7 and x_8 and between x_5 and x_{10}

Denote by F the graph induced by $\{x_1,\ldots,x_{10},y_1,y_2,z_1,z_2\}$. It is easy to check that the vertices y_1,y_2,z_1,z_2 are distinct. Now, as in Case 1, we have already shown that the graph induced on the subset $\{x_1,x_7,x_6,x_5,x_4,x_8,x_9,x_{10},y_1,y_2,z_1,z_2\} \subset V(F)$ cannot exist as a subgraph (see Claim 2.3.8 and Claim 2.3.9). Therefore we may dispel with Case 2b.

Case 2c: a 3-path between x_7 and x_8 and a 4-path between x_6 and x_{10}

Denote by F the graph induced by $\{x_1,\ldots,x_{10},y_1,y_2,z_1,z_2,z_3\}$ (see Figure 2.7).

Claim 2.5.3. *The only edges spanned by F are those spanned by H and the edges of the two paths $x_6z_1z_2z_3x_{10}$ and $x_7y_1y_2x_8$.*

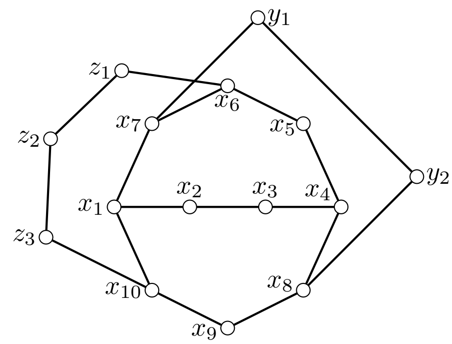

Figure 2.7: Case 2c: the graph F

Proof. First we note that y_1 and y_2 do not have additional neighbors in $\{x_1,\dots,x_{10}\}$. Indeed, by symmetry we assume that y_1 has an additional neighbor in H. The only possible such neighbor is x_2. We reach a contradiction to Corollary 2.2.4 (consider the 7-cycle $(y_1x_2x_3x_4x_5x_6x_7)$ and the vertices x_1 and x_{10}).

We now show that z_1, z_2 and z_3 do not have additional edges into H. Using the fact that H' is well-behaved, the only possible additional neighbor of z_1 is x_4. But then, by replacing x_5 by z_1, we may assume that there is a 3-path from x_5 to x_{10}. This leads to a contradiction, as we have seen in Case 2b. Similarly, the possible additional neighbors of z_3 in H are x_8 and x_2. If z_3 is adjacent to x_8 then, by replacing x_9 by z_3, we may assume that there is a 3-path between x_6 and x_9, contradicting Case 1. If z_3 is adjacent to x_2 we reach a contradiction to Corollary 2.2.4 (x_1 has two neighbors in the 7-cycle $(z_3x_2x_3x_4x_8x_9x_{10})$ while x_7 has none). The possible neighbors of z_2 in H are x_3, x_5 and x_9. But z_2 is not adjacent to x_5 or x_9, because, otherwise, there is a 3-path between x_5 and x_{10} or between x_6 and x_9, contradicting previous cases. Furthermore, z_2 is not adjacent to x_3 because, otherwise, $(z_1z_2x_3x_4x_5x_6)$ is an induced 6-cycle, contradicting Lemma 2.2.2.

Finally, we show that there are no edges between $\{z_1,z_2,z_3\}$ and $\{y_1,y_2\}$. The only such edges that do not create a triangle or pentagon are z_1y_1 and z_2y_2. If z_1 is adjacent to y_1 we reach a contradiction to Corollary 2.2.4 (see $(z_1y_1y_2x_8x_4x_5x_6)$ and the vertices x_1, x_7), and if z_2y_2 is an edge, a contradiction to Lemma 2.2.2 is reached (consider the induced 6-cycle $(z_1z_2y_2y_1x_7x_6)$). This completes the proof of Claim 2.5.3. □

The following claim states that no vertex has more than three neighbors in F. Since $|F| = 15$, this is a contradiction to the minimum degree condition on G by the usual double counting argument, hence the proof of Lemma 2.2.5 in this case follows.

Claim 2.5.4. *No vertex has more than three neighbors in F.*

Proof. Since H' is well-behaved (see Claim 2.5.2; recall that H' is the graph induced by the set $\{x_1,\ldots,x_{10},y_1,y_2\}$) and has maximum degree 3, if there is a vertex u with four neighbors in F, it must be adjacent to at least one of z_1,z_2,z_3. We note that u cannot be adjacent to both z_1 and z_3 because then, by replacing z_2 by u, we may assume that z_2 has an additional edge in F, a contradiction to Claim 2.5.3. It follows that u has one neighbor among z_1,z_2,z_3 and at least three neighbors in H'. Since H' is well-behaved, u is adjacent to all three neighbors of a vertex v in H' of degree three (in H'). But then, by replacing v by u, we may assume that v has an additional edge in F, a contradiction to Claim 2.5.3. □

2.5.2 Case 3: a 4-path between x_6 and x_8

Denote by H' the graph induced by $\{x_1,\ldots,x_{10},y_1,y_2,y_3\}$, and let $H'' = H' \setminus \{x_5,x_7,y_3\}$ (see Figure 2.8).

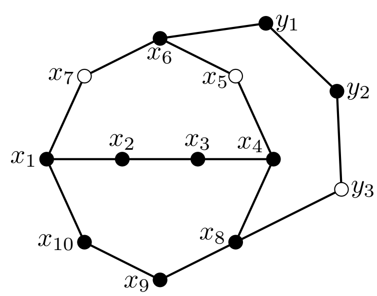

Figure 2.8: Case 3: the graphs H' and H''(marked in black)

Claim 2.5.5. *The only edges in H' are those spanned by H or by the path $x_6y_1y_2y_3x_8$.*

Proof. Suppose that there are additional edges. These must be between $\{y_1, y_2, y_3\}$ and $V(H)$. The only possible neighbor (that is not already accounted for) of y_1 in H' is x_1. But then, by replacing x_7 by y_1, we reach a contradiction to Case 2.

The only possible additional neighbors of y_3 in H are x_3 and x_{10}. If y_3 is adjacent to x_3, then x_4 has two neighbors in $(x_1x_2x_3y_3x_8x_9x_{10})$ whereas x_5 has none, a contradiction to Corollary 2.2.4. If y_3 is adjacent to x_{10} then, by replacing x_9 with y_3, we reduce to Case 1.

The only possible additional neighbors of y_2 in H are x_2, x_7, x_9. If y_2 is adjacent to x_7 or x_9 we reduce to previous cases. Finally, if y_2 is adjacent to x_2 then $(x_6x_7x_1x_2y_2y_1)$ is an induced 6-cycle, a contradiction to Lemma 2.2.2. $\square$

Claim 2.5.6. *No vertex in G has more than two neighbors in H''.*

Proof. Suppose that there is a vertex u in G with three neighbors in H''. Since H is well-behaved (see Claim 2.5.1), u must be a neighbor of either y_1 or y_2.

Suppose first that u is a neighbor of y_1. The other possible neighbors of u in H'' are x_2, x_3, x_9, x_{10}. Out of these four vertices, the only two that may have a common neighbor are x_2 and x_{10}. By considering the 6-cycle $(ux_2x_1x_7x_6y_1)$, it follows that u is adjacent also to x_7, i.e. u is adjacent to x_2, x_7, x_{10}, y_1. By replacing x_1 by u, we may assume that y_1 is adjacent to x_1, a contradiction to Claim 2.5.5.

We may now assume that u is adjacent to y_2. The other possible neighbors of u in H'' are $x_1, x_2, x_3, x_6, x_8, x_{10}$. If u is adjacent to x_6 or x_8, then by replacing y_1 or y_3 by u we see that u cannot have any additional neighbors in H'': otherwise we reach a contradiction to Claim 2.5.5. It follows that u is not adjacent to x_1, because otherwise, $(ux_1x_7x_6y_1y_2)$ is an induced 6-cycle. Similarly, u is not adjacent to x_{10} (see $(x_{10}x_9x_8y_3y_2u)$). This completes the proof of Claim 2.5.6, since the only remaining possible neighbors of u are x_2 and x_3, and these do not have a common neighbor. $\square$

By Claim 2.5.6, we reach a contradiction using the usual double counting argument. This completes the proof of Lemma 2.2.5. □

2.6 The proof of Theorem 2.1.4

In this section we shall finish the proof of Theorem 2.1.4 by combining Theorem 2.2.1 along with some facts we have obtained regarding forbidden substructures in maximal $\{C_3, C_5\}$-free graphs of large minimum degree. First, we prove the following proposition, which records several useful properties of the graphs F_d that we shall need in the sequel.

Proposition 2.6.1. *The following properties of F_d hold.*

1. *Let F be a copy of F_d in a maximal $\{C_3, C_5\}$-free graph G with $\delta(G) > n/5$. Then every vertex in G has either $d-1$ or d neighbors in F.*

2. *Let F be a copy of F_d in a maximal $\{C_3, C_5\}$-free graph G with $\delta(G) > n/5$. Denote the vertices of F by $x_1, \ldots, x_{5d-3}$ and its edges by the pairs $x_i x_j$ for which $|i-j| \equiv 1 \pmod 5$.*

   Then for every vertex u in G there is a vertex x_i in F such that the neighbors of u in F are the neighbors of x_i in F, except at most one of x_{i-1} and x_{i+1}. In particular, F is well-behaved as a subgraph of G.

Proof. We prove Property 1 by induction on d. For $d = 1$ the result is clear (recall that F_1 is an edge). For $d = 2$, the result easily follows from Theorem 2.2.1. So suppose that $d \geq 3$ and the result holds for smaller values of d. Let F be a copy of F_d in G as in the statement of Property 1, denote its vertices and edges as before, and let u be a vertex of G. Assume first that u has $d+1$ neighbors in F. If u has at most one neighbor in some consecutive interval $x_i, \ldots, x_{i+4}$ of five vertices, then u has at least d neighbors in the copy of F_{d-1} induced on $F \setminus \{x_i, \ldots, x_{i+4}\}$, a contradiction to the induction hypothesis. Therefore, u has at least two neighbors in every consecutive interval of five vertices. Suppose, without loss of generality, that u is adjacent to x_1. Then u has at least $1+2(d-2) \geq d$ neighbors (recall that $d \geq 3$) in the copy of F_{d-1} induced on $F \setminus \{x_{5d-7}, \ldots, x_{5d-3}\}$, a contradiction. If u has at most $d-2$ neighbors in F, one of which is, say, x_1, then u has at most $d-3$ neighbors in the copy of F_{d-1} induced on $F \setminus \{x_1, \ldots, x_5\}$, contradicting the induction hypothesis. It follows that u has either $d-1$ or d neighbors in F, as required.

Finally, let us prove Property 2. Let F and G be as before, and suppose that u has $d-1$ neighbors in F. Then one may find five consecutive vertices $x_\ell, \ldots, x_{\ell+4}$ which are not neighbors of u. Let F' be the copy of F_{d-1} given by $F \setminus \{x_\ell, \ldots, x_{\ell+4}\}$. Then by induction there is a vertex x of F' such that u is joined to all neighbors of x in F'. We claim that $x = x_{\ell-1}$ or $x = x_{\ell+5}$. Indeed, note that x must be adjacent to precisely one of $x_{\ell-1}, x_{\ell+5}$ (it cannot be adjacent to both); otherwise, u has no neighbor in the 7-cycle $(x_{\ell-1}x_\ell \ldots x_{\ell+5})$, contradicting Theorem 2.2.1. Suppose, without loss of generality, that x is joined to $x_{\ell-1}$. Since u must have a neighbor in the 7-cycle $(x_\ell x_{\ell+1} \ldots x_{\ell+6})$, u is also adjacent to $x_{\ell+6}$. It follows that $x = x_{\ell+5}$ and u has $d-1$ neighbors in F which are precisely the neighbors of $x_{\ell+5}$, except for $x_{\ell+4}$. Now, suppose that u has precisely d neighbors in F. Then we may find two neighbors of u that are at distance at most four. We claim that this implies there must be two neighbors at distance two apart. Indeed, they cannot be at distance three (this would produce a 5-cycle). So suppose these neighbors are at distance four and suppose they are x_i and x_{i+4}. Then $(ux_{i+4}x_{i+5}x_{i+6}x_i)$ is a 5-cycle in G, a contradiction. Accordingly, we may assume without loss of generality that u is adjacent to both x_2 and x_{5d-3}. Consider the copy of F_{d-1} given by $F \setminus \{x_3, \ldots, x_7\}$ and apply induction. Clearly, we must have u joined to x_7 (u's only possible neighbor in $\{x_3, \ldots, x_7\}$) and the neighborhood of u in F is precisely the neighborhood of x_1 in F. This completes the proof of Property 2. $\square$

We actually prove the following theorem, which clearly implies Theorem 2.1.4. It is the odd-girth 7 analogue of a result of Chen, Jin, and Koh [18] concerning triangle-free graphs of large minimum degree.

Theorem 2.6.2. *Let G be a maximal $\{C_3, C_5\}$-free graph on n vertices with $\delta(G) > n/5$. For every integer $d \geq 2$, if G contains no copy of F_d, then G is homomorphic to F_{d-1}.*

Proof. We shall use induction on d. For $d = 2$ we need to show that if G contains no copy of C_7, then G must be bipartite. Suppose otherwise and let $C = (x_1 \dots x_{2\ell+1})$ be an odd cycle in G of minimal length. Note that, by assumption, $\ell \geq 4$. Also, the minimality of C implies that it is induced. Since the edge x_1x_4 is missing there must be a 4-path connecting x_1 and x_4 (a 2-path is impossible). It follows that there is an odd closed walk of length at most 7 between x_1 and x_4 which contains an odd cycle of length at most 7. Hence G contains a C_7, contrary to our assumption, and so G must indeed be bipartite.

Now fix $d \geq 3$ and suppose the result holds for smaller values of d. Let G be as in the statement of the theorem and suppose it contains no copy of F_d. If G contains no copy of F_{d-1}, then by induction G is homomorphic to F_{d-2}. But F_{d-1} contains F_{d-2} (see Proposition 2.1.3), so we are done. Hence we may assume that G contains a copy of F_{d-1}. Let H be a vertex-maximal blow-up of F_{d-1} in G with vertex classes $X_1, \dots, X_{5d-8}$, where the edges of H are $X_i - X_j$ edges for which $|i-j| \equiv 1 \pmod 5$. Our aim is to show that G is a blow-up of F_{d-1}, or, in other words, that H spans all vertices in G. Note that by Property 1 in Proposition 2.6.1, every vertex in $V(G) \setminus V(H)$ has at most $d-1$ neighbors in F_{d-1}.

Suppose $u \in V(G) \setminus V(H)$ is adjacent to vertices in precisely $d-1$ of the classes of H. Without loss of generality, by Property 2, we may assume that these classes are those in the neighborhood of vertices in X_1, i.e., $X_2, X_7, \dots, X_{5d-8}$, and let $J = \{2, 7, \dots, 5d-8\}$ be the set of indices j such that u has a neighbor in X_j. We claim that u must be adjacent to every vertex in each of these classes, contradicting the assumption that H is a vertex-maximal blow-up in G. Suppose this is not the case. By Property 1, u has a non-neighbor in at most one of the sets X_j with $j \in J$ (indeed, otherwise we find a copy of F_{d-1} in which u has at most $d-3$ neighbors). Furthermore, by Property 2, we may assume that this set is X_2. Let $y \in X_2$ be a neighbor of u and let $z \in X_2$ be a non-neighbor of u.

Owing to the missing edge uz, and by the edge-maximality of G, there must exist a 4-path $uw_1w_2w_3z$ in G between u and z (a 2-path is impossible). Consider the $(5d-3)$-cycle $C = (uw_1w_2w_3zx_1x_{5d-8}\dots x_3y)$, where $x_i \in X_i$ (see Figure 2.9).

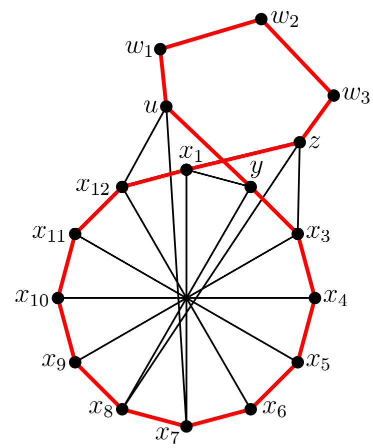

Figure 2.9: The $(5d-3)$-cycle C obtained from u and H, $d = 4$

Our aim is to show that $V(C)$ induces a copy of F_d, contrary to our assumption on G. Relabel the cycle C in order as $(z_0z_1\dots z_{5d-4})$, so that $z_0 = u, z_i = w_i$ for $i = 1,2,3$, $z_4 = z, z_5 = x_1$, $z_i = x_{5d-2-i}$ for $6 \le i \le 5d-5$, and $z_{5d-4} = y$. We must check that all chords of lengths $1+5t$ for $t = 0,\dots d-1$ are present in the graph induced on $V(C)$. Note that all possible chords of these lengths that are not incident with a vertex in $S = \{u,w_1,w_2,w_3\} = \{z_0,z_1,z_2,z_3\}$ are present, since all vertices in $V(C)\setminus S$ are in an appropriate copy of F_{d-1}. So we must check that all possible chords incident with a vertex in S are present. This is summarized in the following claim, where we temporarily revert to the original labelling of C:

Claim 2.6.3. *The following hold:*

- $N(u,V(C)) = \{w_1,y\} \cup \{x_{5\ell+2} : 1 \le \ell \le d-2\}$.

- $N(w_1, V(C)) = \{u, w_2\} \cup \{x_{5\ell+1} : 1 \le \ell \le d-2\}$.
- $N(w_2, V(C)) = \{w_1, w_3\} \cup \{x_{5\ell} : 1 \le \ell \le d-2\}$.
- $N(w_3, V(C)) = \{z, w_2\} \cup \{x_{5\ell-1} : 1 \le \ell \le d-2\}$.

Proof. Observe that the first item is immediate from our choice of u. Fix some ℓ with $1 \le \ell \le d-2$. Note that every vertex in X_2 is joined to $x_{5(\ell-1)+3} = x_{5\ell-2}$. In particular, y and z are joined to $x_{5\ell-2}$. Consider the 12-cycle $C' = (uw_1w_2w_3zx_{5\ell-2}\dots x_{5\ell+2}x_1y)$, with two consecutive diagonals $yx_{5\ell-2}$ and x_1z. Observe that C' gives rise to another 12-cycle $C'' = (uw_1w_2w_3zx_1x_{5\ell+2}x_{5\ell+1}\dots x_{5\ell-2}y)$ with two consecutive diagonals yx_1 and $ux_{5\ell+2}$. By Lemma 2.2.3, either C' or C'' has all of its diagonals present. However, it cannot be C', since u cannot be adjacent to $x_{5\ell-1}$. Therefore, C'' has all diagonals present: $w_1x_{5\ell+1}$, $w_2x_{5\ell}$, and $w_3x_{5\ell-1}$ are edges in G. This completes the proof of Claim 2.6.3. $\square$

It remains to check that Claim 2.6.3 produces chords of the right lengths. We do this for chords incident with w_1; the other cases follow identically. Indeed, $w_1 = z_1$ so we must check that z_1 is joined to $z_{1+(1+5t)}$ for $t = 0, 1, \dots, d-1$. This is obviously true for $t = 0$ and $t = d-1$, so let $1 \le t \le d-2$. Then the above is equivalent to w_1 being joined to $x_{5d-2-(1+(1+5t))} = x_{5(d-t-1)+1}$, where $1 \le d-t-1 \le d-2$, which clearly follows by Claim 2.6.3. Accordingly, there is a copy of F_d in G contrary to our assumption, so u must be adjacent to every vertex in X_j for all $j \in J$. But then we may place u in X_1 and produce a blow-up of F_{d-1} of larger order, which is impossible by our choice of H. It follows that every vertex in $V(G) \setminus V(H)$ is adjacent to vertices in at most $d-2$ of the sets X_i. In fact, by Item 2 of Proposition 2.6.1, it follows that every vertex in $V(G) \setminus V(H)$ is adjacent to precisely $d-2$ of the X_i's.

Before proceeding, let us introduce a bit of notation and terminology. Let $\widetilde{H}$ be the graph with vertex set $\{X_1, \dots, X_{5d-8}\}$, where an edge X_iX_j is present whenever the pair (X_i, X_j) induces a complete bipartite graph in G. As H is a blow-up of F_{d-1}, $\widetilde{H}$ is isomorphic to F_{d-1}. We say that a vertex v is *joined* to a subset $X \subseteq V(G)$ if v is adjacent to every vertex of X.

If a vertex v is joined to vertices in the neighborhood of X_i, then by Property 2 of Proposition 2.6.1 we have that v misses vertices in at most the two sets X_{i-1}, X_{i+1}; by symmetry, we may assume that each such vertex v misses X_{i-1}. Thus the following sets Y_i, where $i = 1, \ldots, 5d-8$, defined below, form a partition of $V(G) \setminus V(H)$ (see Figure 2.10). Note that each of these sets is independent (as G is triangle-free):

$$Y_i = \{u \in V(G) \setminus V(H) : u \text{ is joined to } X_{i+1}, X_{i+6}, \ldots, X_{i+5d-14} \text{ (indices modulo } 5d-8)\}$$

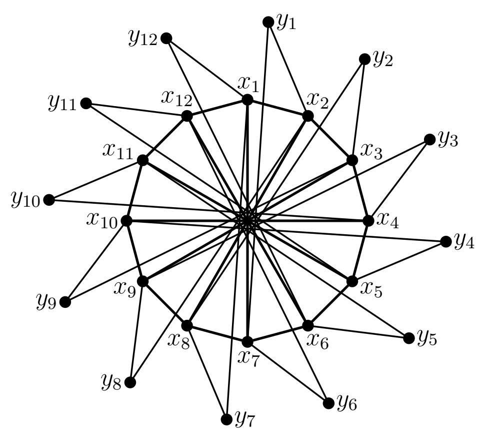

Figure 2.10: The sets X_i and Y_i

The next claim asserts that whether or not there are edges between Y_i and Y_j depends on whether or not there are edges between X_i and X_j. We are then able to 'absorb' Y_i into X_i, for each i.

