4-21-2015

Problems in Extremal and Ramsey Graph Theory

Teeradej Kittipassorn

PROBLEMS IN EXTREMAL AND RAMSEY GRAPH THEORY

by

Teeradej Kittipassorn

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

Major: Mathematical Sciences

The University of Memphis

May 2015

To my parents

ACKNOWLEDGEMENTS

This dissertation cannot exist without the guidance and support of a large number of people. Firstly, I would like to thank Professor Béla Bollobás for this incredible opportunity of being his student. Over the years, I have learnt a lot from him. He has taught me how to tackle problems, how to present results and how to be a better mathematician. Thank you so much for giving me a ton of interesting problems, taking me to many great conferences, introducing me to brilliant mathematicians and showing me beautiful cities. I could not have asked for a better supervisor anywhere in the world.

On this note, I would also like to thank Mrs. Gabriella Bollobás for all her kind-hearted care and love. She treats me like family and is always there for me. All her life advices have taught me how to be a better person. Many thanks for all the delicious food. I have never left without a full stomach.

I thank my examiners: Professors Paul Balister, Anna Kaminska, Imre Leader and Vlado Nikiforov for the comments and suggestions.

During my time here, I am lucky enough to be taught by many incredible combinatorialists whom I would like to thank, including Professors Paul Balister, Vlado Nikiforov, Miklós Ruszinkó, Ervin Győri and Imre Leader.

Thanks to Ms. Tricia Simmons for her guidance with all my problems. She has helped me tremendously through thick and thin.

I would also like to thank my collaborators: Victor Falgas-Ravry, Dániel Korándi, Shoham Letzter, Gábor Mészáros, Bhargav Narayanan and Professor Alexander Scott, for many hours of interesting and insightful discussions.

Thanks to colleagues and friends: Andrew, Andrea, Neal, Dominik, Richard, Tomas, Ago, Stephen, Joon, Merve, Ali, Scott, Jee, McFeely, Karen, Michal, Tom, Nathan, Sebastian, Paul, Bhargav, Shoham, Jonathan, Yury, Julian, Richard, Istvan, Buddhika, Kamil, B, Stefan and Vytautas, for the friendship, support and wonderful memories.

None of this would have been possible without the support of my family, especially my parents and my wife, Put. Thank you for your unconditional love and encouragement throughout my life. You made me the person I am today.

ABSTRACT

Kittipassorn, Teeradej Ph.D. The University of Memphis. May 2015. Problems in Extremal and Ramsey Graph Theory. Major Professor: Béla Bollobás, Ph.D., D.Sc., F.R.S.

This dissertation consists of four chapters dealing with unrelated topics in extremal and Ramsey graph theory.

Chapter 1 is dedicated to the results in the papers with Bhargav P. Narayanan [42, 41]. Given an edge coloring of the complete graph on $\mathbb{N}$, we say that a subset of $\mathbb{N}$ is *exactly m-colored* if exactly m colors appear inside the subset. We answer many of the questions about finding exactly m-colored subgraphs.

In Chapter 2, we present joint work with Béla Bollobás, Bhargav P. Narayanan and Alexander D. Scott [10]. Let us say that a graph is *splittable* if the vertices can be partitioned into two equal halves such that each half induces the same number of edges. The main result is that any graph of order n can be made splittable by deleting at most $o(n)$ vertices. This answers a question of Caro and Yuster.

Chapter 3 is based on joint work with Victor Falgas-Ravry, Dániel Korándi, Shoham Letzter and Bhargav P. Narayanan [27]. A set is *separated* by a collection of its subsets if any two elements of the set can be distinguished using some subset in the collection. We consider a question of separating the edge set of a graph using only paths. We conjecture that every graph of order n admits a separating path system of size linear in n and prove this in certain interesting special cases including random graphs and graphs with linear minimum degree.

Chapter 4 presents results from a joint paper with Gábor Mészáros [40]. A triple of vertices in a graph is a *frustrated triangle* if it induces an odd number of edges. We study the set $F_n \subset [0, \binom{n}{3}]$ of possible number of frustrated triangles $f(G)$ in a graph G on n vertices. Our main result is that F_n contains two interlacing sequences $0 = a_0 \leq b_0 \leq a_1 \leq b_1 \leq \cdots \leq a_m \leq b_m \sim n^{3/2}$ such that $F_n \cap (b_t, a_{t+1}) = \varnothing$ for all t, where the gaps are $|b_t - a_{t+1}| = (n-2) - t(t+1)$ and $|a_t - b_t| = t(t-1)$.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

EXACTLY m-COLORED GRAPHS

1.1 Introduction

A classical result of Ramsey [56] says that when the edges of a complete graph on a countably infinite vertex set are finitely colored, one can always find a complete infinite subgraph all of whose edges have the same color.

Ramsey's theorem has since been generalized in many ways; most of these generalizations are concerned with finding monochromatic substructures in various colored structures. For a survey of many of these generalizations, see the book of Graham, Rothschild and Spencer [33]. Ramsey theory has witnessed many developments over the last fifty years and continues to be an area of active research today; see, for example, [34, 38, 44, 45].

While one is usually concerned with finding monochromatic substructures in various finitely colored structures, two alternative directions are as follows. First, one could study colorings that use infinitely many colors, as was first done by Erdős and Rado [21] and by many others after them. Second, one could look for structures which are colored with exactly m colors for some $m \geq 2$. This was first considered by Erickson [26] and then investigated further by Stacey and Weidl [60]. In this chapter, we shall consider the question of finding structures colored with exactly m colors in colorings that use infinitely many colors.

Our notation is standard. For a set X, denote by $X^{(2)}$ the set of all unordered pairs of elements of X; equivalently, $X^{(2)}$ is the complete graph on the vertex set X. As usual, we write $[n]$ to denote $\{1,\ldots,n\}$, the set of the first n natural numbers. We denote a surjective map f from a set X to another set Y by $f : X \twoheadrightarrow Y$. By a *coloring* of a graph, we mean a coloring of the edges of the graph unless we specify otherwise.

Let $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow \mathcal{C}$ be a surjective coloring of the edges of the complete graph on $\mathbb{N}$ with an arbitrary set of colors $\mathcal{C}$. If the set of colors $\mathcal{C}$ is infinite, we say that Δ is an *infinite-coloring* and if $\mathcal{C}$ is finite, we say that Δ is a *k-coloring* if $|\mathcal{C}| = k$. We say that a subset X of $\mathbb{N}$ is (*exactly*) *m-colored* if $\Delta(X^{(2)})$, the set of values attained by Δ on the edges with both endpoints in X, has size exactly m. We write $\gamma_\Delta(X)$, or $\gamma(X)$ in short, for the size of the set $\Delta(X^{(2)})$; in other words, every set X is $\gamma(X)$-colored.

For a coloring $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow \mathcal{C}$ of the complete graph on $\mathbb{N}$ with an arbitrary set of colors, we define the set

$$\mathcal{G}_\Delta = \{\gamma(X) : X \subset \mathbb{N}\}.$$

In 1999, Stacey and Weidl [60] considered the following question: which natural numbers m are guaranteed to be elements of $\mathcal{G}_\Delta$ for every infinite-coloring Δ? By considering a rainbow coloring Δ of $\mathbb{N}$, we see that unless $m = \binom{n}{2}$ for some $n \geq 2$, m is not guaranteed to be a member of $\mathcal{G}_\Delta$. In the other direction, since an edge is a 1-colored complete graph, $\binom{2}{2} = 1$ is always an element of $\mathcal{G}_\Delta$. They were able to show that $\binom{3}{2} = 3$ is also always an element of $\mathcal{G}_\Delta$ for every infinite-coloring Δ. But for $n \geq 4$, they were unable to decide whether or not there exists an infinite-coloring Δ such that $\binom{n}{2} \notin \mathcal{G}_\Delta$. In particular, they asked if all natural numbers of the form $\binom{n}{2}$ must be contained in $\mathcal{G}_\Delta$ for every infinite-coloring Δ. We give an affirmative answer.

Theorem 1.1.1. *For every infinite-coloring* $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow \mathbb{N}$, *and for every natural number* $n \geq 2$, $\binom{n}{2} \in \mathcal{G}_\Delta$.

Let us now turn to the case of finite coloring. Stacey and Weidl [60] also raised the following question: do there exist natural numbers $m \in \mathbb{N}$ with the property that for all sufficiently large $k \in \mathbb{N}$, $m \in \mathcal{G}_\Delta$ for every k-coloring $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow [k]$? Observe that any such natural number m, assuming one exists, must be of the form $\binom{n}{2}$ or $\binom{n}{2} + 1$ for some natural number $n \geq 2$. One can see this by considering the family of 'small-rainbow colorings' of the complete graph on $\mathbb{N}$ which color all the edges of some finite complete subgraph with distinct colors and all the remaining edges with a single color not used in the finite (rainbow colored) complete subgraph. On the other hand, when m is of the form $\binom{n}{2}$ or $\binom{n}{2}+1$ for some natural number $n \geq 2$, we have the following positive result.

Theorem 1.1.2. *For all $n \in \mathbb{N}$, there exists a natural number $C = C(n)$ such that for any k-coloring $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow [k]$ with $k \geq C$, both $\binom{n}{2}, \binom{n}{2}+1 \in \mathcal{G}_\Delta$.*

We prove Theorem 1.1.1 in Section 1.2 by answering a more general question: when is $\mathcal{G}_\Delta \neq \mathbb{N}$? It turns out that the techniques used to prove Theorem 1.1.1 also allow us to prove a finitary version of the same theorem. In Section 1.3, we present this finitary result and use it prove Theorem 1.1.2 in a slightly stronger form.

In the context of Ramsey theory, one is usually interested in finding 'large' homogeneous structures with certain properties. With this in mind, for a coloring $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow \mathcal{C}$, we define

$$\mathcal{F}_\Delta = \{\gamma(X) : X \subset \mathbb{N} \text{ such that } X \text{ is infinite}\}.$$

When Δ is an infinite-coloring, it might so happen that for each infinite subset X of $\mathbb{N}$, the set $\Delta(X^{(2)})$ is infinite; consequently, it is only really meaningful to study the set $\mathcal{F}_\Delta$ in the case of colorings using finitely many colors. The question of finding m-colored complete infinite subgraphs, was first considered by Erickson [26]. If $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow [k]$ is a k-coloring of the edges of the complete graph on the natural numbers, then clearly $k \in \mathcal{F}_\Delta$ as Δ is surjective, and Ramsey's theorem tells us that $1 \in \mathcal{F}_\Delta$. Erickson [26] noted that a fairly straightforward application of Ramsey's theorem enables one to show that $2 \in \mathcal{F}_\Delta$. Erickson also conjectured that, with the exception of $1, 2$ and k, no other elemets are guaranteed to be in $\mathcal{F}_\Delta$.

Conjecture 1.1.3. *If $k > m > 2$, then there is a k-coloring $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow [k]$ such that $m \notin \mathcal{F}_\Delta$.*

Stacey and Weidl [60] used a probabilistic construction to partially resolve this conjecture, showing for every $m > 2$ that there is a k-coloring Δ such that $m \notin \mathcal{F}_\Delta$ provided k is sufficiently larger than m.

Since an exactly m-colored complete infinite subgraph is not guaranteed to exist, we are naturally led to the question of whether we can find a complete infinite subgraph that is exactly $\hat{m}$-colored for some $\hat{m}$ close to m. We establish the following result which is the best possible up to an additive constant.

Theorem 1.1.4. *For any k-coloring $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow [k]$ and any natural number $m \leq k$, there exists an $\hat{m} \in \mathcal{F}_\Delta$ such that $|m - \hat{m}| \leq \sqrt{m/2} + 1/2$.*

We prove Theorem 1.1.4 and some related results as well as generalizations of these results to uniform hypergraphs in Section 1.4. We then conclude the chapter in Section 1.5 by mentioning some open problems.

1.2 A canonical Ramsey theorem

Our main aim in this section is to establish a canonical Ramsey theory for m-colored graphs which implies Theorem 1.1.1. Canonical Ramsey theory, which originates in a classical paper of Erdős and Rado [21], provides results about colorings which use an arbitrary set of colors. We will need a basic canonical Ramsey theorem proved by Erdős and Rado. To state this result, it will be convenient to introduce some notation. We say that $X \subset \mathbb{N}$ is *rainbow colored* if no two edges with both endpoints in X receive the same color. Also, we say that $X \subset \mathbb{N}$ is *left colored* if for $i, j, k, l \in X$ with $i < j$ and $k < l$, $\Delta(ij) = \Delta(kl)$ if and only if $i = k$, and the definition of *right colored* is analogous; if X is left or right colored, we say, in short, that X is *lexically colored.* With these definitions in place, we can now state the canonical Ramsey theorem of Erdős and Rado [21].

Theorem 1.2.1. *For any coloring $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow \mathcal{C}$, there exists an infinite subset X of $\mathbb{N}$ such that either*

- *X is 1-colored, or*

- *X is rainbow colored, or*

- *X is lexically colored.* $\square$

Here, we shall consider a more general question: when is $\mathcal{G}_\Delta \neq \mathbb{N}$? As remarked above, for an injective coloring Δ, $\mathcal{G}_\Delta = \{\binom{n}{2} : n \geq 2\} \neq \mathbb{N}$. There is another infinite-coloring Δ for which $\mathcal{G}_\Delta \neq \mathbb{N}$ which is slightly less obvious. Given $X \subset \mathbb{N}$, if there is a vertex $v \in X$ such that $X \setminus \{v\}$ is 1-colored and all the edges between v and $X \setminus \{v\}$ have distinct colors (which are also all different from the color of $X \setminus \{v\}$), then we say that X is *star colored* (*with centre* v). It is easy to check (see Figure 1.1) that if $\mathbb{N}$ is star colored by Δ, then $\mathcal{G}_\Delta = \mathbb{N} \setminus \{2\}$.

Our main result, stated below, is that the two colorings described above are, in a sense, the 'canonical' colorings for which $\mathcal{G}_\Delta \neq \mathbb{N}$.

Theorem 1.2.2. *For every infinite-coloring* $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow \mathbb{N}$*, either*

- $\mathcal{G}_\Delta = \mathbb{N}$*, or*

- *there exists an infinite rainbow colored subset of* $\mathbb{N}$*, or*

- *there exists an infinite star colored subset of* $\mathbb{N}$*.*

Theorem 1.1.1 is an immediate consequence of Theorem 1.2.2. We do not prove Theorem 1.2.2 as stated. Instead, it will be more convenient to prove a stronger result which we shall state (and prove) in the next subsection.

1.2.1 Proof of the main theorem

To prove Theorem 1.2.2, it will be more convenient to work with general infinite graphs. By an *infinite graph*, we mean a graph whose vertex set is $\mathbb{N}$ and which has infinitely many edges.

It will be helpful to establish a few notational conveniences. Given an infinite graph G and an infinite-coloring $\Delta : G \twoheadrightarrow \mathbb{N}$ of the edges of G, for a subset X of $\mathbb{N}$, we shall write

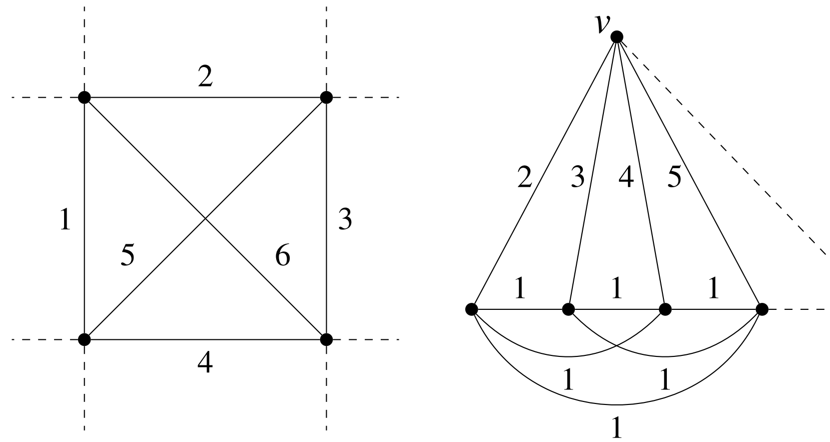

Figure 1.1: A rainbow coloring and a star coloring with centre v.

$\gamma_G(X)$, or $\gamma(X)$ in short, for the number of distinct colors attained by Δ on $G[X]$, the subgraph of G induced by X; if H is a subgraph of G, we write $\gamma_H(\mathrm{X})$ for the the number of distinct colors attained by Δ on $H[X]$. For disjoint subsets X and Y, write $\gamma(X,Y)$ for the number of distinct colors in the induced bipartite subgraph between X and Y in G. Also, for a vertex $v \in \mathbb{N}$, we shall write $\gamma(v)$ for $\gamma(\{v\}, \mathbb{N} \setminus \{v\})$, the number of distinct colors of the edges incident to v in G.

We define the set $\mathcal{G}_\Delta$ for an infinite-coloring $\Delta : G \twoheadrightarrow \mathbb{N}$ of an infinite graph G in the obvious way by setting

$$\mathcal{G}_\Delta = \{\gamma_G(X) : X \subset \mathbb{N}\}.$$

The following result easily implies Theorem 1.2.2.

Theorem 1.2.3. *For every infinite-coloring $\Delta : G \twoheadrightarrow \mathbb{N}$ of an infinite graph G, either*

- $\mathcal{G}_\Delta = \mathbb{N}$, *or*
- *there exists an infinite rainbow colored subset of* $\mathbb{N}$*; or*
- *there exists an infinite star colored subset of* $\mathbb{N}$.

Let us clarify what we mean by a rainbow colored set and a star colored set in a general graph G. We say that X is *rainbow colored in G* if $G[X]$ is a complete subgraph of

G which is rainbow colored. We say that X is *star colored* (*with centre* v) *in* G if there is a vertex $v \in X$ such that $G[X \setminus \{v\}]$ is either an independent set or a 1-colored complete graph, and all the edges between v and $X \setminus \{v\}$ have distinct colors (which are also all different from the color of $G[X \setminus \{v\}]$ in the case where $X \setminus \{v\}$ does not induce an independent set).

For any finite set of colors $\mathcal{S}$, note that if we delete all the edges of an infinite graph G which are colored with a color from $\mathcal{S}$ by an infinite-coloring Δ of the edges of G, the resulting graph H is infinite and the restriction of Δ to H is an infinite-coloring. This makes the statement of Theorem 1.2.3 more amenable to induction than that of Theorem 1.2.2 and motivates the stronger statement of Theorem 1.2.3.

Fix an infinite-coloring $\Delta : G \twoheadrightarrow \mathbb{N}$ of an infinite graph G and note that if we have a partition $X = X_1 \cup X_2 \cup \ldots X_n$ of a subset X of $\mathbb{N}$, then

$$\sum_{1 \le i \le n} \gamma(X_i) + \sum_{1 \le i < j \le n} \gamma(X_i, X_j) \ge \gamma(X).$$

Consequently, if $\gamma(X) = \infty$, then at least one of the terms on the left is infinite; we shall make use of this fact repeatedly.

Next, we state a technical lemma about 'almost bipartite colorings' which will be useful in proving Theorem 1.2.3.

Lemma 1.2.4. *Let G be an infinite graph and suppose that an infinite-coloring $\Delta : G \twoheadrightarrow \mathbb{N}$ of G is such that*

- $\gamma(v) < \infty$ *for all* $v \in \mathbb{N}$, *and*
- *there is a partition of* $\mathbb{N} = A \cup B$ *such that* $\gamma(A) < \infty$, $\gamma(B) < \infty$ *and* $\gamma(A, B) = \infty$.

Then for every natural number m, there exists a subset X of $\mathbb{N}$ such that $X \cap A \neq \varnothing$, $X \cap B \neq \varnothing$ and $\gamma(X) = m$.

Our strategy for proving both Theorem 1.2.3 and Lemma 1.2.4 is to inductively construct a set X for which $\gamma_G(X) = m$. To do this, we shall first delete some edges from G to get a new infinite graph H so that the restriction of Δ to H is also an infinite-coloring. We then inductively find a set Y with $\gamma_H(Y) = l$ for a suitably chosen $l < m$. Finally, we use the deleted edges in conjunction with Y to obtain X.

We first prove Lemma 1.2.4 and then show how to deduce Theorem 1.2.3 from it.

Proof of Lemma 1.2.4. Before we begin, let us note some consequences of our assumptions about the coloring Δ. Since $\gamma(v) < \infty$ for all $v \in \mathbb{N}$ and $\gamma(A,B) = \infty$, both A and B must be infinite. Furthermore, observe that if $\gamma(U) = \infty$ for some $U \subset \mathbb{N}$, then since $\gamma(A) < \infty$ and $\gamma(B) < \infty$, both $U \cap A$ and $U \cap B$ must be infinite.

We proceed by induction on m. The result is trivial for $m = 1$. Assuming the result for all $l < m$, we shall prove the result for m.

Pick an edge uv such that $u \in A$ and $v \in B$ and say that the color of the edge is c. We know that $\gamma(u) < \infty$. We may assume, relabeling colors if necessary, that the colors of the edges incident to u are $1, \dots, \gamma(u)$. Consider the partition

$$\mathbb{N} \setminus \{u\} = U_0 \cup U_1 \cup \cdots \cup U_{\gamma(u)}$$

where U_0 is the set of vertices not adjacent to u in G and for $1 \le i \le \gamma(u)$, U_i is the set of all vertices that are joined to u by an edge of color i. We distinguish the following three cases.

Case 1: $\gamma(U_i) = \infty$ for some $i \neq 0$. We begin by observing (see Figure 1.2) that

$$\gamma(U_i \cap A) + \gamma(U_i \cap B) + \gamma(U_i \cap A, U_i \cap B) \ge \gamma(U_i).$$

Since $\gamma(U_i \cap A) \le \gamma(A) < \infty$ and $\gamma(U_i \cap B) \le \gamma(B) < \infty$, we conclude that $\gamma(U_i \cap A, U_i \cap B) = \infty$.

Let H be the infinite subgraph of $G[U_i]$ obtained by deleting all the edges of $G[U_i]$ of color i. Then there exists, by the induction hypothesis, a subset Y of U_i such that $Y \cap (U_i \cap A) \neq \varnothing$, $Y \cap (U_i \cap B) \neq \varnothing$ and $\gamma_H(Y) = m - 1$. Observe that all the edges

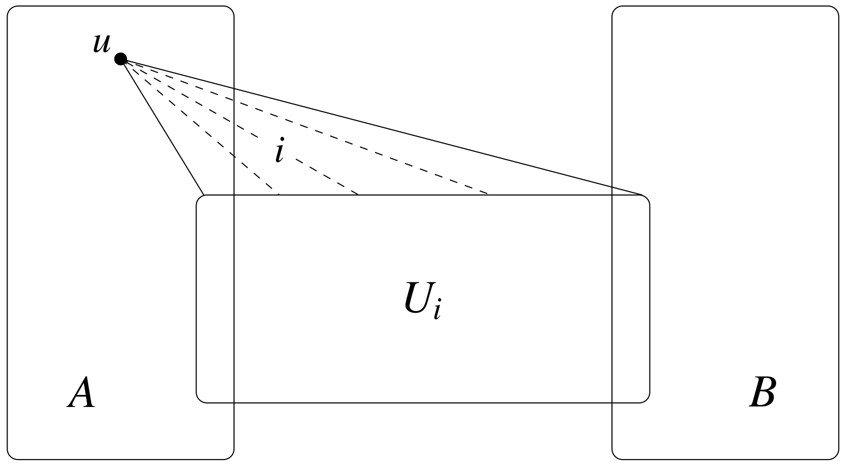

Figure 1.2: Case 1.

between u and $Y \subset U_i$ are colored i in G. Since the color i is not counted by γ_H, we see that $\gamma_G(Y \cup \{u\}) = m$. Therefore $X = Y \cup \{u\}$ is the required subset since $X \cap A \neq \emptyset$ and $X \cap B \neq \emptyset$.

Case 2: $\gamma(U_i, U_j) = \infty$ **for some** $0 < i < j$**.** Let us assume that $\gamma(U_i) < \infty$ and $\gamma(U_j) < \infty$. Observe (see Figure 1.3) that $\gamma(U_i \cap A, U_j \cap A) \leq \gamma(A) < \infty$ and $\gamma(U_i \cap B, U_j \cap B) \leq \gamma(B) < \infty$. So we must either have $\gamma(U_i \cap A, U_j \cap B) = \infty$ or $\gamma(U_i \cap B, U_j \cap A) = \infty$. Without loss of generality, assume that $\gamma(U_i \cap A, U_j \cap B) = \infty$.