Claim 2.6.4. *Let $d \geq 3$ and $1 \leq i, j \leq 5d-8$. If j is such that $X_j \notin N_{\widetilde{H}}(X_i)$, then there are no edges between Y_i and Y_j.*

Proof. Without loss of generality, set $i = 1$. Suppose j is such that $X_j \notin N_{\widetilde{H}}(X_1)$. We may assume that $j \neq 1$, as each Y_i is independent. Then $j = 5l + r$, where $l \in \{0, \dots, d-3\}$ and $r \in \{3,4,5,6\}$. Towards a contradiction, suppose there is an edge $y_1 y_j$ between Y_1 and Y_j. We consider four cases, according to the value of r. Suppose first that $r = 3$. Then we find the following 5-cycle $(y_1 x_2 x_{5l+3} x_{5l+4} y_j)$. If $r = 4$, we find the induced 6-cycle $(y_1 x_2 x_{5l+3} x_{5l+4} x_{5l+5} y_j)$. If $r = 5$, there is, again, an induced 6-cycle $(y_1 x_2 x_1 x_{5l+7} x_{5l+6} y_j)$. Finally, if $r = 6$, there is a 5-cycle $(y_1 x_2 x_{5l+8} x_{5l+7} y_j)$.

For each of the possible values of r, we reached a contradiction by showing that G contains either a 5-cycle or an induced 6-cycle. Claim 2.6.4 follows. $\square$

Let $Z_i = X_i \cup Y_i$. Note that the sets Z_i are independent and they partition $V(G)$. It follows from Claim 2.6.4 that there are no $Z_i - Z_j$ edges if $X_i X_j \notin E(\widetilde{H})$. By maximality of G, all $Z_i - Z_j$ edges are present if $X_i X_j \in E(\widetilde{H})$, implying that G is a blow-up of F_{d-1}. In particular, G is homomorphic to F_{d-1}, as required to complete the proof of Theorem 2.6.2. $\square$

We are then able to establish the following result, as stated in the Introduction, which gives a precise minimum degree condition (depending on d) for forcing a $\{C_3, C_5\}$-free graph to be homomorphic to F_{d-1}.

Proof of Corollary 2.1.5. Note that we may assume that G is maximal $\{C_3, C_5\}$-free. By Theorem 2.6.2, if G is not homomorphic to F_{d-1}, it contains a copy F of F_d. The number of edges between $V(F)$ and $V(G) \setminus V(F)$ is at most $d(n - (5d-3))$, since every vertex in G has at most d neighbors in F, by Proposition 2.6.1. It follows that there is a vertex u in F with at most $\frac{dn}{5d-3} - d$ neighbors outside of F. Since u has d neighbors in F, it follows that u has degree at most $\frac{dn}{5d-3}$, a contradiction to the minimum degree condition. $\square$

We remark that the proof of Theorem 2.1.4 rests heavily upon the fact that if G is an n-vertex maximal $\{C_3, C_5\}$-free graph with $\delta(G) > n/5$, then every vertex of G has a neighbor in every copy of C_7 in G. Let us say that a graph H is *attractive* in another graph

G if every vertex of G has at least one neighbor in every copy of H in G. Then the proof given in this section generalizes in a straightforward way to give the following statement.

Theorem 2.6.5. *Let $k \geq 2$ be an integer and suppose G is an n-vertex graph with odd-girth $2k+1$ such that C_{2k+1} is attractive in G. Then G is homomorphic to F_d^k for some d.*

2.7 Homomorphism thresholds

Recall that, given a family of graphs $\mathscr{H}$, the homomorphism threshold $\delta_{\text{hom}}(\mathscr{H})$ of $\mathscr{H}$ is the infimum of d such that every $\mathscr{H}$-free graph with n vertices and minimum degree at least dn is homomorphic to a bounded $\mathscr{H}$-free graph. In this section we provide the proof of Theorem 2.1.1, which states that the homomorphism threshold of $\{C_3, C_5\}$ is $1/5$. We also prove that $\delta_{\text{hom}}(C_5) \leq 1/5$ by showing that C_5-free graphs of large enough minimum degree are also triangle-free.

Proof of Theorem 2.1.1. Denote $\delta = \delta_{\text{hom}}(\{C_3, C_5\})$. First, we show that $\delta \geq 1/5$. We note that F_d is not homomorphic to a $\{C_3, C_5\}$-free graph H with fewer than $|F_d|$ vertices. Indeed, suppose otherwise. Then two vertices x and y in F_d are mapped to the same vertex u in H. By Proposition 2.1.3, there is a path P of length 1, 3 or 5 between x and y. Clearly, P cannot have length 1 (because the set of vertices mapped to the same vertex is independent). It follows that P has length 3 or 5. This implies that the path P is mapped to a cycle of length 3 or 5, a contradiction. It follows that, for each $d \geq 1$, F_d is a $\{C_3, C_5\}$-free graph with minimum degree at least $|F_d|/5$, which is not homomorphic to a $\{C_3, C_5\}$-free graph on fewer than $|F_d|$ vertices. Hence, indeed, $\delta \geq 1/5$.

It remains to show that $\delta \leq 1/5$. Let $\varepsilon > 0$ be fixed. Suppose that G is a $\{C_3, C_5\}$-free on n vertices and minimum degree at least $(1/5+\varepsilon)n$. Let d be such that $\frac{d}{5d-3} < 1/5+\varepsilon$. Then, by Corollary 2.1.5, G is homomorphic to F_{d-1}. This shows that $\delta \leq 1/5+\varepsilon$. Since ε was arbitrary, we conclude that $\delta \leq 1/5$. $\square$

It would be interesting to determine the homomorphism threshold of C_5. The following lemma enables us to easily obtain an upper bound.

Lemma 2.7.1. *Let G be a C_5-free graph on n vertices with $\delta(G) > n/6+1$. Then G is triangle-free provided n is sufficiently large.*

Before proving Lemma 2.7.1, we use it to prove Corollary 2.1.2, which provides an upper bound on the homomorphism threshold $\delta_{\text{hom}}(C_5)$. We currently do not have any nontrivial lower bound on $\delta_{\text{hom}}(C_5)$.

Proof of Corollary 2.1.2. Suppose that G is a C_5-free graph on n vertices and minimum degree at least $(1/5+\varepsilon)n$ for some fixed $\varepsilon > 0$. Then, by Lemma 2.7.1, G is also triangle-free. It follows from Theorem 2.1.1 that G is homomorphic to a C_5-free (and C_3-free) graph H of order at most $C = C(\varepsilon)$. Hence, indeed, $\delta_{\text{hom}}(C_5) \leq 1/5$. $\square$

We now turn to the proof of Lemma 2.7.1.

Proof of Lemma 2.7.1. We start by showing that every vertex in G is incident with at most 13 *triangular edges* (i.e. edges on triangles). To see this, suppose that u is incident with at least 14 triangular edges. In other words, the neighborhood $N(u)$ of u contains edges that span at least 14 vertices. The following claim implies that there is a set X of seven neighbors of u such that every vertex in X has a neighbor in $N(u) \setminus X$.

Claim 2.7.2. *Let H be a graph with n vertices and no isolated vertices. Then there is a set X of size at least $n/2$ such that every vertex in X has a neighbor outside of X.*

Proof. We note that it suffices to prove the claim under the assumption that H is connected. Indeed, for each component H_i of H, we may pick a set X_i as in the claim, and let X be the union of the X_i's. So now we assume that H is connected. Because of the assumption that there are no isolated vertices, we may assume that $|H| \geq 2$.

Let u be a vertex for which $H \setminus \{u\}$ is connected. Let v be a neighbor of u. Consider the graph $H' = H \setminus \{u, v\}$. Let $H_1, \dots, H_t$ be the connected components of H'. We pick a set X_i for each $i \in [t]$ as follows: if H_i consists of a single vertex x_i, then x_i must be adjacent to v, and we take $X_i = \{x_i\}$; otherwise, if H_i has at least two vertices, then by induction there is a set X_i of size at least $|H_i|/2$ such that every vertex in X_i has a neighbor outside of X_i (but in H_i). Let $X = \bigcup_{i=1}^{t} X_i \cup \{u\}$. It is easy to check that X satisfies the requirements of Claim 2.7.2. □

Let Y be a set of at most seven neighbors of u, which is disjoint from X and satisfies that every vertex in X has a neighbor in Y. Due to the minimum degree condition, we may find two distinct vertices x_1 and x_2 in X that have a common neighbor z outside of $X \cup Y \cup \{u\}$. Let $y \in Y$ be a neighbor of x_1. Then we find the 5-cycle $(x_1 y u x_2 z)$, a contradiction. Thus, indeed, every vertex is incident with at most 13 triangular edges.

We now show that G contains no two triangles that intersect in an edge.

Claim 2.7.3. *G contains no distinct vertices x_1, x_2, x_3, x_4 such that $x_1x_2x_3$ and $x_2x_3x_4$ are triangles.*

Proof. We note that x_1 and x_4 do not have common neighbors (apart from x_2 and x_3). Indeed, suppose that y is such a common neighbor. Then $(y x_1 x_2 x_3 x_4)$ is a 5-cycle. Similarly, x_1 and x_2 do not have a common neighbor. By symmetry, the following pairs do not have common neighbors (outside of x_1, x_2, x_3, x_4): $\{x_1, x_3\}$, $\{x_2, x_4\}$ and $\{x_3, x_4\}$. Finally, x_2 and x_3 have at most 13 common neighbors (because every vertex is incident with at most 13 triangular edges). Denote by N_i the set of neighbors of x_i, that are not in $\{x_1, x_2, x_3, x_4\}$ and are not common neighbors of x_2 and x_3.

The following properties of N_i holds.

- The sets N_i are pairwise disjoint.
- $|N_i| \geq n/6 - 14$ for $i \in [4]$.

- The edges induced by N_i span at most 13 vertices (by Claim 2.7.2).

- There are no edges between N_i and N_j for $i \neq j$.

  Indeed, if $i = 1$ and $j = 2$ then $(y_1 x_1 x_3 x_2 y_2)$ is a 5-cycle, a contradiction. A contradiction may be reached similarly for all other choices of i and j.

- There are no vertices with neighbors in two of the sets N_1, N_2, N_3.

  Indeed, suppose that z is a neighbor of y_1 and y_2 from N_1 and N_2. Then $(z y_1 x_1 x_2 y_2)$ is a 5-cycle, a contradiction.

Denote by M_i the set of neighbors of N_i (apart from $N_i \cup \{x_i\}$. We note that $|M_i| \geq n/6 - 14$. Indeed, every vertex in N_i has at most 13 neighbors in the neighborhood of x_i, thus all but 14 of its neighbors are in M_i. The sets $N_1, N_2, N_3, N_4, M_1, M_2, M_3$ are seven pairwise disjoint sets of size at least $n/6 - 14$, a contradiction. $\square$

Finally, suppose that G contains a triangle $x_1 x_2 x_3$. By Claim 2.7.3, the sets N_i of neighbors of x_i (outside of $\{x_1, x_2, x_3\}$) are disjoint. As in Claim 2.7.3, there are no edges between N_i and N_j for $i \neq j$. Similarly, no two of the sets N_1, N_2, N_3 have a common neighbor. Denote by M_i the sets of neighbors of N_i outside of $N_i \cup \{x_i\}$. Note that $|N_i| > n/6 - 1$ and $|M_i| > n/6$ (by Claim 2.7.2, there is a vertex u in N_i with no neighbors in N_i; all of u's neighbors, apart from x_i, are in M_i). The sets $N_1, N_2, N_3, M_1, M_2, M_3, \{x_1, x_2, x_3\}$ are pairwise disjoint. Thus, their union has size larger than n, a contradiction. It follows that G is triangle-free. $\square$

We remark that the minimum degree condition in Lemma 2.7.1 is best possible, as can be seen by the example depicted in Figure 2.11.

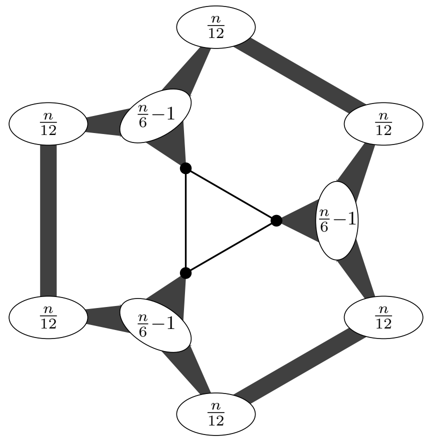

Figure 2.11: A C_5-free but not C_3-free graph with minimum degree $n/6+1$

2.8 Final remarks and open problems

We are able to determine precisely the structure of $\{C_3,C_5\}$-free graphs with high minimum degree, and thereby deduce the value of the homomorphism threshold $\delta_{\text{hom}}(\{C_3,C_5\})$. It would be very interesting to extend this result to $\{C_3,\ldots,C_{2k-1}\}$-free graphs. Recall that, for integers $k\geq 2, d\geq 1$, F_d^k is the graph obtained from a $((2k-1)(d-1)+2)$-cycle by adding all chords joining vertices at distances $j(2k-1)+1$ for $j=0,1,\ldots,d-1$. In light of our Theorem 2.1.4 it is natural to ask whether or not a $\{C_3,\ldots,C_{2k-1}\}$-free graph on n vertices with minimum degree larger than $\frac{n}{2k-1}$ is homomorphic to F_d^k for some d. In particular, is it true that every n-vertex maximal $\{C_3,C_5,\ldots,C_{2k-1}\}$-free graph G with $\delta(G)>\frac{n}{2k-1}$ has the property that every vertex has at least one neighbor in every copy of C_{2k+1} in G?

Rather surprisingly it turns out that this is false when $k\geq 4$ is even, as shown by the following construction due to Oliver Ebsen [21].

Suppose that $k\geq 4$ is even. Starting with a complete graph on 4 vertices, subdivide two independent edges by an additional $2k-6$ vertices and subdivide the remaining four edges by an additional two vertices each. Denote the resulting graph by T_k. It is easy to check that this graph is maximal $\{C_3,\ldots,C_{2k-1}\}$-free. To obtain large minimum degree assign weight 2 to each vertex of the original K_4 and to $k-4$ vertices of the 'long' subdivided edges, and assign weight 1 to the remaining vertices. This may be done in such a way that each vertex has weight 3 in its neighborhood (as k is even). To obtain an unweighted graph of order n simply blow up each vertex with an independent set of size proportional to its weight. Then the resulting graph T_k^* is maximal $\{C_3,\ldots,C_{2k-1}\}$-free and $\delta(T_k^*) = \frac{3n}{6k-4} > \frac{n}{2k-1}$. However, we claim that T_k is not homomorphic to F_d^k, for any d (and therefore no blow-up of T_k is homomorphic to any F_d^k). Indeed, suppose otherwise and let d_0 be minimal such that T_k is homomorphic to $F_{d_0}^k$ (obviously $d_0 \geq 2$). It is easy to check that, for $d \geq 2$ and any vertex v, $F_d^k - \{v\}$ is homomorphic to F_{d-1}^k. Since T_k is not homomorphic to $F_{d_0-1}^k$, it follows that T_k must be a blow-up of $F_{d_0}^k$ with all parts nonempty. Since T_k has precisely $4k$ vertices this implies that $d_0 \leq 3$, and so either T_k is a blow-up of $F_2^k = C_{2k+1}$ or of F_3^k. But no two vertices of T_k have the same neighborhood so the former case is impossible. Moreover, F_3^k has exactly $4k$ vertices as well, but clearly T_k is not isomorphic to F_3^k (as T_k is not regular, for example). It follows that T_k is not homomorphic to any F_d^k, as claimed.

We do not know whether Theorem 2.1.4 extends naturally to $\{C_3,\ldots,C_{2k-1}\}$-free graphs when $k \geq 5$ is odd, and it would be interesting to pursue this line of research further.

Recall that the homomorphism threshold of a family of graphs $\mathscr{H}$ is the infimum of d satisfying that every $\mathscr{H}$-free graph with n vertices and minimum degree at least dn is homomorphic to an $\mathscr{H}$-free graph of bounded order (depending on d but not on n). Despite the above remarks concerning the extension of Theorem 2.1.4 to general odd-girth graphs, we made the conjecture that $\delta_{\text{hom}}(\{C_3,C_5,\ldots,C_{2k-1}\}) = \frac{1}{2k-1}$ for all $k \geq 4$. This conjecture was proved correct by Ebsen and Schacht [20] very recently, during the preparation of this thesis.

We have also obtained an upper bound on $\delta_{\text{hom}}(C_5)$, namely, that it is at most $1/5$. We ask if it is true that $1/5$ is the correct value.

Question 2.8.1. *Is it true that $\delta_{\hom}(C_5) = 1/5$?*

In fact, any nontrivial (i.e. non-zero) lower bound on $\delta_{\hom}(C_5)$ would be interesting. In order to obtain such a lower bound, one would have to find, in particular, a family of graphs that have large minimum degree, are C_5-free and are not 4-colorable (indeed, otherwise, the graphs are homomorphic to K_4, which is clearly C_5-free). Although it is well known that such graphs exist, it seems hard to find explicit examples, especially with the added condition that they are not homomorphic to C_5-free graphs of bounded order.

Finally, what about the colorability of dense graphs of given odd-girth? Recall that the Erdős-Simonovits problem was to determine whether or not all n-vertex triangle-free graphs with $\delta(G) > n/3$ are 3-colorable. As we mentioned in the introduction to this chapter, this problem initiated an intensive amount of research, finally culminating in Brandt and Thomassé's result that all such dense triangle-free graphs are 4-colorable. They proved this by characterizing the structure of these graphs. The 4-colorable structures come from a family of graphs obtained from an induced 6-cycle. Since we proved that induced 6-cycles no longer appear in dense graphs of odd-girth $2k+1$ for $k \geq 3$, we believe the following conjecture might be true.

Conjecture 2.8.2. *Let $k \geq 3$ be an integer and suppose G is an n-vertex graph of odd-girth $2k+1$ and minimum degree larger than $\frac{n}{2k-1}$. Then $\chi(G) \leq 3$.*

Note that by Theorem 2.1.4, this conjecture is true for $k = 3$. However, we do not yet see how to prove Conjecture 2.8.2 for larger values of k.

CHAPTER 3

HIGHLY LINKED TOURNAMENTS WITH LARGE MINIMUM OUT-DEGREE

3.1 Introduction

A graph G is connected if for every two vertices there is path in G joining them. Connectivity problems are some of the most natural in graph theory. Viewing a graph as a communication network, how robust is the network to faulty nodes? In other words, if a set of nodes drop out from the network, does the network remain connected? Formalizing this notion of robustness, we say that a graph is *k-connected* if it remains connected after the removal of any set of $k-1$ vertices. Thus a graph is 1-connected if and only if it is connected. Menger's theorem provides a very useful characterization of k-connected graphs. It states that a graph is k-connected if and only if between any two distinct vertices there exist k internally vertex disjoint paths. This easily implies that if X and Y are any two disjoint sets of k vertices, then there exist k vertex disjoint paths from X to Y, and which are internally disjoint from $X \cup Y$: just form an auxiliary graph by adding two new vertices x and y, and adding all edges between x and X and between y and Y. However, we do not have control over the endpoints of these paths: we cannot specify that a given vertex of X must be joined by one of these paths to a given vertex of Y. This leads to the notion of *k-linkedness*. A graph is *k-linked* if for any two disjoint sets of vertices $\{x_1, \ldots, x_k\}$ and $\{y_1, \ldots, y_k\}$ there are vertex disjoint paths $P_1, \ldots, P_k$ such that P_i joins x_i to y_i for $i = 1, \ldots, k$.

Clearly, k-linkedness is a stronger notion than k-connectivity. But how much stronger is it? Larman and Mani [46] and Jung [40] showed that there is an $f(k)$ such that any $f(k)$-connected graph is k-linked. They used a theorem of Mader [51], which states that any sufficiently dense graph contains a subdivision of a complete graph K_{3k}, and noticed that any $2k$-connected graph containing a subdivided K_{3k} must be k-linked. Their proofs show that $f(k)$ can be taken to be exponential in k. Later, Bollobás and Thomason [14] proved that $f(k) = 22k$ will do.

The definitions for k-connectivity and k-linkedness carry over similarly for directed graphs. A directed graph is *strongly connected* if for any pair of distinct vertices x and y there is a directed path from x to y, and is strongly k-connected if it remains connected upon removal of any set of at most $k-1$ vertices. In the sequel, we shall omit the use of the word 'strongly' with the understanding that we always mean strong connectivity (as opposed to connectivity of the underlying undirected graph). Menger's theorem carries over in the directed case as well and asserts that a directed graph is k-connected if and only if for any two distinct vertices x and y there are k internally vertex disjoint *directed* paths from x to y. A directed graph D is *k-linked* if for any two disjoint sets of vertices $\{x_1, \ldots, x_k\}$ and $\{y_1, \ldots, y_k\}$ there are vertex disjoint directed paths $P_1, \ldots, P_k$ such that P_i has initial vertex x_i and terminal vertex y_i for each $i \in [k]$. Thus, D is 1-linked if and only if it is connected.