If $m \geq 3$, we may assume that the result holds for $m-2$. Let H be the infinite subgraph of $G[(U_i \cap A) \cup (U_j \cap B)]$ obtained by deleting edges of color i and j from $G[(U_i \cap A) \cup (U_j \cap B)]$. Then there exists, by the induction hypothesis, a subset Y of $(U_i \cap A) \cup (U_j \cap B)$ such that $Y \cap (U_i \cap A) \neq \emptyset$, $Y \cap (U_j \cap B) \neq \emptyset$ and $\gamma_H(Y) = m-2$. Since $Y \subset U_i \cup U_j$, all the edges between u and Y in G are colored either i or j and as $Y \cap U_i \neq \emptyset$ and $Y \cap U_j \neq \emptyset$, edges of both colors are present. Since both colors i and j are not counted by γ_H, it follows that $\gamma_G(Y \cup \{u\}) = m$. Clearly, $Y \cap A \neq \emptyset$ and $Y \cap B \neq \emptyset$ and therefore $X = Y \cup \{u\}$ is the required subset.

Now suppose that $m = 2$. Since $\gamma(w) < \infty$ for all $w \in \mathbb{N}$, we can find an infinite matching $M = \{a_1b_1, a_2b_2, \ldots\}$ between $U_i \cap A$ and $U_j \cap B$ in G such that each edge of the matching has a distinct color. If a_k and b_l are not adjacent in G for some $k, l \in \mathbb{N}$, then

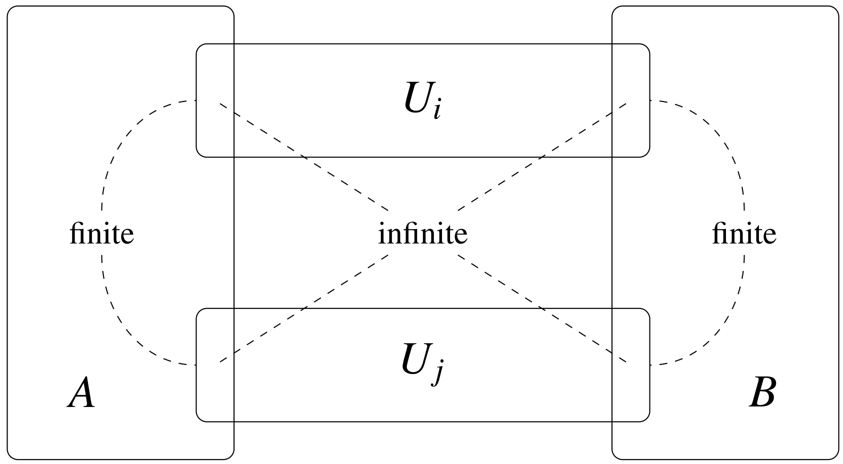

Figure 1.3: Case 2.

$X = \{u, a_k, b_l\}$ is immediately seen to be 2-colored. So we may suppose that for each $k, l \in N$, a_k is adjacent to b_l in G.

Since $\gamma(\{a_1, a_2, \dots\}) < \infty$, it follows from Ramsey's theorem that there exists a subset $\{a'_1, a'_2, \dots\}$ of $\{a_1, a_2, \dots\}$ which either induces an independent set or a 1-colored complete graph. Let a'_k be matched to the vertex b'_k in M and let c_k denote the color of the edge $a'_k b'_k$.

If $\{a'_1, a'_2, \dots\}$ is an independent set in G, then since $\gamma(a'_1) < \infty$, there exist $s, t \in \mathbb{N}$ such that $a'_1 b'_s$ and $a'_1 b'_t$ have the same color, say d. By our choice of M, $c_s \neq c_t$. Hence, at least one of c_s or c_t, say c_s, is not equal to d. Then it is easy to check that $X = \{a'_1, a'_s, b'_s\}$ is the required subset.

If $\{a'_1, a'_2, \dots\}$ induces a complete graph of color d in G, we may assume (by discarding the edge $a'_1 b'_1$ and relabeling the remaining vertices if necessary) that c_1, the color of the edge $a'_1 b'_1$, is not equal to d. Since $\gamma(b'_1) < \infty$, there exist $s, t \in \mathbb{N}$ such that $\Delta(a'_s b'_1) = \Delta(a'_t b'_1)$. If $\Delta(a'_s b'_1) = d$, then we may take $X = \{a'_1, a'_s, b'_1\}$. On the other hand, if $\Delta(a'_s b'_1) \neq d$, then $X = \{a'_s, a'_t, b'_1\}$ is the required subset.

Case 3: $\gamma(U_0, U_i) = \infty$ **for some** $i \neq 0$**.** We argue as we did in Case 2. We may assume that $\gamma(U_0 \cap A, U_i \cap B) = \infty$. Let H be the infinite subgraph of $G[(U_0 \cap A) \cup (U_i \cap B)]$ obtained by deleting all the edges of color i from $G[(U_0 \cap A) \cup (U_i \cap B)]$.

By the induction hypothesis, there exists a subset Y of $(U_0 \cap A) \cup (U_i \cap B)$ such that $Y \cap (U_0 \cap A) \neq \varnothing$, $Y \cap (U_i \cap B) \neq \varnothing$ and $\gamma_H(Y) = m-1$. As before, every edge between u and Y is colored i in G (and u is adjacent to at least one vertex of Y since $Y \cap (U_i \cap B) \neq \varnothing$). Since the color i is not counted by γ_H, it follows that $\gamma_H(Y \cup \{u\}) = m$. Hence, $X = Y \cup \{u\}$ is the required subset.

Case 4: $\gamma(U_0) = \infty$**.** Recall that we chose $u \in A$ and $v \in B$ such that the edge uv has color c. Let V_0 be the set of those vertices of U_0 not adjacent to v in G. Since $\gamma(v) < \infty$, we have a partition of $U_0 \setminus V_0 = V_1 \cup \cdots \cup V_n$, with $n \leq \gamma(v)$, based on the color of the edge joining a given vertex of $U_0 \setminus V_0$ to the vertex v. Applying the same argument as in Cases 1, 2 and 3 to the vertex v, we see that we are done unless $\gamma(V_0) = \infty$.

In this case, we consider the partition $V_0 = (V_0 \cap A) \cup (V_0 \cap B)$. Note that $\gamma(V_0 \cap A) < \infty$, $\gamma(V_0 \cap B) < \infty$ and $\gamma(V_0 \cap A, V_0 \cap B) = \infty$. Let H be the infinite subgraph of $G[V_0]$ obtained by deleting edges of color c from $G[V_0]$.

By the induction hypothesis, there is a subset Y of V_0 such that $\gamma_H(Y) = m-1$. Observe that uv has color c and furthermore, u and v are not adjacent to any of the vertices of Y. Since the color c is not counted by γ_H, we see that $\gamma_G(Y \cup \{u,v\}) = m$. Therefore $X = Y \cup \{u,v\}$ is the required subset since clearly, $X \cap A \neq \varnothing$ and $X \cap B \neq \varnothing$. This completes the proof. □

We are now in a position to deduce Theorem 1.2.3 from Lemma 1.2.4.

Proof of Theorem 1.2.3. Let $\Delta : G \twoheadrightarrow \mathbb{N}$ be an infinite-coloring of an infinite graph G. We shall prove by induction on m that if G contains no infinite rainbow colored or star colored subset, then $m \in \mathcal{G}_\Delta$ for each $m \in \mathbb{N}$. The result is trivial for $m = 1$. Now suppose that $m \geq 2$. We shall inductively find a subset X of $\mathbb{N}$ with $\gamma(X) = m$.

If $\gamma(v) = \infty$ for some vertex $v \in \mathbb{N}$, then we can find an infinite subset $U = \{u_1, u_2, \ldots\}$ of $\mathbb{N}$ such that the edges vu_i and vu_j have distinct colors for all $i \neq j$. Applying Theorem 1.2.1 to the restriction of Δ to $G[U]$ (by coloring non-edges with a new color, for example), we can find an infinite subset $W = \{w_1, w_2, \ldots\}$ of U such that W is either an independent set, 1-colored, rainbow colored or lexically colored. By assumption, W cannot be rainbow colored. If W is either an independent set or 1-colored, it is clear that $W \cup \{v\}$ is star colored with centre v. If W is lexically colored, then it is easy to check that $\mathcal{G}_\Delta = \mathbb{N}$. Indeed, a subset of size $m+1$ is m-colored.

So we may assume that $\gamma(v) < \infty$ for all $v \in \mathbb{N}$. Pick an edge uv of G, and say that the color of the edge is c. We may suppose that the colors of the edges incident to u are $1, \ldots, \gamma(u)$. Consider the partition $\mathbb{N} \setminus \{u\} = U_0 \cup U_1 \cup \cdots \cup U_{\gamma(u)}$ where U_0 is the set of vertices not adjacent to u in G and for $1 \leq i \leq \gamma(u)$, U_i is the set of all vertices that are joined to u by an edge of color i. Since $\gamma(\mathbb{N}) = \infty$, by the pigeonhole principle, we must either have $\gamma(U_i) = \infty$ for some i, or $\gamma(U_i, U_j) = \infty$ for some $i \neq j$. We distinguish the following cases.

Case 1: $\gamma(U_i) < \infty$ **for all** $0 \leq i \leq \gamma(u)$**.** Since $\gamma(\mathbb{N}) = \infty$, it must be the case that $\gamma(U_i, U_j) = \infty$ for some $i \neq j$. Applying Lemma 1.2.4 to the restriction of Δ to $G[U_i \cup U_j]$, we find a subset X of $U_i \cup U_j$ such that $\gamma_G(X) = m$.

Case 2: $\gamma(U_i) = \infty$ **for some** $i \neq 0$**.** Let H be the infinite subgraph of $G[U_i]$ obtained by deleting all the edges of color i from $G[U_i]$. Clearly, $\gamma_H(w) < \infty$ for all $w \in U_i$. So H contains no infinite subset which is rainbow or star colored. By the induction hypothesis, there is a subset Y of U_i such that $\gamma_H(Y) = m-1$. Observe that all the edges between u and $Y \subset U_i$ have color i, and since the color i is not counted by γ_H, we see that $\gamma_G(Y \cup \{u\}) = m$. Therefore $X = Y \cup \{u\}$ is the required subset.

Case 3: $\gamma(U_0) = \infty$**.** Let V_0 be the set of those vertices of U_0 not adjacent to v in G. Since $\gamma(v) < \infty$, we have a partition of $U_0 \setminus V_0 = V_1 \cup \cdots \cup V_n$, with $n \leq \gamma(v)$, based on the color of the edge joining a given vertex of $U_0 \setminus V_0$ to the vertex v. Applying the same argument as in Cases 1 and 2 to the vertex v, we see that we are done unless $\gamma(V_0) = \infty$. In this case, we consider the infinite subgraph H of $G[V_0]$ obtained by deleting all the edges of color c from $G[V_0]$.

The fact that $\gamma_G(w) < \infty$ for all $w \in \mathbb{N}$ implies that $\gamma_H(w) < \infty$ for all $w \in V_0$. So H has no infinite rainbow or star colored subset. By the induction hypothesis, there is a subset Y of V_0 such that $\gamma_H(Y) = m - 1$. Observe that uv has color c and there are no edges between $\{u, v\}$ and $Y \subset V_0 \subset U_0$ in G. Since the color c is not counted by γ_H, it follows that $\gamma_G(Y \cup \{u, v\}) = m$. Therefore $X = Y \cup \{u, v\}$ is the required subset. This completes the proof. □

1.3 Extensions and applications

In this section, we shall first describe a finitary analogue of Theorem 1.2.2. We then use this to prove Theorem 1.1.2.

1.3.1 Finitary extensions

We can prove a version of Theorem 1.2.2 for colorings (of finite or infinite complete graphs) that use only finitely many colors. We say that a set is countable if it is either finite or countably infinite.

Theorem 1.3.1. *For all $n \in \mathbb{N}$, there exists a natural number $K = K(n)$ such that for every k-coloring $\Delta : V^{(2)} \twoheadrightarrow [k]$ of the complete graph on a countable set V with $k \geq K$ colors, either*

- *there is an m-colored complete subgraph for every $m \in [n]$, or*
- *there exists a rainbow colored complete subgraph on n vertices, or*
- *there exists a star colored complete subgraph on n vertices.* □

This result can be proved by arguments similar to those used to prove Theorem 1.2.2. There are two essential differences. First, as opposed to Theorem 1.2.1, we use the following extension of the theorem proved by Erdős and Rado, to colorings of *finite* complete graphs with an arbitrary set of colors.

Theorem 1.3.2. *For every $n \in \mathbb{N}$, and every coloring Δ of the complete graph on a sufficiently large countable set V, there exists a subset X of V of size at least n such that either*

- *X is 1-colored, or*
- *X is rainbow colored, or*
- *X is lexically colored.* □

Second, in the place of Lemma 1.2.4, we use the following finitary analogue which is proved in the same way as the lemma.

Lemma 1.3.3. *For all $m,d \in \mathbb{N}$, there exists a natural number $L = L(m,d)$ with the following property: for every coloring Δ of a graph G on a countable set V such that*

- *$\gamma(v) < d$ for all $v \in V$, and*
- *there is a partition of $V = A \cup B$ such that $\gamma(A) < d$, $\gamma(B) < d$ and $\gamma(A,B) \geq L$,*

there exists a subset X of V such that $X \cap A \neq \varnothing$, $X \cap B \neq \varnothing$ and $\gamma(X) = m$. □

1.3.2 Applications

Theorem 1.1.2 may be deduced from Theorem 1.3.1. Recall that Theorem 1.1.2 says for any natural number $n \in \mathbb{N}$, both $\binom{n}{2}, \binom{n}{2}+1 \in \mathcal{G}_\Delta$ for any coloring Δ of the complete graph on $\mathbb{N}$ using a finite, but sufficiently large number of colors.

Proof of Theorem 1.1.2. We prove two propositions which, taken together, imply the result. The first is an easy corollary of Theorem 1.3.1

Proposition 1.3.4. *For all $n \in \mathbb{N}$, there exists a natural number $C_1 = C_1(n)$ such that for any k-coloring $\Delta : V^{(2)} \twoheadrightarrow [k]$ of the complete graph on a countable set V with $k \geq C_1$ colors, $\binom{n}{2} \in \mathcal{G}_\Delta$.*

Proof. Take $C_1(n) = K(\binom{n}{2})$, where K is as guaranteed by Theorem 1.3.1. $\square$

The next proposition is perhaps not as straightforward.

Proposition 1.3.5. *For all $n \in \mathbb{N}$, there exists a natural number $C_2 = C_2(n)$ with the property that for all $k \geq C_2$, there exists a natural number $D_{k,n}$ such that for any k-coloring $\Delta : V^{(2)} \twoheadrightarrow [k]$ of the complete graph on a countable set V with $k \geq C_2$ colors, $\binom{n}{2} + 1 \in \mathcal{G}_\Delta$, provided $|V| \geq D_{k,n}$.*

Proof. For $n = 2$, it is an easy exercise to check that the result is true with $C_2(2) = 2$ and $D_{k,2} = R(k;k)$ where $R(k;k)$ is the Ramsey number for finding a 1-colored copy of a complete graph on k vertices when using k colors. Indeed, we can find a 1-colored set X of size k in V since $|V| \geq R(k;k)$. Let c be the color inside X. Suppose there is a vertex $y \in V \backslash X$ such that for some $x \in X$, xy has color different from c. Now if there is $x' \in X$ such that $x'y$ has color c then we are done since $\{x, x', y\}$ is 2-colored. On the other hand, if for all $x' \in X$, $x'y$ has color different from c then we can find $x_1, x_2 \in X$ such that $x_1 y$ and $x_2 y$ have distinct colors since $|V| > k - 1$. Hence, $\{x_1, x_2, y\}$ is 2-colored. Now suppose that for all $y \in V \backslash X$, xy has color c for all $x \in X$. Take an edge yy' in $V \backslash X$ with different color from c. We are done since $\{x, y, y'\}$ is 2-colored for any $x \in X$.

For $n \geq 3$, let $s = n^4$. We claim that $C_2(n) = K(s)$ will do, where K is the constant guaranteed by Theorem 1.3.1. For $k \geq C_2(n)$, we take $D_{k,n} = k^s + s + 1$. Now, suppose that $\Delta : V^{(2)} \twoheadrightarrow [k]$ is a k-coloring and $|V| \geq D_{k,n}$. Then, by our choice of $C_2(n)$, either

- there is an m-colored complete subgraph for every $m \in [s]$, or
- there exists a rainbow colored complete subgraph on s vertices, or
- there exists a star colored complete subgraph on s vertices.

Note that a star colored complete subgraph on s vertices contains an m-colored complete subgraph for $2 < m \leq s$. Since $2 < \binom{n}{2} + 1 \leq s$, we are done unless there exists a rainbow colored complete subgraph on s vertices. Hence, suppose that the complete subgraph on the vertex set $S = \{u_1, u_2, \dots, u_s\}$ is rainbow colored. For each $x \in V \setminus S$, there are k^s possible values for the s-tuple $(\Delta(xu_1), \Delta(xu_2), \dots, \Delta(xu_s))$. Since, $|V \setminus S| \geq D_{k,n} - s > k^s$, we can find vertices $x, y \in V \setminus S$ such that

$$(\Delta(xu_1), \Delta(xu_2), \dots, \Delta(xu_s)) = (\Delta(yu_1), \Delta(yu_2), \dots, \Delta(yu_s)).$$

We claim that there is a subset $T \subset S$ of size $t = n^2$ such that for all $u \in T$, $\Delta(xu) \notin \Delta(T^{(2)})$. Assume for the sake of contradiction that for every subset $T \subset S$ of size t, there exists at least one vertex $u \in T$ such that $\Delta(xu) \in \Delta(T^{(2)})$. Consider the set

$$A = \{(u, T) : u \in T \subset S, |T| = t, \Delta(xu) \in \Delta(T^{(2)})\}.$$

By our assumption, for each $T \subset S$ of size t, there is at least one $u \in T$ such that $(u, T) \in A$, and so $|A| \geq \binom{s}{t}$. As S is rainbow colored, there is at most one edge ab in $S^{(2)}$ of color $\Delta(xu)$ for each $u \in S$. If (u, T) is in A, then we must have $a, b \in T$. So for each $u \in S$, there are at most $\binom{s-2}{t-2}$ sets T such that $(u, T) \in A$. Thus, $|A| \leq s\binom{s-2}{t-2}$. Combining these two inequalities for $|A|$, we get

$$\binom{s}{t} \leq |A| \leq s\binom{s-2}{t-2}.$$

This means that $t(t-1) \geq s-1$, contradicting the fact that $s = t^2$.

Hence, there is indeed a subset T of S of size $t = n^2$ such that $\Delta(xu) \notin \Delta(T^{(2)})$ for all $u \in T$. Let $Q = \{\Delta(xu) : u \in T\}$. If $|Q| < n$, then as $|T| = n^2$, there are vertices $v_1, v_2, \dots, v_n$ in T such that

$$\Delta(xv_1) = \Delta(xv_2) = \dots = \Delta(xv_n).$$

Since this color $\Delta(xv_1)$ is not an element of $\Delta(T^{(2)})$, we conclude that the set

$\{x, v_1, v_2, \ldots, v_n\}$ is $(\binom{n}{2}+1)$-colored.

So we may assume that $|Q| \geq n$. Then there is a subset $U \subset T$ of size n such that the colors $\Delta(xu)$ are distinct for all $u \in U$. Since $U \subset T$, the color $\Delta(xu)$ is not an element of $\Delta(U^{(2)})$ for each $u \in U$. We hence conclude that $U \cup \{x\}$ is rainbow colored.

Recall that there is a vertex $y \neq x$ in $V \setminus S$ such that $\Delta(xu) = \Delta(yu)$ for all $u \in S$. Since at most one edge e in $(U \cup \{x\})^{(2)}$ is colored with the same color as the edge xy, by removing the endpoint of e which lies in U if necessary, we can find a subset U' of U of size $n-1$ such that $\Delta(xy)$ is not an element of $\Delta((U' \cup \{x\})^{(2)})$. Then $U' \cup \{x, y\}$ is $(\binom{n}{2}+1)$-colored since $U' \cup \{x\}$ and $U' \cup \{y\}$ are rainbow colored sets of size n using the same set of colors. □

It is easy to see that, taken together, Corollary 1.3.4 and Theorem 1.3.5 imply Theorem 1.1.2. □

The following corollary of Lemma 1.3.3 about finding m-colored complete bipartite subgraphs might be of independent interest.

Corollary 1.3.6. *For all $m \in \mathbb{N}$, there exists a natural number $B = B(m)$ such that if $\Delta : U \times V \twoheadrightarrow [k]$ is a k-coloring of the complete bipartite graph between two countable sets U and V with $k \geq B$ colors, then there exist $X \subset U$ and $Y \subset V$ such that the complete bipartite subgraph between by X and Y is m-colored.*

Proof. It is easy to verify that it suffices to to take $B(m) = L(m,m)$, where L is the constant guaranteed by Lemma 1.3.3. □

1.4 Approximations to m-colored hypergraphs

In this section, we prove Theorem 1.1.4. In fact, we shall do much more; we study the generalizations of the theorem to uniform hypergraphs.

Following Erdős, for a set X, we write $X^{(r)}$ for the family of all subsets of X of cardinality r; equivalently, $X^{(r)}$ is the complete r-uniform hypergraph on the vertex set X.

Let $\Delta : \mathbb{N}^{(r)} \twoheadrightarrow [k]$ be a surjective k-coloring of the edges of the complete r-uniform hypergraph on the natural numbers. As before, we say that a subset $X \subset \mathbb{N}$ is (*exactly*) *m-colored* if $\Delta(X^{(r)})$, the set of values attained by Δ on the edges induced by X, has size exactly m. The definitions of γ_Δ and $\mathcal{F}_\Delta$ are analogous.

We shall prove the following generalization of Theorem 1.1.4.

Theorem 1.4.1. *Fix a positive integer $r \geq 2$. For any k-coloring $\Delta : \mathbb{N}^{(r)} \twoheadrightarrow [k]$ and any natural number $m \leq k$, there exists an $\hat{m} \in \mathcal{F}_\Delta$ such that*

$$|m - \hat{m}| \leq c_r m^{1-1/r} + O(m^{1-2/r})$$

where $c_r = r/(2(r!)^{1/r})$.

Theorem 1.4.1 is tight up to the $O(m^{1-2/r})$ term. To see this, let $k = \binom{n}{r} + 1$ for some $n \in \mathbb{N}$. We consider the 'small-rainbow coloring' Δ which colors all the edges induced by $[n]$ with $\binom{n}{r}$ distinct colors and all the remaining edges with the one color that has not been used so far. In this case, we see that $\mathcal{F}_\Delta = \{\binom{i}{r} + 1 : i \leq n\}$. Now let $m = \lfloor (\binom{l}{r} + \binom{l+1}{r} + 2)/2 \rfloor$ for some natural number l such that $l < n$. It is not difficult to check that $|m - \hat{m}| \geq \lfloor \binom{l}{r-1}/2 \rfloor$ for each $\hat{m} \in \mathcal{F}_\Delta$; also, clearly $\lfloor \binom{l}{r-1}/2 \rfloor = c_r m^{1-1/r} + O(m^{1-2/r})$.

In the case of graphs where $r = 2$, Theorem 1.4.1 tells us that for any finite coloring of the edges of the complete graph on $\mathbb{N}$ with m or more colors, there is an exactly $\hat{m}$-colored complete infinite subgraph for some $\hat{m}$ satisfying $|m - \hat{m}| \leq \sqrt{m/2} + O(1)$; a careful analysis of the proof of Theorem 1.4.1 in this case allows us to replace the $O(1)$ term with an explicit constant, $1/2$.

We know from Theorem 1.4.1 that $\mathcal{F}_\Delta$ cannot contain very large gaps. Another natural question we are led to ask is if there are any sets, and in particular, intervals that $\mathcal{F}_\Delta$ is guaranteed to intersect. Making this more precise, Narayanan [51] conjectured that the small-rainbow coloring described above is extremal for graphs in the following sense.

Conjecture 1.4.2. *Let $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow [k]$ be a k-coloring of the complete graph on $\mathbb{N}$ and suppose n is a natural number such that $k > \binom{n}{2} + 1$. Then $\mathcal{F}_\Delta \cap (\binom{n}{2} + 1, \binom{n+1}{2} + 1] \neq \varnothing$.*

In this section, we shall prove this conjecture. There are two natural generalizations of this conjecture to r-uniform hypergraphs which are equivalent to Conjecture 1.4.2 in the case of graphs.

The first comes from considering small-rainbow colorings; indeed we can ask whether $\mathcal{F}_\Delta \cap I_{r,n} \neq \varnothing$ when $k > \binom{n}{r} + 1$ where $I_{r,n}$ is the interval $(\binom{n}{r} + 1, \binom{n+1}{r} + 1]$.

The second comes from considering a different family of colorings which we call 'small-set colorings'. Let $k = \sum_{i=0}^{r} \binom{n}{i}$ and consider the surjective k-coloring Δ of $\mathbb{N}^{(r)}$ defined by $\Delta(e) = e \cap [n]$. Note that in this case, $\mathcal{F}_\Delta = \{\sum_{i=0}^{r} \binom{j}{i} : j \leq n\}$. Consequently, we can ask whether $\mathcal{F}_\Delta \cap J_{r,n} \neq \varnothing$ when $k > \sum_{i=0}^{r} \binom{n-1}{i}$ where $J_{r,n}$ is the interval $(\sum_{i=0}^{r} \binom{n-1}{i}, \sum_{i=0}^{r} \binom{n}{i}]$.

Note that both these questions are identical when $r = 2$. Indeed $\binom{n}{2} + \binom{n}{1} + \binom{n}{0} = \binom{n+1}{2} + 1$ and so $I_{2,n} = J_{2,n}$.

We shall demonstrate that the correct generalization is the former. We shall first prove that the answer to the first question is in the affirmative, provided n is sufficiently large.

Theorem 1.4.3. *For every $r \geq 2$, there exists a natural number $n_r \geq r - 1$ such that for any natural number $n \geq n_r$ and any k-coloring $\Delta : \mathbb{N}^{(r)} \twoheadrightarrow [k]$ with $k > \binom{n}{r} + 1$, $\mathcal{F}_\Delta \cap I_{r,n} \neq \varnothing$.*

Using a result of Baranyai [6] on factorizations of uniform hypergraphs, we shall exhibit an infinite family of colorings that answer the second question negatively for every $r \geq 3$.

Theorem 1.4.4. *For every $r \geq 3$, there exist infinitely many values of n for which there exists a k-coloring $\Delta : \mathbb{N}^{(r)} \twoheadrightarrow [k]$ with $k > \sum_{i=0}^{r} \binom{n-1}{i}$ such that $\mathcal{F}_\Delta \cap J_{r,n} = \varnothing$.*

1.4.1 Proofs of the main results

We start with the following lemma which we shall later use to prove both Theorems 1.4.1 and 1.4.3.

Lemma 1.4.5. *Let $m \geq 2$ be an element of $\mathcal{F}_\Delta$. Then there exists a natural number $a = a(m,\Delta)$ such that*

- $\sum_{i=0}^{r} \binom{a}{i} \geq m$*, and*
- $\mathcal{F}_\Delta \cap [m-x, m) \neq \varnothing$ *where* $x = \min(\sum_{i=0}^{r-1} \binom{a-1}{i}, r(m-1)/a)$.

Futhermore, if

$$m = \sum_{i=t+1}^{r} \binom{a}{i} + s + 1$$

for some $s \geq 0$ and $0 \leq t+1 \leq r$, then

$$\mathcal{F}_\Delta \cap \left[\sum_{i=t+1}^{r} \binom{a-1}{i} + \left(1 - \frac{t}{a}\right) s + 1, m \right) \neq \varnothing.$$

Proof. We start by establishing the following claim.

Claim 1.4.6. *There is an infinite m-colored set $X \subset \mathbb{N}$ with a finite subset $A \subset X$ such that*

(i) the color of every edge of X is determined by its intersection with A, i.e., if $e_1 \cap A = e_2 \cap A$, then $\Delta(e_1) = \Delta(e_2)$, and

(ii) $\gamma(X \setminus \{v\}) < m$ for all $v \in A$.

Proof. To see this, let $W \subset \mathbb{N}$ be an infinite m-colored set. For each color $c \in \Delta(W^{(r)})$, pick an edge e_c in W of color c and let $A = \bigcup_c e_c$ be the set of vertices incident to these edges. So $A \subset W$ is a finite m-colored set. Let $A_1, A_2, \ldots, A_l$ be an enumeration of subsets of A of size at most r. Note that this is the complete list of possible intersections of an edge with A. We now define a descending sequence of infinite sets $B_0 \supset B_1 \supset \cdots \supset B_l$ as follows. Let $B_0 = W \setminus A$. Having defined the infinite set B_{i-1}, we induce a coloring of the

$(r-|A_i|)$-tuples T of B_{i-1}, by giving T the color of the edge $A_i \cup T$. By Ramsey's theorem, there is an infinite monochromatic subset $B_i \subset B_{i-1}$ with respect to this induced coloring, and so the edges of $A \cup B_i$, whose intersection with A is A_i, have the same color.

Hence, $X = A \cup B_l$ is an infinite m-colored set satisfying property (i). Now, if we have a vertex $v \in A$ such that $\gamma(X \setminus \{v\}) = m$, we delete v from A. We repeat this until we are left with an m-colored set X satisfying (i) and (ii). $\square$

Let X and A be as guaranteed by Claim 1.4.6. Note that A is nonempty since $m \geq 2$. We shall prove the lemma with $a(m,\Delta) = |A|$. From the structure of X and A, we note that $\sum_{i=0}^{r} \binom{a}{i} \geq m$. That

$$\mathcal{F}_\Delta \cap \left[m - \min\left(\sum_{i=0}^{r-1} \binom{a-1}{i}, \frac{r(m-1)}{a} \right), m \right) \neq \varnothing$$

is a consequence of the following claim.

Claim 1.4.7. *There exist infinite sets $X_1, X_2 \subset X$ such that $m - \sum_{i=0}^{r-1} \binom{a-1}{i} \leq \gamma(X_1) < m$ and $m - r(m-1)/a \leq \gamma(X_2) < m$.*

Proof. Let $X_1 = X \setminus \{v\}$ for any $v \in A$. We know from Claim 1.4.6 that $\gamma(X_1) < m$. We shall now prove that $\gamma(X_1) \geq m - \sum_{i=0}^{r-1} \binom{a-1}{i}$; that is, the number of colors lost by removing v from X is at most $\sum_{i=0}^{r-1} \binom{a-1}{i}$. Since the color of an edge is determined by its intersection with A, the number of colors lost is at most the number of subsets of A containing v of size at most r, which is precisely $\sum_{i=0}^{r-1} \binom{a-1}{i}$.

Next, we shall prove that there is a subset $X_2 \subset X$ such that $m - r(m-1)/a \leq \gamma(X_2) < m$. Let $A = \{v_1, v_2, \ldots, v_a\}$ and let

$$C_i = \Delta\left(X^{(r)}\right) \setminus \Delta\left((X \setminus \{v_i\})^{(r)}\right)$$

be the set of colors lost by removing v_i from X; since $\gamma(X \setminus \{v_i\}) < m$ for all $v_i \in A$, it follows that $C_i \neq \varnothing$. For each color $c \in \Delta(X^{(r)})$, pick an edge e_c of color c, and let

$A_c = e_c \cap A$; in particular, we take $A_{c_\varnothing} = \varnothing$, where $c_\varnothing$ is the color corresponding to an empty intersection with A. Since every edge of color $c \in C_i$ contains v_i,

$$\sum_{i=1}^{a} |C_i| \leq \sum_{c \neq c_\varnothing} |A_c| \leq r(m-1),$$

and so there exists an i such that $0 < |C_i| \leq r(m-1)/a$; the claim follows by taking $X_2 = X \setminus \{v_i\}$. □

We finish the proof of the lemma by establishing the following claim.

Claim 1.4.8. *If we can write $m = \sum_{i=t+1}^{r} \binom{a}{i} + s + 1$ for some $s \geq 0$ and $0 \leq t+1 \leq r$, then*

$$\mathcal{F}_\Delta \cap \left[\sum_{i=t+1}^{r} \binom{a-1}{i} + \left(1 - \frac{t}{a}\right) s + 1, m \right) \neq \varnothing.$$

Proof. As in the proof of Claim 1.4.7, for each color $c \in \Delta(X^{(r)})$, pick an edge e_c of color c, and let $A_c = e_c \cap A$; in particular, let $A_{c_\varnothing} = \varnothing$. We know from Claim 1.4.6 that edges of X of distinct colors cannot have the same intersection with A. Consequently, all the A_c are distinct subsets of A, each of size at most r. Hence,

$$\sum_{c \neq c_\varnothing} |A_c| \leq \sum (\text{sizes of } m-1 \text{ largest subsets of } A) \leq \sum_{i=t+1}^{r} i \binom{a}{i} + ts.$$

Arguing as in the proof of Claim 1.4.7, we conclude that there exists a vertex $v \in A$ such that the number of colors lost by removing v from X is at most $(\sum_{i=t+1}^{r} i \binom{a}{i} + ts)/a$. Therefore

$$\begin{aligned} \gamma(X \setminus \{v\}) &\geq m - \frac{1}{a} \left(\sum_{i=t+1}^{r} i \binom{a}{i} + ts \right) \\ &= m - \left(\sum_{i=t+1}^{r} \binom{a-1}{i-1} + \frac{ts}{a} \right) \\ &= \sum_{i=t+1}^{r} \binom{a-1}{i} + \left(1 - \frac{t}{a}\right) s + 1, \end{aligned}$$

and so

$$\mathcal{F}_\Delta \cap \left[\sum_{i=t+1}^{r} \binom{a-1}{i} + \left(1 - \frac{t}{a}\right) s + 1, m \right) \neq \varnothing. \qquad \square$$

The lemma follows from Claims 1.4.6, 1.4.7 and 1.4.8. We are done. $\square$

Having established Lemma 1.4.5, it is easy to deduce both Theorem 1.4.1 and 1.4.3 from the lemma.

Proof of Theorem 1.4.1. Let $t = m + c_r m^{1-1/r}$. We may assume that $m > r^r/r!$ since otherwise $m = O(1)$ and there is nothing to prove. Also, if $t \geq k$, then the result follows easily by taking $\hat{m} = k$ so we may assume that $t < k$. Let $\hat{t}$ be the smallest element of $\mathcal{F}_\Delta$ greater than t. Applying Lemma 1.4.5 to $\hat{t}$, we find an $\hat{m} \in \mathcal{F}_\Delta$ such that $\hat{m} \leq t$ and

$$\hat{m} \geq \hat{t} - \min\left(\sum_{i=0}^{r-1} \binom{a-1}{i}, \frac{r(\hat{t}-1)}{a} \right)$$

for some natural number a. Now if $a \geq (r!m)^{1/r} > r$, then

$$\hat{m} \geq \hat{t} - \frac{r(\hat{t}-1)}{a} \geq \hat{t}\left(1 - \frac{r}{a}\right) \geq t\left(1 - \frac{r}{a}\right)$$

$$\geq (m + c_r m^{1-1/r}) \left(1 - \frac{r}{(r!m)^{1/r}}\right) = m + c_r m^{1-1/r} - \frac{r}{(r!)^{1/r}} m^{1-1/r} - O(m^{1-2/r})$$

and so it follows that $\hat{m} \geq m - c_r m^{1-1/r} - O(m^{1-2/r})$. If $a < (r!m)^{1/r}$ on the other hand, then using the fact that

$$\begin{aligned} \hat{m} &\geq \hat{t} - \sum_{i=0}^{r-1} \binom{a-1}{i} \\ &\geq t - \frac{a^{r-1}}{(r-1)!} - O(a^{r-2}) \\ &\geq m + c_r m^{1-1/r} - \frac{(r!m)^{1-1/r}}{(r-1)!} - O(m^{1-2/r}), \end{aligned}$$

it follows once again that $\hat{m} \geq m - c_r m^{1-1/r} - O(m^{1-2/r})$. $\square$

Proof of Theorem 1.4.3. If $k \leq \binom{n+1}{r} + 1$, we are done since $k \in \mathcal{F}_\Delta$. So suppose that $k > \binom{n+1}{r} + 1$. Let m be the smallest element of $\mathcal{F}_\Delta$ such that $m > \binom{n+1}{r} + 1$; hence, $\mathcal{F}_\Delta \cap (\binom{n+1}{r} + 1, m) = \varnothing$. Now, since $m \geq 2$, there exists by Lemma 1.4.5, a natural number a such that

$$\mathcal{F}_\Delta \cap \left[m - \frac{r(m-1)}{a}, \binom{n+1}{r} + 1\right] \neq \varnothing.$$

To prove the theorem, it is sufficient to show that $m - r(m-1)/a > \binom{n}{r} + 1$. We know from Lemma 1.4.5 that $\sum_{i=0}^{r} \binom{a}{i} \geq m > \binom{n+1}{r} + 1$. If n is sufficiently large, we must have $a \geq n$ since

$$\begin{aligned}
\binom{n+1}{r} &= \binom{n}{r} + \binom{n}{r-1} \\
&= \binom{n-1}{r} + \binom{n-1}{r-1} + \binom{n}{r-1} \\
&= \sum_{i=0}^{r} \binom{n-1}{i} + \binom{n}{r-1} - O(n^{r-2}).
\end{aligned}$$

If $a \geq n+1$, then

$$\begin{aligned}
m - \frac{r(m-1)}{a} &= (m-1)\left(1 - \frac{r}{a}\right) + 1 \\
&> \binom{n+1}{r}\left(1 - \frac{r}{n+1}\right) + 1 \\
&= \binom{n}{r} + 1
\end{aligned}$$

since $m > \binom{n+1}{r} + 1$ and $n \geq r - 1$.

We now deal with the case $a = n$. First, we write $m = \binom{n}{r} + \binom{n}{r-1} + s + 1$. Since $m > \binom{n+1}{r} + 1$ and $\binom{n}{r} + \binom{n}{r-1} = \binom{n+1}{r}$, we see that $s > 0$. By Lemma 1.4.5, it follows that

$$\mathcal{F}_\Delta \cap \left[\binom{n}{r} + \left(1 - \frac{r-2}{n}\right)s + 1, m\right) \neq \varnothing.$$

Since $n \geq r - 1$ and $s > 0$, the result follows. $\square$

A careful inspection of the proof of Theorem 1.4.3 shows that when $r=2$, the statement holds for all $n \in \mathbb{N}$. Indeed, if $a \leq n$ then since $m > \binom{n+1}{2}+1$, we have $m-a > \binom{n+1}{2}+1-n = \binom{n}{2}+1$ as required. We hence obtain a proof of Conjecture 1.4.2. By constructing a sequence of highly structured subgraphs, Narayanan [51] proved that for any k-coloring $\Delta : \mathbb{N}^{(2)} \twoheadrightarrow [k]$ with $k \geq \binom{n}{2}+1$ for some natural number n, $|\mathcal{F}_\Delta| \geq n$; our proof of Conjecture 1.4.2 gives a short proof of this lower bound. Theorem 1.4.3 also yields a generalization of this lower bound for r-uniform hypergraphs, albeit with a constant additive error term (which depends on r).

We now turn to the proof of Theorem 1.4.4. We will need a result of Baranyai's [6] which states that the set of edges of the complete r-uniform hypergraph on l vertices can be partitioned into perfect matchings when $r \mid l$.

Proof of Theorem 1.4.4. We shall show that if n is sufficiently large and $(r-1) \mid (n+1)$, then there is a surjective k-coloring Δ of $\mathbb{N}^{(r)}$ with $k > \sum_{i=0}^{r} \binom{n-1}{i}$ and $\mathcal{F}_\Delta \cap J_{r,n} = \varnothing$. We shall define a coloring of $\mathbb{N}^{(r)}$ such that the color of an edge e is determined by its intersection with a set A of size $n+1$, say $A = [n+1]$. Let $\mathcal{B}$ be the family of all subsets of A of size at most r. For $B \in \mathcal{B}$, we denote the color assigned to all the edges e such that $e \cap A = B$ by c_B.

To define our coloring, we shall construct a partition $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ with $\varnothing \in \mathcal{B}_2$. Then for every $B \in \mathcal{B}_2$, we set c_B to be equal to $c_\varnothing$. Finally, we take the colors c_B for $B \in \mathcal{B}_1$ to all be distinct and different from $c_\varnothing$. Hence, the number of colors used is $k = |\mathcal{B}_1|+1$. It remains to construct this partition of $\mathcal{B}$.

Since $(r-1) \mid (n+1)$, by Baranyai's theorem there exists an ordering

$$B_1, B_2, \dots, B_{\binom{n+1}{r-1}}$$

of the subsets of A of size $r-1$ such that for all $0 \le t \le \binom{n}{r-2}$, the family

$$\left\{B_{\binom{n+1}{r-1}t+1}, B_{\binom{n+1}{r-1}t+2}, \ldots, B_{\binom{n+1}{r-1}(t+1)}\right\}$$

is a perfect matching. Let $\mathcal{B}_1 = \{B_1, B_2, \ldots, B_s\} \cup \{B \in \mathcal{B} : |B| = r\}$, where

$$s = \sum_{i=0}^{r} \binom{n}{i} - \binom{n+1}{r} = \sum_{i=0}^{r-2} \binom{n}{i};$$

our coloring is well defined because $0 \le s \le \binom{n+1}{r-1}$ for all sufficiently large n. Observe that

$$k = |\mathcal{B}_1| + 1 = \binom{n+1}{r} + s + 1 = \sum_{i=0}^{r} \binom{n}{i} + 1.$$

We shall show that the second largest element of $\mathcal{F}_\Delta$ is at most $\sum_{i=0}^{r} \binom{n-1}{i}$. Note that any $X \subset \mathbb{N}$ with $\gamma(X) < k$ cannot contain A. As before, let C_i be the set of colors lost by removing $i \in A$ from $\mathbb{N}$, i.e.,

$$C_i = \Delta\left(\mathbb{N}^{(r)}\right) \setminus \Delta\left((\mathbb{N} \setminus \{i\})^{(r)}\right).$$

We shall complete the proof by showing that $k - |C_i| \le \sum_{i=0}^{r} \binom{n-1}{i}$ for all $i \in A$.

Note that our construction ensures that $||C_i| - |C_j|| \le 1$ for all $i, j \in A$. Now, observe that

$$\sum_{i=1}^{n+1} |C_i| = \sum_{B \in \mathcal{B}_1} |B| = r\binom{n+1}{r} + (r-1)s,$$

and so $|C_i| \ge (r\binom{n+1}{r} + (r-1)s)/(n+1) - 1$ for all $i \in A$. It is then easily verified using Pascal's identity that when $r \ge 4$ and n is sufficiently large,

$$\begin{aligned} k - |C_i| &\le \left(\binom{n+1}{r} + s + 1\right) - \frac{1}{n+1}\left(r\binom{n+1}{r} + (r-1)s\right) + 1 \\ &= \left(\binom{n+1}{r} + s + 1\right) - \left(\binom{n}{r-1} + \left(\frac{r-1}{n+1}\right)s\right) + 1 \end{aligned}$$

$$
\begin{aligned}
&= \binom{n}{r} + \left(1 - \frac{r-1}{n+1}\right) s + 2 \\
&= \binom{n}{r} + \left(1 - \frac{r-1}{n+1}\right) \sum_{i=0}^{r-2} \binom{n}{i} + 2 \\
&\leq \sum_{i=0}^{r} \binom{n-1}{i}
\end{aligned}
$$

since

$$
\begin{aligned}
\sum_{i=0}^{r} \binom{n-1}{i} - \binom{n}{r} - \left(1 - \frac{r-1}{n+1}\right) \sum_{i=0}^{r-2} \binom{n}{i} - 2 \\
&= \sum_{i=0}^{r-2} \binom{n-1}{i} - \left(1 - \frac{r-1}{n+1}\right) \sum_{i=0}^{r-2} \binom{n}{i} - 2 \\
&= \frac{r-1}{n+1} \sum_{i=0}^{r-2} \binom{n}{i} - \sum_{i=0}^{r-2} \binom{n-1}{i-1} - 2 \\
&= (r-1)\frac{n^{r-3}}{(r-2)!} - \frac{n^{r-3}}{(r-3)!} + O(n^{r-4}) \\
&= \frac{n^{r-3}}{(r-2)!} + O(n^{r-4}).
\end{aligned}
$$

When $r = 3$, it is easy to check that $s = n+1$ and so s is divisible by $(n+1)/(r-1) = (n+1)/2$. Consequently, in this case, $|C_i| = |C_j|$ for $i, j \in A$. Hence,

$$
\begin{aligned}
k - |C_i| &\leq \left(\binom{n+1}{3} + s + 1\right) - \frac{1}{n+1}\left(r\binom{n+1}{3} + 2s\right) \\
&= \binom{n}{3} + \left(1 - \frac{2}{n+1}\right)(n+1) + 1 \\
&= \sum_{i=0}^{3} \binom{n-1}{i}.
\end{aligned}
$$

This completes the proof. □

1.5 Open problems

We conclude by a few questions that would merit further study. First, the problem of characterizing the sets $\mathcal{F}_\Delta$ and $\mathcal{G}_\Delta$ for both infinite and finite colorings is quite interesting; while we have taken a few steps towards this in this chapter, the full question is still far from being resolved.

Second, it would be reasonable to generalize Theorem 1.2.2 for r-uniform hypergraphs. However, even in the case of $\mathbb{N}^{(3)}$, it is not immediately clear to us what the canonical structures analogous to the rainbow colored and star colored complete graphs should be.

To state the next problem, let us define

$$\psi_r(k) = \min_{\Delta:\mathbb{N}^{(r)} \twoheadrightarrow [k]} |\mathcal{F}_\Delta|.$$

A consequence of Theorem 1.4.3 is that $\psi_r(k) \geq (r!k)^{1/r} - O(1)$. Turning to the question of upper bounds for ψ_r, the small-rainbow coloring shows that the lower bound that we get from Theorem 1.4.3 is tight infinitely often, i.e., when k is of the form $\binom{n}{r} + 1$ for some $n \in \mathbb{N}$. When k is not of this form, there are two obvious ways of generalising the small-rainbow coloring: we could replace the rainbow colored clique in our construction either with a disjoint union of cliques or with a clique along with a pendant vertex attached to some subset of the vertices of the clique. However, both these obvious generalizations of the small-rainbow coloring fail to give us good upper bounds for $\psi_r(k)$ for a general $k \in \mathbb{N}$. Narayanan [51] proved using rainbow colorings of complete bipartite graphs that

$$\psi_2(k) = O\left(\frac{k}{(\log\log k)^{\delta}(\log\log\log k)^{3/2}}\right)$$

for almost all natural numbers k and some absolute constant $\delta > 0$. The same construction can be extended to show that $\psi_r(k) = o(k)$ for almost all natural numbers k. It would be very interesting to decide if, in fact, $\psi_r(k) = o(k)$ for all $k \in \mathbb{N}$.

CHAPTER 2

DISJOINT INDUCED SUBGRAPHS OF THE SAME ORDER AND SIZE

2.1 Introduction

Given a graph G, can we guarantee that G contains two large, vertex-disjoint copies of the same graph? It follows from Ramsey's theorem that any graph on n vertices contains two vertex-disjoint isomorphic induced subgraphs on $\Omega(\log n)$ vertices, either complete or empty; by considering a random graph on n vertices, it is easy to check that this is also best possible up to constant factors.

What if, rather than asking for two isomorphic subgraphs, we ask for two subgraphs that are the same with respect to one or more graph parameters? Caro and Yuster [17] considered the question of finding two vertex-disjoint subgraphs of a given graph of the same order which induce the same number of edges. For a graph G, let $f(G)$ be the largest integer k such that there are two vertex-disjoint subgraphs of G each on k vertices, both inducing the same number of edges and let $f(n)$ be the minimum value of $f(G)$ taken over all graphs on n vertices. Trivially, $f(n) \leq \lfloor n/2 \rfloor$; also, as shown by Ben-Eliezer and Krivelevich [7], equality holds (with high probability) for the Erdős-Rényi random graphs $G(n,p)$ for all $0 \leq p \leq 1$.

There is a large gap between the best known upper and lower bounds for $f(n)$. From below, one can easily show using the pigeonhole principle that $f(n) = \Omega(n^{1/3})$. As observed by Caro and Yuster, it is possible to improve this to $f(n) = \Omega(n^{1/2})$ using a well known result of Lovász determining the chromatic number of Kneser graphs. By considering a carefully constructed disjoint union of cliques, each on an odd number of vertices, Caro and Yuster showed that $f(n) \leq n/2 - \Omega(\log\log n)$.

As expected, one can say more about $f(G)$ when G belongs to certain special graph classes. For example, Axenovich, Martin and Ueckerdt [4] showed that $f(G) \geq \lceil n/2 \rceil - 1$ when G is a forest; this is clearly best possible. Indeed, it is possible to get quite close to the trivial upper bound of $n/2$ when we restrict our attention to sparse graphs. In their paper, Caro and Yuster showed, for any fixed $\alpha > 0$, that if G is a graph on n vertices, then $f(G) \geq n/2 - o(n)$ provided G has at most $n^{2-\alpha}$ edges (or non-edges). Axenovich, Martin and Ueckerdt [4] later showed that the same holds for graphs with at most $o(n^2/(\log n)^2)$ edges.

Our main aim in this chapter is to narrow considerably the gap between the best known upper and lower bounds for $f(n)$, and thereby answer a question of Caro and Yuster [17].

Theorem 2.1.1. *For every $\varepsilon > 0$, there exists a natural number $N = N(\varepsilon)$ such that for any graph G on $n > N$ vertices, $f(G) \geq n/2 - \varepsilon n$. Consequently,*

$$n/2 - o(n) \leq f(n) \leq n/2 - \Omega(\log\log n).$$

We remark that much research has been done on the family of induced subgraphs of a graph. For example, call a graph *k-universal* if it contains every graph of order k as an induced subgraph. Very crudely, if G is a k-universal graph with n vertices, then

$$\binom{n}{k} \geq \frac{2^{\binom{k}{2}}}{k!},$$

and so $n \geq 2^{(k-1)/2}$. As remarked in [13], almost all graphs with $k^2 2^{k/2}$ vertices are k-universal, and the Paley graphs come close to providing examples which are almost as good. Hajnal conjectured that if a graph only has a 'small' number of distinct (non-isomorphic) induced subgraphs, then it contains a trivial (complete or empty) subgraph with linearly many vertices. This was proved, shortly after the conjecture was made, by Alon and Bollobás [2], and Erdős and Hajnal [23], the latter in a stronger form. In [2] only a few parameters, like order, size and maximal degree, were used to distinguish non-isomorphic graphs.