It turns out that directed graphs exhibit quite different behavior than undirected graphs with respect to the relations they bear between connectivity and linkedness. Indeed, Thomassen [64] constructed directed graphs with arbitrarily large connectivity which are not even 2-linked. Since large connectivity does not necessarily imply linkedness for general directed graphs, it is natural to consider the situation for a restricted class of directed graphs, namely, tournaments. A *tournament* on n vertices is a directed graph formed by orienting each edge of the complete graph K_n. In this line of research, Thomassen [62] proved that there is a $g(k)$ such that every $g(k)$-connected tournament is k-linked, with $g(k) = Ck!$, for some absolute constant C. Greatly improving Thomassen's bound on $g(k)$, Kühn, Lapinskas, Osthus, and Patel [45] showed that one may take $g(k) = 10^4 k \log k$ and still ensure k-linkedness. They went on to conjecture that $g(k)$ may be taken to be linear in k. Pokrovskiy [58] resolved this conjecture by showing that any $452k$-connected tournament is k-linked. Except for small k, an optimal bound for $g(k)$ is not known. Bang-Jensen [10] showed that any 5-connected tournament is 2-linked, and there exists a family of 4-connected tournaments which are not 2-linked. For general k, it is not difficult to construct $(2k-2)$-connected tournaments with high minimum out and in-degree, which are not k-linked. For example, consider the blow-up of a directed triangle with vertex sets A,B,C, such that $|C| = 2k-2$ and $|A|,|B| \geq 2k$. Direct the edges from A to B, from B to C, and from C to A; inside each set, orient the edges in any way. Note that this tournament is $(2k-2)$-connected: if we do not remove one of A,B, or C, then the resulting tournament is still a blow-up of a directed triangle, and hence is still strongly connected. To see that this tournament is not k-linked, just split C into two sets C_1,C_2 of size $k-1$, and let $X = \{b\} \cup C_1$ for some $b \in B$ and $Y = \{a\} \cup C_2$ for some $a \in A$. Then we cannot link X to Y, since there is no way to get from b to a without using a vertex from C.

Going back to undirected graphs for a moment, if some density conditions are assumed on the graph, then Bollobás and Thomason's $22k$ can be taken all the way down to $2k$. This is due to the result of Mader [51] mentioned earlier, that a graph with sufficiently large average degree contains a subdivision of a complete graph of order $3k$. Indeed, if S denotes our subdivision of K_{3k}, and X and Y are the k-sets we want to link, then by Menger's theorem there are $2k$ pairwise vertex disjoint paths from $X \cup Y$ to a subset U of the branch vertices of S of size $2k$. Moreover, these paths are internally vertex disjoint from U. Then, provided these paths are chosen to minimize the number of edges outside of S, it is possible to show that one can link X to Y utilizing the k unused branch vertices of S.

Note that $2k$ is close to the theoretical minimum connectivity in any k-linked graph: any k-linked graph must be $(2k-1)$-connected. Supposing otherwise, let W be a vertex cut of size $2k-2$ and arbitrarily partition W into two sets W_1,W_2 of size $k-1$. Since $G-W$ consists of at least two nonempty components C_1,C_2, pick a vertex $c_1 \in C_1$ and $c_2 \in C_2$. Then let $X = W_1 \cup \{c_1\}$ and $Y = W_2 \cup \{c_2\}$. Clearly, there is no way to get from c_1 to c_2 without using a vertex of $W_1 \cup W_2$.

Recently, Thomas and Wollan [61] showed that any $2k$-connected graph with average degree at least $10k$ is k-linked, greatly reducing the bound on the required average degree (which was initially exponential in k). Motivated by this result, Pokrovskiy [58] conjectured that a similar phenomenon should occur for tournaments with a natural 'density' condition: high minimum out-degree and in-degree. In particular, he conjectured that there is a function $f : \mathbb{N} \to \mathbb{N}$ such that any $2k$-connected tournament with minimum out and in-degree at least $f(k)$ is k-linked. Here is our main result, which makes progress on this conjecture.

Theorem 3.1.1. *For every positive integer k there exists $f(k)$ such that every $4k$-connected tournament T with $\delta^+(T) \geq f(k)$ is k-linked.*

Note that we do not assume any lower bound on the minimum in-degree. Moreover, we remark our $f(k)$ in the above theorem can be taken to be $2^{2^{ck^4}}$, for an absolute constant c. It would be nice to determine the smallest function f such that this theorem holds; this is related to determining the smallest function d for which the next theorem (Theorem 3.1.2) holds.

Recall that the complete directed graph $\overrightarrow{K}_k$ is the directed graph on k vertices where, for every pair x, y of distinct vertices, both xy and yx are present. In order to prove Theorem 3.1.1 we shall show that large minimum out-degree allows us to embed subdivisions of the complete directed graph $\overrightarrow{K}_k$. As mentioned above, Mader showed that for any positive integer k there is $g(k)$ such that any graph with average degree at least $g(k)$ contains a subdivision of K_k. The following theorem can be viewed as an analogue of Mader's result for tournaments, replacing 'average degree' with 'minimum out-degree', and may be of independent interest.

Theorem 3.1.2. *For any positive integer k there exists a $d(k)$ such that the following holds. If T is a tournament with $\delta^+(T) \geq d(k)$, then T contains a subdivision of $\overrightarrow{K}_k$.*

We prove this theorem with $d(k) = 2^{2^{Ck^2}}$ for some absolute constant C, and leave the determination of the smallest possible function d as open problem. This theorem does not hold if we replace T by a general digraph, as was shown by Mader [52] (in fact, he showed that it need not hold even if we also assume large minimum in-degree). This fact also follows from a result of Thomassen [63], who showed that for every integer n there exist digraphs on n vertices with minimum out-degree at least $\frac{1}{2}\log n$ which do not contain an even directed cycle. But since any subdivision of $\overrightarrow{K}_3$ must contain an even directed cycle, these digraphs do not contain any subdivision of a complete directed graph.

In order to prove Theorem 3.1.1, we shall need a little more than Theorem 3.1.2. Roughly speaking, we shall first embed in T a subdivided $\overrightarrow{K}_k$, and then attach a few additional paths to it (see Section 3.2).

3.1.1 Organization and Notation

The remainder of this chapter is organized as follows. In Section 3.2, we shall prove Theorem 3.1.2 which allows us to embed subdivisions of a complete directed graph and related structures in tournaments with high minimum out-degree. In Section 3.3, we shall prove one preparatory lemma and then finish our proof of Theorem 3.1.1. Our final section concludes with some open problems.

Our notation is standard. Thus, for a directed graph D we use $N^+(x), N^-(x), d^+(x)$, and $d^-(x)$ to denote the out-neighborhood, in-neighborhood, out-degree, and in-degree of a vertex x, respectively. We use $\delta^+(D)$ to denote the minimum out-degree of D. A directed path $P = x_1 \dots x_\ell$ in D is a sequence of distinct vertices such that $x_i x_{i+1}$ is an edge for every $i = 1, \dots, \ell - 1$. We call x_1 the *initial vertex* and x_ℓ the *terminal vertex* of P. The length of P is the number of its directed edges. We say that P is *internally disjoint* from some subset $X \subset V(D)$ if $\ell \geq 3$ and $\{x_2, \dots, x_{\ell-1}\} \cap X = \varnothing$. If A and B are subsets of $V(D)$, then we shall write $A \to B$ if every edge with one endpoint in A and the other endpoint in B is directed from A to B. Lastly, if $\mathscr{P}$ is a family of directed paths in a digraph, then we use $\bigcup \mathscr{P}$ to denote the set $\bigcup_{P\in\mathscr{P}} V(P)$.

3.2 Proof of Theorem 3.1.2

The first proofs of the result that graphs with sufficiently large connectivity are k-linked use a result of Mader, which allows one to embed a subdivision of a complete graph in a graph with sufficiently large average degree. Our proof of Theorem 3.1.1 follows a similar strategy. In order to proceed, we need a directed analogue of Mader's result for tournaments: we prove this in the present section. We shall use the following simple lemma of Lichiardopol [47] (independently rediscovered by Havet and Lidický [36]). We include the short proof for convenience of the reader.

Lemma 3.2.1. *Every tournament with minimum out-degree at least k has a subtournament with minimum out-degree k and order at most $3k^2$.*

Proof. Let T be a tournament with minimum out-degree at least k, and let T' be a vertex-minimal subtournament of T such that $\delta^+(T') \geq k$. Denote by L the collection of vertices in T' with out-degree k in T', and let $|T'| = t$ and $|L| = \ell$. By minimality, for every vertex $v \in T'$ we have $\delta^+(T' \setminus \{v\}) \leq k-1$. Hence, every vertex in $T' \setminus L$ has an in-neighbor in L, and so there are at least $t-\ell$ edges from L to $T' \setminus L$. On the other hand, the number of such edges is exactly

$$\ell k - \binom{\ell}{2},$$

and so $t - \ell \leq \ell k - \ell^2/2 + \ell/2$. It follows that

$$\ell^2 - \ell(2k+3) + 2t \leq 0,$$

implying the bound $(2k+3)^2 - 8t \geq 0$. In other words, $t \leq \frac{1}{8}(2k+3)^2$, so since t must be an integer we get $t \leq \frac{1}{8}((2k+3)^2 - 1) = k^2/2 + 3k/2 + 1 \leq 3k^2$, as required.

$\square$

We are now ready to prove Theorem 3.1.2. In the following, for a positive integer k and nonnegative integer $m \leq 2\binom{k}{2}$, an *m-partial* $\overrightarrow{K}_k$ is any spanning subdigraph of $\overrightarrow{K}_k$ with precisely m directed edges present. Our proof shows how to find a subdivision of $\overrightarrow{K}_k$ by inductively finding subdivisions of m-partial $\overrightarrow{K}_k$'s for each $m \leq 2\binom{k}{2}$.

Proof of Theorem 3.1.2. For a positive integer k and nonnegative integer $m \leq 2\binom{k}{2}$, let $d(k,m)$ denote the smallest positive integer such that any tournament with $\delta^+(T) \geq d(k,m)$ contains a subdivision of an m-partial complete directed graph on k vertices. We shall show that if $m < 2\binom{k}{2}$, then $d(k,m+1) \leq 7d(k,m)^2$. We use induction on k, and for each fixed k, induction on m. For $k = 1$ there is nothing to show and we can take $d(1,0) = 1$. So let us assume $k \geq 2$ is given and that we can embed a subdivision of an m-partial $\overrightarrow{K}_k$ in any tournament with minimum out-degree at least $d(k,m)$, and let T be a tournament with $\delta^+(T) \geq 7d(k,m)^2$.

Claim 3.2.2. *We may assume that there is a subdivision of an m-partial $\overrightarrow{K}_k$ contained in the out-neighborhood of some vertex of T, and which spans at most $3d(k,m)^2$ vertices.*

Proof. Since certainly we have $\delta^+(T) \geq d(k,m)$, by Lemma 3.2.1 we may find a subtournament T' of size at most $3d(k,m)^2$ and with minimum out-degree at least $d(k,m)$. By induction we may embed in T' a subdivision of an m-partial $\overrightarrow{K}_k$. Denote this subdivision by K. We wish to add a missing directed edge, say xy. In other words, we must find a directed path from x to y in T such this path is internally disjoint from $V(K)$. Let $T'' = T \setminus T'$ and partition it into strongly connected subtournaments $T'' = S_1 \cup \cdots \cup S_\ell$ such that $S_i \to S_j$ for all $1 \leq i < j \leq \ell$ (unless, of course, T'' itself is strongly connected). Observe that since $d^+(x) \geq 7d(k,m)^2$ and $|T'| \leq 3d(k,m)^2$, we have that x has an out-neighbor in T''. Therefore, if some vertex of S_ℓ is joined to y we are done, as we can find a directed path from x to y outside of T'. So we may assume that $S_\ell \subseteq N^+(y)$. Now, as $|T'| \le 3d(k,m)^2$ and no vertex of S_ℓ is joined to any vertex of S_i for $i < \ell$, we have that

$$\delta^+(S_\ell) \ge 7d(k,m)^2 - 3d(k,m)^2 \ge d(k,m),$$

Applying Lemma 3.2.1 to S_ℓ, we find a subtournament $S \subseteq S_\ell$ such that $\delta^+(S) \ge d(k,m)$ and with size at most $3d(k,m)^2$. It follows by induction that we may embed a subdivision of an m-partial $\overrightarrow{K}_k$ in S. But since $S \subseteq S_\ell \subseteq N^+(y)$ and $|S| \le 3d(k,m)^2$, the claim holds. $\square$

By Claim 3.2.2, choose a vertex z with the smallest possible minimum out-degree satisfying the property that there is a subdivision of an m-partial $\overrightarrow{K}_k$ contained in $N^+(z)$ spanning at most $3d(k,m)^2$ vertices. Denote by N the out-neighborhood of z and K_z the subdivision with $K_z \subseteq N$. We wish to add one more directed edge to this subdivision, say uv with $u,v \in K_z$. From N remove all vertices of K_z except for u and v and call this set N'. If $T[N']$ is strongly connected then we are done; otherwise, partition $T[N']$ into strongly connected subtournaments, say $T[N'] = S'_1 \cup \cdots \cup S'_t$ where $S'_i \to S'_j$ for all $1 \le i < j \le t$. Suppose that some vertex $w \in S'_t$ is joined to a vertex $w' \in N^-(z)$. Then since there is a directed path P from u to w in $T[N']$ we have that $uPww'zv$ is a directed path from u to v which avoids $K_z \setminus \{u,v\}$. Hence we may assume that every vertex of $N^-(z)$ dominates S'_t. But then, since $|K_z| \le 3d(k,m)^2$ and there are no edges from S'_t to S'_i for $i < t$, one has that $\delta^+(S'_t) \ge 7d(k,m)^2 - 3d(k,m)^2 = 4d(k,m)^2$. So we can repeat the argument in Claim 3.2.2 to S'_t with minimum out-degree $4d(k,m)^2$ instead of $7d(k,m)^2$ (observe that we need $4d(k,m)^2 - 3d(k,m)^2 \ge d(k,m)$ to hold, which is clearly true). Accordingly, there is a vertex $q \in S'_t$ such that $N^+(q)$ contains a subdivision of an m-partial $\overrightarrow{K}_k$ spanning at most $3d(k,m)^2$ vertices. However, since $\bigcup_{i<t} S'_i \neq \varnothing$ (as $T[N']$ is not strongly connected), and q is not joined to any vertex of $\bigcup_{i<t} S'_i \cup N^-(z)$, we have $d^+(q) < d^+(z)$, a contradiction to the minimality of z. This completes the proof of Theorem 3.1.2, as we may take $d(k) = d(k, 2\binom{k}{2})$.

$\square$

We now need to embed a slightly more complicated structure in T. In particular, we shall need to attach a few special paths to our subdivided complete directed graph. Say a subdivision $\mathscr{S}$ is *minimal* in a tournament T if all of its paths have minimal length. This implies that every path in $\mathscr{S}$ is *backwards transitive*: if $x_1 \ldots x_t$ is a path in $\mathscr{S}$ between branch vertices, then $x_i x_j \notin E(T)$ whenever $i \in [t-2]$ and $i < j+1$. Let $\mathscr{K}_r^{\min}$ denote a minimal subdivision of a $\overrightarrow{K}_r$. Since any subdivision of $\overrightarrow{K}_r$ contains a minimal subdivision, Theorem 3.1.2 allows us to find a $\mathscr{K}_r^{\min}$ in tournaments with sufficiently large out-degree. If U denotes the set of branch vertices of this subdivision, then for every $u, v \in U$, $\mathscr{K}_r^{\min}$ consists of directed paths P_{uv}, P_{vu} going from u to v and from v to u, respectively. Since T is a tournament and $\mathscr{K}_r^{\min}$ is minimal, precisely one of these paths is a directed edge.

Now we define our augmented subdivision, denoted by $\mathscr{K}_r^*$, as follows. Let $\mathscr{K}$ denote a copy of $\mathscr{K}_r^{\min}$ in T. The *branch vertices* of $\mathscr{K}_r^*$ are precisely the branch vertices of $\mathscr{K}$; denote this set by U. We form $\mathscr{K}_r^*$ by adding a collection $\mathscr{L}$ of special 'loop' paths in the following manner. For each pair $u, v \in U$, if, say, P_{uv} is the path between u and v in $\mathscr{K}$ of length at least two, then each of u and v has an associated directed path from $\mathscr{L}$: one directed path L_{uv}^u going from the second vertex of P_{uv} to u, and another directed path L_{uv}^v going from v to the penultimate vertex of P_{uv}; we require that these paths are internally disjoint from $V(\mathscr{K})$. We also impose that the paths in $\mathscr{L}$ are minimal and hence backwards transitive. For $u \in U$, we let $\mathscr{L}_u$ denote the collection of paths in $\mathscr{L}$ which contain u. Note that $\mathscr{K}_r^*$ and $\mathscr{K}_r^{\min}$ really denote *families* of subdigraphs which depend on the underlying tournament T. When we speak of 'a $\mathscr{K}_r^*$' we really mean 'a member of $\mathscr{K}_r^*$ in T'; we hope this usage of notation does not cause confusion, but we think that it is simpler. Now the proof of the existence of a $\mathscr{K}_r^*$ follows exactly in the same way as the proof of Theorem 3.1.2, namely by induction on the number of 'loops'. We state it as a corollary and provide only a sketch of the proof.

Corollary 3.2.3. *For any positive integer k there exists a $d^*(k)$ such that the following holds. If T is a tournament with $\delta^+(T) \geq d^*(k)$, then T contains a $\mathscr{K}_k^*$.*

Proof sketch. Similarly as in Theorem 3.1.2, for a positive integer k and nonnegative integer $m \leq 2\binom{k}{2}$, an *m-partial* $\mathscr{K}_k^*$ is any minimal subdivision of $\overrightarrow{K}_k$ with precisely m loop paths present. Let $d^*(k,m)$ denote the smallest positive integer such that any tournament with $\delta^+(T) \geq d^*(k,m)$ contains a subdivision of an m-partial $\mathscr{K}_k^*$. We show, as before, that if $m < 2\binom{k}{2}$, then $d^*(k,m+1) \leq 7d^*(k,m)^2$. For $k = 1$ there is nothing to show and we can take $d^*(1,0) = 1$. So assume $k \geq 2$ is given. Then $d^*(2,0)$ exists by Theorem 3.1.2 (i.e., we can embed a subdivision of $\overrightarrow{K}_2$ which contains a minimal such subdivision). Thus let $m \geq 1$ and suppose we can embed an m-partial $\mathscr{K}_k^*$ in any tournament with minimum out-degree at least $d^*(k,m)$. Let T be a tournament with $\delta^+(T) \geq 7d^*(k,m)^2$. Then the same proof used to show Theorem 3.1.2 gives that we may attach one more loop path, which we may assume has minimal length. Therefore we can embed an $(m+1)$-partial $\mathscr{K}_k^*$ in T, as claimed. $\square$

3.3 Proof of the main theorem

In this section we finish the proof of Theorem 3.1.1. The structure of the proof is as follows. First, assuming the minimum degree of our tournament is sufficiently large, we shall embed in T a copy $\mathscr{S}$ of $\mathscr{K}_r^*$ where $r = r(k)$ is sufficiently large. If $x_1, \ldots, x_k$, $y_1, \ldots, y_k$ are the vertices we want to link, then we shall show that there exists a collection of k directed paths going from the x_i's to the branch vertices of $\mathscr{S}$, and a collection of k directed paths going from the branch vertices of $\mathscr{S}$ to the y_i's, all of these paths being pairwise vertex disjoint. Here we only use the assumption that T is $4k$-connected (see

Lemma 3.3.1 below). Finally, we show that, provided one chooses these paths appropriately, one can link each x_i to y_i by rerouting the paths through $\mathscr{S}$. The rerouting step is more complicated than one might expect, and we shall see that we do need the slightly richer structure $\mathscr{K}_r^*$ rather than just a subdivided complete directed graph.

We need a small bit of terminology first before proceeding. If X and Y are two disjoint sets of vertices in a directed graph, then we say that there is an *out-matching* (resp., *in-matching*) of X to Y if there is a matching from X into Y such that all matching edges are directed from X to Y (resp., directed from Y to X).

Lemma 3.3.1. *Let T be a $4k$-connected tournament. Suppose $A, B \subset V(T)$ are two disjoint subsets of size k, and let $L \subset V(T)$ be a set of $4k$ vertices disjoint from $A \cup B$. Then there are k directed paths from A to L, and k directed paths from L to B, all these paths pairwise vertex disjoint and internally disjoint from L.*

Proof. Choose two disjoint subsets W_A, W_B disjoint from $A \cup B \cup L$ with maximum size subject to the following properties:

- Every vertex in W_A has at least $2k$ out-neighbors in L, and every vertex in W_B has at least $2k$ in-neighbors in L.
- There is an in-matching $\mathscr{M}_A$ from W_A to A, and an out-matching $\mathscr{M}_B$ from W_B to B.