Erdős and Hajnal [24] then went much further: they realized that forbidding a single graph as an induced subgraph severely constrains the structure of a graph. More precisely, they made the major conjecture that there is a positive constant $\gamma(H)$ for every graph H such that if a graph of order n does not contain H as an induced subgraph, then the graph contains a trivial subgraph with at least $n^{\gamma(H)}$ vertices. In spite of all the work on this conjecture (for a small sample, see [18, 30, 54]) we are very far from the desired bound.

Let us finally mention another interesting line of research about finding disjoint isomorphic (not necessarily induced) subgraphs. Jacobson and Schönheim (see [25, 46]) independently raised the question of finding edge-disjoint isomorphic subgraphs. Improving on results of Erdős, Pyber and Pach [25], it has been shown by Lee, Loh and Sudakov [46] that every graph on m edges contains a pair of edge-disjoint isomorphic subgraphs with at least $\Omega((m \log m)^{2/3})$ edges and that this is also best possible up to a multiplicative constant.

The rest of this chapter is organised as follows. We give an overview of our approach in Section 2.3, and then fill in the details and prove Theorem 2.1.1 in Section 2.4. There are many natural questions about induced subgraphs which are close to Theorem 2.1.1 in spirit; we conclude in Section 2.5 by mentioning some of these.

2.2 Preliminaries

Our objective in this section is to establish some notational conveniences and collect together for easy reference, some simple propositions that we shall make use of when proving our main result.

2.2.1 Notation

It will be convenient to establish some notation for working with sets of pairs. A *pair* (x, y) will always mean an unordered pair with $x \neq y$, and a *collection* of pairs $\mathcal{P}$ will always mean a set of disjoint pairs; for example, $\mathcal{P} = \{(1,2),(3,4)\}$ is a collection of pairs, but $\mathcal{Q} = \{(1,2),(2,3)\}$ is not. For a collection of pairs denoted by $\mathcal{P}$, we shall write P for the underlying ground set of elements, i.e., $P = \bigcup_{(x,y)\in\mathcal{P}}\{x,y\}$; in other words, we reserve the corresponding upper case letter for the ground set. We shall say that two collections of pairs $\mathcal{P}$ and $\mathcal{Q}$ are disjoint if $P \cap Q = \varnothing$; to wit, the collections $\mathcal{P}_1 = \{(1,2),(3,4)\}$ and $\mathcal{Q}_1 = \{(5,6),(7,8)\}$ are disjoint, while the collections $\mathcal{P}_2 = \{(1,2),(3,4)\}$ and $\mathcal{Q}_2 = \{(1,3),(2,4)\}$ are not.

As usual, given a graph $G = (V,E)$, we write $\deg(v)$ and $\Gamma(v)$ respectively for the degree and for the neighborhood of a vertex v in G. For a subset $U \subset V$, we write $G[U]$ for the subgraph induced by U, $e(U)$ for the number of edges of $G[U]$, and $\deg(U)$ for the sum of the degrees (in G) of the vertices of U. Given two disjoint subsets $A,B \subset V$, we write $e(A,B)$ for the number of edges with one endpoint each in A and B.

We shall also use less common terminology and notation. Thus, for any two vertices $x,y \in V$, we write $\delta(x,y)$ for the *degree difference* between x and y, namely the quantity $|\deg(x) - \deg(y)|$. We say that two vertices x and y *disagree on a vertex* $v \neq x,y$ if v is adjacent to exactly one of x and y; otherwise x and y *agree* on v. For any two vertices $x,y \in V$, the *difference neighborhood* $\Gamma(x,y) = (\Gamma(x)\Delta\Gamma(y))\setminus\{x,y\}$ of x and y is the set of vertices $v \neq x,y$ on which x and y disagree; we write $\Delta(x,y)$ for the size of the difference neighborhood, so that $\delta(x,y) \leq \Delta(x,y)$. If two vertices x and y agree on every vertex $v \neq x,y$, we say that the pair (x,y) is a *clone pair*. When the graph G in question is not clear from the context, we shall, for example, write $\delta(x,y,G)$ to denote the degree difference between x and y in G.

We say that a graph G is *splittable* if there is a partition $V = A \cup B$ of its vertex set into two sets A and B of equal size with $e(A) = e(B)$; in this case, we call (A,B) a *splitting* of G. Note that $e(A) = e(B)$ if and only if $\deg(A) = \deg(B)$, since $\deg(A) = 2e(A) + e(A,B)$.

Our conventions for asymptotic notation are largely standard; however, we feel obliged to point out that we write $o_{k\to\infty}(1)$ to denote a function (of k) that goes to 0 as

$k \to \infty$, and that when we write, say $\Omega_k(.)$, we mean that the constant suppressed by the asymptotic notation is allowed to depend on (but is completely determined by) the parameter k. For the sake of clarity of presentation, we systematically omit floors and ceilings whenever they are not crucial.

2.2.2 Preliminary observations

We shall make use of the following simple observation repeatedly when constructing a splitting.

Proposition 2.2.1. *Given positive real numbers $x_1, x_2, \ldots, x_t$ in the interval $[a,b]$ with $0 \le a \le b$, we may, for every $y \in [-ta, ta]$, choose signs $\zeta_i \in \{-1, +1\}$ such that $|y + \sum \zeta_i x_i| \le b$.* □

The following first moment bound will prove useful; it is easily checked that the bound is the best possible.

Proposition 2.2.2. *Let X be a random variable such that $X \le N$ and $\mathbb{E}[X] \ge Np$. Then*

$$\mathbb{P}\left(X \ge \frac{\mathbb{E}[X]}{2}\right) \ge \frac{p}{2-p}.$$

Proof. Let $q = \mathbb{P}\left(X \ge \frac{\mathbb{E}[X]}{2}\right)$. Since $X \le N$, we have $\mathbb{E}[X] \le \frac{\mathbb{E}[X]}{2}(1-q) + Nq$ and hence

$$q \ge \frac{\mathbb{E}[X]}{2N - \mathbb{E}[X]} \ge \frac{Np}{2N - Np} = \frac{p}{2-p}$$

as required. □

We will also need the following two easy propositions.

Proposition 2.2.3. *Given $x_1, x_2, \ldots, x_t$ in the interval $[0,a]$, a positive real b and a natural number N, it is possible to find $\lfloor t/N \rfloor - \lceil a/b \rceil$ disjoint subsets of $\{x_1, x_2, \ldots, x_t\}$, each of size N, such that $|x_i - x_j| \le b$ for any x_i and x_j belonging to the same subset.*

Proof. Suppose that $x_1 \le x_2 \le \cdots \le x_t$. Let $i_0 = 1$ and define i_j to be the smallest index such that $x_{i_j} > x_{i_{j-1}} + b$ and consider the sets $S_j = \{x_{i_j}, x_{i_j+1}, \ldots, x_{i_{j+1}-1}\}$. Since $x_1 \ge 0$ and $x_t \le a$, there are at most $\lceil a/b \rceil$ such sets. Now, by discarding at most N numbers from each S_j if necessary, we can assume that N divides $|S_j|$ for each j. We now partition each S_j into subsets of size N. Clearly, $|x_i - x_j| \le b$ for any x_i and x_j belonging to the same subset. The number of elements we have discarded is at most $N\lceil a/b \rceil$. So the number of subsets of size N we are left with is at least $\lfloor t/N \rfloor - \lceil a/b \rceil$. □

Remark. We shall often apply Proposition 2.2.3 to the degrees of a subset of vertices of a graph; we consequently obtain disjoint groups of vertices such that the degree difference of any two vertices in the same group is suitably bounded.

Proposition 2.2.4. *Let x, y and z be three vertices and U some subset of vertices of a graph G. Then some two of the vertices x, y and z disagree on at most two thirds of the vertices of U.*

Proof. Any vertex $v \in U$ belongs to at most two of the three difference neighborhoods $\Gamma(x,y)$, $\Gamma(y,z)$ and $\Gamma(z,x)$. The claim follows by averaging. □

2.2.3 Binomial random variables

We will need some easily proven statements about binomial random variables. We collect these here. As usual, for a random variable with distribution $\mathrm{Bin}(N,p)$, we write $\mu(= Np)$ for its mean and $\sigma^2(= Np(1-p))$ for its variance.

The first proposition we shall require is an easy consequence of the fact that $e^{-2x} \le 1 - x \le e^{-x}$ for all $0 \le x \le 1/2$.

Proposition 2.2.5. *Let X be a random variable with distribution* $\mathrm{Bin}(N,p)$*, with $p \le 1/2$. Then for any $k \ge 1$,*

$$\exp(-2\mu)(\mu/k)^k \le \mathbb{P}(X = k) \le \exp(-\mu)(2e\mu/k)^k.$$

Also, $\exp(-2\mu) \leq \mathbb{P}(X=0) \leq \exp(-\mu)$. □

We shall make use of the following standard concentration result which first appeared in a paper of Bernstein and was later rediscovered by Chernoff and Hoeffding; see [3] for example.

Proposition 2.2.6. *Let X be a random variable with distribution* $\mathrm{Bin}(N,p)$. *Then*

$$\mathbb{P}(|X-Np|>t) \leq \exp\left(\frac{-t^2}{N/2+2t/3}\right). \qquad \square$$

Proposition 2.2.7. *Let X be a random variable with distribution* $\mathrm{Bin}(N,p)$. *Then*

$$\mathbb{P}(X \textit{ is even}) = \frac{1}{2}(1+(1-2p)^N). \qquad \square$$

Proposition 2.2.8. *Let X_1 and X_2 be two independent random variables both with distribution* $\mathrm{Bin}(N,p)$. *Then*

$$\mathbb{P}(X_1 = X_2) = o_{\sigma\to\infty}(1).$$

In particular, when $p \leq 1/2$, $\mathbb{P}(X_1 = X_2) = o_{\mu\to\infty}(1)$. □

Proposition 2.2.9. *Let X_1 and X_2 be two independent random variables with distributions* $\mathrm{Bin}(N_1,p)$ *and* $\mathrm{Bin}(N_2,p)$ *respectively, with $p \geq 1/2$. Then*

$$\mathbb{P}(|X_1-X_2| < |N_1-N_2|^{1/3}) = o_{|N_1-N_2|\to\infty}(1). \qquad \square$$

Proposition 2.2.10. *Let X_1 and X_2 be two independent random variables with distributions* $\mathrm{Bin}(N_1,p)$ *and* $\mathrm{Bin}(N_2,p)$ *respectively, with $p \geq 1/2$. Suppose $N_1 \leq N$, $N_2 \leq N$ and $|N_1-N_2| \leq cN^{1/2}$ for some absolute constant c. Then*

$$\mathbb{P}(|X_1-X_2| > N^{2/3}) = O\left(\exp\left(\frac{-N^{1/3}}{5}\right)\right).$$

Proof. Since $|X_1 - X_2| \leq |X_1 - pN_1| + |pN_1 - pN_2| + |X_2 - pN_2|$, we have

$$\begin{aligned}&\mathbb{P}(|X_1 - X_2| > N^{2/3})\\ &\quad \leq \mathbb{P}(|X_1 - pN_1| > \frac{1}{3}N^{2/3}) + \mathbb{P}(|pN_1 - pN_2| > \frac{1}{3}N^{2/3}) + \mathbb{P}(|X_2 - pN_2| > \frac{1}{3}N^{2/3}).\end{aligned}$$

It is sufficient to show that each term on the right hand side is $O\left(\exp\left(\frac{-N^{1/3}}{5}\right)\right)$. This is clear for the middle term since $|N_1 - N_2| \leq cN^{1/2}$. For the first and last term, we apply Propposition 2.2.6,

$$\mathbb{P}(|X_i - pN_i| > \frac{1}{3}N^{2/3}) \leq \exp\left(\frac{-\frac{1}{9}N^{4/3}}{\frac{1}{2}N + \frac{2}{9}N^{2/3}}\right) \leq \exp\left(\frac{-N^{1/3}}{5}\right)$$

for sufficiently large n. □

2.3 Overview of our strategy

To prove Theorem 2.1.1, we need to show that if $\varepsilon > 0$ and n is sufficiently large, then any graph G on n vertices contains two disjoint subsets of vertices of the same size, each of cardinality at least $(1/2 - \varepsilon)n$, which induce the same number of edges. Equivalently, we need to show that it is possible to transform G into a splittable graph by deleting at most $2\varepsilon n$ vertices from G. Recall that a graph is splittable if and only if there is a partition of its vertex set into two sets of equal size such that the sums of the degrees of the vertices in the two sets are equal.

We shall show that there is a probability $0 < p \leq \varepsilon$ (depending on G) such that if we delete vertices from G with probability p, then the resulting graph H is splittable with positive probability.

To show that this random subgraph H is splittable, we shall exhibit a large collection of 'gadgets' in H. Given $0 \leq a \leq b$, by an $[a,b]$-*gadget*, we mean a pair of vertices (x,y) such that $a \leq \delta(x,y) \leq b$; a gadget, in other words, is just a pair of vertices whose degree difference we can control.

Once we have found sufficiently many suitable gadgets in H, we construct a splitting of H as follows: we use Proposition 2.2.1 to decide, one by one for each gadget, which way round to assign the vertices of the gadget to the sides of the splitting. The following lemma makes this idea precise.

Lemma 2.3.1. *Let H be a graph on an even number of vertices and suppose that we can partition $V(H)$ into disjoint collections of pairs $\mathcal{P}_1, \mathcal{P}_2, \dots, \mathcal{P}_k$ such that the pairs in $\mathcal{P}_i$ are $[a_i, b_i]$-gadgets, where $0 \le a_1 \le b_1$ and $0 < a_i \le b_i$ for $2 \le i \le k$. If $b_{i-1} \le a_i |\mathcal{P}_i|$ for each $2 \le i \le k$, then $V(H)$ can be partitioned into two sets A, B of the same size such that $|\deg(A) - \deg(B)| \le b_k$. In particular, if $b_k = 1$, then H is splittable.*

Proof. We show by induction on i that it is possible to partition the vertices of the gadgets in $\mathcal{P}_1, \dots, \mathcal{P}_i$ into two sets A_i and B_i of equal size such that $|\deg(A_i) - \deg(B_i)| \le b_i$. The lemma follows by taking $A = A_k$ and $B = B_k$.

We set $b_0 = 0$ and $A_0 = B_0 = \varnothing$ and so the claim is trivially true when $i = 0$. So suppose that $i \ge 1$ and that we have constructed A_{i-1} and B_{i-1}. Denote the $[a_i, b_i]$-gadgets in $\mathcal{P}_i$ by (x_j, y_j), where $\deg(x_j) \ge \deg(y_j)$ for $1 \le j \le |\mathcal{P}_i|$. Using the fact that $b_{i-1} \le a_i |\mathcal{P}_i|$, it follows from Proposition 2.2.1 that there is a choice of signs $\zeta_j \in \{-1, +1\}$ for $1 \le j \le |\mathcal{P}_i|$ such that

$$\Big|(\deg(A) - \deg(B)) + \sum_j \zeta_j \delta(x_j, y_j)\Big| \le b_i.$$

Given ζ_j as above, we construct A_i and B_i from A_{i-1} and B_{i-1} as follows: for each $1 \le j \le |\mathcal{P}_i|$, we add x_j to A_{i-1} and y_j to B_{i-1} if $\zeta_j = 1$, and y_j to A_{i-1} and x_j to B_{i-1} if $\zeta_j = -1$. The claim follows.

If $b_k = 1$, notice that we have a partition of $V(H)$ into two sets A and B of equal size such that $|\deg(A) - \deg(B)| \le 1$. As $\deg(A) + \deg(B)$ is the sum of all the vertex degrees, we conclude that $\deg(A) = \deg(B)$ since $\deg(A) - \deg(B)$ must be even. $\square$

Lemma 2.3.1 tells us that a graph is splittable if we can find the right gadgets in the graph. The majority of the work in proving Theorem 2.1.1 is in showing that it is possible to find a good collection of gadgets.

2.4 Proof of the main result

We now try and make the intuition presented in Section 2.3 precise. We shall show that if $\varepsilon > 0$ and n is sufficiently large, it is possible to transform any graph G on n vertices into a splittable graph by deleting at most $2\varepsilon n$ vertices from G. Before we begin, we remark that the various constants suppressed by the asymptotic notation throughout the proof are allowed to depend on ε. We shall use $c_1, c_2, \dots$ to represent small constants depending on ε and $C_1, C_2, \dots$ for large constants depending on ε. All our estimates will hold when n is sufficiently large.

Proof of Theorem 2.1.1. Let $\varepsilon > 0$ be fixed. By deleting an arbitrary vertex of G if necessary, assume that $n = |V(G)|$ is even. Let $\beta = \beta(\varepsilon)$ be a small constant whose value we shall fix at the end of the argument in Case 1.

Call a pair of vertices (x, y) a 'large' pair if $\delta(x, y) \in [n^{1/3}, \beta n]$. Let $c_1 = \varepsilon/2$. We distinguish two cases depending on how many disjoint large pairs we can find in G. We first deal with the case when G contains many disjoint large pairs.

Case 1: G contains $c_1 n$ disjoint large pairs of vertices. In this case, we shall show that G has an induced subgraph H of even order on at least $(1 - 2\varepsilon)n$ vertices that contains

- a collection $\mathcal{S}_H$ of $[1, 1]$-gadgets of size $\Omega(n/\log n)$,
- a collection $\mathcal{M}_H$ of $[1, n^{2/3}]$-gadgets of size at least $2\beta n$, and
- a collection $\mathcal{L}_H$ of $[n^{1/9}, 2\beta n]$-gadgets of size $\Omega(n)$

such that the collections $\mathcal{S}_H$, $\mathcal{M}_H$, and $\mathcal{L}_H$ are disjoint. It is straightforward to check that such a graph H is splittable using Lemma 2.3.1. Indeed, pair up the vertices

$V(H) \setminus (L_H \cup M_H \cup S_H)$ arbitrarily; any such pair is a $[0,n]$-gadget and so we have a partition of $V(H)$ into disjoint collections of $[0,n]$-gadgets, $[n^{1/9}, 2\beta n]$-gadgets, $[1, n^{2/3}]$-gadgets and $[1,1]$-gadgets. The sizes of these collections satisfy the conditions of Lemma 2.3.1 if n is sufficiently large and it follows that H is splittable.

We shall now show that G does indeed contain such an induced subgraph H. We shall construct H by deleting vertices from G at random.

To avoid notational clutter, in the rest of the argument in Case 1, we shall write *large-gadget* for an $[n^{1/9}, 2\beta n]$-gadget, *medium-gadget* for a $[1, n^{2/3}]$-gadget and *one-gadget* for a $[1,1]$-gadget.

Let $\mathcal{L}$ be a collection of $c_1 n$ large pairs of vertices of G. The pairs in $\mathcal{L}$ will be the candidates for the large-gadgets we hope to find in H. Our next task is to find a large collection $\mathcal{M}$ of 'medium' pairs and a reasonably large collection $\mathcal{S}$ of 'small' pairs; the collections $\mathcal{M}$ and $\mathcal{S}$ will provide the candidate pairs for the medium-gadgets and one-gadgets that we would like to find in H.

Now, $|V \setminus L| = (1-2c_1)n$; recall that in our notation, L denotes the underlying ground set of $\mathcal{L}$. If we find more than $(1/2-\varepsilon)n$ disjoint clone pairs (x,y) in $G[V \setminus L]$, we are done. Indeed, we can delete all the other ($\leq 2\varepsilon n$) vertices not in any of these clone pairs to get a splittable graph: we split this graph by assigning different vertices of each clone pair to different halves of the partition. So we may assume that we can find a set $V' \subset V \setminus L$ of vertices of G such that any two vertices of V' disagree on some vertex of $V \setminus L$ and $|V'| \geq (2\varepsilon - 2c_1)n \geq \varepsilon n$.

Let $C_1 = 4/\varepsilon$ and let $c_2 = \varepsilon/12$. We now apply Proposition 2.2.3 to the degrees of the vertices of V'; by our choice of C_1 and c_2, we see that we can find $c_2 n$ disjoint groups of three vertices from V' such that $\delta(x,y) \leq C_1$ for any two vertices x and y in the same group. By Proposition 2.2.4, from each of these triples, we may choose a pair of vertices (x,y) such that $\Delta(x,y) \leq 2n/3$. Write $\mathcal{P}$ for this collection of $c_2 n$ pairs.

For $0 \leq i \leq \log n - 1$, let $\mathcal{P}_i$ be the collection of those pairs (x,y) in $\mathcal{P}$ such that

$\Delta(x,y) \in [2^i, 2^{i+1})$. There are two possibilities that we need to consider. It might be that no collection $\mathcal{P}_i$ contains too many pairs; we deal with this case next. The case where one of these collections contains many pairs is easier; we deal with this scenario later with a modification of the argument that follows.

Let $C_2 \geq 4$ be a (large) constant depending on ε; we shall fix the value of C_2 later in the proof at the end of Case 1A. Also, let $c_3 = c_2/3C_2 \leq c_2/12$.

Case 1A: None of the collections $\mathcal{P}_0, \mathcal{P}_1, \dots, \mathcal{P}_{\log n - 1}$ contains $c_3 n$ pairs. It is clear that at least one of the collections $\mathcal{P}_0, \mathcal{P}_1, \dots, \mathcal{P}_{\log n-1}$ contains at least $c_2 n/\log n$ pairs. Let k be the smallest index such that $|\mathcal{P}_k| \geq c_3 n/\log n$ and let us define our collection of small pairs $\mathcal{S}$ by setting $\mathcal{S} = \mathcal{P}_k$. We now define our collection of medium pairs $\mathcal{M}$ by setting

$$\mathcal{M} = \mathcal{P}_{k+C_2} \cup \cdots \cup \mathcal{P}_{\log n - 1}.$$

Since k is minimal and $c_3 \leq c_2/12$, we see that $|\mathcal{M}| \geq c_2 n/2$.

We shall now restrict our attention to the collections $\mathcal{S}$, $\mathcal{M}$ and $\mathcal{L}$; note that they are disjoint. We shall make use of the following facts about these collections.

- $\mathcal{S}$ contains $c_3 n/\log n$ pairs of vertices (x,y) such that $\delta(x,y) \leq C_1$, $\Delta(x,y) \in [2^k, 2^{k+1})$, and $\Delta(x,y) \leq 2n/3$.
- $\mathcal{M}$ contains $c_2 n/2$ pairs of vertices (x,y) such that $\delta(x,y) \leq C_1$, and $\Delta(x,y) \geq 2^{k+C_2}$.
- $\mathcal{L}$ contains $c_1 n$ pairs of vertices (x,y) with $\delta(x,y) \in [n^{1/3}, \beta n]$.
- For any pair of vertices (x,y) in $\mathcal{S}$ or $\mathcal{M}$, there exists at least once vertex in $V \setminus L$ on which x and y disagree.

We are now in a position to describe how we intend to construct a splittable graph from G. We shall delete vertices from G independently with a fixed probability. We shall show that with positive probability, many of the small pairs from $\mathcal{S}$ form one-gadgets in the resulting graph, many of the medium pairs from $\mathcal{M}$ form medium-gadgets, and many of the large pairs from $\mathcal{L}$ form large-gadgets in the resulting graph.

Fix $p = \min\{\varepsilon, 2^{-k}\}$. We now delete vertices from G independently with probability p. Let H be the resulting graph. We shall show that with probability $\Omega(1)$, the graph H is splittable and contains at least $(1-2\varepsilon)n$ vertices; this clearly implies the result we are trying to prove.

Note that for a graph to be splittable, it must necessarily contain an even number of vertices. With this in mind, let $\mathscr{E}$ be the event that an even number of vertices have been deleted, in other words, $\mathscr{E}$ is the event that $|V(H)|$ is even. By Proposition 2.2.7, we see that $\mathbb{P}(\mathscr{E}) \geq 1/2$. We now analyse what happens to the degree differences of the pairs in $\mathcal{S}$, $\mathcal{M}$ and $\mathcal{L}$ in the graph H.

ONE-GADGETS. We first show that many of the pairs in $\mathcal{S}$ form one-gadgets in H.

Lemma 2.4.1. *For any pair* $(x,y) \in \mathcal{S}$,

$$\mathbb{P}((x,y) \textit{ is a one-gadget in } H \mid \mathscr{E}) \geq f(\varepsilon, C_1) > 0.$$

The crucial fact about Lemma 2.4.1 is that the lower bound on the probability is independent of C_2.

Proof of Lemma 2.4.1. Let $A = \Gamma(x) \setminus (\Gamma(y) \cup \{y\})$ and $B = \Gamma(y) \setminus (\Gamma(x) \cup \{x\})$. Thus, $\delta(x,y) = ||A| - |B||$ and $\Delta(x,y) = |A| + |B|$. Note that since x and y disagree on at least one vertex of $V \setminus L$, it cannot be the case that both A and B are empty. Suppose without loss of generality that $|A| \geq |B|$ and that in particular, $A \neq \varnothing$.

Let E_1 be the event that both x and y are not deleted, E_2 the event that no vertices are deleted from B, E_3 the event that exactly $|\delta(x,y) - 1|$ vertices are deleted from A, and E_4 the event that the number of vertices deleted from $V \setminus (A \cup B \cup \{x,y\})$ has the same parity as $|\delta(x,y) - 1|$. It is obvious that the family $\{E_1, E_2, E_3, E_4\}$ is independent and it is easy to check that if $E_1 \cap E_2 \cap E_3 \cap E_4$ holds then (x,y) is a 1-gadget. Hence,

$$\mathbb{P}((x,y) \text{ is a 1-gadget in } H \,|\, \mathscr{E}) \geq \prod_{i=1}^{4} \mathbb{P}(E_i).$$

To complete our proof of the claim, we shall bound the factors on the right one by one. Clearly, $\mathbb{P}(E_1) \geq (1-\varepsilon)^2$.

We trivially have $|A|,|B| \leq 2^{k+1}$. Furthermore $|A|,|B| \geq 2^{k-1} - C_1/2$, since $0 \leq \delta(x,y) \leq C_1$. Also, we know that $\varepsilon 2^{-k} \leq p \leq 2^{-k}$. To bound $\mathbb{P}(E_2)$, first note that $p|B| \leq 2$. Now, $\mathbb{P}(E_2) = \mathbb{P}(\mathrm{Bin}(|B|,p) = 0)$ and so, by Proposition 2.2.5, $\mathbb{P}(E_2) \geq \exp(-4)$.

We now bound $\mathbb{P}(E_3)$. Clearly, $p|A| \leq 2$. If $2^k \geq 2C_1$, then $|A| \geq 2^{k-2}$ and so $p|A| \geq \varepsilon/4$. If $2^k \leq 2C_1$, then $p \geq \varepsilon 2^{-k} \geq \varepsilon/2C_1$ and so $p|A| \geq \varepsilon/2C_1$ since $|A| \geq 1$. Consequently,

$$\min\{\varepsilon/4, \varepsilon/2C_1\} \leq p|A| \leq 2.$$

Now, $\mathbb{P}(E_3) = \mathbb{P}(\mathrm{Bin}(|A|,p) = |\delta(x,y) - 1|)$. Using the above estimates for $p|A|$ and the fact that $0 \leq \delta(x,y) \leq C_1$ in Proposition 2.2.5, we see that $\mathbb{P}(E_3) = \Omega_{\varepsilon,C_1}(1)$.

Finally, since $\Delta(x,y) \leq 2n/3$, it follows that $|V \setminus (A \cup B)| \geq n/3$ and hence by Proposition 2.2.7, $\mathbb{P}(E_4) \geq 1/6$ for sufficiently large n. The claim follows. □

From Lemma 2.4.1 and Proposition 2.2.2 we see that, conditional on $\mathscr{E}$, the number of one-gadgets in H from $\mathcal{S}$ is $\Omega(n/\log n)$ with probability at least $f(\varepsilon, C_1)/2$; furthermore, and crucially, we note that this lower bound on the probability is independent of C_2.

MEDIUM-GADGETS. We next shift our attention to the pairs in $\mathcal{M}$.

Lemma 2.4.2. *For any pair $(x,y) \in \mathcal{M}$,*

$$\mathbb{P}(1 \leq \delta(x,y,H) \leq n^{2/3} \,|\, x,y \in V(H)) = 1 - o_{C_2\to\infty}(1) - o(1).$$

Proof. Let $N_1 = |\Gamma(x) \setminus (\Gamma(y) \cup \{y\})|$ and let $N_2 = |\Gamma(y) \setminus (\Gamma(x) \cup \{x\})|$ and suppose without loss of generality that $N_1 \geq N_2$. Note that $\delta(x,y) = |N_1 - N_2| \leq C_1$. Let X_1 and X_2 be independent random variables with distributions $\mathrm{Bin}(N_1, 1-p)$ and $\mathrm{Bin}(N_2, 1-p)$ respectively. Observe that $\delta(x,y,H)$ has the same distribution as $|X_1 - X_2|$.

We condition on $x, y \in V(H)$. Let E_1 be the event that $\delta(x,y,H) = 0$. Clearly, $\mathbb{P}(E_1) = \mathbb{P}(X_1 = X_2)$. Let E_2 denote the event that $\delta(x,y,H) \geq n^{2/3}$. It is enough to show that $\mathbb{P}(E_1 \cup E_2) = o_{C_2 \to \infty}(1) + o(1)$.

For any fixed values of p and N_2, it is not hard to check that $\mathbb{P}(X_1 = X_2)$ attains its maximum when $N_1 = N_2$. Thus $\mathbb{P}(E_1)$ is bounded above by the probability of two independent random variables with the distribution $\mathrm{Bin}(N_2, 1-p)$, or equivalently $\mathrm{Bin}(N_2, p)$, being equal. Now, $N_2 \geq 2^{k+C_2-1} - C_1/2$ and $p \geq \varepsilon 2^{-k}$. So, $pN_2 \geq \varepsilon 2^{C_2-1} - 2^{-k+1}$ which, since $k \geq 0$, means that $pN_2 \geq \varepsilon 2^{C_2-1} - 2$. As ε is fixed, we note that pN_2 can be made arbitrarily large by choosing C_2 large enough. Since $p \leq 1/2$, by Proposition 2.2.8, we see that $\mathbb{P}(E_1) = o_{C_2 \to \infty}(1)$.

Clearly $\mathbb{P}(E_2) = \mathbb{P}(|X_1 - X_2| \geq n^{2/3})$. Applying Proposition 2.2.10 to X_1 and X_2, we conclude that $\mathbb{P}(E_2) = O(\exp(-n^{1/3}/5))$. $\square$

Let $\mathcal{M}'$ be the collection of those pairs $(x,y) \in \mathcal{M}$ such that both x and y survive in H. Since the family of events $\{x, y \in V(H)\}$ is a family of mutually independent events for different pairs $(x,y) \in \mathcal{M}$ and since $\mathbb{P}(x, y \in V(H)) \geq (1-\varepsilon)^2$, it follows from Proposition 2.2.6 that $\mathbb{P}(|\mathcal{M}'| < (1-\varepsilon)^2|\mathcal{M}|/2) = \exp(-\Omega(n))$.

Consequently, from Lemma 2.4.2, it follows that for any pair $(x,y) \in \mathcal{M}$,

$$\mathbb{P}\left(1 \leq \delta(x,y,H) \leq n^{2/3} \,\middle|\, (x,y) \in \mathcal{M}',\, |\mathcal{M}'| > \frac{(1-\varepsilon)^2|\mathcal{M}|}{2}\right) = 1 - o_{C_2 \to \infty}(1) - o(1).$$

Thus by Proposition 2.2.2, conditional on $|\mathcal{M}'| > (1-\varepsilon)^2|\mathcal{M}|/2$, the number of medium-gadgets from $\mathcal{M}'$ in H is at least $|\mathcal{M}'|/3$ with probability $1 - o_{C_2 \to \infty}(1) - o(1)$. Thus, the number of medium-gadgets in H is at least $(1-\varepsilon)^2|\mathcal{M}|/6$ with probability

$(1-\exp(-\Omega(n)))(1-o_{C_2\to\infty}(1)-o(1))$, which is still $1-o_{C_2\to\infty}(1)-o(1)$.

Thus, conditional on the event $\mathscr{E}$, the number of medium-gadgets in H from $\mathcal{M}$ is $\Omega(n)$ with probability $1-o_{C_2\to\infty}(1)-o(1)$.

LARGE-GADGETS. We finally consider the pairs of vertices in $\mathcal{L}$. Recall that every pair $(x,y)\in\mathcal{L}$ is such that $\delta(x,y)\in[n^{1/3},\beta n]$ where β is a (small) constant whose value we have yet to fix. (Indeed, the value of β has so far played no role in our calculations.)

Lemma 2.4.3. *For any pair $(x,y)\in\mathcal{L}$,*

$$\mathbb{P}(n^{1/9}\le\delta(x,y,H)\le 2\beta n\,|\,x,y\in V(H))=1-o(1).$$

Proof. We condition on $x,y\in V(H)$. Let E_1 be the event that $\delta(x,y,H)<n^{1/9}$. Since $\delta(x,y)\ge n^{1/3}$, it follows immediately from Proposition 2.2.9 that $\mathbb{P}(E_1)=o(1)$.

Let E_2 be the event that $\delta(x,y,H)>2\beta n$. Let $A=\Gamma(x)\setminus(\Gamma(y)\cup\{y\})$ and $B=\Gamma(y)\setminus(\Gamma(x)\cup\{x\})$, and let X_1 and X_2 be random variables that denote the the number of vertices from A and B respectively which survive in H. Clearly, the distributions of X_1 and X_2 are $\mathrm{Bin}(|A|,1-p)$ and $\mathrm{Bin}(|B|,1-p)$ respectively.

If E_2 were to occur, i.e., it were the case that $|X_1-X_2|>2\beta n$, then this would imply that either $|X_1-(1-p)|A||\ge\beta n/2$ or $|X_2-(1-p)|B||\ge\beta n/2$, since $(1-p)||A|-|B||\le\delta(x,y)\le\beta n$. It follows that $\mathbb{P}(E_2)=o(1)$ since the probability of either of the above two possibilities is $\exp(-\Omega(n))$ by Proposition 2.2.6. □

Arguing as in the case of medium-gadgets, we see from Lemma 2.4.3 that conditional on the event $\mathscr{E}$, the number of large-gadgets in H from $\mathcal{L}$ is $\Omega(n)$ with probability $1-o(1)$.

CONSTRUCTING A SPLITTING. We now have a reasonably clear picture of what the degree differences in H of the pairs of vertices in $\mathcal{S}$, $\mathcal{M}$ and $\mathcal{L}$ look like. In summary, conditional on $\mathscr{E}$, we have demonstrated that in H, we can find

- a collection $\mathcal{S}_H$ of $\Omega(n/\log n)$ one-gadgets with probability $f(\varepsilon,C_1)/2$,

- a collection $\mathcal{M}_H$ of $\Omega(n)$ medium-gadgets with probability $1 - o_{C_2 \to \infty}(1) - o(1)$, and
- a collection $\mathcal{L}_H$ of $\Omega(n)$ large-gadgets with probability $1 - o(1)$

such that the collections $\mathcal{S}_H, \mathcal{M}_H$ and $\mathcal{L}_H$ are disjoint.

Thus by choosing C_2 to be a sufficiently large constant depending on ε, by the union bound, we find all of the above with probability $\Omega(1)$ conditional on $\mathscr{E}$, provided n is sufficiently large. Also, the expected number of vertices deleted is at most εn and so by Proposition 2.2.6, the probability that we have deleted more than $2\varepsilon n$ vertices is $\exp(-\Omega(n))$.

Consequently, we see that H, with probability $\Omega(1)$, has the aforementioned collections of gadgets, and furthermore, also has an even number of vertices and at least $(1-2\varepsilon)n$ vertices. We are done if we can guarantee that $2\beta n \leq |\mathcal{M}_H|$; this is possible if we choose $\beta = \beta(\varepsilon)$ to be a suitably small constant because $|\mathcal{M}_H| = \Omega(n)$.

We now consider the case where one of the sets $\mathcal{P}_i$ contains many pairs.

Case 1B: One of the sets $\mathcal{P}_0, \mathcal{P}_1, \ldots, \mathcal{P}_{\log n - 1}$ contains $c_3 n$ pairs. This case is easier to deal with than the former. We shall argue exactly as before; however we shall have no need of medium-gadgets and it will suffice to consider one-gadgets and large-gadgets alone.

Let k be any index such that $|\mathcal{P}_k| \geq c_3 n$ (while we chose k to be minimal previously, any index k such that $|\mathcal{P}_k| \geq c_3 n$ will do in this case). As before, we set $p = \min\{\varepsilon, 2^{-k}\}$ and $\mathcal{S} = \mathcal{P}_k$. We now delete vertices from G independently with probability p. Let H be the resulting graph; as before, we condition on deleting an even number of vertices. We claim that H is splittable with probability $\Omega(1)$.

It is not hard to check that Lemma 2.4.1 and Lemma 2.4.3 hold in this case as well. We conclude that we can delete an even number of vertices from G to obtain a graph H with $|V(H)| \geq (1-2\varepsilon)n$ in such a way that in H, we can find

- a collection $\mathcal{S}_H$ of $\Omega(n)$ one-gadgets, and

- a collection $\mathcal{L}_H$ of $\Omega(n)$ large-gadgets

such that $\mathcal{S}_H$ and $\mathcal{L}_H$ are disjoint. As before, it follows from Lemma 2.3.1 that H is splittable when n is sufficiently large provided $2\beta n \leq |\mathcal{S}_H|$; this is possible if we choose $\beta = \beta(\varepsilon)$ to be a suitably small constant because $|\mathcal{S}_H| = \Omega(n)$.

Thus, for sufficiently small β (chosen so as to satisfy the conditions from both Case 1A and 1B), we see that we are done if G contains many disjoint large pairs. Note that we have now fixed the value of β. We now deal with the case G does not contain many disjoint large pairs.

Case 2: G does not contain $c_1 n$ disjoint large pairs. In this case, we shall show that G has an induced subgraph H of even order on at least $(1-2\varepsilon)n$ vertices such that $V(H)$ may be partitioned into

- a collection $\mathcal{S}_H$ of $[1,1]$-gadgets of size $\Omega(n/\log n)$, and
- a collection $\mathcal{M}_H$ of $[0,n^{2/3}]$-gadgets.

In the rest of the argument in Case 2, we shall, as before, call $[1,1]$-gadgets *one-gadgets* and we denote $[0,n^{2/3}]$-gadgets (as opposed to $[1,n^{2/3}]$-gadgets as we did earlier) by *medium-gadgets*.

It is easily seen from Lemma 2.3.1 that H is splittable if n is sufficiently large. We construct our splitting by starting with the pairs in $\mathcal{M}_H$ - we can use these pairs to construct a partition such that sums of the degrees of the vertices of the two halves of the partition differ by at most $n^{2/3}$. We then use the the pairs in $\mathcal{S}_H$ to reduce the difference to at most one; we are done by parity considerations.

We now show how to find such a subgraph H. We start by describing how to find pairs of vertices which will be the candidates for the medium-gadgets we hope to find in H.

Let $\mathcal{L}$ be a maximal collection of large pairs in G. Note that since $\mathcal{L}$ is maximal, we have either $\delta(x,y) < n^{1/3}$ or $\delta(x,y) > \beta n$ for any two vertices $x,y \in V \setminus L$. As $\beta n > 2n^{1/3}$ for all sufficiently large n, there is a partition $V \setminus L = K_1 \cup K_2 \cup \cdots \cup K_m$ into 'clumps' K_i

with $m \leq 1/\beta$ in such a way that $\delta(x,y) < n^{1/3}$ for any $x, y \in K_i$ and $\delta(x,y) > \beta n$ if $x \in K_i$ and $y \in K_j$ with $i \neq j$.

We ignore the way in which vertices are originally paired in $\mathcal{L}$ and focus on the ground set L. By Proposition 2.2.3, we can find from L, at least $|L|/2 - n^{1/2}$ disjoint pairs (x,y) such that $\delta(x,y) \leq n^{1/2}$; call this collection of pairs Q.

Let F be the graph obtained from G as follows. Delete every vertex of $L \setminus Q$. Delete one vertex from every clump K which contains an odd number of vertices. Having done this, delete a clump K (i.e., delete all the vertices of K) if $|K| \leq n^{1/2}$.

Note that the vertex set of F consists of the surviving clumps, each of which has even size and cardinality at least $n^{1/2}$, and the (possibly empty) set of pairs Q. Since we had at most $1/\beta$ clumps initially, we have deleted $O(n^{1/2})$ vertices of total from G to obtain F. Hence, note that for any two vertices $x, y \in V(F)$, $|\delta(x,y,F) - \delta(x,y,G)| = O(n^{1/2})$. Hence, if either x and y both belong to the same (surviving) clump or if the pair (x,y) is in Q, then $\delta(x,y,F) = O(n^{1/2})$. Let us say that two vertices $x, y \in V(F)$ are *proximate* if either both x and y belong to the same clump in F or if $(x,y) \in Q$; these proximate pairs of vertices will be our candidates for medium-gadgets in H.

We now show how to find pairs of vertices which will be the candidates for the one-gadgets we hope to find in H. We shall henceforth work with F as opposed to G. We shall write V for $V(F)$ and all degrees and degree differences, unless specified otherwise, will be with respect to F.

Since $|\mathcal{L}| \leq c_1 n = \varepsilon n/2$ and since we have only deleted $O(n^{1/2})$ vertices so far, note that $|V \setminus Q| \geq (1 - 3\varepsilon/2)n$ for n sufficiently large.

If we find at least $(1/2 - \varepsilon)n$ disjoint clone pairs (x,y) in $F[V \setminus Q]$, we are done. So we may assume that we can find a set $V' \subset V \setminus Q$ of vertices of F with $|V'| \geq (2\varepsilon - 3\varepsilon/2)n = \varepsilon n/2$ such that any two vertices of V' disagree on some vertex in $V \setminus Q$.

We claim that if C_3 is sufficiently large (as a function of β), then we can find from any subset of C_3 vertices of V', two vertices x and y such that for each clump K, the number of vertices of K on which x and y disagree is at most $2|K|/3$. To see this, suppose that we have found C_3 vertices such that any two of them x and y disagree on more than two thirds of some clump $K_{x,y}$. Applying Ramsey's theorem to the complete graph on these C_3 vertices where the edge between x and y is labelled by the clump $K_{x,y}$, we see that we can find a monochromatic triangle provided C_3 is large enough. But by Proposition 2.2.4, out of any three vertices, at least two disagree on at most two thirds of the vertices of K. We have a contradiction.

Choose C_3 as described above and set $C_4 = 4C_3/\varepsilon$ and $c_4 = \beta/2C_4$. By Proposition 2.2.3, we can find from V', at least n/C_4 disjoint groups of size C_3 such that that $\delta(x,y) \leq C_4$ for any two vertices x and y in the same group. From each of these n/C_4 groups of size C_3, choose a pair of vertices (x,y) such that x and y disagree on at most two thirds of every clump. Choose a clump K^* such that at least a β fraction of these pairs (x,y) are such that x and y disagree on at least one vertex in K^*; this is possible because any two vertices of V' disagree on $V(F) \setminus Q$ and consequently, on at least one clump and furthermore, there are at most $1/\beta$ clumps. Let $\mathcal{P}$ be this collection of pairs which all disagree on at least one vertex in K^*; clearly $|\mathcal{P}| \geq \beta n/C_4 = 2c_4 n$.

We shall proceed as in Case 1 by pigeonholing the pairs in $\mathcal{P}$ into different boxes based on the size of their difference neighborhoods, but with one important difference. Note that while *any two* vertices in the same clump have a small ($O(n^{1/2})$) degree difference, we can only guarantee that two vertices of Q have small ($O(n^{1/2})$) degree difference if the pair belongs to $\mathcal{Q}$. Consequently, when we later delete vertices at random, we shall either delete both vertices of a pair in $\mathcal{Q}$ or retain both; hence we shall treat a pair of vertices in $\mathcal{Q}$ as a single vertex when it comes to pigeonholing the pairs in $\mathcal{P}$. This is made precise below.

Let $F_{\mathcal{Q}}$ be the multigraph without loops obtained from F by contracting every pair (x,y) in $\mathcal{Q}$ (we ignore the loops that might arise). Note that there are at most two parallel edges between any two vertices of F_Q. In F_Q, we say that two vertices x and y disagree on a vertex $v \neq x, y$ if the number of edges between v and x is not equal to the number of edges between v and y. For $0 \leq i \leq \log n - 1$, let $\mathcal{P}_i$ be the collection of those pairs (x, y) in $\mathcal{P}$ such that $\Delta(x, y, F_Q) \in [2^i, 2^{i+1})$ where $\Delta(x, y, F_Q)$ is the number of vertices of F_Q on which x and y disagree.

As before, let k be any index such that $|\mathcal{P}_k| \geq 2c_4 n / \log n$; take $\mathcal{S} = \mathcal{P}_k$ and set $p = \min\{\varepsilon, 2^{-k}\}$.

In summary, $\mathcal{S}$ consists of pairs (x, y) such that

- x and y disagree on at most two thirds of every clump,
- x and y disagree on at least one vertex of K^*,
- $\delta(x, y) \leq C_4$, and
- $\Delta(x, y, F_Q) \in [2^k, 2^{k+1})$.

Furthermore, since $\delta(x, y) \leq C_4 = o(n^{1/2})$ for any $(x, y) \in \mathcal{S}$, both members of any pair in $\mathcal{S}$ must belong to the same clump.

Consider the partition $\mathcal{S} = \mathcal{S}_o \cup \mathcal{S}_e$ where $\mathcal{S}_o$ is the set of those pairs $(x, y) \in \mathcal{S}$ such that $\delta(x, y)$ is odd. Recall that $|\mathcal{S}| \geq 2c_4 n / \log n$ and so one of $\mathcal{S}_o$ or $\mathcal{S}_e$ contains more than $c_4 n / \log n$ pairs. At this point, we need slightly different arguments depending on whether we have more pairs with odd degree difference or even degree difference in $\mathcal{S}$.

Case 2A: $\mathcal{S}$ contains many odd pairs. We first consider the case where $|\mathcal{S}_o| \geq c_4 n / \log n$. We shall delete vertices from F as follows. We pick vertices of F_Q independently with probability p. For every vertex of F_Q that we pick, we delete (as appropriate) either the corresponding vertex or both vertices of the corresponding pair of vertices from Q in F_Q. Let H be the resulting graph. Our aim is to show that H is splittable with probability $\Omega(1)$.

Earlier, we conditioned on deleting an even number of vertices from G. In this case, we need a little more. Let $\mathcal{E}^*$ be the event that an even number of vertices were deleted from each clump. By Proposition 2.2.7, we see that $\mathbb{P}(\mathscr{E}^*) \geq (1/2)^{1/\beta}$. Note that a consequence of $\mathscr{E}^*$ is that $|V(H)|$ is even.

ONE-GADGETS. First, we shall show that many of the pairs in $\mathcal{S}_o$ become one-gadgets in H.

Lemma 2.4.4. *For any pair* $(x,y) \in \mathcal{S}_o$*,*

$$\mathbb{P}((x,y) \textit{ is a one-gadget in } H \mid \mathscr{E}^*) = \Omega(1).$$

Proof. In F_Q, let A be the set of those vertices $v \neq x,y$ such that number of edges between v and x is more than the number of edges between v and y and let B be defined analogously by interchanging x and y. Let $A = A_1 \cup A_2$ where A_1 and A_2 are respectively those vertices v in A such that the number of edges between v and x is one, respectively two, more than the number of edges from v to y; define B_1 and B_2 analogously.

The proof follows that of Lemma 2.4.1. Clearly,

$$2^k \leq |A_1| + |A_2| + |B_1| + |B_2| < 2^{k+1}.$$

Furthermore, $\delta(x,y) = ||A_1| + 2|A_2| - |B_1| - 2|B_2||$ and so,

$$-C_4 \leq |A_1| + 2|A_2| - |B_1| - 2|B_2| \leq C_4.$$

Using the above two inequalities, it is not hard to check that

$$\max\{|A_1|,|A_2|\}, \max\{|B_1|,|B_2|\} \geq 2^{k-3} - C_4/4.$$

Since $\delta(x,y)$ is odd, suppose without loss of generality that $\deg(x) > \deg(y)$. Let E_1 be the event that both x and y are not picked to be deleted, E_2 the event that no vertices are picked from B, E_3 the event that $X_1 + 2X_2 = \delta(x,y) - 1$ where X_1 and X_2 are the number of vertices picked from A_1 and A_2 respectively, and E_4 the event that the number of vertices picked from $K \setminus (A \cup B \cup \{x,y\})$ has the same parity as the number of vertices picked from $K \cap (A \cup B \cup \{x,y\})$ for every clump K. The collection of events $\{E_1, E_2, E_3\}$ is clearly independent, and it is easy to check that

$$\mathbb{P}((x,y) \text{ is a one-gadget} \mid \mathscr{E}^*) \geq \mathbb{P}(E_1)\mathbb{P}(E_2)\mathbb{P}(E_3)\mathbb{P}(E_4|E_1,E_2,E_3).$$

Clearly, $\mathbb{P}(E_1) \geq (1-\varepsilon)^2$. As in Lemma 2.4.1, note that $p|B| \leq 2$ and so, by Proposition 2.2.5, $\mathbb{P}(E_2) \geq \exp(-4)$.