We shall assume, without loss of generality, that $|W_A| \le |W_B|$. Let A' denote the set of $|W_A|$ vertices in A that are incident with an edge of $\mathscr{M}_A$, and let $A'' = A \setminus A'$. Let B', B'' denote the analogous sets of vertices in B. As T is $4k$-connected, we can find pairwise vertex disjoint directed paths from some $k - |W_B|$ vertices of L to B'' avoiding $A \cup W_A \cup B' \cup W_B$. Choose a collection of such paths $\mathscr{P}$ which minimizes $|\bigcup \mathscr{P}|$, and subject to that, maximizes the number of paths whose second vertex has at least $2k$ in-neighbors in L. Partition $\mathscr{P}$ into sets $\mathscr{P}', \mathscr{P}''$ where the former denotes the collection of paths in $\mathscr{P}$ whose second vertex has at least $2k$ in-neighbors in L, and the latter denotes the collection of remaining paths. Denote by X' the set of all second and third vertices on paths in $\mathscr{P}'$, and denote by X'' the set of all first and second vertices on paths in $\mathscr{P}''$. Consider the set $Y := A' \cup W_A \cup X' \cup X'' \cup B \cup W_B$ and note that we can bound the size of Y as

$$|Y| \leq 2|W_A| + 3(k - |W_B|) + 2|W_B|.$$

We shall now find $k - |W_A|$ disjoint directed paths from the vertices in A'' to some subset of L, avoiding Y. This is possible since T is $4k$-connected and

$$\begin{aligned} 4k - |Y| &\geq 4k - (2|W_A| + 3(k - |W_B|) + 2|W_B|) \\ &= k - 2|W_A| + |W_B| \geq k - |W_A|, \end{aligned}$$

where the last inequality holds since we are assuming that $|W_A| \leq |W_B|$. Therefore, choose a collection $\mathscr{Q}$ of pairwise disjoint directed paths from A'' to L avoiding Y with $|\bigcup \mathscr{Q}|$ as small as possible. We claim that these new paths do not intersect any path from $\mathscr{P}$:

Claim 3.3.2. *No path from $\mathscr{Q}$ intersects a path from $\mathscr{P}$.*

Proof. Suppose that some path $Q \in \mathscr{Q}$ intersects a path $P \in \mathscr{P}$. Let $P = x_1 \dots x_s$ and $Q = y_1 \dots y_t$, and let $L_A = (\bigcup \mathscr{Q}) \cap L$ and similarly $L_B = (\bigcup \mathscr{P}) \cap L$. We shall consider two cases, according to whether $P \in \mathscr{P}'$ or $P \in \mathscr{P}''$. Suppose first the former holds, and let y_i ($i \geq 2$) be the first vertex of Q that intersects P. We may assume that $y_i \neq x_1$; indeed, if $y_i = x_1$, then $|L_A \cup L_B| \leq 2k - 1$, and since $P \in \mathscr{P}'$, we have that x_2 has at least $2k$ in-neighbors in L. Therefore, we may choose some in-neighbor x' disjoint from $L_A \cup L_B$ and replace P with $P' := x'x_2 \dots x_s$. Moreover, since the paths in $\mathscr{Q}$ avoid $\{x_2, x_3\}$ we may assume that $y_i = x_4$. Consider y_{i-1} and pick any vertex $z \in L \setminus (L_A \cup L_B)$. If $y_{i-1}z \in E(T)$, then we may replace Q with the shorter directed path $y_1 \dots y_{i-1}z$, contradicting the minimality of $|\bigcup \mathscr{Q}|$. So we have $zy_{i-1} \in E(T)$. But then as long as $i \geq 3$ we may replace P with the shorter path $zy_{i-1}x_4 \dots x_s$, contradicting the initial minimal choice of $|\bigcup \mathscr{P}|$. It remains to consider when $i=2$. In this case, $zy_2 \notin E(T)$ for every $z \in L \setminus (L_A \cup L_B)$, since otherwise we can replace P with a shorter directed path. Thus y_2 has at least $2k$ out-neighbors in L, and we can add $y_1 y_2$ to the matching $\mathscr{M}_A$, a contradiction to the maximality of this matching. It follows that $P \cap Q = \varnothing$ for $P \in \mathscr{P}'$.

So let us assume that $P \in \mathscr{P}''$. Since the paths in $\mathscr{Q}$ avoid $\{x_1, x_2\}$, we may assume in this case that $y_i = x_3$. The same argument as in the previous paragraph shows that we may assume $i \geq 3$ (otherwise, we obtain a larger matching than $\mathscr{M}_A$). Also, as before, if $z \in L \setminus (L_A \cup L_B)$, then $y_{i-1} z \notin E(T)$; otherwise we can replace Q with the shorter path $y_1 \ldots y_{i-1} z$. Hence y_{i-1} has at least $|L| - |L_A \cup L_B| \geq 2k$ in-neighbors in L. Choose one of these in-neighbors u (disjoint from $L_A \cup L_B$) and consider the path $P^* := u y_{i-1} x_3 \ldots x_s$. Then P^* has the same length as P and its second vertex has at least $2k$ in-neighbors in L, so we could replace P with P^*, contradicting the maximality of $\mathscr{P}'$. Therefore, we must have $P \cap Q = \varnothing$, and the proof of Claim 3.3.2 is complete.

□

Armed with Claim 3.3.2, the proof of Lemma 3.3.1 is essentially complete. Indeed, every vertex in W_A has at least $2k$ out-neighbors in L, and so each of these vertices has at least

$$2k - |L_A \cup L_B| = |W_A| + |W_B|,$$

out-neighbors in $L \setminus (L_A \cup L_B)$. So for each vertex in W_A we may select a distinct out-neighbor in $L \setminus (L_A \cup L_B)$. Then every vertex in W_B has at least $|W_B|$ in-neighbors from the remaining vertices of L, so we can pick a distinct in-neighbor for every vertex of W_B. The paths of length 2 using vertices of $W_A \cup W_B$ together with $\mathscr{P}$ and $\mathscr{Q}$ form the required collection of paths. □

We can now finish the proof of Theorem 3.1.1.

Proof of Theorem 3.1.1. Let $k \geq 2$ be an integer and let $f(k) := d^*(12k^2) + 2k$, where $d^* : \mathbb{N} \to \mathbb{N}$ is the function provided by Corollary 3.2.3. Suppose that T is a $4k$-connected tournament with minimum out-degree at least $f(k)$, and let $X = \{x_1, \ldots, x_k\}$, $Y = \{y_1, \ldots, y_k\}$ be two disjoint k-sets of vertices. We wish to find pairwise vertex disjoint directed paths going from x_i to y_i for each $i \in [k]$. Remove $X \cup Y$ from T; the tournament induced on $V(T) \setminus (X \cup Y)$ has minimum out-degree at least $d^*(12k^2)$, so by Corollary 3.2.3 we may embed in T a $\mathscr{K}^*_{12k^2}$ disjoint from $X \cup Y$. Denote this subdivision by $\mathscr{S}$. We shall use the same notation as in Section 3.2, namely, U denotes the branch vertices of $\mathscr{S}$, $\mathscr{K}$ denotes the underlying minimal subdivision of $\overrightarrow{K}_{12k^2}$ composed of minimal paths P_{uv}, P_{vu} for every pair of branch vertices $u, v \in U$, and $\mathscr{L}$ denotes the collection of minimal paths attached to $\mathscr{K}$. We call a *path of* $\mathscr{S}$ any path P_{uv} between branch vertices of length at least 2, and any member of $\mathscr{L}$. We consider the following edges to belong to the structure $\mathscr{S}$:

- The edges belonging to paths in $\mathscr{K}$, except the paths of length one.
- The edges belonging to paths in $\mathscr{L}$.
- For every pair $u, v \in U$, every edge in T between $\{u, v\}$ and $V(P_{uv}) \cup V(P_{vu})$.
- For every $u \in U$, every edge in T between u and $\bigcup \mathscr{L}_u$.

We denote the set of edges of $\mathscr{S}$ by $E(\mathscr{S})$. For example, whenever we speak of distances in $\mathscr{S}$, we insist that they are computed using only these directed edges. Let $\mathscr{P}$ and $\mathscr{Q}$ be any two collections of pairwise disjoint directed paths such that every path in $\mathscr{P}$ goes from U to Y, every path in $\mathscr{Q}$ goes from X to U, and all of these paths are internally vertex disjoint from U; by Lemma 3.3.1, such collections exist. We say that a pair $(u, x) \in U \times V(\mathscr{S})$ is at *in-distance d in $\mathscr{S}$* if d is the smallest integer such that there is a directed path P' of length d using only edges of $\mathscr{S}$, and such that P' goes from u to x. We shall also sometimes say that x has in-distance d in $\mathscr{S}$ from u. Similarly, we say that

$(u,x) \in U \times V(\mathscr{S})$ is at *out-distance d in $\mathscr{S}$* if d is the smallest integer such that there is a directed path Q' of length d using only edges of $\mathscr{S}$, and such that Q' goes from x to u in $\mathscr{S}$; we shall also sometimes say that x has out-distance d in $\mathscr{S}$ from u. We denote in-distance by $d^{\text{in}}(u,x)$ and out-distance by $d^{\text{out}}(u,x)$ (where we have suppressed the dependence on $\mathscr{S}$).

Observation 3.3.3. *Let $x \in V(\mathscr{S}) \setminus U$. Then x is at in-distance (or out-distance) at least 3 from every vertex of U, except possibly the branch vertex (or vertices) belonging to the path of $\mathscr{S}$ containing x.*

Proof. If $x \in V(\mathscr{S}) \setminus U$, then either $x \in P_{uv}$ for some $u, v \in U$ or $x \in L^u_{uv} \in \mathscr{L}_u$ (or possibly both). Let $w \in U \setminus \{u,v\}$. In order to get from w to x using only edges of $\mathscr{S}$, we must first reach either u or v. However, recall that the single edge paths in $\mathscr{K}$ are *not* edges of $\mathscr{S}$, so the path from w to u or v in $\mathscr{S}$ has length at least 2. Therefore, x has in-distance at least 3 from w, as required. A symmetric argument shows that the observation remains true with 'out-distance' instead of 'in-distance'.

□

In the following, we shall always assume that any family $\mathscr{F}$ of directed paths in T between $X \cup Y$ and U are internally disjoint from U. We also denote by $U_{\mathscr{F}}$ the set $U \cap (\bigcup \mathscr{F})$. Our first claim asserts that we may assume the paths in one of the collections $\mathscr{P}$, $\mathscr{Q}$ contains few vertices which are 'close' in $\mathscr{S}$ to a vertex in U.

Lemma 3.3.4. *We may choose either $\mathscr{P}$ or $\mathscr{Q}$ such that there are at most $8k^2+4k$ vertices $u \in U \setminus U_{\mathscr{P}}$ (resp., $U \setminus U_{\mathscr{Q}}$) with $d^{in}(u,x) \le 2$ (resp., $d^{out}(u,x) \le 2$) for some $x \in \bigcup \mathscr{P} \setminus U_{\mathscr{P}}$ (resp., for some $x \in \bigcup \mathscr{Q} \setminus U_{\mathscr{Q}}$).*

Proof. Apply Lemma 3.3.1 with $A = X$, $B = Y$, and $L = U$. Using the proof and notation of Lemma 3.3.1, assume that $|W_X| \le |W_Y|$. Then recall that we may choose the paths from U to Y first minimally (with respect to the number of vertices used) upon the removal of

$W_X \cup W_Y$, a set of at most $2k$ vertices. Recall also that each such path which uses a vertex of $W_X \cup W_Y$ has length two. Suppose there is a set $U' \subset U \setminus U_{\mathscr{P}}$ of more than $8k^2+4k$ vertices such that for every $u \in U'$ there is $x \in \bigcup \mathscr{P} \setminus U_{\mathscr{P}}$ with $d^{\text{in}}(u,x) \leq 2$. We claim that this contradicts minimality. Indeed, by pigeonhole there is a set $U'_0 \subset U'$ of size more than $8k+4$, and a path $P \in \mathscr{P}$ such that for each $u \in U'_0$ there is some $x \in P$ with $d^{\text{in}}(u,x) \leq 2$. From Observation 3.3.3, it follows that for each interior vertex v of P there are at most two vertices of U'_0 that are at in-distance 2 from v. Therefore P must have more than two edges so does not intersect $W_X \cup W_Y$. For each vertex $u \in U'_0$, pick some vertex $v_u \in P$ at in-distance exactly 2 from u, and denote by D the set containing all such vertices v_u. Note that P contains at most one vertex at in-distance 1 from a vertex in $U \setminus U_{\mathscr{P}}$, as otherwise we may reroute P and obtain a shorter path avoiding $W_X \cup W_Y$. Using Observation 3.3.3 again, there is a set D' of at least $\frac{1}{2}(8k+4) = 4k+2$ vertices in D corresponding to distinct vertices of U'_0. Let $P = p_0 \dots p_\ell$, where $p_0 \in U$ and $p_\ell \in X$, $F := D' \setminus \{p_1, p_2\}$. For each $p_j \in F$, we may choose vertex disjoint directed paths $u_j m_j p_j$ of length 2 in $\mathscr{S}$, where $u_j \in U'_0$. Accordingly, there are at least $4k$ 'middle vertices' m_j, at least $2k$ of which are disjoint from $W_X \cup W_Y$; let M denote the set of middle vertices disjoint from $W_X \cup W_Y$. Now, suppose some $m_j \in M$ does not intersect any path in $\mathscr{P}$. Then we may replace P with the path $u_j m_j p_j P$, which is shorter and still avoids $W_X \cup W_Y$, a contradiction. Thus, each middle vertex in M belongs to some member of $\mathscr{P}$ and so by pigeonhole there is a path P' which contains at least two vertices of M. But both of these vertices are at in-distance 1 from a vertex in $U \setminus U_{\mathscr{P}}$, which, as noted before, is a contradiction. Hence at most $8k^2+4k$ vertices in $U \setminus U_{\mathscr{P}}$ have the stated property, as claimed. A symmetric argument shows that we may choose $\mathscr{Q}$ with the stated property in the event that $|W_Y| \leq |W_X|$. This completes the proof of the lemma. □

Suppose $\mathscr{F}$ is a collection of pairwise disjoint directed paths from U to Y (internally disjoint from U), and let $P = p_0 \dots p_t$ be any path in $\mathscr{F}$. We call the pairs (p_0, p_1) and (p_0, p_2) *trivial* if they have in-distance at most 2 in $\mathscr{S}$; any other pair with in-distance at most 2 is *nontrivial*. For a subset $U' \subseteq U$ we shall say that $\mathscr{F}$ is U'-*good* if no nontrivial pair of vertices from $U' \times (\bigcup \mathscr{F} \setminus U_{\mathscr{F}})$ is at in-distance at most 2 in $\mathscr{S}$. In particular, each path $P \in \mathscr{F}$ intersects U' in at most one vertex, namely its initial vertex. Suppose that $\mathscr{F}$ satisfies the property stated in Lemma 3.3.4. Then we have the following:

Claim 3.3.5. *There exists a subset $U' \subset U \setminus U_{\mathscr{F}}$ of size at least $2k$ such that $\mathscr{F}$ is U'-good.*

Proof. This follows immediately from the previous lemma. Indeed, remove from U every vertex in $U_{\mathscr{F}}$ and every vertex in $U \setminus U_{\mathscr{F}}$ at in-distance at most 2 in $\mathscr{S}$ from some vertex of $\bigcup \mathscr{F} \setminus U_{\mathscr{F}}$; let U' denote the remaining set of vertices. By Lemma 3.3.4, we have removed at most $8k^2 + 5k$ vertices. As $|U| = 12k^2$ we have $|U'| \geq 12k^2 - (8k^2 + 5k) \geq 2k$, since $k \geq 2$. Clearly $\mathscr{F}$ is U'-good. $\square$

We shall assume without loss of generality that we may choose the paths from U to Y with the property stated in Lemma 3.3.4. So the previous two claims show that we may find collections of vertex disjoint directed paths $\mathscr{P}, \mathscr{Q}$ which are internally disjoint from U and such that the paths in $\mathscr{P}$ go from U to Y, the paths in $\mathscr{Q}$ go from X to U, and $\mathscr{P}$ is U'-good for some $U' \subset U \setminus U_{\mathscr{P}}$ with $|U'| \geq 2k$. Conditioned on this, we assume that $\mathscr{P} \cup \mathscr{Q}$ minimizes the number of edges outside of $\mathscr{S}$, and again conditioned on this, we take such a pair with $|\bigcup \mathscr{P}| + |\bigcup \mathscr{Q}|$ as small as possible. Let $U'' = U' \setminus U_{\mathscr{Q}}$ so that $|U''| \geq k$ and it is disjoint from $U_{\mathscr{P}} \cup U_{\mathscr{Q}}$; we may assume that $U'' = \{u_1, \ldots, u_k\}$ has precisely k elements. We now show that one can reroute the paths in $\mathscr{P} \cup \mathscr{Q}$ through U'' in order to create the desired paths linking x_i to y_i for each $i \in [k]$. Let $U_{\mathscr{P}} = \{z_1, \ldots, z_k\}$ and $U_{\mathscr{Q}} = \{w_1, \ldots, w_k\}$ so that z_i is the initial vertex in U of the path $P_i \in \mathscr{P}$ with terminal vertex $y_i \in Y$, and w_i is the terminal vertex in U of the path $Q_i \in \mathscr{Q}$ with initial vertex $x_i \in X$. Recall that for every pair of branch vertices $u, v \in U$, P_{uv} and P_{vu} denotes the path in $\mathscr{K}$ from u to v, and from v to u, respectively. The following sequence of claims show that we can control intersections of paths in $\mathscr{P} \cup \mathscr{Q}$ with appropriate paths in $\mathscr{S}$ in order to link each x_i to y_i.

Claim 3.3.6. *Suppose some path $Q \in \mathcal{Q}$ intersects $L^{u_i}_{w_i u_i} \in \mathcal{L}_{u_i}$, for some $i \in [k]$. Let z be the first vertex of $L^{u_i}_{w_i u_i}$ in the intersection. Then one of the following holds: z is the terminal vertex of $L^{u_i}_{w_i u_i}$ and $z \in Q_i$, or z is the second vertex of $L^{u_i}_{w_i u_i}$.*

Proof. Suppose z is not the second vertex of $L^{u_i}_{w_i u_i}$. If z is an interior point of $L^{u_i}_{w_i u_i}$, then $zu_i \in E(T)$ by minimality of the path $L^{u_i}_{w_i u_i}$. Note that if Q has an edge which is not in $E(\mathcal{S})$ after z then we have a contradiction: indeed replacing Q with Qzu_i yields a collection of paths with fewer edges outside of $E(\mathcal{S})$. Otherwise, $Q = Q_i$ and it must use at least 2 edges after z, so we obtain a contradiction to the minimality of $|\bigcup \mathcal{P}| + |\bigcup \mathcal{Q}|$ by rerouting the path as before. Therefore, z must be the terminal vertex of $L^{u_i}_{w_i u_i}$. Finally, z must belong to Q_i, otherwise we may similarly reroute Q through u_i, decreasing the number of edges used outside $E(\mathcal{S})$. $\square$

Claim 3.3.7. *No path in $\mathcal{P}$ intersects $P_{w_i u_i}$. Moreover, if q_i denotes the last vertex in $P_{w_i u_i}$ which occurs as the intersection of some path in $\mathcal{Q}$, then $q_i \in Q_i$.*

Proof. No path in $\mathcal{P}$ intersects $\{u_i, w_i\}$, so it suffices to show that no such path intersects the interior of $P_{w_i u_i}$. Therefore, we may assume that $P_{w_i u_i}$ has length at least 2. Suppose first that some $P \in \mathcal{P}$ contains a vertex v in the interior. Note that v must be the penultimate vertex of $P_{w_i u_i}$. Otherwise, $u_i v \in E(T) \cap E(\mathcal{S})$ by the minimality of the subdivision $\mathcal{K}$, and this contradicts the fact that $\mathcal{P}$ is U'-good. Consider the loop path $L = L^{u_i}_{w_i u_i} \in \mathcal{L}_{u_i}$ at u_i ending at v, and recall that the edges of L are edges of $\mathcal{S}$. Let z be the first vertex in $L^{u_i}_{w_i u_i}$ belonging to some path $P' \in \mathcal{P}$: such a vertex and path exist since we may take $z = v$ and $P' = P$. Let L' be the initial segment of the path $L^{u_i}_{w_i u_i}$ ending at z.

Suppose first that no path in $Q \in \mathcal{Q}$ intersects L', and replace P with $P'' = u_i L' z P'$. Since P' cannot intersect u_i or w_i it must have an edge which is not in $E(\mathcal{S})$ before z. It follows that P'' has fewer edges outside of $\mathcal{S}$. This is a contradiction to our choice of $\mathcal{P} \cup \mathcal{Q}$, provided $\mathcal{P}'' := (\mathcal{P} \setminus \{P'\}) \cup \{P''\}$ is U'-good. To see this, observe that any vertex of $L \setminus \{v\}$ is at in-distance at least 3 from w_i. Moreover, if $w_i \in U'$, and $z = v$ (and hence $P' = P$), then z is also at in-distance at least 3 from w_i. Accordingly, if $w_i \in U'$, then every vertex of P'' is still at in-distance at least 3 from w_i. By the minimality of L, every vertex in the interior of L (except the second) is directed towards u_i; thus, the only vertices at in-distance at most 2 from u_i are the second and third vertices of L, say x and y, respectively. But the pairs (u_i, x) and (u_i, y) are trivial pairs, and thus do not contradict U'-goodness. Lastly, by Observation 3.3.3 every vertex of P'' (except possibly u_i) is at in-distance at least 3 from every vertex of $U' \setminus \{u_i, w_i\}$. It follows that $\mathscr{P}''$ is U'-good, which is a contradiction to our choice of $\mathscr{P} \cup \mathscr{Q}$.

On the other hand, if some path $Q' \in \mathscr{Q}$ intersects L' in some vertex r, then by Claim 3.3.6 r must the second vertex of $L^{u_i}_{w_i u_i}$. Note that by U'-goodness, no path in $\mathscr{P}$ contains the third vertex r_1 of $L^{u_i}_{w_i u_i}$, hence we can replace Q' by $Q' r r_1 u_i$ thus decreasing the number of edges outside $E(\mathscr{S})$. Therefore we conclude that no path in $\mathscr{P}$ can intersect $P_{w_i u_i}$. Let us now show the second part of the claim. Suppose that $q_i \in Q_j$ for some $j \neq i$. Since Q_j must avoid $\{u_i, w_i\}$ it contains an edge which is not in $E(\mathscr{S})$ after q_i. Replace Q_j with $Q' = Q_j v P_{w_i u_i}$. Then by the previous paragraph, no path in $\mathscr{P}$ intersects Q' and the resulting collection of paths has fewer edges outside of $\mathscr{S}$, a contradiction. This completes the proof of the claim.

□

It remains to establish the analogous claims for the path $P_{u_i z_i}$, namely that intersections of paths in $\mathscr{P} \cup \mathscr{Q}$ with $P_{u_i z_i}$ and $L^{u_i}_{u_i z_i}$ behave as one expects. The arguments are similar to those in the previous two claims. Theorem 3.1.1 will then be an immediate consequence.

Claim 3.3.8. *For every $i \in [k]$, no path in $\mathscr{P}$ intersects $L^{u_i}_{u_i z_i} \in \mathscr{L}_{u_i}$.*

Proof. Suppose some $P \in \mathscr{P}$ intersects $L^{u_i}_{u_i z_i}$ in a vertex z. Then z cannot be the first vertex of $L^{u_i}_{u_i z_i}$, as this would contradict the fact that $\mathscr{P}$ is U'-good. Therefore, if z' denotes the vertex preceding z in $L^{u_i}_{u_i z_i}$, then by the minimality of paths in $\mathscr{L}$, we have $u_i z' \in E(T) \cap E(\mathscr{S})$. But then z is at in-distance 2 from u_i, contradicting U'-goodness.

$\square$

Claim 3.3.9. *Let p_i denote the first vertex in $P_{u_iz_i}$ which occurs as the intersection of some path in $\mathscr{P}$. Then no path in $\mathscr{Q}$ intersects $P_{u_iz_i}$ and $p_i \in P_i$.*

Proof. As before, it suffices to show that no path in $\mathscr{Q}$ intersects the interior of $P_{u_iz_i}$, so we may assume that $P_{u_iz_i}$ has length at least 2. Suppose some $Q \in \mathscr{Q}$ intersects the interior of $P_{u_iz_i}$ at v. Note that since Q does not meet $\{u_i, z_i\}$, it must leave $\mathscr{S}$ at some time after v. If v is not the second vertex of $P_{u_iz_i}$, then $vu_i \in E(T) \cap E(\mathscr{S})$, and so we may replace Q with Qvu_i. This path has fewer edges outside of $\mathscr{S}$ than Q, and this contradicts our minimal choice of $\mathscr{P} \cup \mathscr{Q}$. If v is the second vertex, then let $L = L^{u_i}_{u_iz_i} \in \mathscr{L}_{u_i}$ be the loop path at u_i directed from v to u_i. Let z be the last vertex of L which occurs as the intersection of some path $Q' \in \mathscr{Q}$ (z and Q' exist since we may take $z = v$ and $Q' = Q$), and let L' be the subpath of L from z to u_i. By Claim 3.3.8, no path in $\mathscr{P}$ intersects L', so replace Q' with $Q'zL'u_i$. Again, the edges of L' are in $E(\mathscr{S})$ so this path has fewer edges outside $\mathscr{S}$ than Q', a contradiction. It follows that no path in $\mathscr{Q}$ intersects $P_{u_iz_i}$ as claimed. For the second part of the claim, suppose that $p_i \in P_j$ for some $j \neq i$. Then P_j avoids $\{u_i, z_i\}$ and therefore leaves $\mathscr{S}$ at some time before p_i. Now, no path in $\mathscr{P} \cup \mathscr{Q}$ intersects the interior of the subpath $P_{u_iz_i}p_i$ so replace P_j with $P' = P_{u_iz_i}p_iP_j$. This path has fewer edges outside of $\mathscr{S}$. We claim that $\mathscr{P}' = (\mathscr{P} \setminus \{P_j\}) \cup \{P'\}$ is U'-good. Indeed, note that since $\mathscr{P}$ is U'-good, the subpath $P_{u_iz_i}p_i$ has length at least 3. Also, for every $v \in P_{u_iz_i}$ we have that $vu_i \in E(T)$ by the minimality of $\mathscr{K}$. So the only pairs at in-distance at most 2 in $U' \times (\bigcup \mathscr{P}' \setminus U_{\mathscr{P}'})$ are the trivial pairs (u_i, x) and (u_i, y), where x, y are the second and third vertices, respectively, of $P_{u_iz_i}$. But these pairs, by definition, do not contradict U'-goodness. It follows that $j = i$, and the claim is proved.

$\square$

By Claims 3.3.7 and 3.3.9, the directed paths $Q_iq_iP_{w_iu_i}u_iP_{u_iw_i}p_iP_i$, for each $i \in [k]$, are pairwise vertex disjoint and link x_i to y_i. This completes the proof of Theorem 3.1.1.

$\square$

3.4 Final remarks and open problems

The most obvious open problem is to reduce our bound of $4k$ on the connectivity in Theorem 3.1.1. We remark that an improvement on the connectivity bound in Lemma 3.3.1 translates directly into a better bound in Theorem 3.1.1. Unfortunately, we could not go beyond $4k$. Furthermore, Lemma 3.3.1 does *not* hold if we replace $4k$ with anything smaller than $3k$. The following construction, of a $(3k-1)$-connected tournament T where Lemma 3.3.1 fails, was communicated to us by Kamil Popielarz. Suppose $V(T) = [n]$ and partition $V(T)$ into disjoint sets A, S, B, L, where $L = V(T) \setminus (A \cup S \cup B)$, and $|A| = |B| = k$, $|S| = 2k-1$. Direct the edges from L to A; from B to L; from A to S and from S to B; and from A to B. Inside L we place a balanced blow-up of a directed triangle. That is, equitably partition L into sets L_1, L_2, L_3 with directed edges $L_1 \to L_2$, $L_2 \to L_3$, $L_3 \to L_1$, and inside each of the L_i's we orient the edges arbitrarily. Now, join every vertex in S to all of L_1 and join every vertex of L_2 to all of S. Finally, orient the edges between S and L_3, and the edges inside A, B, and S, arbitrarily.

Provided n is sufficiently large (depending on k), it is not hard to show that T is $(3k-1)$-connected. Indeed, suppose n is large enough so that $|L_i| \geq 3k-1$ for $i = 1,2,3$, and let K be a vertex cut of size $3k-2$. Note that removing K from T does not destroy any of the L_i. Also, observe that $T[A \cup S \cup L_1]$, $T[B \cup S \cup L_2]$, $T[A \cup B \cup L]$ are blow-ups of directed triangles and hence strongly connected. Furthermore, $T[L \cup S]$ is strongly connected. From these facts it is easy to see that the only way to disconnect T is to remove either $A \cup S$ or $B \cup S$, which are both sets of size $3k-1$. Hence $T \setminus K$ is connected, showing that T is $(3k-1)$-connected.

Now, we cannot get from A to L (disjointly from B) without using vertices of S. Similarly, we cannot get from L to B (disjointly from A) without using vertices of S. As

$|S| = 2k-1$, any path system as in Lemma 3.3.1 will not be pairwise disjoint. Accordingly, Lemma 3.3.1 fails for this tournament, T. We remark that a slight modification of this construction yields a tournament which additionally has large minimum in and out-degree. Notice, however, that in this example there is an easy way to link the vertices from A to B. Accordingly, our approach to proving Theorem 3.1.1 might still be able to be extended to prove Pokrovskiy's conjecture.

Aside from improving our bound of $4k$ on the connectivity and resolving completely Pokrovskiy's conjecture, there are a few other open problems of interest. For example, what is the smallest function $d(k)$ such that Theorem 3.1.2 holds?

Problem 3.4.1. *Determine the smallest function $d : \mathbb{N} \to \mathbb{N}$ such that any tournament T with $\delta^+(T) \geq d(k)$ contains a subdivision of the complete directed graph $\overrightarrow{K}_k$.*

Note that our proof gives a doubly exponential bound on $d(k)$. Indeed, it is easy to check that $d(k) \leq 2^{2^{Ck^2}}$ for an absolute constant C. Furthermore, one can obtain a quadratic lower bound on $d(k)$: any subdivision of $\overrightarrow{K}_k$ must contain at least $3\binom{k}{2}$ vertices. So an appropriate balanced blow-up of a directed triangle on, say, $k^2/2$ vertices contains no subdivision of $\overrightarrow{K}_k$, yet has minimum out-degree quadratic in k. Accordingly, there is quite a large gap in our understanding of the function $d(k)$.

Finally, while the conclusion of Theorem 3.1.2 does not hold if we replace T with a general digraph, can we embed subdivisions of *acyclic* digraphs in digraphs of large minimum out-degree? We end by recalling the following beautiful conjecture of Mader [52] from 1985.

Conjecture 3.4.2. *For every positive integer k, there exists a function $f(k)$ such that every digraph with minimum out-degree at least $f(k)$ contains a subdivision of the transitive tournament of order k.*

Of course, since every acyclic digraph is contained in the transitive tournament of the same order, this conjecture (if true) would give an affirmative answer to the preceding question. Since large minimum out-degree (and in-degree) is not enough to embed subdivisions of a complete directed graph, it is natural to wonder whether some other parameter might allow us to do so. For example, if $\kappa(D)$ denotes the strong connectivity of a digraph, is it true that for every k there is an $f(k)$ such that any digraph D with $\kappa(D) \geq f(k)$ contains a subdivision of $\overrightarrow{K}_k$? (see Problem 16 in [1]). The answer to this question is still unknown.

CHAPTER 4

DISJOINT PAIRS IN SET SYSTEMS WITH RESTRICTED INTERSECTION

4.1 Introduction

Before getting into some mathematics let us recall a bit of notation, most of which is standard. Let $[n]$ denote the set $\{1,\ldots,n\}$, $\mathscr{P}[n]$ denote the power set of $[n]$, and let $[n]^{(r)}$ denote the collection of all subsets of $[n]$ of size r. For two set systems $\mathscr{A}$, $\mathscr{B} \subset \mathscr{P}[n]$, we let $d(\mathscr{A},\mathscr{B})$ denote the number of *disjoint pairs*; that is, the number of pairs $(A,B) \in \mathscr{A} \times \mathscr{B}$ with $A \cap B = \varnothing$. Similarly, for a set system $\mathscr{F} \subset \mathscr{P}[n]$ we let $d(\mathscr{F})$ denote the number of disjoint pairs in $\mathscr{F}$. Accordingly, $d(\mathscr{F}) = \frac{1}{2}d(\mathscr{F},\mathscr{F})$ (unless, of course, $\varnothing \in \mathscr{F}$, in which case $d(\mathscr{F}) = \frac{1}{2}(d(\mathscr{F},\mathscr{F}) - 1)$). We are interested in the maximum number of disjoint pairs a set system $\mathscr{F}$ can have under certain restrictions on the possible intersection sizes of elements of $\mathscr{F}$. For a set L of nonnegative integers, a set system $\mathscr{F}$ is said to be *L-intersecting* if $|F_1 \cap F_2| \in L$ for all distinct $F_1, F_2 \in \mathscr{F}$. Similarly, a pair of set systems $(\mathscr{A},\mathscr{B})$ is *L-cross-intersecting* if $|A \cap B| \in L$ whenever $A \in \mathscr{A}, B \in \mathscr{B}$. When $L = \{t,\ldots,n\}$ we say $\mathscr{F}$ is *t-intersecting*, and when $t = 1$ we shall simply say $\mathscr{F}$ is *intersecting*. Finally, if $L = [n] \setminus \{t\}$, we shall say that $\mathscr{F}$ (resp., $(\mathscr{A},\mathscr{B})$) is *t-avoiding* (resp., *t-cross-avoiding*).

4.1.1 Background

The problem of bounding the size of a set system under certain intersection restrictions has a central place in Extremal Set Theory. We shall not give a full account of such problems, but only touch upon some results that are particularly relevant for our purposes (for a broader account we refer the interested reader to the recent survey of Frankl and Tokushige [31]). The Erdős-Ko-Rado Theorem [24] is perhaps the most foundational result in this area, determining the maximum size of an intersecting r-uniform set system.

More precisely, if $n \geq 2r$ and $\mathscr{F} \subset [n]^{(r)}$ is an intersecting set system, then $|\mathscr{F}| \leq \binom{n-1}{r-1}$, and moreover, if $n > 2r$, then equality holds only when $\mathscr{F}$ consists of all r-sets containing a fixed element of the ground set. Numerous extensions and variations have been addressed over the years. Perhaps most notably, The Complete Intersection Theorem of Ahlswede and Khachatrian [3] determines the maximum size of a t-intersecting set system $\mathscr{F} \subset [n]^{(r)}$ for all values of n. In the non-uniform case, Katona [43] showed that any $(t+1)$-intersecting set system $\mathscr{F}$ satisfies

$$|\mathscr{F}| \leq |\mathscr{F}(n,t)|,$$

where $\mathscr{F}(n,t)$ is $\{A : |A| \geq \frac{n+t+1}{2}\}$ if $n+t$ is odd, or $\{A : |A \cap ([n] \setminus \{1\})| \geq \frac{n+t}{2}\}$ if $n+t$ is even. Trivially, if a set system is $(t+1)$-intersecting then it is also t-avoiding. Erdős asked what happens when we weaken the condition that all $F_1, F_2 \in \mathscr{F}$ satisfy $|F_1 \cap F_2| > t$ to $|F_1 \cap F_2| \neq t$. Frankl and Füredi [27] answered this question, showing that when $n \geq n_0(t)$ we recover the same asymptotic solution as in Katona's theorem. In particular, letting

$$\mathscr{F}^*(n,t) = \mathscr{F}(n,t) \cup [n]^{(\leq t-1)},$$

they showed that as long as $n \geq n_0(t)$ and $\mathscr{F} \subset \mathscr{P}([n])$ is t-avoiding, then $|\mathscr{F}| \leq |\mathscr{F}^*(n,t)|$.

In this paper, instead of focusing on the *size* of set systems with imposed intersection conditions, we are interested in the maximum number of disjoint pairs they can have. Alon and Frankl [5] addressed the problem of determining the maximum number of disjoint pairs in a set system of fixed size. Obviously, we always have $d(\mathscr{F}) < |\mathscr{F}|^2$, but for large families they showed that this bound is far off: if $\mathscr{F}$ has size $m = 2^{n/2+\delta n}$, then $d(\mathscr{F}) < m^{2-\delta^2/2}$. Problems concerning the minimum number of disjoint pairs in set systems have been studied by Ahlswede [2], Frankl [26], Bollobás and Leader [12], and Das, Gan, and Sudakov [19].

What happens to the maximum number of disjoint pairs if we impose the condition that $\mathscr{F}$ forbids a single intersection size? Our first result provides an upper bound for the maximum number for disjoint pairs in t-avoiding set systems for any $t \geq 1$.

Theorem 4.1.1. *Let n,t be positive integers with $t \leq n$ and suppose that $\mathscr{F} \subset \mathscr{P}[n]$ is t-avoiding. Then*

$$d(\mathscr{F}) \leq \frac{1}{2}\left(\sum_{k=0}^{t-1}\binom{n}{k}2^{n-k}-1\right).$$

Note that the number of disjoint pairs in $\mathscr{F}^*(n,t)$ is at least (assuming for simplicity that $n+t$ is odd)

$$\begin{aligned}\sum_{k=0}^{t-1}\binom{n}{k}\cdot\sum_{j=(n+t+1)/2}^{n-k}\binom{n-k}{j} &= \sum_{k=0}^{t-1}\binom{n}{k}(1-o(1))2^{n-k-1}\\ &= (1-o(1))\frac{1}{2}\sum_{k=0}^{t-1}\binom{n}{k}2^{n-k},\end{aligned}$$

as $n\to\infty$. Therefore, for large n the upper bound we obtain in Theorem 4.1.1 is essentially best possible. We conjecture that $\mathscr{F}^*(n,t)$ in fact maximizes the number of disjoint pairs for t-avoiding set systems (see Section 4.4). We shall actually prove a 'two-family' version which bounds the number of disjoint pairs in a pair $(\mathscr{A},\mathscr{B})$ of t-cross-avoiding set systems. In particular, Theorem 4.1.1 immediately follows from the following result.

Theorem 4.1.2. *Let n,t be positive integers with $t \leq n$ and suppose that $(\mathscr{A},\mathscr{B}) \subset \mathscr{P}[n]\times\mathscr{P}[n]$ is a pair of t-cross-avoiding set systems. Then*

$$d(\mathscr{A},\mathscr{B}) \leq \sum_{k=0}^{t-1}\binom{n}{k}2^{n-k}.$$

We remark that this is a generalization of a result in [41], where the case $t=1$ was established. We are also able to classify the extremal examples for Theorem 4.1.2. Namely, if $t=1$, then equality occurs if and only if $\mathscr{A}=\mathscr{P}(S), \mathscr{B}=\mathscr{P}([n]\setminus S)$ for some subset $S\subseteq[n]$, and if $t\geq 2$, equality holds if and only if $\mathscr{A}=[n]^{(\leq t-1)}, \mathscr{B}=\mathscr{P}[n]$.

Note that Theorem 4.1.2 has the following immediate corollary.

Corollary 4.1.3. *Let L be a set of s nonnegative integers and suppose that $(\mathscr{A},\mathscr{B})$ is a pair of L-cross-intersecting set systems. Then*

$$d(\mathscr{A},\mathscr{B}) \leq \sum_{k=0}^{s-1} \binom{n}{k} 2^{n-k},$$

with equality if and only if $L = \{0,\dots,s-1\}$.

Of course, the most trivial bound upper bound on $d(\mathscr{A},\mathscr{B})$ is given by the product $|\mathscr{A}||\mathscr{B}|$, and the problem of bounding $|\mathscr{A}||\mathscr{B}|$ for L-cross-intersecting $(\mathscr{A},\mathscr{B})$ has been studied before. For example, Keevash and Sudakov [44] proved that if L is a set of s nonnegative integers and n is sufficiently large (depending on s), then $|\mathscr{A}||\mathscr{B}| \leq \sum_{k=0}^{s-1} \binom{n}{k} 2^n$ for any L-cross-intersecting pair $(\mathscr{A},\mathscr{B})$ in $\mathscr{P}[n] \times \mathscr{P}[n]$, with equality if and only if $L = \{0,\dots,s-1\}$. Therefore, the same example that maximizes the number of disjoint pairs in Corollary 4.1.3 maximizes the product $|\mathscr{A}||\mathscr{B}|$, when n is sufficiently large. It is still unknown whether this bound holds for every s and n. The only general upper bound was given by Sgall [59]. In contrast, note that in Corollary 4.1.3, our bound holds for all s and n.

Motivated by Theorem 4.1.2, it is natural to ask what happens to the parameter $d(\mathscr{A},\mathscr{B})$ when we impose that $\mathscr{A},\mathscr{B} \subset [n]^{(r)}$ are both *uniform*. Here it turns out that avoiding an intersection is not that much of a restriction, at least when r is fixed and n is large. Consider the following family of examples.

Example 4.1.4. *For integers $r \geq 1$, $s \geq 0$ and a non-empty proper subset $X \subset [n]$ let $\mathscr{F}_{X,s} = \{F \in [n]^{(r)} : |F \cap X| \geq r-s\}$. For a positive integer $t \leq r$ and nonnegative integers a,b with $a+b \leq t-1$, consider the pair $\left(\mathscr{F}_{X,a}, \mathscr{F}_{X^c,b}\right)$. It is easy to see that this pair is t-cross-avoiding (in fact, it is $\{0,\dots,t-1\}$-cross-intersecting). Intuitively, the number of disjoint pairs should be maximized when $a = \lfloor \frac{t}{2} \rfloor$ and $b = \lfloor \frac{t-1}{2} \rfloor$ are as equal as possible.*

It is easy to see that $d\left(\mathscr{F}_{X,\lfloor\frac{t}{2}\rfloor}, \mathscr{F}_{X^c,\lfloor\frac{t-1}{2}\rfloor}\right) = \Theta_{r,t}\left(n^{2r}\right)$ and similarly $|\mathscr{F}_{X,\lfloor\frac{t}{2}\rfloor}||\mathscr{F}_{X^c,\lfloor\frac{t-1}{2}\rfloor}| = \Theta_{r,t}\left(n^{2r}\right)$, when $|X| \sim cn$ for some constant $c \in (0,1)$. While we began our investigation by considering maximizing the number of disjoint pairs, this example suggests that the problem of determining the maximum number of disjoint pairs in a t-cross-avoiding pair $(\mathscr{A},\mathscr{B})$ of r-uniform set systems is roughly equivalent to determining the maximum of the product $|\mathscr{A}||\mathscr{B}|$ when n is large and r,t remain fixed. In other words, good upper bounds on $|\mathscr{A}||\mathscr{B}|$ translate into good upper bounds on $d(\mathscr{A},\mathscr{B})$. To formalize this we shall introduce two functions. Let

- $d(n,r,t) = \max\{d(\mathscr{A},\mathscr{B}) : (\mathscr{A},\mathscr{B}) \subset [n]^{(r)} \times [n]^{(r)}$ is t-cross-avoiding$\}$,
- $p(n,r,t) = \max\{|\mathscr{A}||\mathscr{B}| : (\mathscr{A},\mathscr{B}) \subset [n]^{(r)} \times [n]^{(r)}$ is t-cross-avoiding$\}$.

We prove the following theorem, which states that these two functions are asymptotically equivalent. Here, and in the sequel, we assume that r and t are fixed and $n \to \infty$.

Theorem 4.1.5. *Let $r \geq t \geq 1$ be integers. Then*

$$p(n,r,t) = (1+o(1))d(n,r,t),$$

as $n \to \infty$.