We now bound $\mathbb{P}(E_3)$. First suppose that $2^{k-3} - C_4/4 > C_4$. Recall that $\delta(x,y)$ is odd. If $|A_2| \geq |A_1|$, we consider the event that $(\delta(x,y)-1)/2$ vertices are picked from A_2 and no vertices are picked from A_1 in F_Q; as in Lemma 2.4.1, we see that $p|A_2| = \Theta(1)$ and so this event occurs with probability $\Omega(1)$. Hence E_3 occurs with probability $\Omega(1)$. If $|A_1| > |A_2|$, we consider the event that $\delta(x,y)-1$ vertices are picked from $|A_1|$ and no vertices are picked from A_2 and note that this event occurs with probability $\Omega(1)$ and hence E_3 occurs with probability $\Omega(1)$.

If, on the other hand, $2^{k-3} - C_4/4 \leq C_4$, then clearly $k = \Theta(1)$ and hence $p, |A_1|, |A_2|$ are all $\Theta(1)$. In this case, we consider the event that $t = \min\{(\delta(x,y)-1)/2, |A_2|\}$ vertices are picked from A_2 and $\delta(x,y)-1-2t$ vertices are picked from A_1. Now, $|A_1| + 2|A_2| \geq \delta(x,y)$ since we assumed that $\deg(x) > \deg(y)$ and so $|A_1| \geq \delta(x,y) - 1 - 2t$. Also, as noted above, $p, |A_1|, |A_2|$ are all $\Theta(1)$. So this event occurs with probability $\Omega(1)$. Hence the event E_3 occurs with probability $\Omega(1)$.

Since x and y disagree on at most two thirds of every clump and since every clump has size at least $n^{1/2}$, it follows from Proposition 2.2.7 that $\mathbb{P}(E_4|E_1,E_2,E_3) \geq (1/6)^{1/\beta}$ for sufficiently large n. $\square$

Let $\mathcal{S}_H$ be the set of pairs from $\mathcal{S}_o$ that form one-gadgets in H. From Lemma 2.4.4 and Proposition 2.2.2, we see that conditional on $\mathscr{E}^*$, $|\mathcal{S}_H| \geq \mathbb{E}[|\mathcal{S}_H|]/2 = \Omega(n/\log n)$ with probability $\Omega(1)$.

MEDIUM-GADGETS. We now show that the degree difference of any pair of vertices which are proximate in F cannot become too large in H.

Lemma 2.4.5. *Conditional on $\mathscr{E}^*$ and $|\mathcal{S}_H| \geq \mathbb{E}[|\mathcal{S}_H|]/2$, the probability that there exist $x, y \in V(H)$ which are proximate in F and satisfy $\delta(x,y,H) > n^{2/3}$ is $o(1)$.*

Proof. Recall that for any two vertices x and y which are proximate in F, $\delta(x,y) = O(n^{1/2})$. For such a pair of vertices x and y, note by Proposition 2.2.10 that

$$\mathbb{P}(\delta(x,y,H) > n^{2/3} \mid x, y \in V(H)) = O(\exp(-n^{1/3}/5)).$$

Consequently, since we have conditioned on an event with probability $\Omega(1)$, the probability that there exist some vertices $x, y \in V(H)$ such that x and y are proximate and $\delta(x,y,H) > n^{2/3}$ is $O(n^2 \exp(-n^{1/3}/5)) = o(1)$. □

CONSTRUCTING A SPLITTING. We now describe how to construct a splitting of H. Let $\mathcal{Q}_H$ be the set of pairs from $\mathcal{Q}$ that survive in H. For a clump K in F, let K_H denote the set $(K \setminus S_H) \cap V(H)$. Clearly $V(H)$ is the disjoint union of S_H, $\mathcal{Q}_H$ and the clumps K_H. Note that conditional on $\mathscr{E}^*$, the size of K_H is even for every clump K since both members of any pair in $\mathcal{S}_H$ must necessarily belong to the same clump. Since each K_H has even cardinality, we may group the vertices of each K_H into pairs. Pair up the vertices in each K_H arbitrarily; let $\mathcal{M}_H$ be the collection consisting of these pairs and the pairs in $\mathcal{Q}_H$. Clearly, every pair of vertices in $\mathcal{M}_H$ are proximate in F and by Lemma 2.4.5, the probability that there exists some pair $(x,y) \in \mathcal{M}_H$ with $\delta(x,y,H) > n^{2/3}$ is $o(1)$.

The expected number of vertices deleted from F is at most εn and the number of vertices deleted from G to obtain F is $O(n^{1/2})$. Hence, by Proposition 2.2.6, the probability that we have deleted more than $2\varepsilon n$ vertices from G is $\exp(-\Omega(n))$.

We conclude that there exists an induced subgraph H of G such that $|V(H)| \geq (1-2\varepsilon)n$, and with the further property that $V(H)$ may be partitioned into

- a collection $\mathcal{S}_H$ of one-gadgets of size $\Omega(n/\log n)$, and
- a collection $\mathcal{M}_H$ of medium-gadgets.

It follows from Lemma 2.3.1 that H is splittable and we are done.

Case 2B: $\mathcal{S}$ contains many even pairs. Now we consider the case where $|\mathcal{S}_e| \geq c_4 n/\log n$.

Note that since we intend to delete either both vertices of a pair in Q or neither, it might be the case that it is impossible to make the parity of the degree difference of a pair in $\mathcal{S}_e$ odd in H. Consequently, in this case, we will need to work with $[2,2]$-gadgets, or *two-gadgets* for short, in addition to one-gadgets. With the exception of this slight change of tack to account for parity considerations, the argument is quite similar to the one in the previous case, and we only sketch it.

Let c_5 be a (small) constant depending on ε; the value of c_5 will be chosen later, following the statement of Lemma 2.4.6.

Recall that every pair of vertices in $\mathcal{S}_e$ disagree on some vertex in the clump K^*. Suppose there exists a vertex $v \in K^*$ such that $c_5 n/\log n$ pairs from $\mathcal{S}_e$ all disagree on v. In this case, we may complete the proof as follows. Let $\mathcal{S}_v \subset \mathcal{S}_e$ be the collection of pairs in $\mathcal{S}_e$ that disagree on v. We shall delete vertices from F as follows. We first delete v and then delete one other vertex uniformly at random from K^*. Following this, we proceed as before by picking vertices of F_Q independently with probability p and then deleting the corresponding vertices or pairs of vertices from Q in F. Let H be the resulting graph. Note that when we delete v, the degree difference of every pair in $\mathcal{S}_v$ changes parity and becomes odd. When we then delete another vertex uniformly at random from K^*, the parity of the degree difference of a pair in $\mathcal{S}_v$ is unaltered with probability at least $1/3$ since every pair in $\mathcal{S}$ disagree on at most two thirds of any clump. Arguing as in Lemma 2.4.4, for any pair in $\mathcal{S}_v$, we see that the probability that this pair forms a one-gadget in H, conditional on deleting an even number of vertices from every clump, is

$\Omega(1)$ (albeit with a smaller constant than in Case 2A). Since $|\mathcal{S}_v| \geq c_5 n/\log n$, we can conclude the proof exactly as in the case where $\mathcal{S}$ contains many odd pairs.

Hence we may assume that for every vertex $v \in K^*$, the number of pairs in $\mathcal{S}_e$ that disagree on v is at most $c_5 n/\log n$. We delete vertices from F as before by picking vertices of F_Q independently with probability p and then deleting the corresponding vertices or pairs of vertices from Q in G. Let H be the resulting graph.

As before, let $\mathcal{E}^*$ be the event that an even number of vertices were deleted from each clump. The proof of Lemma 2.4.4, with minor modifications for the change in parity, yields a proof of the following lemma.

Lemma 2.4.6. *For any* $(x,y) \in \mathcal{S}_e$, $\mathbb{P}((x,y)\,\textit{is a two-gadget in}\ H \,|\, \mathcal{E}^*) = \Omega(1)$. □

Let $\mathcal{S}_H$ be the collection of pairs from $\mathcal{S}_e$ that form two-gadgets in H. From Lemma 2.4.6 and Proposition 2.2.2, we see that there exists a small positive constant c_6 such that, conditional on $\mathcal{E}^*$, $|\mathcal{S}_H| \geq c_6 n/\log n$ with probability $\Omega(1)$. Let us now fix $c_5 = c_6/4$.

CONSTRUCTING A SPLITTING. As before, let Q_H be the collection of pairs from Q that survive in H and for each clump K in F, let us write K_H for the set $(K \setminus S_H) \cap V(H)$.

We have shown that with probability $\Omega(1)$, the graph H is such that

(i) $|K_H|$ is even for every clump K, and

(ii) $|\mathcal{S}_H| \geq c_6 n/\log n$.

Consider any pair $(x,y) \in \mathcal{S}_H$ and note that in F, x and y disagree on at most two thirds of any clump; in particular, x and y agree on at least a third of K^*. Consequently, the probability that x and y disagree on every vertex of K^*_H is $\exp(-\Omega(n^{1/2}))$. Hence, with probability $1-o(1)$, for every $(x,y) \in \mathcal{S}_H$, there exists some vertex in K^*_H on which x and y agree.

Next, it follows from Lemma 2.4.5 that with probability $1-o(1)$, any two vertices $x,y \in V(H)$ which are proximate satisfy $\delta(x,y,H) \leq n^{2/3}$. Finally, the probability that we have deleted more that $2\varepsilon n - 2$ vertices of total from G is, by Proposition 2.2.6, $\exp(-\Omega(n))$. It follows that with probability $\Omega(1)$, the graph H, in addition to possessing the aforementioned properties, also has the following properties.

(iii) For every $(x,y) \in \mathcal{S}_H$, there exists some vertex in K_H^* on which x and y agree.

(iv) For any $x,y \in V(H)$ such that x and y are proximate in F, $\delta(x,y,H) \le n^{2/3}$.

(v) $|V(H)| \ge (1-2\varepsilon)n + 2$.

With a view of making the graph H splittable, we alter H as follows. Fix a pair $(x^*,y^*) \in \mathcal{S}_H$ and a vertex $v \in K^*$ on which x^* and y^* disagree. We know that there is a vertex $u \in K_H^*$ on which x^* and y^* agree. Delete u from H. If $v \in V(H)$, delete v from H and if $v \notin V(H)$, add v back. After these alterations, note that H still has an even number of vertices. Note also that now, $|V(H)| \ge (1-2\varepsilon)n$ and $\delta(x^*,y^*,H) \in \{1,3\}$.

Before we altered H, at most $c_5 n/\log n$ pairs in $\mathcal{S}_H$ disagreed on any vertex in K^*; the alterations above change the degree differences of at most $2c_5 n/\log n = c_6 n/2\log n$ pairs in $\mathcal{S}_H$. Hence, H contains a collection $\mathcal{S}_H$ of least $c_6 n/2\log n - 1$ pairs of vertices (x,y) such that $\delta(x,y,H) = 2$ and a pair (x^*,y^*) such that $\delta(x^*,y^*,H) \in \{1,3\}$. Furthermore, all the vertices of $V(H) \setminus (S_H \cup \{x^*,y^*\})$ may be grouped into pairs (x,y) such that $\delta(x,y,H) \le n^{2/3} + 2$; let $\mathcal{M}_H$ denote this collection of pairs.

It is now easy to check that H is splittable using the argument used to prove Lemma 2.3.1. Indeed, we can use pairs in $\mathcal{M}_H$ to construct a partition such that sums of the degrees of the vertices of the two halves of the partition differ by at most $n^{2/3} + 2$. For n sufficiently large, we can then reduce the difference to at most two by using all but one of the pairs in $\mathcal{S}_H$. Finally, using the one remaining pair in $\mathcal{S}_H$ and the pair (x^*,y^*), we can reduce the difference to at most one; we are done constructing a splitting of H by parity considerations. This completes the proof of Theorem 2.1.1. $\square$

2.5 Open problems

We have shown that $f(n) \geq n/2 - o(n)$. In fact, it should be possible to read out a bound of $f(n) \geq n/2 - n/(\log\log n)^c$ from our proof for some absolute constant $c > 0$; we chose not to include a proof of this fact to keep the presentation simple, and because we do not believe that such a bound is close to the truth. While we have managed to pin down f up to its first order term, there is still a large gap between the upper and lower bounds for $n/2 - f(n)$.

Problem 2.5.1. *What is the asymptotic behaviour of $n/2 - f(n)$?*

We know that $n/2 - f(n) = \Omega(\log\log n)$ and $n/2 - f(n) = o(n)$; we suspect that the truth lies closer to the lower bound and that in particular, $n/2 - f(n) = o(n^{\varepsilon})$ for every $\varepsilon > 0$. Indeed, it is not inconceivable that $n/2 - f(n) = \Theta(\log n)$.

It is natural to generalize the problem to the case where we have more than one type of edge, or ask for more than two disjoint subgraphs. For any $r, l \in \mathbb{N}$, given an edge coloring Δ of the complete graph on n vertices with r colors, let $g(\Delta)$ be the largest integer k for which we can find l disjoint subsets $V_1, V_2, \dots, V_l$ of $[n]$, each of cardinality k, such that for each $1 \leq i \leq r$, the number of edges induced by V_j of color i is the same for every $1 \leq j \leq l$. Let $g(n, r, l)$ be the minimum value of $g(\Delta)$ taken over all edge colorings of the complete graph on n vertices. In particular, note that $g(n, 2, 2) = f(n)$. We conjecture that $g(n, r, 2) = n/2 - o(n)$ and more generally, ask the following question.

Problem 2.5.2. *For $r, l \in \mathbb{N}$, what is the asymptotic behaviour of $g(n, r, l)$?*

Finally, we mention a question about digraphs that we find particularly appealing. Given a digraph D on n vertices, let $h(D)$ denote the largest integer k for which there exist disjoint subsets $A, B \subset V$ such that $|A| = |B| = k$ and the number of directed edges from A to B is equal to the number of directed edges from B to A. Let $h(n)$ be the minimum value of $h(D)$ taken over all digraphs on n vertices.

Problem 2.5.3. *Determine* $h(n)$.

CHAPTER 3

SEPARATING PATH SYSTEMS

3.1 Introduction

Given a set S, we say that a family $\mathcal{F}$ of subsets of S *separates* a pair of distinct elements $x,y \in S$ if there exists a set $A \in \mathcal{F}$ which contains exactly one of x and y. If $\mathcal{F}$ separates all pairs of distinct elements of S, we say that $\mathcal{F}$ is a *separating system* of S.

The study of separating systems was initiated by Rényi [58] in 1961. It is essentially trivial that the minimal size of a separating system of an n-element set is $\lceil \log_2 n \rceil$. However, the question of finding the minimal size of a separating system becomes much more interesting when one imposes restrictions on the elements of the separating system. For example, separating systems with restrictions on the cardinalities of their members have been studied by Katona [39], Wegener [66], Ramsay and Roberts [55] and Kündgen, Mubayi and Tetali [43], amongst others. Stronger notions of separation as well as other extremal questions about separating systems have also been studied; see, for example, the papers of Spencer [59], Hansel [35], and Bollobás and Scott [11].

Another interesting direction involves imposing a graph structure on the underlying ground set and imposing graph theoretic restrictions on the separating family (see, for instance, [16, 12]). In this chapter, we investigate the question of separating the edges of a graph using paths. Given a graph $G = (V,E)$, we say that a family $\mathcal{P}$ of subsets of the edge set $E(G)$ is a *separating path system of* G if $\mathcal{P}$ separates $E(G)$ and every element of $\mathcal{P}$ is a path of G. The analogous question of separating the vertices of a graph with paths has also been studied; we refer the reader to [29] for details.

Separating path systems arise naturally in the context of network design (see, for instance, [1, 36, 61]). We are presented with a communication network with one (and at most one) defective link and our goal is to identify this link. Of course, one could test

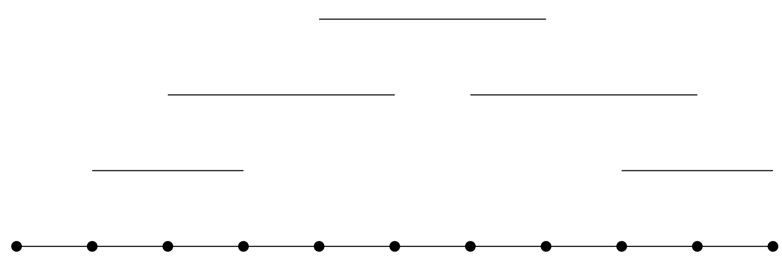

Figure 3.1: A path on 11 vertices and a separating path system with 5 paths.

every link, but this is not very efficient; can we do better? A natural test to perform is to send a message between a pair of nodes along a predetermined path; if the message does not reach its intended destination, we conclude that the defective link lies on this path. If we model the communication network as a graph, a fixed set of such tests succeeds in locating any defective link if and only if the corresponding family of paths is a separating path system of the underlying graph. We are naturally led to the following question: what is the size of a minimal separating path system of a given graph?

For a graph G, let $f(G)$ be the size of a minimal separating path system of G. As a separating path system of G is also a separating system of $E(G)$, it follows that $f(G) \geq \lceil \log_2 |E(G)| \rceil$. In particular, for any connected n-vertex graph G, $f(G) = \Omega(\log n)$. With a little work, we can construct graphs that come close to matching this bound. Let L_n be the *ladder* of order $2n$, that is, the Cartesian product of a path of length $n-1$ with a single edge. Given any subset A of $[n-1]$, there is (see Figure 3.2) a natural way of mapping A to a path P_A in L_n. With this, it is an easy exercise to establish that $f(L_n) = O(\log n)$; indeed, the top and bottom paths split the edges into three groups: top, bottom and vertical edges. The collection of paths P_A separates pairs of edges on the top path and also pairs of edges on the bottom path. Lastly, we can separate vertical edges by taking paths that correspond naturally to subsets of those edges.

A more interesting problem is to determine $f(n)$, the *maximum* of $f(G)$ taken over all n-vertex graphs; this question was raised by Gyula Katona in August 2013 at the 5$^{\text{th}}$

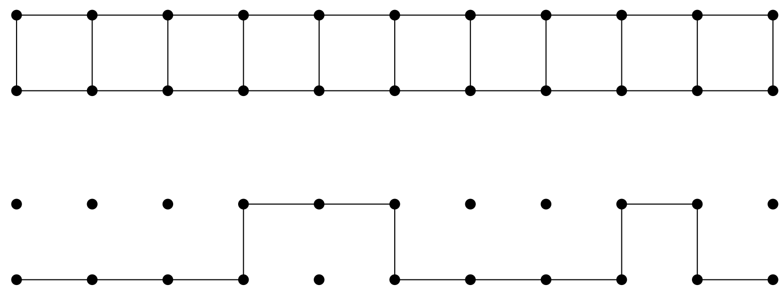

Figure 3.2: The graph L_{11} and the path P_A corresponding to the subset $A = \{4,5,9\}$.

Emléktábla Workshop in Budapest.

Clearly, at most one edge of a graph can be left uncovered by the paths of a separating path system of the graph; it is thus unsurprising that the question of building small separating path systems is closely related to the well-studied question of covering a graph with paths. It would be remiss not to remind the reader of a beautiful conjecture of Gallai which asserts that every connected graph on n vertices can be decomposed into $\lfloor (n+1)/2 \rfloor$ paths. The following fundamental result of Lovász [49], which provides support for Gallai's conjecture, will prove useful; here and elsewhere, by a decomposition of a graph we mean a covering of its edges with edge disjoint subgraphs.

Theorem 3.1.1. *Every n-vertex graph can be decomposed into at most $n/2$ paths and cycles. Consequently, every n-vertex graph can be decomposed into at most n paths.* □

Let G be any graph on n vertices and let $E_1, E_2, \ldots, E_k$ be a separating system of the edge set $E(G)$ where $k = \lceil \log_2 |E(G)| \rceil \leq 2\log_2 n$. Let G_i be the subgraph of G induced by the edges of E_i. By Theorem 3.1.1, each G_i may be decomposed into at most n paths. Putting these together, we get a separating path system of G of cardinality at most kn. Consequently, we note that $f(n) \leq 2n\log_2 n$.

To bound $f(n)$ from below, let us consider K_n, the complete graph on n vertices. Suppose that we have a separating path system $\mathcal{P}$ of K_n with k paths. Note that at most one edge of K_n goes uncovered by the paths of $\mathcal{P}$ and furthermore, at most k edges of K_n

belong to exactly one path of $\mathcal{P}$. Since any path of K_n has length at most $n-1$, we deduce that

$$k(n-1) \geq 1+k+2\left(\binom{n}{2}-k-1\right)$$

or equivalently, $k \geq n-1-1/n$. Thus, we note that $f(n) \geq n-1$. We believe that the lower bound, rather than the upper bound, is closer to the truth; we make the following conjecture.

Conjecture 3.1.2. *There exists an absolute constant C such that for every graph G on n vertices, $f(G) \leq Cn$.*

Let us remark that it is not inconceivable that $f(n) = (1+o(1))n$ and Conjecture 3.1.2 is true for every $C > 1$. In this chapter, we shall prove Conjecture 3.1.2 in certain special cases. Our first result establishes the conjecture for graphs of linear minimum degree.

Theorem 3.1.3. *Let $c > 0$ be fixed. Every graph G on n vertices with minimum degree at least cn has a separating path system of cardinality at most $122n/c^2$ for all sufficiently large n.*

Building upon the ideas used to prove Theorem 3.1.3, we shall prove Conjecture 3.1.2 for the Erdős-Rényi random graphs using the fact that these graphs have good connectivity properties.

Theorem 3.1.4. *For any probability $p = p(n)$, with high probability, the random graph $G(n, p)$ has a separating path system of size at most $48n$.*

Note in particular that Theorem 3.1.4 implies that Conjecture 3.1.2 is true for almost all n-vertex graphs with $C = 48$. Using Theorem 3.1.3, we shall also establish the conjecture for a class of dense graphs, which includes quasi-random graphs (in the sense of Chung, Graham and Wilson [19] and Thomason [63]) as a subclass.

Theorem 3.1.5. *Let $c > 0$ be fixed and let G be a graph on n vertices such that every subset $U \subseteq V(G)$ of size at least $\sqrt{n}$ spans at least $c|U|^2$ edges. Then $f(G) \le 638n/c^3$ for all sufficiently large n.*

The above results are far from best possible but we make no attempt to optimise our bounds since it seems unlikely that our methods will yield the best possible constants. However in the case of trees, we are able to obtain tight bounds.

Theorem 3.1.6. *Let T be a tree on $n \ge 4$ vertices. Then*

$$\left\lceil \frac{n+1}{3} \right\rceil \le f(T) \le \left\lfloor \frac{2(n-1)}{3} \right\rfloor.$$

Furthermore, these bounds are best possible.

We use standard graph theoretic notions and notation and refer the reader to [8] for terms and notation not defined here. We shall also make use of some well known results about random graphs without proof; see [9] for details.

The rest of this chapter is organised as follows. In the next section, we describe a general strategy that we adopt to prove Theorems 3.1.3 and 3.1.4. We then prove Theorems 3.1.3 and 3.1.4 in Sections 3.3 and 3.4 respectively. Section 3.5 is devoted to the proof of Theorem 3.1.5. We then prove Theorem 3.1.6 in Section 3.6. We conclude the chapter in Section 3.7 with a discussion of related questions and problems. For the sake of clarity, we systematically omit floors and ceilings in Sections 3.3, 3.4 and 3.5.

3.2 A general strategy

Theorems 3.1.3 and 3.1.4 are proved similarly, using the following strategy. Let G_1 and G_2 be subgraphs of G which partition the edge set of G. First, we decompose the edges of G_1 into at most $3n$ matchings $M_1, \ldots, M_{3n}$ as follows. Initially, each M_i is empty; we add the edges of G_1 one by one to a suitably chosen matching M_i. By Theorem 3.1.1, there exists a path decomposition of G_1 into (at most) n paths $P_1, \ldots, P_n$. Given an edge $e = xy \in E(G_1)$, let j be such that $e \in P_j$. We add e to a matching M_i which contains no edge of P_j and no edge incident to x or y. As the length of P_j is at most $n-1$ and there are at most $2n$ edges incident to either x or y, this process is well defined; indeed, we can always find a matching M_i satisfying the required conditions. Note that we have ensured that $|M_i \cap P_j| \leq 1$ for each $1 \leq i \leq 3n$ and $1 \leq j \leq n$.

Next, for each $1 \leq i \leq 3n$, we find a covering of $E(M_i)$ with paths using edges from $E(G_2) \cup E(M_i)$. These covering paths together with the paths $P_1, \ldots, P_n$ separate the edges of G_1 from each other and from the edges of G_2. To check this, consider an edge $e \in E(G_1)$ such that $e \in P_j$ and $e \in M_i$ and note that P_j separates e from every edge of G_2 as well as each edge of $E(G_1) \setminus P_j$, while the path covering M_i separates e from every other edge of P_j since $|M_i \cap P_j| \leq 1$. Repeating this process with the roles of G_1 and G_2 reversed, we obtain a separating path system of G.

In order to prove the existence of a small separating path system, we shall partition the graph G into G_1 and G_2 in a way that will enable us to keep the cardinalities of the above coverings small.