In view of Theorem 4.1.5, it is perhaps more natural to provide upper bounds for the function $p(n,r,t)$ in the context of trying to obtain upper bounds for $d(n,r,t)$. The function $p(n,r,t)$ has been investigated before by Frankl and Rödl [30] when r and t are both linear in n. When $n \geq n_0(r,t)$, the problem of determining $p(n,r,t)$ can be viewed as the cross-analogue of a problem resolved by Frankl and Füredi [28]. They showed, in particular, that if n is sufficiently large and $\mathscr{F} \subset [n]^{(r)}$ is t-avoiding, then the family consisting of all r-sets containing a fixed $(t+1)$-set is optimal. Now, note that we may assume that $t < r$, as trivially $p(n,r,r) = \frac{1}{4}\binom{n}{r}^2$. We make progress in determining $p(n,r,t)$ in the first two cases, $t = 1$ and $t = 2$.

Theorem 4.1.6. *Let $r \geq 2$ be an integer. There exists $n_0 = n_0(r)$ such that if $n > n_0$ and $(\mathscr{A}, \mathscr{B})$ is a pair of 1-cross-avoiding r-uniform set systems, then*

$$|\mathscr{A}||\mathscr{B}| \leq \binom{\lfloor n/2 \rfloor}{r}\binom{\lceil n/2 \rceil}{r}.$$

This result is clearly tight: just consider the pair $(\mathscr{F}_{X,0}, \mathscr{F}_{X^c,0})$ where $X \subset [n]$ has size $\lfloor n/2 \rfloor$. It is also tight for the problem of maximizing $d(\mathscr{A}, \mathscr{B})$.

Our last theorem gives an asymptotically tight upper bound for $p(n,r,2)$.

Theorem 4.1.7. *Suppose $r \geq 3$ and let $(\mathscr{A}, \mathscr{B})$ be a pair of 2-cross-avoiding r-uniform set systems. Then*

$$|\mathscr{A}||\mathscr{B}| \leq (\gamma_r + o(1))\binom{n}{r}^2,$$

where $\gamma_r = \max_{\alpha \in [0,1]}\{\alpha^r(1-\alpha)^r + r\alpha^{r+1}(1-\alpha)^{r-1}\}$.

The pair $(\mathscr{F}_{X,1}, \mathscr{F}_{X^c,0})$ with $|X| = \alpha n$, where $\alpha \in [0,1]$ gives the maximum value γ_r above, shows that this upper bound is asymptotically optimal. Indeed, one need only apply the inequality

$$\binom{\theta x}{r} \leq \theta^r \binom{x}{r},$$

which is valid for all $\theta \in (0,1]$ with $\theta x > r$. Moreover, using Theorem 4.1.5, we have that $p(n,r,2) = (\gamma_r + o(1))\binom{n}{r}^2$ and $d(n,r,2) = (\gamma_r + o(1))\binom{n}{r}^2$. Notice that in the case of both Theorem 4.1.6 and Theorem 4.1.7, pairs $\left(\mathscr{F}_{X,a}, \mathscr{F}_{X^c,b}\right)$ where a and b are as equal as possible are optimal. We conjecture that this phenomenon persists for higher forbidden intersection sizes (see Section 4.4). The asymmetry suggested by the construction for the 2-cross-avoiding case may indicate, however, that this problem is difficult. For example, the value of $\alpha \in [0,1]$ giving γ_3 is $\frac{1}{8} + \frac{\sqrt{17}}{8} \approx .6404$.

4.1.2 Organization and Notation

The remainder of this chapter is organized as follows. In Section 4.2 we prove Theorem 4.1.2, which implies Theorem 4.1.1. In Section 4.3, we shall prove Theorem 4.1.5, Theorem 4.1.6, and Theorem 4.1.7. In the final section, we shall state some open problems.

Our notation is standard. For a set X we let $X^{(r)}$ (resp., $X^{(\leq r)}$) denote the collection of all r-element subsets of X (resp., subsets of X of size at most r). Any set system $\mathscr{F} \subset X^{(r)}$ is said to be *r-uniform* and its elements are *r-sets*. For $\mathscr{F} \subset \mathscr{P}[n]$ and $T \subset [n]$ we let $\mathscr{F}(T)$ denote the collection of sets in $\mathscr{F}$ that contain T. When $T = \{x\}$ is a singleton we shall simply write $\mathscr{F}(x)$.

4.2 Disjoint pairs in t-avoiding set systems

Our aim in this section is to establish Theorem 4.1.2, which we restate for convenience.

Theorem 4.1.2. *Let n,t be positive integers with $t \leq n$ and suppose that $(\mathscr{A},\mathscr{B}) \subset \mathscr{P}[n] \times \mathscr{P}[n]$ is a pair of t-cross-avoiding set systems. Then*

$$d(\mathscr{A},\mathscr{B}) \leq \sum_{k=0}^{t-1} \binom{n}{k} 2^{n-k}.$$

Let us point out one fact before giving a proof of the theorem. Note that if we let $f(n,t) = \sum_{k=0}^{t-1} \binom{n}{k} 2^{n-k}$, then f satisfies the recurrence

$$f(n,t) = 2f(n-1,t) + f(n-1,t-1),$$

for natural numbers $n,t \geq 1$.

Proof. We shall apply induction on n and t. The base case $t = 0$ holds trivially for every value of n. Therefore, we fix $t > 0$ and assume the theorem holds for $t' < t$ (and every value of n), and we may suppose the theorem holds for $t' = t$ and all $n' < n$. We aim to show it holds for $t' = t$ and $n' = n$.

To do so, suppose that $(\mathscr{A}, \mathscr{B}) \subset \mathscr{P}[n] \times \mathscr{P}[n]$ is t-cross-avoiding. We shall split $\mathscr{A}$ and $\mathscr{B}$ into certain subfamilies. More specifically, let $\mathscr{A}_n = \{A \in \mathscr{A} : n \in A\}$ and $\mathscr{A}_0 = \{A \in \mathscr{A} : n \notin \mathscr{A}\}$, and define $\mathscr{B}_n$ and $\mathscr{B}_0$ analogously. We further identify three subfamilies of $\mathscr{A}_n$, namely,

- $\mathscr{A}_n^* = \{A \in \mathscr{A}_n : A \setminus \{n\} \in \mathscr{A}\}$,
- $\mathscr{A}_n^{t+1} = \{A \in \mathscr{A}_n : \exists B \in \mathscr{B}_n \text{ with } |A \cap B| = t+1\}$, and
- $\mathscr{X} = \mathscr{A}_n \setminus (\mathscr{A}_n^* \cup \mathscr{A}_n^{t+1})$.

We define similarly the corresponding subfamilies $\mathscr{B}_n^*, \mathscr{B}_n^{t+1}$, and $\mathscr{Y} = \mathscr{B}_n \setminus (\mathscr{B}_n^* \cup \mathscr{B}_n^{t+1})$ of $\mathscr{B}_n$. Note that the subfamilies defined above actually partition $\mathscr{A}_n$ and $\mathscr{B}_n$. Indeed, suppose $A \in \mathscr{A}_n^* \cap \mathscr{A}_n^{t+1}$. Then there exists $B \in \mathscr{B}_n$ such that $|A \cap B| = t+1$. But we also have that $A \setminus \{n\} \in \mathscr{A}$ and then $|A \setminus \{n\} \cap B| = t$, a contradiction. The same argument shows that $\mathscr{B}_n^*$ and $\mathscr{B}_n^{t+1}$ are disjoint.

For a subset $A \subset [n]$, a family $\mathscr{F} \subset \mathscr{P}[n]$, and $i \in [n]$ let $D_i(A) = A \setminus \{i\}$ and

$$D_i(\mathscr{F}) = \{D_i(A) : A \in \mathscr{F}\}.$$

To reduce clutter we shall simply write D for D_n. Our aim is to apply D to a suitable pair of families and apply induction. Indeed, consider the pairs

$$(\mathscr{A}_0 \cup D(\mathscr{X} \cup \mathscr{A}_n^{t+1}), \mathscr{B}_0 \cup D(\mathscr{Y})),$$

and

$$(\mathscr{A}_0 \cup D(\mathscr{X}), \mathscr{B}_0 \cup D(\mathscr{Y} \cup \mathscr{B}_n^{t+1})).$$

Of course, each of the families in these pairs belongs to $\mathscr{P}[n-1]$. We also need that the above pairs are t-cross-avoiding, which we formulate in the following claim.

Claim 4.2.1. *$(\mathscr{A}_0 \cup D(\mathscr{X} \cup \mathscr{A}_n^{t+1}), \mathscr{B}_0 \cup D(\mathscr{Y}))$ and $(\mathscr{A}_0 \cup D(\mathscr{X}), \mathscr{B}_0 \cup D(\mathscr{Y} \cup \mathscr{B}_n^{t+1}))$ are t-cross-avoiding pairs of set systems.*

Proof. We only prove that the first pair is t-cross-avoiding. The second follows by a similar argument. By way of contradiction, suppose there exists $A \in \mathscr{A}_0 \cup D(\mathscr{X} \cup \mathscr{A}_n^{t+1})$ and $B \in \mathscr{B}_0 \cup D(\mathscr{Y})$ such that $|A \cap B| = t$. Clearly, either $B \in \mathscr{B}$ or $B \cup \{n\} \in \mathscr{B}$. If $A \in \mathscr{A}_0$, then $|A \cap B| = |A \cap (B \cup \{n\})| = t$, which is a contradiction. So we may assume that $A \cup \{n\} \in \mathscr{A}$ and similarly $B \cup \{n\} \in \mathscr{B}$. Hence $|A \cup \{n\} \cap B \cup \{n\}| = t+1$ which would imply $B \cup \{n\} \in \mathscr{B}_n^{t+1}$, which is again a contradiction. This completes the proof. $\square$

Our second claim exhibits a pair of subfamilies that are, in fact, $(t-1)$-cross-avoiding.

Claim 4.2.2. *The pair of set systems $(D(\mathscr{A}_n^*), D(\mathscr{B}_n^*))$ is $(t-1)$-cross-avoiding in $\mathscr{P}[n-1] \times \mathscr{P}[n-1]$.*

Indeed, suppose there is $A \in D(\mathscr{A}_n^*)$ and $B \in D(\mathscr{B}_n^*)$ such that $|A \cap B| = t-1$. But since $A' = A \cup \{n\} \in \mathscr{A}_n$ and $B' = B \cup \{n\} \in \mathscr{B}_n$, we have that $|A' \cap B'| = t$, a contradiction.

We shall now count the disjoint pairs (A, B) with $A \in \mathscr{A}$ and $B \in \mathscr{B}$ in such a way that every such pair gets counted except those disjoint pairs in $(D(\mathscr{A}_n^*), \mathscr{B}_n^*)$. The following lemma summarizes this, from which our theorem follows easily. Before stating it we shall rename some families in order to make the statement cleaner. Let

- $(\mathscr{A}_0 \cup D(\mathscr{X} \cup \mathscr{A}_n^{t+1}), \mathscr{B}_0 \cup D(\mathscr{Y})) = (\mathscr{F}_1, \mathscr{F}_2)$, and
- $(\mathscr{A}_0 \cup D(\mathscr{X}), \mathscr{B}_0 \cup D(\mathscr{Y} \cup \mathscr{B}_n^{t+1})) = (\mathscr{F}_3, \mathscr{F}_4)$.

With this in mind we shall prove the following.

Lemma 4.2.3. $d(\mathscr{F}_1,\mathscr{F}_2)+d(\mathscr{F}_3,\mathscr{F}_4)\geq d(\mathscr{A},\mathscr{B})-d(D(\mathscr{A}_n^*),D(\mathscr{B}_n^*))$.

Proof. Let us see how the left-hand side $d(\mathscr{F}_1,\mathscr{F}_2)+d(\mathscr{F}_3,\mathscr{F}_4)$ counts disjoint pairs. Note that it counts every disjoint pair in $(\mathscr{A}_n^{t+1}\cup\mathscr{X},\mathscr{B}_0)$ and $(\mathscr{A}_0,\mathscr{B}_n^{t+1}\cup\mathscr{Y})$ once (it may count more; namely, disjoint pairs in $(D(\mathscr{X}),D(\mathscr{Y}))$ that do not exist in $(\mathscr{A},\mathscr{B})$). Furthermore, it counts disjoint pairs in $(\mathscr{A}_0,\mathscr{B}_0)$ twice. Such pairs between $\mathscr{A}_0$ and $\mathscr{B}_0$ can be broken up into the following three types:

- those in $(D(\mathscr{A}_n^*),D(\mathscr{B}_n^*))$;
- those in $(D(\mathscr{A}_n^*),\mathscr{B}_0\setminus D(\mathscr{B}_n^*))$;
- those in $(\mathscr{A}_0\setminus D(\mathscr{A}_n^*),D(\mathscr{B}_n^*))$.

The remaining disjoint pairs to be counted are those in $(\mathscr{A}_n^*,\mathscr{B}_0)$ and $(\mathscr{A}_0,\mathscr{B}_n^*)$. Since

$$d(D(\mathscr{A}_n^*),\mathscr{B}_0\setminus D(\mathscr{B}_n^*))=d(\mathscr{A}_n^*,\mathscr{B}_0\setminus D(\mathscr{B}_n^*)),$$

and, similarly, $d(\mathscr{A}_0\setminus D(\mathscr{A}_n^*),D(\mathscr{B}_n^*))=d(\mathscr{A}_0\setminus D(\mathscr{A}_n^*),\mathscr{B}_n^*)$, we have that the disjoint pairs in $(\mathscr{A}_n^*,\mathscr{B}_0\setminus D(\mathscr{B}_n^*))$ and $(\mathscr{A}_0\setminus D(\mathscr{A}_n^*),\mathscr{B}_n^*)$ get counted when we count those disjoint pairs in $(\mathscr{A}_0,\mathscr{B}_0)$. Furthermore, since $d(D(\mathscr{A}_n^*),D(\mathscr{B}_n^*))=d(\mathscr{A}_n^*,D(\mathscr{B}_n^*))$, the disjoint pairs in $(\mathscr{A}_n^*,D(\mathscr{B}_n^*))$ also get counted whenever we count pairs in $(\mathscr{A}_0,\mathscr{B}_0)$. As $d(\mathscr{F}_1,\mathscr{F}_2)+d(\mathscr{F}_3,\mathscr{F}_4)$ counts the disjoint pairs in $(\mathscr{A}_0,\mathscr{B}_0)$ twice we can equivalently say that it counts

- disjoint pairs in $(\mathscr{A}_0,\mathscr{B}_0)$ once;
- disjoint pairs in $(\mathscr{A}_0\setminus D(\mathscr{A}_n^*),\mathscr{B}_n^*)$ once;
- disjoint pairs in $(\mathscr{A}_n^*,\mathscr{B}_0\setminus D(\mathscr{B}_n^*))$ once;

- disjoint pairs in $(\mathscr{A}_n^*, D(\mathscr{B}_n^*))$ once.

Thus the only disjoint pairs not counted are those in $(D(\mathscr{A}_n^*), \mathscr{B}_n^*)$, and since $d(D(\mathscr{A}_n^*), \mathscr{B}_n^*) = d(D(\mathscr{A}_n^*), D(\mathscr{B}_n^*))$ we have that

$$d(\mathscr{F}_1, \mathscr{F}_2) + d(\mathscr{F}_3, \mathscr{F}_4) \geq d(\mathscr{A}, \mathscr{B}) - d(D(\mathscr{A}_n^*), D(\mathscr{B}_n^*)),$$

as claimed.

□

Theorem 4.1.2 now follows easily from Lemma 4.2.3. Indeed, by Claim 4.2.1, $(\mathscr{F}_1, \mathscr{F}_2)$ and $(\mathscr{F}_3, \mathscr{F}_4)$ are both t-cross-avoiding in $\mathscr{P}[n-1] \times \mathscr{P}[n-1]$, so by induction we have $d(\mathscr{F}_1, \mathscr{F}_2) \leq f(n-1,t)$ and $d(\mathscr{F}_3, \mathscr{F}_4) \leq f(n-1,t)$. By Claim 4.2.2, $(D(\mathscr{A}_n^*), D(\mathscr{B}_n^*)) \subset \mathscr{P}[n-1] \times \mathscr{P}[n-1]$ is $(t-1)$-cross-avoiding, and so $d(D(\mathscr{A}_n^*), D(\mathscr{B}_n^*)) \leq f(n-1,t-1)$. Therefore, by Lemma 4.2.3 and using the recurrence for f, we have

$$d(\mathscr{A}, \mathscr{B}) \leq 2f(n-1,t) + f(n-1,t-1) = f(n,t),$$

as claimed. □

To end this section, let us classify the extremal examples occurring in Theorem 4.1.2. We must break the analysis up into two cases, when $t = 1$ and when $t > 1$, as the extremal behavior is different. We consider first the case $t > 1$.

- $t > 1$

Observe that when $n = t$ equality is trivially only attained when the families are $(\mathscr{P}[n] \setminus \{[n]\}, \mathscr{P}[n]) = ([n]^{(\leq n-1)}, \mathscr{P}[n])$. We may assume now that $n > t$. From the proof of Theorem 4.1.2, both pairs $(\mathscr{F}_1, \mathscr{F}_2)$ and $(\mathscr{F}_3, \mathscr{F}_4)$ must satisfy $d(\mathscr{F}_1, \mathscr{F}_2) = d(\mathscr{F}_3, \mathscr{F}_4) = f(n-1,t)$. By induction on n, we may assume without loss of generality that $\mathscr{A}_0 \cup D(\mathscr{X} \cup \mathscr{A}_n^{t+1}) = \mathscr{P}[n-1]$ and $\mathscr{B}_0 \cup D(\mathscr{Y}) = [n-1]^{(\leq t-1)}$. Since $\varnothing \in \mathscr{A}_0$ (as $t \geq 1$) and, by the definition of $\mathscr{Y}$, for any element $B \in \mathscr{Y}$, $B \setminus \{n\}$ can be added to $\mathscr{B}_0$ implying that $\mathscr{Y}$ is empty. We then have that $\mathscr{B} = \mathscr{B}_0 \cup \mathscr{B}_n^* \cup \mathscr{B}_n^{t+1}$ and $\mathscr{B}_0 = [n-1]^{(\leq t-1)}$. Similarly we must have that $\mathscr{X}$ is empty and so $\mathscr{A} = \mathscr{A}_0 \cup \mathscr{A}_n^* \cup \mathscr{A}_n^{t+1}$ and $\mathscr{A}_0 \cup D(\mathscr{A}_n^{t+1}) = \mathscr{P}[n-1]$. Moreover, we must have that $d(\mathscr{F}_3, \mathscr{F}_4) = d(\mathscr{A}_0, \mathscr{B}_0 \cup D(\mathscr{B}_n^{t+1})) = f(n-1,t)$ and again by induction, either $\mathscr{A}_0 = \mathscr{P}[n-1]$ and $\mathscr{B}_0 \cup D(\mathscr{B}_n^{t+1}) = [n-1]^{(\leq t-1)}$ or $\mathscr{A}_0 = [n-1]^{(\leq t-1)}$ and $\mathscr{B}_0 \cup D(\mathscr{B}_n^{t+1}) = \mathscr{P}[n-1]$. We split our analysis into two parts according to whether the former or latter case holds.

(i). Suppose the latter case holds. Then any set $A \in [n-1]^{(t)}$ must be of the form $D(A')$ for some $A' \in \mathscr{A}_n^{t+1}$ and similarly of the form $D(B')$ for some $B' \in \mathscr{B}_n^{t+1}$. If $n > t+1$ then we reach an immediate contradiction as we can find two elements $A, B \in [n-1]^{(t)}$ with $|A \cap B| = t-1$ which would imply $|(A \cup \{n\}) \cap (B \cup \{n\})| = t$. So suppose that $n = t+1$. Now, neither $\mathscr{A}_n^{t+1}$ nor $\mathscr{B}_n^{t+1}$ can be empty. For if $\mathscr{A}_n^{t+1} = \varnothing$, then $\mathscr{P}[t] = \mathscr{A}_0 = [t]^{(\leq t-1)}$, which is a contradiction. Similarly, $\mathscr{B}_n^{t+1} \neq \varnothing$. Then $\mathscr{A}_n^{t+1} = \mathscr{B}_n^{t+1} = \{[n]\}$, and so no set $A \in [n]^{(t)}$ can belong to either $\mathscr{A}$ or $\mathscr{B}$. It follows that the only sets that can belong to $\mathscr{A}_n^*$ are of the form $A \cup \{n\}$ for some $A \in \mathscr{A}_0$ with $|A| \leq t-2$ (and similarly for the sets in $\mathscr{B}_n^*$). Accordingly, $\mathscr{A} = \mathscr{B} = [n]^{(\leq t-1)} \cup \{[n]\}$, which is impossible as the number of disjoint pairs is smaller than $f(n,t)$.