3.3 Graphs of linear minimum degree

Proof of Theorem 3.1.3. Let G be a graph on n vertices with minimum degree at least cn, for some fixed $0 < c < 1$. It is easy to decompose G into two disjoint subgraphs G_1 and G_2 in such a way that both subgraphs have minimum degree at least $cn/3$. Indeed, one way to do this is to define G_1 by randomly selecting each edge of G with probability $1/2$ and to take G_2 to be the complement of G_1 in G, i.e., $V(G_2) = V(G)$ and $E(G_2) = E(G) \setminus E(G_1)$; the minimum degree conditions follow from the standard estimates for the tail of the binomial distribution.

Following the strategy described in Section 3.2, let $P_1, \ldots, P_n$ be a path decomposition of G_1 and let $M_1, \ldots, M_{3n}$ be a decomposition of G_1 into matchings such that the intersection $M_i \cap P_j$ contains at most one edge for each i and j.

Define a graph H on $V(G)$ as follows: two distinct vertices $x, y \in V(G)$ are adjacent in H if they have at least $c^2n/24$ common neighbors in G_2. Note that H has no independent set of size $4/c$. Indeed, if $A \subseteq V(G)$ is an independent set in H of size $k = 4/c$, then writing $\Gamma(.)$ to denote vertex neighborhoods, we have

$$\begin{aligned} n = |V(G)| &\geq \sum_{x \in A} |\Gamma(x, G_2)| - \sum_{x \neq y \in A} |\Gamma(x, G_2) \cap \Gamma(y, G_2)| \\ &> \frac{kcn}{3} - \frac{k^2c^2n}{48} = (4/3 - 1/3)n = n \end{aligned}$$

which is a contradiction.

For each $1 \leq i \leq 3n$, define a sequence of paths in $E(M_i) \cup E(H)$ as follows. Color the edges of M_i blue and the edges of H red; note that there may be edges colored both red and blue. Let $Q_{i,1}$ be a longest path alternating between blue and red edges and starting with a blue edge. Having defined $Q_{i,1}, \ldots, Q_{i,j-1}$, we set

$$E_{i,j} = E(M_i) \setminus (E(Q_{i,1}) \cup \cdots \cup E(Q_{i,j-1})).$$

If $E_{i,j} = \varnothing$, we stop. If not, let $Q_{i,j}$ be a longest path alternating between blue edges from $E_{i,j}$ and red edges, starting with a blue edge. Note that we might reuse red edges in this process, but not blue edges.

Since each $Q_{i,j}$ is a longest path, the starting vertices of the paths $Q_{i,j}$ form an independent set in H. Thus for each $1 \leq i \leq 3n$, we have at most $4/c$ such paths $Q_{i,j}$ and consequently at most $12n/c$ paths in total. Note that every edge of G_1 appears exactly once in one of these $12n/c$ paths as a blue edge. Thus the sum of the lengths of these paths $Q_{i,j}$ is at most $2|E(G_1)| \leq n^2$. We split each of the paths $Q_{i,j}$ into paths of length $c^2n/48$, where we allow one of the subpaths to have length less than $c^2n/48$. We thus obtain at most $n^2/(c^2n/48) + 12n/c \leq 60n/c^2$ red-blue paths. Note that for every red edge

xy, the vertices x and y have at least $c^2n/24$ common neighbors in G_2. Consequently, we can transform all the red-blue paths into simple paths in G: we replace every red edge with a path of length two in G_2 with the same endpoints. We can do this because the number of common neighbors in G_2 of the ends of a red edge is at least twice the length of the original red-blue path. The family consisting of these paths and the paths $P_1, \ldots, P_n$ separates the edges of G_1 and has size at most $60n/c^2 + n \leq 61n/c^2$.

By repeating the above process with the roles of G_1 and G_2 reversed, we obtain a separating path system of G of size at most $122n/c^2$. □

3.4 Random graphs

Proof of Theorem 3.1.4. We use different arguments for different ranges of the edge probability.

Case 1: $p \geq 10\log n/n$**.** Let G be a copy of $G(n,2p)$, where $p \geq 5\log n/n$. We define graphs G_1 and G_2 on the vertex set of G as follows. We construct G_1 by randomly selecting each edge of G with probability $1/2$ and we take G_2 to be the complement of G_1 in G; clearly, G_1 and G_2 are edge-disjoint copies of $G(n,p)$.

The following lemma is easily proved using the standard estimates for the tail of the binomial distribution.

Lemma 3.4.1. *Let $p \geq 5\log n/n$. Then with high probability, the following assertions hold.*

- $n^2p/4 \leq |E(G(n,p))| \leq n^2p$.
- $G(n,p)$ *has minimum degree at least* $np/5$.
- $G(n,p)$ *is* $np/10$-*connected.*

We will need the notion of a *k-linked graph*: a graph is said to be *k-linked* if it has at least $2k$ vertices and for every sequence of $2k$ distinct vertices $u_1, \ldots, u_k, v_1, \ldots, v_k$, there exist vertex disjoint paths $P_1,\ldots,P_k$ such that the endpoints of P_i are u_i and v_i. Bollobás and Thomason [14] showed that every $22k$-connected graph is k-linked. This was later improved by Thomas and Wollan [62], who proved that every $2k$-connected graph on n vertices with at least $5kn$ edges is k-linked. From the latter result and Lemma 3.4.1, we conclude that with high probability, both G_1 and G_2 are $np/20$-linked.

Following the strategy described in Section 3.2, we find a decomposition of G_1 into paths $P_1,\ldots,P_n$ and a decomposition of G_1 into matchings $M_1,\ldots,M_{3n}$ such that the intersection $M_i \cap P_j$ contains at most one edge for each i and j.

We decompose each matching M_i into submatchings of size at most $np/20$. Since G_1 has at most $n^2 p$ edges, we thus obtain at most $23n$ different matchings $M'_1,\ldots,M'_{23n}$. Now since G_2 is $np/20$-linked, we can complete each such matching M'_i into a path using the edges of G_2. These paths along with $P_1,\ldots,P_n$ constitute a separating path system of G_1 of size at most $24n$. Reversing the roles of G_1 and G_2, we obtain a set of $24n$ paths separating the edges of G_2. The union of these two families of paths is a separating path system of G of cardinality at most $48n$.

Case 2: $p \leq 10/n$**.** In this case, with high probability, $G(n,p)$ has at most $20n$ edges and so the edges of G constitute a separating path system of size at most $20n$.

Case 3: $10/n \leq p \leq 10\log n/n$**.** We begin by collecting together some useful properties of sparse random graphs. First, let us establish some notation: given a graph G, write $B_i(v) = B_i(v,G)$ for the set of vertices at (graph-)distance at most i from v and let $\Gamma_i(v) = \Gamma_i(v,G) = B_i(v) \setminus B_{i-1}(v)$. The following lemma is somewhat technical; we defer its proof to the end of the section.

Lemma 3.4.2. *Let $10 \leq d \leq 10\log n$. Then with high probability, the following assertions hold for $G = G(n,d/n)$.*

(i) G has at most dn edges.

(ii) $|\Gamma_i(x)| \leq (2d)^i \log n$ for every $x \in V(G)$ and $i \leq n$.

(iii) *Every set of $i \le \sqrt{n}$ vertices spans at most $2i$ edges. Furthermore, every set of $i \le 10 \log\log n$ vertices spans at most i edges.*

(iv) *If G' is a subgraph of G with minimum degree at least 10, then $|\Gamma_i(x, G')| \ge 2^i$ for every $x \in V(G')$ and every $1 \le i \le 10 \log\log n$.*

(v) *Let G' be a subgraph of G obtained by deleting at most $20d \log n$ vertices and edges and let $l = 3 \log\log n$. For every pair of vertices $x, y \in V(G')$ such that $|B_l(x, G')|, |B_l(y, G')| \ge (\log n)^3$, there is a path of length at most $2 \log n$ between x and y in G'.*

The *k-core* of a graph is its largest induced subgraph with minimum degree at least k. Let H be the 15-core of $G = G(n, p)$ and let $d = np$. By Theorem 3.1.1, we can decompose H into n paths. Since by Lemma 3.4.2(i) there are at most dn edges of G, we can decompose these n paths into at most $2n$ subpaths $Q_1, \ldots, Q_{2n}$, each of which has length at most d.

Let $l = 3 \log\log n$. We shall define a collection of at most $2n$ matchings in H of size d each using the paths $Q_1, \ldots, Q_{2n}$. Each of these matchings will consist of d edges $e_1, \ldots, e_d$ chosen from some d distinct paths $Q_{i_1}, \ldots, Q_{i_d}$ which have the additional property that for every $j \ne j'$ and every $x \in V(Q_{i_j})$ and $x' \in V(Q_{i_{j'}})$ we have $B_l(x) \cap B_l(x') = \varnothing$.

We begin with a collection of paths $R_1, \ldots, R_{2n}$ which we modify as we go along. Initially we set $R_i = Q_i$ for every i. We define our collection of matchings in H in a sequence of rounds.

At the beginning of a round, if we have fewer than $2\sqrt{n}$ non-empty paths R_i, we stop. Otherwise, we select d of the R_i (in a way we specify below), remove the initial edge from each of these paths and use these d removed edges to form a matching of size d. To choose our d paths $R_{i_1}, \ldots, R_{i_d}$ we proceed as follows. Let R_{i_1} be any non-empty path. Now, assume that we have chosen $R_{i_1}, \ldots, R_{i_{t-1}}$, where $t \le d$. Let $N_t = \bigcup_x B_{2l+1}(x)$, where the union is taken over all $x \in V(Q_{i_1}) \cup \cdots \cup V(Q_{i_{t-1}})$. From Lemma 3.4.2(ii), we see that

$$|N_t| < (t-1)d(2d)^{2l+2}\log n < (2d)^{2l+4}\log n < \sqrt{n}.$$

Thus by Lemma 3.4.2(iii), N_t spans at most $2\sqrt{n}$ edges. Since we started the round with more than $2\sqrt{n}$ non-empty paths, there is a path which contains no edge induced by the vertex set N_t; let R_{i_t} be any such a path. We repeat the procedure until the d paths $R_{i_1}, R_{i_2}, \ldots, R_{i_d}$ have been obtained. Clearly the matchings defined by this process are disjoint and of size d, so there are at most n of them; denote them by $M_1, \ldots, M_n$.

In Lemma 3.4.3 (stated below), we show that for each such matching M_i, there is a path containing $E(M_i)$ and avoiding, for every $e \in E(M_i)$, the other edges of the path $Q \in \{Q_1, Q_2 \ldots, Q_{2n}\}$ containing e.

We then obtain a separating system of size at most $19n$ by taking the union of the following families of paths.

- the edges $E(G) \setminus E(H)$ of which there are at most $15n$,
- the paths $Q_1, \ldots, Q_{2n}$,
- the edges of H which are not covered by the matchings $M_1, \ldots, M_n$ of which there are at most $2d\sqrt{n} \leq n$, and
- the set of n paths promised by Lemma 3.4.3.

We now state and prove Lemma 3.4.3.

Lemma 3.4.3. *Let $G = G(n, p)$ be a graph satisfying Lemma 3.4.2. Let $S_1, \ldots, S_d$ be vertex-disjoint paths of length at most d in the 15-core H of G. Set $l = 3\log\log n$ and assume that $B_l(x) \cap B_l(y) = \varnothing$ for every $x \in V(S_i)$ and $y \in V(S_j)$ with $i \neq j$. For each i, select an edge $e_i = x_i y_i$ from S_i, and set $M = \{e_1, e_2, \ldots, e_d\}$. Then there exists a path in G containing all the edges of M and no other edge from $\bigcup_{1 \leq i \leq d} E(S_i)$.*

Proof. Write $E' = (\bigcup_{1\le i\le d} E(S_i)) \setminus E(M)$, let G_0 be the graph on $V(G)$ with edge set $E(G) \setminus E'$, and let G_1 be the graph obtained from G_0 by deleting x_1. Consider the graph H_1 on the vertex set $V(H) \setminus \{x_1\}$ with edge set $E(H) \cap E(G_1)$. Note that H_1 has minimum degree at least 12, since by removing the vertex-disjoint paths $S_1, \dots, S_d$ and the vertex x_1 we decrease vertex degrees in H by at most 3. Thus by Lemma 3.4.2(iv), $|B_l(v, G_1)| \ge (\log n)^3$ for every $v \in V(M)$.

We define vertex-disjoint paths $P_1, \dots, P_{d-1}$ of size at most $2\log n$ as follows. Suppose that we have already defined the paths $P_1, P_2, \dots, P_{i-1}$ for some $i < d$. Set $G_i = G_1 \setminus \bigcup_{1\le j<i} V(P_j)$ and let P_i be a shortest path in G_i connecting y_i to a vertex from $\bigcup_{i+1\le j\le d}\{x_j, y_j\}$. Relabelling the remaining vertices and edges if necessary, assume that this path connects y_i to x_{i+1}.

We shall show by induction that P_i has length at most $2\log n$. Assume that we have defined $P_1, \dots, P_{i-1}$. By the inductive hypothesis, note that we may assume that G_i is obtained by removing at most $2d\log n$ vertices and at most $d^2 \le 10d\log n$ edges from G. Consequently, the bound on the length of P_i would follow from Lemma 3.4.2(v) by showing that $|B_l(y_i, G_i)|, |B_l(x_{i+1}, G_i)| \ge (\log n)^3$.

First, we claim that $B_l(x_{i+1}, G_i) = B_l(x_{i+1}, G_1)$. To see this, first note that for every $j \le i-1$, the sets $B_l(y_j, G)$ and $B_l(x_{i+1}, G)$ are disjoint and consequently, so are $B_l(y_j, G_j)$ and $B_l(x_{i+1}, G_j)$. Since P_j is a shortest path from y_j to $\bigcup_{j+1\le k\le d}\{x_k, y_k\}$, it follows that $V(P_j) \cap B_l(x_{i+1}, G_j) = \varnothing$. Hence, $B_l(x_{i+1}, G_{j+1}) = B_l(x_{i+1}, G_j)$ for every $j \le i-1$. Therefore $|B_l(x_{i+1}, G_i)| \ge (\log n)^3$ and notice that by the same argument, $B_l(y_i, G_{i-1}) = B_l(y_i, G_1)$.

Next, by the minimality of P_{i-1}, it is clear that the set $V(P_{i-1}) \cap B_l(y_i, G_{i-1})$ is contained in the set V'_{i-1} of the last $l+1$ vertices of P_{i-1}. Let H_i be the subgraph of H_1 induced by the vertex subset $V(H_1) \setminus V'_{i-1}$. We deduce from Lemma 3.4.2(iii) that no vertex of G_1 has more than two neighbors in V'_{i-1} and so H_i has minimum degree at least 10. By Lemma 3.4.2(iv) we then have $|B_l(y_i, H_i)| \ge (\log n)^3$. Since

$V(H_i) \cap B_l(y_i, G_1) \subseteq V(G_i) \cap B_l(y_i, G_1)$, it follows that $B_l(y_i, H_i) \subseteq B_l(y_i, G_i)$. Hence, $|B_l(y_i, G_i)| \geq (\log n)^3$ and Lemma 3.4.3 follows by Lemma 3.4.2(v). $\square$

We now complete the proof of Theorem 3.1.4 by proving Lemma 3.4.2.

Proof of Lemma 3.4.2. Parts (i) and (ii) of Lemma 3.4.2 follow easily from the standard Chernoff-type bounds for the tails of binomial random variables. Part (iii) is also routine: the probability that a given set of i vertices induces k or more edges is at most $\binom{i^2}{k}(d/n)^k$ and a straightforward union bound over all sets of i vertices establishes both the claimed statements.

To prove part (iv), we assume that G satisfies parts (ii) and (iii). Let G' be a subgraph of G with minimum degree at least 10. Let $x \in V(G')$ and write $\Gamma_i = \Gamma_i(x, G')$ and $B_i = B_i(x, G')$.

Claim 3.4.4. *$|\Gamma_i| \geq 2|B_{i-1}|$ for $1 \leq i \leq 10 \log\log n$.*

Proof. By part (ii), $|\Gamma_i| \leq (2d)^i \log n$ for $1 \leq i \leq 10 \log\log n$ and so, $|B_i| \leq 2(2d)^{i+1} \log n \leq \sqrt{n}$. So by part (iii), B_i spans at most $2|B_i|$ edges for every $1 \leq i \leq 10 \log\log n$. Since every vertex of B_{i-1} has degree at least 10 in G' and B_{i-1} spans at most $2|B_{i-1}|$ edges, there are at least $6|B_{i-1}|$ edges from B_{i-1} to Γ_i. As B_{i-1} is connected, B_i must span at least $7|B_{i-1}| - 1$ edges. Since B_i spans at most $2|B_i|$ edges, this implies that $|B_i| \geq (7|B_{i-1}| - 1)/2 \geq 3|B_{i-1}|$, i.e., $|\Gamma_i| \geq 2|B_{i-1}|$. $\square$

Claim 3.4.4 implies in particular that $|\Gamma_i| \geq 2^i$ for $i \leq 10 \log\log n$, proving part (iv). In order to prove part (v), we will need the following.

Claim 3.4.5. *Let $l = 3 \log\log n$. Let G' be a graph obtained from G by removing at most $20d \log n$ vertices and edges and let x be a vertex of G' satisfying $|B_l(x, G')| \geq (\log n)^3$. Then with high probability, for every such G' and x, there exists an $i < \log n$ such that $|\Gamma_i(x, G')| \geq n/2d$.*

Proof. Write $\Gamma_i = \Gamma_i(x, G')$ and $B_i = B_i(x, G')$, and let V' and E' be the set of vertices and edges removed from G to obtain G'. Note that the assumption on x implies in particular that there exists a $k \le l$ such that $|\Gamma_k| \ge (\log n)^{5/2}$. By part (ii), with high probability we also have $|B_k| = o(n/d)$.

We show that with high probability, for every G' and x as above and $i \ge k$, either $|\Gamma_{i+1}| \ge (d/2)|\Gamma_i|$ or $|\Gamma_i| \ge (n/2d)$. Note that this would prove Claim 3.4.5.

Conditional on $|\Gamma_i| \le n/2d$ and on $|\Gamma_{j+1}| \ge d|\Gamma_j|/2$ for $k \le j < i$, we shall bound from above the probability that $|\Gamma_{i+1}| \le d|\Gamma_i|/2$. Write $A_i = V(G') \setminus (\Gamma_i \cup V' \cup V(E'))$, and let A'_i be the set of those vertices of A_i that are adjacent to some vertex of Γ_i. It follows that since $|\Gamma_i| \le n/2d$ and $|V'|, |E'| \le 20d \log n$, $|A_i| \ge 9n/10$ for all sufficiently large n. We shall estimate the probability that $|A'_i| \le (d/2)|\Gamma_i|$, conditional on $|\Gamma_i| \ge (\log n)^{5/2}$ and $|A_i| \ge 9n/10$. The probability that a particular vertex of A_i is adjacent to some vertex of Γ_i is

$$1 - (1 - d/n)^{|\Gamma_i|} \ge \frac{d|\Gamma_i|}{n} - \frac{1}{2}\left(\frac{d|\Gamma_i|}{n}\right)^2 \ge \frac{3d|\Gamma_i|}{4n}$$

since no edge adjacent to any vertex of A_i is deleted in passing to G' from G. Thus the expected size of A'_i is at least $27d|\Gamma_i|/40$. By appealing to the standard bounds for the tail of a binomial random variable, we see that

$$\begin{aligned}\mathbb{P}\big[\,|A'_i| \le \frac{d}{2}|\Gamma_i|\,\big] &\le \mathbb{P}\Big[\,|A'_i| \le \frac{3}{4}\mathbb{E}|A'_i|\,\Big] \\ &\le \exp\left(-\mathbb{E}|A'_i|/32\right) \\ &\le \exp\left(-d(\log n)^{5/2}/100\right).\end{aligned}$$

Since we have $2^{O(d(\log n)^2)}$ choices for G', x and i, this implies that $|A'_i| \ge (d/2)|\Gamma_i|$ with high probability, as required. □

We now complete the proof of part (v) of Lemma 3.4.2. Using Claim 3.4.5, we can find $s, t < \log n$ such that $|\Gamma_s(x, G')|, |\Gamma_t(y, G')| \ge n/2d$. If $B_s(x, G') \cap B_t(y, G') \ne \varnothing$, the assertion of part (v) follows. Otherwise, note that the probability that there are no edges between $\Gamma_s(x,G')$ and $\Gamma_t(y,G')$ is at most $(1-d/n)^{n^2/4d^2-20d\log n} \leq e^{-n/5d}$. Since we have $2^{O(d(\log n)^2)}$ choices for x, y and G', this implies that with high probability, the assertion of part (v) holds. □

We have established Lemma 3.4.2, thus completing the proof of Theorem 3.1.4. □

3.5 Dense graphs

Proof of Theorem 3.1.5. Let $c > 0$ and let G be a graph on n vertices such that for every $k \geq \sqrt{n}$, every set of k vertices spans at least ck^2 edges.

We define a sequence of subgraphs $G = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_{l-1}$ and a related sequence of graphs $H_1, H_2, \ldots, H_l$ as follows. Start by setting $G_0 = G$. If $|V(G_{i-1})| \leq \sqrt{n}$, we stop and take $H_i = G_{i-1}$. Otherwise, we take H_i to be the $(c|G_{i-1}|/2)$-core of G_{i-1} and define G_i to be the graph induced by $V(G_{i-1}) \setminus V(H_i)$. Note that the sets $V(H_i)$ form a partition of $V(G)$.

Let us write g_i and h_i respectively for the number of vertices of G_i and H_i. It is well known that the k-core of a graph can be found by removing vertices of degree at most $k-1$, in arbitrary order, until no such vertices exist. So the number of edges removed from G_{i-1} to obtain its $(cg_{i-1}/2)$-core is at most $cg_{i-1}^2/2$. Thus, at least $cg_{i-1}^2/2$ edges remain, so $h_i \geq \sqrt{c}g_{i-1} \geq cg_{i-1} \geq cg_i$.

We first separate the internal edges of the graphs H_i. Note that H_i has minimum degree at least $ch_i/2$. So we conclude from Theorem 3.1.3 that H_i has a separating system of size at most $488h_i/c^2$ for every $1 \leq i < l$. Also, since $V(H_l) \leq \sqrt{n}$, we may separate the edges of H_l trivially (by adding each edge individually to our separating path system); this contributes at most n paths. Since the graphs H_i are pairwise vertex disjoint, we may separate the internal edges of the H_i using at most $488n/c^2$ paths.

It remains to separate the crossing edges between the H_i. For $1 \leq i < l$, let E_i be the set of edges of the form xy where $x \in V(H_i)$ and $y \in V(G_i)$, and let E_i' be the set of such edges xy where y has at least 3 neighbors in H_i. Note that every edge of G not contained in any of the H_i is contained in one of the E_i.

We define a g_i-edge-colored multigraph F_i on the vertex set of H_i as follows. If $v \in G_i$ has at least three neighbors in H_i, say $x_1, \ldots, x_k$, we add the edges $x_1x_2, x_2x_3, \ldots, x_kx_1$ to F_i and color these edges with the color v; in other words, we add a v-colored cycle through the neighbors of v. Note that the degree of every vertex of F_i (as a multigraph) is at most $2g_i$ and every color class contains at most h_i edges. Since each edge has at most $4g_i + h_i \leq 5h_i/c$ edges which are either incident to it or from the same color class, we can, as in Section 3.2, decompose F_i into at most $5h_i/c$ rainbow matchings $M_1, \ldots, M_{5h_i/c}$ where by a rainbow matching, we mean a matching containing at most one edge from each color class.

We now construct another sequence of rainbow matchings decomposing F_i with the following property. Denote by $e_1, \ldots, e_m$ the edges of M_j and let $v_1, \ldots, v_m$ be their respective color classes. Let α_k and β_k be the two neighbors of e_k in the cycle whose edges have color v_k. In our second sequence of rainbow matchings, we would like the matching containing e_k to avoid e_t, α_t and β_t for every $t \neq k$. Since each edge has to avoid at most $4g_i + h_i + 3h_i \leq 8h_i/c$ other edges, we can find such a decomposition into at most $8h_i/c$ matchings, say, $M_{5h_i/c+1}, \ldots, M_{13h_i/c}$.

We now mimic the proof of Theorem 3.1.3. Let us define a graph H_i' on $V(H_i)$ where we join two vertices if they have more than $c^2h_i/24$ common neighbors in H_i. For each $1 \leq j \leq 13h_i/c$, we can find a collection of at most $4/c$ paths whose edges alternate between those of M_j and H_i' which cover each edge of M_j exactly once; we obtain $52h_i/c^2$ such paths in total. We divide these paths into subpaths of length at most $c^2h_i/48$ each, resulting in a collection of at most $96h_i/c^3 + 52h_i/c^2$ paths. Each such path can be transformed into a path in G by replacing every edge from H_i' with a suitably chosen path of length two in H_i and every colored edge $e = xy$ from M_j with the path xvy where v is the color of e. Since the matchings are rainbow matchings, these paths are guaranteed to be simple. It is easy to see that the collection of $96h_i/c^3 + 52h_i/c^2$ paths defined above separates E_i'.