(ii). Suppose the former case holds, that $\mathscr{A}_0 = \mathscr{P}[n-1]$ and $\mathscr{B}_0 \cup D(\mathscr{B}_n^{t+1}) = [n-1]^{(\leq t-1)}$. Since we cannot have an element $A \in \mathscr{A}_n^{t+1}$ and $D(A) \in A_0$, we must have $A_n^{t+1} = \varnothing$ and, analogously, $B_n^{t+1} = \varnothing$. It follows that $\mathscr{A} = \mathscr{A}_n^* \cup \mathscr{P}[n-1]$ and $\mathscr{B} = \mathscr{B}_n^* \cup [n-1]^{(\leq t-1)}$. Let us first deal with the case $t = 2$. If $\mathscr{B}_n^*$ contains no sets of the form $\{i,n\}$, then $\mathscr{B}_n^* = \{\{n\}\}$ and we are done. Our aim is to show that if $\mathscr{B}_n^*$ contains a 2-set, then the number of disjoint pairs is strictly smaller than $f(n,2)$. So suppose, by way of contradiction, that $\mathscr{B}_n^*$ contains sets $\{i_1,n\}, \ldots, \{i_l,n\}$ for some $i_1, \ldots, i_l \in [n-1]$. It follows that $\mathscr{A}_n^*$ can consist of only sets containing n and avoiding $i_1,\dots,i_l$. Therefore, we may assume $|\mathscr{A}_n^*| = 2^{n-1-l}$. The number of disjoint pairs between $\mathscr{P}[n-1]$ and $\mathscr{B}$ is $2^{n-1}+(n-1)2^{n-2}+2^{n-1}+l2^{n-2} = 2^n+(n-1+l)2^{n-2}$. The number of disjoint pairs between $\mathscr{A}_n^*$ and $\mathscr{B}$ is $(l+1)2^{n-1-l}+(n-1-l)2^{n-2-l}$. Since $f(n,2) = 2^n + n2^{n-1}$ we have to check that

$$(n-1+l)2^{n-2}+(l+1)2^{n-1-l}+(n-1-l)2^{n-2-l} < n2^{n-1}, \tag{4.1}$$

for $1 \le l \le n-1$. It is easy to check that (4.1) holds for $l = 1,2$ (bearing in mind that we may assume $n > 2$). Further, (4.1) is equivalent to $n > \frac{2^l(l-1)+l+1}{2^l-1}$, which is true since $\frac{2^l(l-1)+l+1}{2^l-1} \le l$ for $l \ge 3$, and also since $l < n$. Accordingly, $\mathscr{B}_n^*$ contains no 2-sets, and so the proof is complete for $t = 2$.

Finally, we see in the proof of Theorem 4.1.2 that in order to have equality, it must hold that $d(D(\mathscr{A}_n^*), D(\mathscr{B}_n^*)) = f(n-1,t-1)$. By induction on n and t ($t = 2$ being the base case), we have that $D(\mathscr{A}_n^*) = \mathscr{P}[n-1]$ and $D(\mathscr{B}_n^*) = [n-1]^{(\le t-2)}$. So, since $\mathscr{A} = \mathscr{A}_n^* \cup \mathscr{P}[n-1]$ and $\mathscr{B} = \mathscr{B}_n^* \cup [n-1]^{(\le t-1)}$, it follows that $\mathscr{A} = \mathscr{P}[n]$ and $\mathscr{B} = [n]^{(\le t-1)}$ as required.

- $t = 1$

We claim that equality holds only if $\mathscr{A} = \mathscr{P}(S), \mathscr{B} = \mathscr{P}([n]\setminus S)$ for some $S \subseteq [n]$. This is certainly true for $n = 1$. Let $n \ge 2$ and suppose the result holds for smaller values of n. As before, since both pairs $(\mathscr{F}_1,\mathscr{F}_2)$ and $(\mathscr{F}_3,\mathscr{F}_4)$ must satisfy $d(\mathscr{F}_1,\mathscr{F}_2) = d(\mathscr{F}_3,\mathscr{F}_4) = f(n-1,1)$, by induction on n, we may assume $\mathscr{F}_1 = \mathscr{A}_0 \cup D(\mathscr{X} \cup \mathscr{A}_n^2) = \mathscr{P}(W)$ for some $W \subseteq [n-1]$ and $\mathscr{F}_2 = \mathscr{B}_0 \cup D(\mathscr{Y}) = \mathscr{P}([n-1]\setminus W)$. Similarly $\mathscr{F}_3 = \mathscr{A}_0 \cup D(\mathscr{X}) = \mathscr{P}(W')$ and $\mathscr{F}_4 = \mathscr{B}_0 \cup D(\mathscr{Y} \cup \mathscr{B}_n^2) = \mathscr{P}([n-1]\setminus W')$. Note that as before we may assume $\mathscr{X}$ and $\mathscr{Y}$ are empty.

Clearly we have that $W' \subseteq W$ and we shall show they actually must be equal. Suppose first that $|W \setminus W'| \geq 2$ and let i_1, i_2 be two distint elements in $W \setminus W'$. By definition, the sets $\{i_1\}, \{i_2\}$ belong to $\mathscr{A}_0 \cup D(\mathscr{A}_n^2)$ and to $\mathscr{B}_0 \cup D(\mathscr{B}_n^2)$. But this implies both $\{i_1, n\}, \{i_2, n\}$ belong to $\mathscr{A}_n^2$ and to $\mathscr{B}_n^2$, which is a contradiction since we generate a cross-intersection of size 1. So we may assume that $W \setminus W' = \{i\}$, which implies $\{i, n\}$ belongs to $\mathscr{A}_n^2$ and to $\mathscr{B}_n^2$. Note that both $\mathscr{A}_n^*, \mathscr{B}_n^*$ are empty. Indeed, for any element $A \in \mathscr{A}_n^*$ (or $\mathscr{B}_n^*$), the set $A \setminus \{n\}$ belongs to $\mathscr{A}_0$(or $\mathscr{B}_0$) and therefore $A \setminus \{n\} \subseteq W'$ (or $A \setminus \{n\} \subseteq [n-1] \setminus W$). In any case, $A \cap \{i, n\} = \{n\}$, which is impossible. We must then have that $\mathscr{A} = \mathscr{P}(W') \cup (\{i, n\} \vee \mathscr{P}(W'))$ and $\mathscr{B} = \mathscr{P}([n-1] \setminus W) \cup (\{i, n\} \vee \mathscr{P}([n-1] \setminus W))$, for some $W \in \mathscr{P}([n-1])$ and $i \in [n-1]$ with $W' = W \setminus \{i\}$ (as usual, for a set A and a family $\mathscr{F}$, $A \vee \mathscr{F} := \{A \cup F : F \in \mathscr{F}\}$). A simple calculation shows there are exactly $2^{n-2} + 2^{n-1} < 2^n$ disjoint pairs in $(\mathscr{A}, \mathscr{B})$, a contradiction. It follows that $W = W'$. Hence $\mathscr{A}_n^2$ and $\mathscr{B}_n^2$ must be empty. Clearly at most one of the sets $\mathscr{A}_n^*, \mathscr{B}_n^*$ can be non-empty, and our result follows.

4.3 Disjoint pairs in uniform set systems

Our aim in this section is prove Theorems 4.1.5, 4.1.6, and 4.1.7. We first prove Theorem 4.1.5 which provides a relation between the maximum number of disjoint pairs and the maximum size of the product of two t-cross-avoiding r-uniform set systems. Recall that, for positive integers $t \leq r$ we have defined $d(n, r, t)$ to be the maximum of $d(\mathscr{A}, \mathscr{B})$ over all t-cross-avoiding r-uniform $(\mathscr{A}, \mathscr{B})$ on the ground set $[n]$. Analogously, we have defined $p(n, r, t)$ to be the maximum of the product $|\mathscr{A}||\mathscr{B}|$ over all such pairs of set systems. To these two functions we add a third:

$$p^*(n, r, t) := \max\{|\mathscr{A}||\mathscr{B}| : (\mathscr{A}, \mathscr{B}) \subset [n]^{(r)} \times [n]^{(r)} \text{ is } \{0, \ldots, t-1\}\text{-cross-intersecting}\}.$$

Clearly, $p^*(n, r, t) \leq p(n, r, t)$. In order to prove Theorem 4.1.5, we first show that $p(n, r, t) \sim p^*(n, r, t)$ as $n \to \infty$. First, let us recall a notion that will be useful in the proof.

Let $\mathscr{F}$ be a family of subsets of $[n]$. A *delta-system* in $\mathscr{F}$ of size s with core C is a collection of sets $F_1, \ldots, F_s \in \mathscr{F}$ such that for every $i \neq j$, $F_i \cap F_j = \cap_{k=1}^{s} F_k = C$. We shall prove the following.

Lemma 4.3.1. *Let t, r be positive integers with $t \leq r$. Then*

$$p(n,r,t) \leq p^*(n,r,t) + C_{r,t} n^{2r-1},$$

for some constant $C_{r,t}$ depending on r and t.

Proof. Let $(\mathscr{A}, \mathscr{B})$ be a t-avoiding pair of r-uniform families with $|\mathscr{A}||\mathscr{B}| = p(n,r,t)$. We say that a t-set $T \subset [n]$ is *$\mathscr{A}$-good* (resp., *$\mathscr{B}$-good*) if there exists a delta-system in $\mathscr{A}$ (resp., $\mathscr{B}$) of size at least $r-t+1$ with core T. Observe that if T is $\mathscr{A}$-good, then no set in $\mathscr{B}$ contains T (the symmetric claim holds if T is $\mathscr{B}$-good). Indeed, suppose otherwise that some $B \in \mathscr{B}$ contains T. Let $\Delta \subset \mathscr{A}$ be the corresponding delta-system with core T, so that $|\Delta| \geq r-t+1$. Then $B \setminus T$ has size $r-t$ and accordingly there exists $A \in \Delta$ such that

$$(A \setminus T) \cap (B \setminus T) = \varnothing.$$

It follows that $|A \cap B| = t$, a contradiction.

Let $\mathscr{T}$ be the collection of t-sets which are neither $\mathscr{A}$-good nor $\mathscr{B}$-good and let

$$\mathscr{A}_0 = \bigcup_{T \in \mathscr{T}} \mathscr{A}(T) \quad \text{and} \quad \mathscr{B}_0 = \bigcup_{T \in \mathscr{T}} \mathscr{B}(T).$$

We claim that the subfamilies $\mathscr{A}_0$ and $\mathscr{B}_0$ are small. Indeed, suppose $T \in \mathscr{T}$. Then any maximum-sized delta-system $\Delta \subset \mathscr{A}$ with core T has size $|\Delta| \leq r-t$. It follows that any set in $\mathscr{A}(T)$ must non-trivially intersect a set in Δ outside of T. Therefore, somewhat crudely, we may bound

$$|\mathscr{A}(T)| \leq 2(r-t)^2 \binom{n}{r-t-1},$$

and the same bound holds for $|\mathscr{B}(T)|$. Accordingly, $|\mathscr{A}_0|, |\mathscr{B}_0| \leq c_{r,t}n^{r-1}$ for some constant $c_{r,t}$, depending only on r and t. Now, let

$$\mathscr{A}' = \mathscr{A} \setminus \mathscr{A}_0 \quad \text{and} \quad \mathscr{B}' = \mathscr{B} \setminus \mathscr{B}_0,$$

and note that the pair $(\mathscr{A}', \mathscr{B}')$ is $\{0, \ldots, t-1\}$-intersecting, for if $A' \in \mathscr{A}'$ and $B' \in \mathscr{B}'$ intersect in t points, then this t-set is both $\mathscr{A}$-good and $\mathscr{B}$-good, which is impossible. Finally, we see that

$$\begin{aligned} p^*(n,r,t) \geq |\mathscr{A}'||\mathscr{B}'| &= (|\mathscr{A}| - |\mathscr{A}_0|)(|\mathscr{B}| - |\mathscr{B}_0|) \\ &\geq |\mathscr{A}||\mathscr{B}| - O_{r,t}\left(n^{2r-1}\right) \\ &= p(n,r,t) - O_{r,t}\left(n^{2r-1}\right), \end{aligned}$$

completing the proof.

□

With Lemma 4.3.1 in mind we can now complete the proof of Theorem 4.1.5, which asserts that the functions $p(n,r,t)$ and $d(n,r,t)$ are essentially equivalent as $n \to \infty$.

Proof of Theorem 4.1.5. First note that $p^*(n,r,t) \leq d(n,r,t) + C_{r,t}n^{2r-1}$ for some constant $C_{r,t}$ depending on r, t. Indeed, if $(\mathscr{A}, \mathscr{B})$ is $\{0, \ldots, t-1\}$-cross-intersecting with $|\mathscr{A}||\mathscr{B}| = p^*(n,r,t)$, then we can count

$$|\mathscr{A}||\mathscr{B}| = d(\mathscr{A}, \mathscr{B}) + \sum_{\substack{A \in \mathscr{A}, B \in \mathscr{B} \\ A \cap B \neq \varnothing}} 1.$$

Now, for each element $A \in \mathscr{A}$ there are at most $2^t \binom{n-r}{r-1}$ sets in $[n]^{(r)}$ which have non-empty intersection with A. Hence, the second summand on the right-hand side is bounded by $|\mathscr{A}|2^t \binom{n-r}{r-1} \leq 2^t \binom{n}{r}\binom{n-r}{r-1} \leq C_{r,t}n^{2r-1}$.

Now, applying Lemma 4.3.1 we see that

$$p(n,r,t) \leq d(n,r,t) + c_{r,t}n^{2r-1},$$

for some constant $c_{r,t}$ depending on r,t. Example 4.1.4 shows that $d(n,r,t) = \Omega_{r,t}\left(n^{2r}\right)$, and so the result holds as claimed.

$\square$

In the next two subsections we shall shift our focus to proving upper bounds for $p(n,r,t)$ in the first two cases $t = 1,2$. When $t = 1$, the extremal example exhibits some symmetry (in particular, both families have the same size). This symmetry disappears when $t = 2$, indicating that the problem of bounding $p(n,r,t)$ for general t could be quite challenging.

4.3.1 Forbidding an intersection of size 1

It is very easy to give an upper bound for $p^*(n,r,1)$, and so, by Lemma 4.3.1, this translates to an asymptotic upper bound for $p(n,r,1)$. Indeed, if $\mathscr{A},\mathscr{B} \subset [n]^{(r)}$ are $\{0\}$-cross-intersecting, then rather trivially $(\bigcup_{A\in\mathscr{A}} A) \cap (\bigcup_{B\in\mathscr{B}} B) = \varnothing$, so we may assume that $\mathscr{A} = X^{(r)}$ and $\mathscr{B} = ([n]\setminus X)^{(r)}$ for some set $X \subset [n]$. If $|X| = x$, then we have

$$|\mathscr{A}||\mathscr{B}| = \binom{x}{r}\binom{n-x}{r}.$$

It can be shown that the function $f(x) = \binom{x}{r}\binom{n-x}{r}$ is concave for $r-1 \leq x \leq n-r+1$. Hence, we have

$$f(x) = \frac{1}{2}(f(x)+f(n-x)) \leq f(n/2).$$

Therefore,

$$p(n,r,1) = (1+o(1))\binom{n/2}{r}^2.$$

However, in this case we are able to remove the error term and prove that $p(n,r,1) \leq \binom{\lfloor n/2 \rfloor}{r}\binom{\lceil n/2 \rceil}{r}$, for n sufficiently large compared with r.

Theorem 4.1.6. *Let $r \geq 2$ be an integer. There exists $n_0 = n_0(r)$ such that if $n > n_0$ and $(\mathscr{A}, \mathscr{B})$ is a pair of* 1*-cross-avoiding r-uniform set systems, then*

$$|\mathscr{A}||\mathscr{B}| \leq \binom{\lfloor n/2 \rfloor}{r}\binom{\lceil n/2 \rceil}{r}.$$

Proof. Suppose that $\mathscr{A}, \mathscr{B}$ are 1-avoiding and maximize $|\mathscr{A}||\mathscr{B}|$, and suppose without loss of generality that $|\mathscr{A}| \geq \binom{\lfloor n/2 \rfloor}{r}$. As in the proof of Lemma 4.3.1, we give a reduction via delta-systems. More precisely, recall that we say $x \in [n]$ is $\mathscr{A}$*-good* (resp., $\mathscr{B}$*-good*) if there exists a delta-system in $\mathscr{A}$ (resp., $\mathscr{B}$) of size at least r with core $\{x\}$. Let X and Y denote the set of $\mathscr{A}$-good and $\mathscr{B}$-good points, respectively, and observe that $A \cap Y = \varnothing$ for every $A \in \mathscr{A}$ and $B \cap X = \varnothing$ for every $B \in \mathscr{B}$. We therefore obtain a partition $[n] = X \cup Y \cup Z$ where Z denotes the set of points which are neither $\mathscr{A}$-good nor $\mathscr{B}$-good. We may also assume that $X^{(r)} \subset \mathscr{A}$ and $Y^{(r)} \subset \mathscr{B}$. It also follows from the proof of Lemma 4.3.1 that, if $\mathscr{A}_0 := \{A \in \mathscr{A} : A \cap Z \neq \varnothing\}$ and $\mathscr{B}_0 := \{B \in \mathscr{B} : B \cap Z \neq \varnothing\}$, then $|\mathscr{A}_0|, |\mathscr{B}_0| \leq 2(r-1)^2\binom{n}{r-2}n \leq \frac{2(r-1)^2}{(r-2)!}n^{r-1}$. Let us write $|\mathscr{A}| = \binom{x}{r} + |\mathscr{A}_0|$ and $|\mathscr{B}| = \binom{y}{r} + |\mathscr{B}_0|$, where $x := |X|$ and $y := |Y|$, so

$$|\mathscr{A}||\mathscr{B}| = \binom{x}{r}\binom{y}{r} + |\mathscr{A}_0|\binom{y}{r} + |\mathscr{B}_0|\binom{x}{r} + |\mathscr{A}_0||\mathscr{B}_0|.$$

The rest of the proof will be broken into two claims. The first claim asserts that we may assume that the size of Y is large (i.e., linear in n). The second claim states that, under the assumption that $\mathscr{A}, \mathscr{B}$ maximize $|\mathscr{A}||\mathscr{B}|$, no point of $[n]$ can be neither $\mathscr{A}$-good nor $\mathscr{B}$-good. We therefore obtain the structural information that $\mathscr{A} = X^{(r)}$ and $\mathscr{B} = Y^{(r)}$.

Claim 4.3.2. *We may assume that $y \geq \beta n$, where $\beta = \beta(r) = \frac{(r!)^{2/r}}{200^{1/r} 4 r^2}$ (as long as n is sufficiently large).*

Proof. Put $c_r = \frac{2(r-1)^2}{(r-2)!}$, let β be as above, and suppose that $y < \beta n$. Using the fact that $|\mathscr{A}_0|, |\mathscr{B}_0| \leq c_r n^{r-1}$ and crudely bounding $\binom{x}{r} \leq \binom{n}{r}$, we have that

$$\begin{aligned}|\mathscr{A}||\mathscr{B}| &\leq \binom{n}{r}\binom{\beta n}{r} + 2c_r n^{2r-1} + c_r^2 n^{2r-2} \\ &\leq \frac{\beta^r}{(r!)^2} n^{2r} + 3c_r n^{2r-1},\end{aligned}$$

where in the first line we have used the monotonicity of the function $z \mapsto \binom{z}{r}$ (for $z \geq r-1$) and the inequality $\binom{\theta n}{r} \leq \theta^r \binom{n}{r}$, valid for any $\theta \in (0,1)$ with $\theta n > r$. Assuming that $n \geq 600 c_r 4^r r^{2r}$ we have that $3c_r/n \leq 1/200 4^r r^{2r}$, and therefore by our assumption on β

$$|\mathscr{A}||\mathscr{B}| < \frac{1}{100}\frac{n^{2r}}{4^r r^{2r}} \leq \frac{1}{100}\binom{n/2}{r}^2 \leq \binom{\lfloor n/2 \rfloor}{r}\binom{\lceil n/2 \rceil}{r},$$

completing the proof of Claim 4.3.2. □

The proof of Theorem 4.1.6 will be nearly finished once we establish that the set Z of points which are neither $\mathscr{A}$-good nor $\mathscr{B}$-good is empty. Our second claim asserts just this.

Claim 4.3.3. $Z = \varnothing$.

Proof. Suppose to the contrary that there is some $x \in Z$. Form a new pair $\mathscr{A}', \mathscr{B}'$ of 1-avoiding families in the following way. First, create $\mathscr{A}'$ by removing all sets of $\mathscr{A}$ that contain x. We are then free to add to $\mathscr{B}$ all sets of the form $B \cup \{x\}$ where $B \subset Y$ is a subset of size $r-1$ (note that as long as $\mathscr{A}(x) \neq \varnothing$, none of these sets originally belonged to $\mathscr{B}$ as otherwise there would be a cross-intersection of size 1). It follows that $\mathscr{A}', \mathscr{B}'$ are 1-avoiding and

$$|\mathscr{A}'||\mathscr{B}'| = (|\mathscr{A}| - |\mathscr{A}(x)|)\left(|\mathscr{B}| + \binom{y}{r-1}\right)$$

$$= |\mathscr{A}||\mathscr{B}| + |\mathscr{A}|\binom{y}{r-1} - |\mathscr{A}(x)||\mathscr{B}| - |\mathscr{A}(x)|\binom{y}{r-1},$$

so if $|\mathscr{A}| > \left(|\mathscr{B}|\binom{y}{r-1}^{-1} + 1\right)|\mathscr{A}(x)|$, then we reach a contradiction to the maximality of $|\mathscr{A}||\mathscr{B}|$. But by Claim 4.3.2 we have $\binom{y}{r-1} \geq \frac{\beta^{r-1}}{(r-1)^{r-1}} n^{r-1}$ and so the right-hand side is at most

$$c_r n^{r-2}\left(1 + \frac{c_r(r-1)^{r-1}}{\beta^{r-1} r!} n\right) \leq \frac{2c_r^2(r-1)^{r-1}}{\beta^{r-1} r!} n^{r-1} \tag{4.2}$$

Now, as long as $n > \frac{1002^{r+1} c_r^2 r^r (r-1)^{r-1}}{\beta^{r-1} r!}$, the right-hand side of (4.2) is strictly less than

$$\frac{n^r}{1002^r r^r} \leq \frac{1}{100}\binom{n/2}{r} \leq \binom{\lfloor n/2 \rfloor}{r} \leq |\mathscr{A}|,$$

and the proof of Claim 4.3.3 is complete.

□

Since $Z = \varnothing$ it follows that $\mathscr{A} = X^{(r)}$ and $\mathscr{B} = Y^{(r)}$. Accordingly, $|\mathscr{A}||\mathscr{B}| = \binom{x}{r}\binom{n-x}{r}$, where $x = |X|$ is an integer. This product is maximized when x and $n-x$ are as equal as possible, so since $x \geq \lfloor n/2 \rfloor$ we must have $x = \lfloor n/2 \rfloor$, establishing Theorem 4.1.6.

□

We have not made an attempt to optimize the value of $n_0(r)$ in our proof. The value we obtain is exponential in r; on the other hand, n_0 must be exponential in r for such a result to hold. To see this, consider the following example $(\mathscr{A}_0, \mathscr{B}_0)$ where $\mathscr{A}_0 = \mathscr{B}_0 = \{A \in [n]^{(r)} : 1, 2 \in A\}$. This pair is certainly 1-cross-avoiding and

$$|\mathscr{A}_0||\mathscr{B}_0| = \binom{n-2}{r-2}^2$$

It is not difficult to check that $\binom{n-2}{r-2}^2$ is larger than $\binom{n/2}{r}^2$ whenever $n < (r-1)4^{r/4}$. Thus the above star construction dominates whenever n is small, and it would be interesting to investigate our problem in this regime of n.

4.3.2 Forbidding an intersection of size 2

The extremal example showing that Theorem 4.1.6 is tight is symmetric in the sense that both families in the pair have the same size. We shall see now that this kind of symmetry is lost when forbidding a cross-intersection of size 2. However, in view of our reduction via Lemma 4.3.1, Theorem 4.1.7 will follow quite easily from a result of Huang, Linial, Naves, Peled and Sudakov [37], and independently in a weaker form by Frankl, Kato, Katona and Tokushige [29]. In order to state this result we need to introduce some notation. Following the first set of authors, for a k-vertex graph H and an n-vertex graph G let $\mathrm{Ind}(H;G)$ denote the collection of induced copies of H in G. The *induced H-density* in G is defined as

$$d(H;G) = \frac{|\mathrm{Ind}(H;G)|}{\binom{n}{k}}.$$

Theorem 4.3.4. *Let $r,s \geq 2$ be integers and suppose that $d\left(\overline{K}_r;G\right) \geq p$ where G is an n-vertex graph and $0 \leq p \leq 1$. Let q be the unique root of $q^r + rq^{r-1}(1-q) = p$ in $[0,1]$. Then $d\left(K_s;G\right) \leq M_{r,s,p} + o(1)$, where*

$$M_{r,s,p} := \max\{(1-p^{1/r})^s + sp^{1/r}(1-p^{1/r})^{s-1}, (1-q)^s\}.$$

After these preparations, Theorem 4.1.7 easily follows.

Proof of Theorem 4.1.7. Let $(\mathscr{A},\mathscr{B})$ be a pair of r-uniform families. By Lemma 4.3.1, we may assume that $(\mathscr{A},\mathscr{B})$ is $\{0,1\}$-cross-intersecting. Thus, the pair $(\mathscr{A},\mathscr{B})$ gives rise to a red-blue coloring of the edges of K_n such that every r-set in $\mathscr{A}$ induces a red copy of K_r, and every r-set in $\mathscr{B}$ induces a blue copy of K_r. We may assume that $|\mathscr{A}| = \alpha^r\binom{n}{r}$ for some $\alpha \in (0,1)$. Then in Theorem 4.3.4 we may take $G = K_n$, $r = s$, and $p = \alpha^r$. It follows that

$$|\mathscr{B}| \le (M_{r,r,\alpha^r} + o(1))\binom{n}{r},$$

and hence

$$|\mathscr{A}||\mathscr{B}| \le (\gamma_r + o(1))\binom{n}{r}^2,$$

where $\gamma_r = \max_{\alpha\in[0,1]}\{\alpha^r M_{r,r,\alpha^r}\} = \max_{\alpha\in[0,1]}\{\alpha^r(1-\alpha)^r + r\alpha^{r+1}(1-\alpha)^{r-1}\}$.

□

Accordingly, from Theorem 4.1.7 we get that $p^*(n,r,2) \le p(n,r,2) \le (\gamma_r + o(1))\binom{n}{r}^2$ as $n \to \infty$, and hence the same is true for $d(n,r,2)$ by Theorem 4.1.5. This bound is asymptotically tight for these problems by considering the pair $(\mathscr{F}_{X,1}, \mathscr{F}_{X^c,0})$ where $|X| = \alpha n$, and $\alpha \in [0,1]$ yields the maximum value of γ_r, as above.

4.4 Final remarks and open problems

We have addressed a variety of problems concerning the maximum number of disjoint pairs in set systems with certain intersection conditions. Many problems remain open. For example, Theorem 4.1.1 shows that the family $\mathscr{F}^*(n,t)$ (see Section 4.1) that maximizes the size of t-avoiding set systems for n sufficiently large also is asymptotically optimal for maximizing the number of disjoint pairs. We conjecture that $\mathscr{F}^*(n,t)$ indeed maximizes the number of disjoint pairs among all t-avoiding set systems, for n sufficiently large.

Conjecture 4.4.1. *For every integer $t \ge 1$ there exists an integer $n_0 = n_0(t)$ such that the following holds. If $n \ge n_0$ and $\mathscr{F} \subset \mathscr{P}[n]$ is t-avoiding, then*

$$d(\mathscr{F}) \le d\left(\mathscr{F}^*(n,t)\right).$$

We also considered the analogue of Theorem 4.1.2 when both set systems are r-uniform, and we introduced three functions $d(n,r,t)$, $p(n,r,t)$, and $p^*(n,r,t)$, each of which turned out to be asymptotically equivalent (see Section 4.3). Further, we made progress in determining $p(n,r,1)$ (and hence also $d(n,r,1)$ and $p^*(n,r,1)$) for n large, and also $p(n,r,2)$, asymptotically. The extremal constructions for all three of these problems turned out to be of the form $\left(\mathscr{F}_{X,a},\mathscr{F}_{X^c,b}\right)$, for suitable $X \subset [n]$ and nonnegative integers a,b, as equal as possible. We conjecture that this phenomenon persists for all $t < r$.

Conjecture 4.4.2. *Let r and t be positive integers with $t < r$. Then there exist nonnegative integers a,b and $X \subset [n]$ such that*

$$p(n,r,t) = (1+o(1))|\mathscr{F}_{X,a}||\mathscr{F}_{X^c,b}|.$$

By Theorem 4.1.5, Conjecture 4.4.2 would imply that $d(n,r,t) = (1+o(1))d\left(\mathscr{F}_{X,a},\mathscr{F}_{X^c,b}\right)$. Note that by Lemma 4.3.1 we may pass from a t-cross-avoiding pair to a $\{0,\ldots,t-1\}$-cross-intersecting pair of set systems when attempting to prove Conjecture 4.4.2. When $t = 2k+1$ is odd we expect that the product $|\mathscr{F}_{X,a}||\mathscr{F}_{X^c,b}|$ is maximized with the most symmetric parameters: $|X| = \lfloor n/2 \rfloor$ and $a = b = k$ (recall that when $t = 1$, the pair $(\mathscr{F}_{X,0},\mathscr{F}_{X^c,0})$ is best possible). Hence, we expect the extremal construction to exhibit some symmetry when t is odd. On the other hand, when t is even we expect the extremal construction to be asymmetric, as evidenced by the optimal configuration in Theorem 4.1.7. Note that in order to deal with this asymmetry, we relied on a result of Huang, Linial, Naves, Peled and Sudakov [37], concerning densities of red and blue cliques in 2-edge-colorings of the complete graph. One way of tackling Conjecture 4.4.2 might be to give a suitable hypergraph generalization of their result. In particular, consider the case $t = 3$. If $\mathscr{A},\mathscr{B}$ are cross-3-avoiding r-uniform families, then we may pass to a cross-$\{0,1,2\}$-intersecting pair (incurring a $o(1)$ error term). Thus, our families of sets give rise to a red-blue coloring of the edges of the complete 3-uniform hypergraph $K_n^{(3)}$ where each r-set of $\mathscr{A}$ is a red copy of $K_r^{(3)}$ and each r-set of $\mathscr{B}$ is a blue copy of $K_r^{(3)}$. Our question, then, reduces to the following one: which 3-uniform hypergraph $\mathscr{H}$ on $[n]$ maximizes the product $d(K_r^{(3)};\mathscr{H})d(\overline{K}_r^{(3)};\mathscr{H})$? The answer to this question should be the following hypergraph (or its complement): split the ground set $[n] = X \cup Y$ as evenly as possible and let $\mathscr{H}$ consist of $X^{(3)}$ together with all triples which intersect Y in exactly 1 point.

Let us close by mentioning a connection to isoperimetric problems. We believe that the pairs $\left(\mathscr{F}_{X,a}, \mathscr{F}_{X^c,b}\right)$ with $a+b \le t-1$ as equal as possible should be optimal for maximizing $p^*(n,r,t)$. For simplicity, let us specialize to the case when $r=3$ and $t=2$ (this case has a pleasant interpretation as the maximum product of monochromatic triangles in a 2-edge-coloring of K_n). Thus, if $(\mathscr{A},\mathscr{B})$ is a pair of 3-uniform $\{0,1\}$-intersecting hypergraphs and n is sufficiently large, is it true that the exact bound $|\mathscr{A}||\mathscr{B}| \le \gamma\binom{n}{3}^2$ holds, where $\gamma = \gamma_3 = \max_{\alpha\in[0,1]}\{\alpha^3(1-\alpha)^3 + 3\alpha^4(1-\alpha)^2\}$? One way of establishing this might be to prove a lower bound on the *lower-upper shadow*. Recall that the *lower shadow* of a set system $\mathscr{F} \subset [n]^{(r)}$, denoted $\partial\mathscr{F}$, is the set $\{A \in [n]^{(r-1)} : A \subset F, \text{ for some } F \in \mathscr{F}\}$. The upper shadow is defined similarly, and denoted $\partial^+\mathscr{F}$.

Question 4.4.3. *Suppose that $\mathscr{A} \subset [n]^{(3)}$ with $|\mathscr{A}| = \binom{x}{3}$ for some real number $x \ge 3$. Is it true that*

$$|\partial^+\partial\mathscr{A}| \ge \binom{x}{3} + \binom{x}{2}(n-x)?$$

Let $(\mathscr{A},\mathscr{B})$ be a pair of $\{0,1\}$-intersecting 3-uniform set systems and write $|\mathscr{A}| = \binom{x}{3}$ for some real $x \ge 3$. If Question 4.4.3 is true, then, since $\mathscr{B} \subset (\partial^+\partial\mathscr{A})^c$, we have that $|\mathscr{B}| \le \binom{n-x}{3} + \binom{n-x}{2}x$, and hence $|\mathscr{A}||\mathscr{B}| \le \binom{x}{3}\left(\binom{n-x}{3} + \binom{n-x}{2}x\right)$. Setting $x = \alpha n$, using the inequality $\binom{\alpha n}{3} \le \alpha^3\binom{n}{3}$, and optimizing yields $|\mathscr{A}||\mathscr{B}| \le \gamma\binom{n}{3}^2$. We finally remark that Question 4.4.3 is related to several stronger (and most likely difficult) conjectures made by Bollobás and Leader [13] concerning minimizing mixed shadows.

BIBLIOGRAPHY

[1] P. Aboulker, N. Cohen, F. Havet, W. Lochet, P.F.S Moura, and S. Thomassé. Subdivisions in digraphs of large out-degree or large dichromatic number. *Preprint, arXiv:1610.00876*.

[2] R. Ahlswede. Simple hypergraphs with maximal number of adjacent pairs of edges. *J. Combinatorial Theory Series B.*, 28:164–167, 1980.

[3] R. Ahlswede and L.H. Khachatrian. The complete intersection theorem for systems of finite sets. *European J. Combin.*, 18:125–136, 1997.

[4] P. Allen, J. Böttcher, S. Griffiths, Y. Kohayakawa, and R. Morris. The chromatic thresholds of graphs. *Adv. Math.*, 235:261–295, 2013.

[5] N. Alon and P. Frankl. The maximum number of disjoint pairs in a family of subsets. *Graphs and Combinatorics*, 1:13–21, 1985.

[6] K. Amin, J. Faudree, R. J. Gould, and E. Sidorowicz. On the non-$(p-1)$-partite K_p-free graphs. *Discuss. Math. Graph Theory*, 33(1):9–23, 2013.

[7] B. Andrásfai. Über ein Extremalproblem der Graphentheorie. *Acta Math. Acad. Sci. Hungar.*, 13:443–455, 1962.

[8] B. Andrásfai, P. Erdős, and V.T. Sós. On the connection between chromatic number, maximal clique and minimal degree of a graph. *Discrete Math.*, 8:205–218, 1974.

[9] B. Andrásfai, P. Erdős, and V. T. Sós. On the connection between chromatic number, maximal clique and minimal degree of a graph. *Discr. Math.*, 8:205–218, 1974.

[10] J. Bang-Jensen. On the 2-linkage problem for semicomplete digraphs. *Ann. Discrete Math.*, 41:23–38, 1989.

[11] B. Bollobás. *Modern Graph Theory*, volume 184 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1998.

[12] B. Bollobás and I. Leader. Set systems with few disjoint pairs. *Combinatorica*, 23:559–570, 2003.

[13] B. Bollobás and I. Leader. Isoperimetric problems for r-sets. *Combinatorics, Probability and Computing*, 13:277–279, 2004.

[14] B. Bollobás and A. Thomason. Highly linked graphs. *Combinatorica*, 16(3):313–320, 1996.

[15] S. Brandt and E. Ribe-Baumann. European Conference on Combinatorics, Graph Theory and Applications (EuroComb 2009) Graphs of odd girth 7 with large degree. *Electron. Notes Discr. Math.*, 34:89–93, 2009.

[16] S. Brandt and S. Thomassé. Dense triangle-free graphs are four colorable: a solution to the Erdős-Simonovits problem. *To appear*.

[17] A.E. Brouwer. Some lotto numbers from an extension of Turán's theorem. *Afdeling Zuivere Wiskunde [Department of Pure Mathematics]*, 152, 1981.

[18] C. C. Chen, G. P. Jin, and K. M. Koh. Triangle-free graphs with large degree. *Combinatorics, Probability and Computing*, 6(4):381–396, 1997.

[19] S. Das, W. Gan, and B. Sudakov. The minimum number of disjoint pairs in set systems and related problems. *Combinatorica*, 36:623–660, 2016.

[20] O. Ebsen and M. Schacht. Homomorphism thresholds for odd cycles. *Preprint, arXiv:1712.07026*.

[21] O. Ebsen and M. Schacht. Personal communication.

[22] P. Erdős. Some recent results on extremal problems in graph theory (Results). In P. Rosenstiehl, editor, *Theory of Graphs*, pages 118–123. Gordon and Breach, New York, Dunod, Paris, 1967.

[23] P. Erdős. On some new inequalities concerning extremal properties of graphs. In P. Erdős and G. Katona, editors, *Theory of Graphs*, pages 77–81. Academic Press, New York, 1968.

[24] P. Erdős, C. Ko, and R. Rado. Intersection theorems for systems of finite sets. *Quart. J. Math. Oxford Ser.*, 12:313–320, 1961.

[25] P. Erdős and M. Simonovits. On a valence problem in extremal graph theory. *Discrete Math.*, 5:323–334, 1973.

[26] P. Frankl. On the minimum number of disjoint pairs in a family of finite sets,. *J. Combinatorial Theory Series A*, 22:249–251, 1977.

[27] P. Frankl and Z. Füredi. On hypergraphs without two edges intersecting in a given number of vertices. *J. Combinatorial Theory Series A*, 36:230–236, 1984.

[28] P. Frankl and Z. Füredi. Forbidding just one intersection. *J. Combinatorial Theory Series A*, 39:160–176, 1985.

[29] P. Frankl, M. Kato, G.O.H. Katona, and N. Tokushige. Two-colorings with many monochromatic cliques in both colors. *J. Combinatorial Theory Series B*, 103:415–427, 2016.

[30] P. Frankl and V. Rödl. Forbidden intersections. *Transactions of the American Mathematical Society*, 300(1):259–286, 1987.

[31] P. Frankl and N. Tokushige. Invitation to intersection problems for finite sets. *J. Combinatorial Theory Series A*, 144:157–211, 2016.

[32] W. Goddard and J. Lyle. Dense graphs with small clique number. *J. Graph Theory*, 66(4):319–331, 2011.

[33] R. Häggkvist. Odd cycles of specified length in nonbipartite graphs. In *Graph theory (Cambridge, 1981)*, volume 62 of *North-Holland Math. Stud.*, pages 89–99. North-Holland, Amsterdam-New York, 1982.

[34] R. Häggkvist and G. Jin. Graphs with odd girth at least seven and high minimum degree. *Graphs Combin.*, 14(4):351–362, 1998.

[35] D. Hanson and B. Toft. k-saturated graphs of chromatic number at least k. *Ars. Combin.*, 31:159–164, 1991.

[36] F. Havet and B. Lidický. Splitting a tournament into two subtournaments with given minimum outdegree. *Preprint*, 2014.

[37] H. Huang, N. Linial, H. Naves, Y. Peled, and B. Sudakov. On the densities of cliques and independent sets in graphs. *Combinatorica*, 36:493–512, 2016.

[38] G. Jin. Triangle-free graphs with high minimal degrees. *Combinatorics, Probability and Computing*, 2(4):479–490, 1993.

[39] G. Jin. Triangle-free four-chromatic graphs. *Discrete Math.*, 145(1-3):151–170, 1995.

[40] H.A. Jung. Verallgemeinerung des n-fachen Zusammenhangs für Graphen. *Math. Ann.*, 187:95–103, 1970.

[41] V. Kaibel and S. Weltge. A short proof that the extension complexity of the correlation polytope grows exponentially. *Discrete and Computational Geometry*, 53:397–401, 2015.

[42] M. Kang and O. Pikhurko. Maximum K_{r+1}-free graphs which are not r-partite. *Mat. Stud.*, 24(1):12–20, 2005.

[43] G. O. H. Katona. Intersection theorems for systems of finite sets. *Acta Math. Acad. Sci. Hungar.*, 15:329–337, 1964.

[44] P. Keevash and B. Sudakov. On a restricted cross-intersection problem. *J. Combinatorial Theory Series A*, 113:1536–1542, 2006.

[45] D. Kühn, J. Lapinskas, D. Osthus, and V. Patel. Proof of a conjecture of Thomassen on hamilton cycles in highly connected tournaments. *Proc. Lond. Math. Soc.*, 109:733–762, 2014.

[46] D.G. Larman and P. Mani. On the existence of certain configurations within graphs and the 1-skeletons of polytopes. *Proc. Lond. Math. Soc.*, 20:144–160, 1974.

[47] N. Lichiardopol. Vertex-disjoint subtournaments of prescribed minimum outdegree or minimum semidegree: proof for tournaments of a conjecture of Stiebitz. *Int. J. Comb.*, 2012:9 pp., 2012.

[48] T. Łuczak. On the structure of triangle-free graphs of large minimum degree. *Combinatorica*, 26(4):489–493, 2006.

[49] T. Łuczak and S. Thomassé. Coloring dense graphs via VC-dimension. *Submitted, arXiv:1007.1670*.

[50] J. Lyle. On the chromatic number of H-free graphs of large minimum degree. *Graphs and Combinatorics*, 27(5):741–754, 2011.

[51] W. Mader. Homomorphieeigenschaften und mittlere Kantendichte von Graphen. *Math. Annalen*, 174:265–268, 1967.

[52] W. Mader. Degree and local connectivity in digraphs. *Combinatorica*, 5(2):161–165, 1985.

[53] S. Messuti and M. Schacht. On the structure of graphs with given odd girth and large minimum degree. *J. Graph Theory*, 80(1):69–81, 2015.

[54] V. Nikiforov. Chromatic number and minimum degree of K_r-free graphs. *Preprint, arXiv:1001.2070*.

[55] V. Nikiforov. Some new results in extremal graph theory. In *Surveys in combinatorics 2011, London Math. Soc. Lecture Note Ser., 392*, pages 141–181. Cambridge Univ. Press, Cambridge, 2011.

[56] V. Nikiforov and C. C. Rousseau. Large generalized books are p-good. *J. Combinatorial Theory Ser. B*, 92:85–97, 2004.

[57] H. Oberkampf and M. Schacht. On the structure of dense graphs with fixed clique number. *Preprint, arXiv:1602.02302*.

[58] A. Pokrovskiy. Highly linked tournaments. *J. Combinatorial Theory Series B*, 115:339–347, 2015.

[59] J. Sgall. Bounds on pairs of families with restricted intersections. *Combinatorica*, 19:555–566, 1999.

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

[61] R. Thomas and P. Wollan. An improved extremal function for graph linkages. *European J. Combin.*, 26:309–324, 2005.

[62] C. Thomassen. Connectivity in tournaments. In *Graph Theory and Combinatorics, a Volume in Honour of Paul Erdős*, pages 305–313. Academic Press, London, 1984.

[63] C. Thomassen. Even cycles in directed graphs. *Europ. J. Combin.*, 6:85–89, 1985.

[64] C. Thomassen. Note on highly connected non-2-linked digraphs. *Combinatorica*, 11:393–395, 1991.

[65] C. Thomassen. On the chromatic number of triangle-free graphs of large minimum degree. *Combinatorica*, 22(4):591–596, 2002.

[66] C. Thomassen. On the chromatic number of pentagon-free graphs of large minimum degree. *Combinatorica*, 27(2):241–243, 2007.

[67] P. Turán. On an extremal problem in graph theory(in Hungarian). *Math. Fiz. Lapok*, 48:436–452, 1941.

[68] M. Tyomkyn and A. J. Uzzell. Strong Turán stability. *Electronic Journal of Combinatorics*, 22(3), 2015.

[69] Mykhaylo Tyomkyn and Andrew J. Uzzell. Strong turán stability. *Electronic Notes in Discrete Mathematics*, 49:433–440, 2015.

[70] A. A. Zykov. On some properties of linear complexes. *Mat. Sbornik N.S.*, 24(66):163–188, 1949.