It remains to separate the edges of $E_i \setminus E_i'$ for $1 \leq i < l$. Note that there are at most $2(g_1 + \cdots + g_l) \leq 2(h_1 + \cdots + h_l)/c \leq 2n/c$ such edges; we add each such edge to our separating path system.

It is easy to check that we have constructed a separating path system of G of cardinality at most $488n/c^2 + 96n/c^3 + 52n/c^2 + 2n/c \leq 638n/c^3$. The result follows. □

3.6 Trees

We begin by collecting together a few simple observations into the following lemma.

Lemma 3.6.1. *Let T be a tree on $n \geq 3$ vertices, and let $\mathcal{P}$ be a separating path system of T. Then the following assertions hold.*

(i) With the exception of at most one leaf, every leaf of T is an endpoint of a path in $\mathcal{P}$.

(ii) If a path in $\mathcal{P}$ has two leaves u and v as its endpoints, then there must be at least one path in $\mathcal{P}$ which has exactly one of u and v as an endpoint.

(iii) Every vertex of degree two in T is an endpoint of a path in $\mathcal{P}$.

Proof. Clearly, a leaf must be an endpoint of any path through it. Since $\mathcal{P}$ separates $E(T)$, there is at most one edge of T which is not covered by any path in $\mathcal{P}$. As $n \geq 3$, T does not consist of a single edge and thus at most one leaf of T is visited by no path in $\mathcal{P}$. This establishes part (i).

Suppose that we have a path $P \in \mathcal{P}$ having two leaves $u, v \in V(T)$ as its endpoints. Let e_u and e_v be the edges incident to u and v respectively. Since $\mathcal{P}$ separates $E(T)$, there must be some path $P' \in \mathcal{P}$ containing exactly one of e_u and e_v. This establishes part (ii).

Suppose that v is a vertex of degree two in T; let e_1 and e_2 be the two edges of T incident to v. Since $\mathcal{P}$ separates $E(T)$, there must be some path $P \in \mathcal{P}$ containing exactly one of e_1 and e_2. Since v has degree two, it must be an endpoint of this path P. This establishes part (iii). □

We split the proof of Theorem 3.1.6 into two parts.

Proof of the lower bound in Theorem 3.1.6. Let T be a tree on $n \geq 4$ vertices and let $\mathcal{P}$ be a separating system of T. We shall show that $|\mathcal{P}| \geq \lceil (n+1)/3 \rceil$.

First, suppose that there is a leaf v which is not the endpoint of any path in $\mathcal{P}$. Let e_v be the edge of T incident to v. Since $\mathcal{P}$ separates $E(T)$, e_v is the unique edge of T not covered by any path in $\mathcal{P}$. Delete v from T to obtain a tree T' on $n-1 \geq 3$ vertices.

The family $\mathcal{P}$ both covers and separates $E(T')$. From Lemma 3.6.1, we note that every leaf and every vertex of degree two of T' is the endpoint of at least one path from $\mathcal{P}$. Furthermore, we know that if a path from $\mathcal{P}$ has a pair of leaves for its endpoints, then at least one of those leaves is the endpoint of at least one other path from $\mathcal{P}$.

Let d_1 and d_2 denote the number of leaves and degree two vertices of T'. We claim that $\mathcal{P}$ contains at least $(2d_1 + d_2)/3$ paths. To see this, start by placing a red token on every leaf and a blue token on every vertex of degree two in T'. We then iterate through the paths of $\mathcal{P}$ in some order and in each iteration, we remove whatever tokens there are at the endpoints of the current path. If both the tokens removed are red, then we know that both the endpoints, say u and v, of the current path are leaves and that at least one of them, say u, is the endpoint of a different path; we then place a blue token on u. Writing R and B respectively for the number of red and blue tokens remaining on the tree, we see that the quantity $2R + B$ does not decrease by more than three in any iteration. Since $\mathcal{P}$ is a separating path system, all the tokens must have been removed by the end of the procedure. It follows that

$$|\mathcal{P}| \geq \frac{2d_1 + d_2}{3}.$$

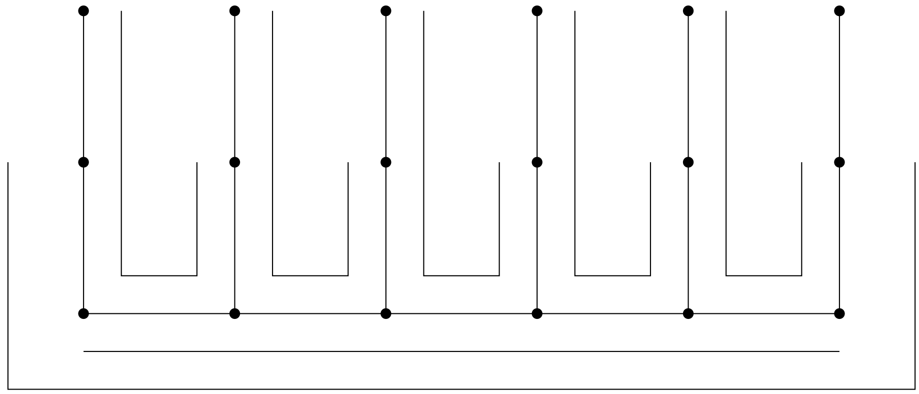

Figure 3.3: A hair-comb of order 18 and a separating system of 7 paths.

Now, note that

$$2e(T') = 2(n-2) \geq d_1 + 2d_2 + 3(n-1-d_1-d_2),$$

which we can rearrange to get

$$\frac{2d_1 + d_2}{3} \geq \frac{n+1}{3}.$$

Taken together, these inequalities show that $|\mathcal{P}| \geq (n+1)/3$.

If on the other hand every leaf of T is the endpoint of some path from $\mathcal{P}$, then by repeating the argument above with T instead of T', we find that $|\mathcal{P}| \geq (n+2)/3$. We know from Lemma 3.6.1 that these are the only two possibilities; consequently, we are done.

To see that this lower bound is best possible, consider the family of *hair combs*, where the hair comb of order $3n$ is obtained by starting with a *spine* consisting of a path of length $n-1$ and then attaching a path of length two to each vertex of the spine. It is an easy exercise to show that this lower bound is tight for hair combs. (See Figure 3.3 for an example of an optimal separating path system.) □

We now turn our attention to the second part of the proof of Theorem 3.1.6.

Proof of the upper bound in Theorem 3.1.6. We shall show by induction on $n = |V(T)|$ that $f(T) \leq \lfloor 2(n-1)/3 \rfloor$.

There is, up to isomorphism, only one tree of order n for each of $n = 1, 2, 3$, namely the path of length $n-1$. It is trivial to check that the claim holds for these trees.

Let T be a tree of order $n > 3$. If T is a path, then it is easy to show using Lemma 3.6.1 that $f(T) \geq \lceil (n-1)/2 \rceil$ and it is easy (see Figure 3.1) to construct a separating path system matching this bound. Since $\lceil (n-1)/2 \rceil \leq \lfloor 2(n-1)/3 \rfloor$ for all $n \geq 4$, we may suppose that T is not a path; hence T must contain at least one vertex u with three distinct neighbors, say v_1, v_2 and v_3. Contract the edges uv_1, uv_2 and uv_3 to obtain a new tree T' on $n-3$ vertices.

We find a separating path system $\mathcal{P}'$ of T' of size at most $2(n-4)/3$. We may think of $\mathcal{P}'$ as a family of paths of T since paths in T' map to paths in T in a natural way: a path in T' is lifted up to a path in T with the same endpoints (where we identify the vertex resulting from the contraction of u, v_1, v_2 and v_3 with u). Consider the family

$$\mathcal{P} = \mathcal{P}' \cup \{v_1uv_2, v_2uv_3\}.$$

Since $\mathcal{P}'$ separates $E(T')$, it readily follows that $\mathcal{P}'$, when viewed as a family of paths of T, separates $E(T) \setminus \{uv_1, uv_2, uv_3\}$. The two paths v_1uv_2 and v_2uv_3 then separate uv_1, uv_2 and uv_3 from each other and from the rest of $E(T)$. Thus,

$$|\mathcal{P}| \leq \frac{2(n-4)}{3} + 2 = \frac{2(n-1)}{3}.$$

We are done by induction.

To see that this upper bound is best possible, consider the family of *stars*, where the star of order n consists of a single internal vertex joined to $n-1$ leaves. By mimicking the proof of the lower bound using Lemma 3.6.1, it is an easy exercise to verify that the upper bound is tight for stars. □

3.7 Open problems

There remain a number of interesting questions which merit investigation. While the main open problem of course is to establish that $f(n) = O(n)$, there are many other attractive related extremal questions. For instance, it would be interesting to determine the value of $f(K_n)$ exactly; one can also ask the same question for the the d-dimensional hypercube Q_d. It is easy to cover Q_d with $d-1$ ladders and so $f(Q_d) = O(d^2)$. On the other hand, we know from the information theoretic lower bound that $f(Q_d) = \Omega(d)$. It would be interesting to nail down the exact value of $f(Q_d)$.

A different question, though of a similar flavour, raised by Bondy [15] and answered by Li [47], is that of finding *perfect path double covers*, i.e., a set of paths of a graph such that each edge of the graph belongs to exactly two of the paths and each vertex of the graph is an endpoint of exactly two of the paths. We suspect that the tools developed to tackle this problem and its variants might prove useful in attempting to establish that $f(n) = O(n)$.

CHAPTER 4

FRUSTRATED TRIANGLES

4.1 Introduction

Given a graph G on n vertices, let $t_i = t_i(G)$ denote the number of triples of vertices in G inducing i edges for $i = 0, 1, 2, 3$. One of the earliest and best-known results on colorings is due to Goodman [31], who showed that $t_0 + t_3$ is asymptotically minimised by the random graph. Goodman [32] also conjectured the maximum of $t_0 + t_3$ amongst the graphs with a given number of edges, which was later proved by Olpp [53]. More recently, Linial and Morgenstern [48] showed that every sequence of graphs with $t_0 + t_3$ asymptotically minimal is 3-universal. Hefetz and Tyomkyn [37] then proved that such sequences are 4-universal, but not necessarily 5-universal, and moreover that any sufficiently large graph H can be avoided by such a sequence.

The minimum number of triangles in a graph with a given number of edges has also been widely investigated. Erdős [22] conjectured that a graph with $\lfloor \frac{n^2}{4} \rfloor + k$ edges contains at least $k\lfloor \frac{n}{2} \rfloor$ triangles if $k < \frac{n}{2}$, which was later proved by Lovász and Simonovits [50]. More recently, Rozborov [57] determined completely asymptotically the minimum number of triangles in a graph with a given number of edges using flag algebras. Some bounds for other combinations of t_0, t_1, t_2, t_3 were also given in [32, 52], while similar results for three-colored graphs have recently been proved in [5, 20].

In this chapter, we are interested in another natural quantity $t_1 + t_3$. This study is further motivated by a phenomenon occurred in several fields of physics called *geometrical frustration* (see [28, 64, 65]). For example, suppose each vertex of a graph is a spin which can take only two values, say up and down, and each edge of the graph is either ferromagnetic (meaning the spins on its end points prefer to be aligned), or anti-ferromagnetic (the spins prefer to be in opposite directions). Now consider a cycle in the graph, and observe that there is no choice of the values of the spins satisfying the preference of every edge of the cycle if and only if there are an odd number of anti-ferromagnetic edges.

We say that a triple of vertices of a graph is a *frustrated triangle* if it induces an odd number of edges, i.e. either it contains exactly one edge, or it is a triangle. We shall write $f(G) = t_1 + t_3$ for the number of frustrated triangles in a graph G. Our aim is to study the set of possible number of frustrated triangles in a graph with n vertices,

$$F_n = \{f(G) : G \text{ is a graph on } n \text{ vertices}\}.$$

We remark that this study has a similar flavor to those in Chapter 1 where the object of interest is the set of possible number of colors appearing in a complete subgraph of the complete graph on $\mathbb{N}$.

Clearly, $F_n \subset [0, \binom{n}{3}]$. Note also that F_n is symmetric about the midpoint $\frac{1}{2}\binom{n}{3}$, i.e. $x \in F_n$ iff $\binom{n}{3} - x \in F_n$. This is because a triple is frustrated in G iff it is not frustrated in the complement graph $\overline{G}$, and so $f(G) + f(\overline{G}) = \binom{n}{3}$.

One might expect F_n to be the whole interval $[0, \binom{n}{3}]$. However, this is surprisingly very far from the truth. Our main result specifies subintervals of $[0, n^{3/2}]$ that are forbidden from F_n, and moreover, we characterise the graphs G with $f(G) \in [0, n^{3/2}]$.

Before we state the main theorem, let us introduce the following two sequences which play an important role throughout the chapter. For $0 \le t \le n-1$, let $a_t = a_t^{(n)}$ be the number of frustrated triangles in a graph on n vertices containing t edges forming a star. For $0 \le t \le \frac{n}{2}$, let $b_t = b_t^{(n)}$ be the number of frustrated triangles in a graph on n vertices containing t edges forming a matching. It is immediate that

$$a_t = t(n-t-1) \text{ and } b_t = t(n-2).$$

Let $t_{\max} = \max\{t : b_t < a_{t+1}\}$ be the last t such that the intervals $[a_t, b_t]$ and $[a_{t+1}, b_{t+1}]$

are disjoint. It is easy to check that

$$t_{\max} = \max\{t : t(t+1) < n-2\} = \left\lceil \sqrt{n - \frac{7}{4}} - \frac{3}{2} \right\rceil \sim \sqrt{n}$$

exists for $n \geq 3$, and

$$0 = a_0 = b_0 < a_1 = b_1 < a_2 < b_2 < \cdots < a_{t_{max}} < b_{t_{max}} < a_{t_{max}+1} \sim n^{3/2}.$$

We can now state the main result.

Theorem 4.1.1. *Let G be a graph on $n \geq 3$ vertices.*

(i) If $f(G) < a_{t_{max}+1}$ then $f(G) \in [a_t, b_t]$ for a unique $t \leq t_{max}$.

(ii) If $t \leq t_{max}$ then $f(G) \in [a_t, b_t]$ iff G can be obtained from a complete bipartite graph on n vertices by flipping exactly t pairs of vertices.

Here, to *flip* a pair of vertices uv means to change its edge/nonedge status, i.e. if uv is an edge, make it a nonedge, and vice versa. We remark that Theorem 4.1.1 part (i) is equivalent to the statement: $f(G) \notin (b_t, a_{t+1})$ for all $t \geq 0$. Note also that the intervals have lengths

$$a_{t+1} - b_t = (n-2) - t(t+1) \text{ and } b_t - a_t = t(t-1).$$

Theorem 4.1.1 is complemented by the following result. Observe that $a_{t_{max}+1} = (t_{max}+1)(n - t_{max} - 2) \sim n^{3/2}$ and so Theorem 4.1.1 deals with the case $f(G) \lesssim n^{3/2}$. Since F_n is symmetric, we automatically have a corresponding result for $f(G) \gtrsim \binom{n}{3} - n^{3/2}$. On the other hand, we prove that every number, up to parity condition, in the large central part of $[0, \binom{n}{3}]$ is realisable as $f(G)$ for some G. Note that if n is even, then $f(G)$ must also be even, since adding an edge to a graph changes the 'frustration status' of exactly $n-2$ triples.

Theorem 4.1.2. *(i) If n is even and sufficiently large then F_n contains every even integer between $n^{3/2}+O(n^{5/4})$ and $\binom{n}{3}-(n^{3/2}+O(n^{5/4}))$, and this is best possible up to the second order term.*

(ii) If n is odd and sufficiently large then F_n contains every integer between $\sqrt{2}n^{3/2}+O(n^{5/4})$ and $\binom{n}{3}-(\sqrt{2}n^{3/2}+O(n^{5/4}))$, and this is best possible up to the second order term.

Let us change the direction and turn to the following related natural question. Given the number of vertices and the number of edges, which graphs maximise/minimise the number of frustrated triangles? The method we develop to prove Theorem 4.1.1 allows us to partially answer this question. Before we state the result, we shall define some necessary notations. For $0 \le x \le \frac{n}{2}$, let $c_x^{(n)} = x(n-x)$ be the number of edges in the complete bipartite graph $K_{x,n-x}$. For an integer e, let $g(e) = \min\{|e-c_x| : 0 \le x \le \frac{n}{2}\}$ be the distance from e to the closest element of the sequence (c_x). We are able to determine the minimal graphs when the number of edges is at most $\frac{n^2}{4}$.

Theorem 4.1.3. *If G is a graph with n vertices and e edges then $f(G) \ge a_{g(e)}$. Moreover, this bound can be achieved when $e \le \lfloor \frac{n^2}{4} \rfloor + \lfloor \frac{n-1}{2} \rfloor - 1$: in this case the extremal graphs are obtained from a complete bipartite graph by deleting or adding $g(e)$ edges forming a star.*

The rest of this chapter is organised as follows. In Section 4.2, we present some preliminary results for the readers to get familiar with frustrated triangles and to motivate the definitions and ideas used to prove the main results. Sections 4.3 and 4.4 are devoted to the proofs of Theorems 4.1.1 and 4.1.2 respectively. In Section 4.5, we describe an application of Theorems 4.1.1. We conclude the chapter in Section 4.6 with some open problems.

4.2 Preliminaries

We shall start with a coffee time problem which is a special case of Theorem 4.1.1. By considering the empty and the complete graphs, we see that 0 and $\binom{n}{3}$ are always in F_n. A natural question is that, what is the first nonzero element of F_n? If we consider a graph with only one edge uv, the frustrated triangles are those triples containing both u, v; therefore there are $n-2$ frustrated triangles. It turns out that $n-2$ is the answer. Before we give the proof, let us introduce the *flipping operation* which is an important idea for dealing with frustrated triangles.

Recall that to flip a pair of vertices is to change its edge/nonedge status. For a vertex v of a graph G, let G_v denote the graph obtained from G by flipping the pairs uv for all $u \in G \backslash \{v\}$ (see Figure 4.1), i.e.
$V(G_v) = V(G)$ and $E(G_v) = E(G) \cup \{uv : uv \notin E(G)\} - \{uv : uv \in E(G)\}$.

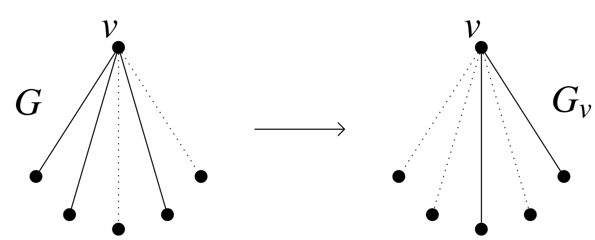

Figure 4.1: The flipping operation on vertex v.

By *flipping* v, we mean an operation of changing G to G_v. The readers should be warned to note the difference between flipping a pair of vertices and flipping a vertex. We then have the following easy but useful lemma.

Lemma 4.2.1. *For any graph G, $f(G)$ is preserved under the flipping operation, i.e. $f(G_v) = f(G)$ for any vertex $v \in G$.*

Proof. More is true: $\{x, y, z\}$ is frustrated in G_v iff it is in G. This is obvious if $v \notin \{x, y, z\}$. If $v \in \{x, y, z\}$ then exactly two pairs of $\{x, y, z\}$ were flipped; therefore the parity of the number of edges induced by $\{x, y, z\}$ stays the same. □

Before we show that $n-2$ is the first nonzero element of F_n, it is useful to answer the following question. What are the graphs with no frustrated triangles?

Proposition 4.2.2. *For any graph G, $f(G)=0$ iff G is a complete bipartite graph.*

Proof. It is clear that a complete bipartite graph contains no frustrated triangles. Conversely, given a vertex v of a graph G with $f(G)=0$. Then there is no edge in $G[\Gamma(v)]$ where $\Gamma(v)$ is the neighborhood of v; otherwise it would form a frustrated triangle with v. Similarly, there is no edge in the nonneighborhood of v. Also, if vx is an edge and vy is a nonedge, then xy has to be an edge. Therefore, G is a complete bipartite graph with parts $\Gamma(v)$ and $V\backslash\Gamma(v)$. □

We are now ready to show that there is no graph on n vertices with number of frustrated triangles strictly between 0 and $n-2$.

Proposition 4.2.3. *For all $n\in\mathbb{N}$, $F_n\cap(0,n-2)=\varnothing$.*

Proof. We apply induction on n. There is nothing to check for $n=3$. Let G be a graph on n vertices. Our aim is to show that $f(G)\notin(0,n-2)$. Without loss of generality, we may assume that G has an isolated vertex v. This is because we can flip each neighbor of v to make it a nonneighbor while preserving the number of frustrated triangles in G by Lemma 4.2.1. Let $G'=G-v$ be the graph obtained from G by deleting v. Since v is an isolated vertex, we have

$$f(G)=f(G')+e(G').$$

By the induction hypothesis, $f(G')\notin(0,n-3)$. We shall distinguish two cases.

Case 1: $f(G')\geq n-3$

If $e(G')\geq 1$ then $f(G)=f(G')+e(G')\geq(n-3)+1\geq n-2$ as required. If $e(G')=0$ then G is empty and so $f(G)=0$ as required.

Case 2: $f(G')=0$

By Proposition 4.2.2, G' is complete bipartite and so $e(G') = x(n-1-x)$ for some $0 \le x \le n-1$. We are done if G' is empty. If G' is not empty, then $e(G')$ is minimised when $x = 1$ or $n-2$, i.e. $f(G) = e(G') \ge n-2$ as required. □

We have just proved that $f(G)$ cannot lie in the gap between the intervals $[a_0, b_0] = \{0\}$ and $[a_1, b_1] = \{n-2\}$ which is the first case of Theorem 4.1.1 part (i). The equation $f(G) = f(G') + e(G')$ in the proof suggests that, in order to understand the possible number of frustrated triangles, we should understand the possible number of edges in a graph with a given number of frustrated triangles. In fact, we will have an analogue of Proposition 4.2.2 for $f(G) \lesssim n^{3/2}$, i.e. we will not only know the possible number of edges, but we will also know the possible structure of the graph (see Theorem 4.1.1 part (ii)).

Let us now consider the converse of Lemma 4.2.1. We write $G \sim H$ if G can be obtained from a graph H by a sequence of vertex flippings. Clearly, $\sim$ is an equivalence relation. Observe that the complete bipartite graphs on n vertices form an equivalence class. Indeed, let G be a complete bipartite graph with parts A, B and let $v \in A$. Then flipping v is equivalent to moving v across from A to B, i.e. G_v is the complete bipartite graph with parts $A \backslash \{v\}$ and $B \cup \{v\}$. Therefore, another way to state Proposition 4.2.2 is: for any graph G on n vertices, $f(G) = f(E_n)$ iff $G \sim E_n$ where E_n is the empty graph on n vertices.

This shows that the converse of Lemma 4.2.1 is true in the case $f(G) = 0$. Does it hold in general? That is, given two graphs with the same number of frustrated triangles, can we always obtain one from the other by a sequence of vertex flippings? Unfortunately, this is false. As we can see from the proof of Lemma 4.2.1, if $G \sim H$ then not only do we have $f(G) = f(H)$ but there is also a bijection $\phi : V(G) \to V(H)$ such that $\{u, v, w\}$ is frustrated iff $\{\phi(u), \phi(v), \phi(w)\}$ is frustrated. It is easy to see that this is also sufficient.

Proposition 4.2.4. *Let G and H be graphs on n vertices. The following statements are equivalent.*

(i) $G \sim H$.

(ii) *There is a bijection* $\phi : V(G) \to V(H)$ *such that* $\{u, v, w\}$ *is frustrated in G iff* $\{\phi(u), \phi(v), \phi(w)\}$ *is frustrated in* H.

Proof. The implication $(i) \Rightarrow (ii)$ follows from the proof of Lemma 4.2.1.

To prove $(ii) \Rightarrow (i)$, suppose ϕ is a bijection satisfying (ii). We say that a pair of vertices ab of G and a pair of vertices xy of H *agree* if both ab and xy are edges, or both ab and xy are nonedges. Let v be a vertex of G and let $U = \{u \in V(G) \backslash \{v\} : uv \text{ and } \phi(u)\phi(v) \text{ disagree}\}$. Let G' be the graph obtained from G by flipping each vertex of U. We claim that G' and H are isomorphic. Clearly, uv in G' and $\phi(u)\phi(v)$ in H agree for all $u \in V(G') \backslash \{v\}$. Furthermore, by the proof of Lemma 4.2.1, ϕ still satisfies (ii) after replacing G with G'. Hence, for any $u, w \in V(G') \backslash \{v\}$, the graphs $G'[u, v, w]$ and $H[\phi(u), \phi(v), \phi(w)]$ have the same 'frustration status'. Since two of the pairs agree, the third pairs uw and $\phi(u)\phi(w)$ must also agree. Therefore, G' and H are isomorphic, and so H can be obtained from a G by a sequence of vertex flippings. □

Now we shall give an explicit counterexample to the converse of Lemma 4.2.1. Let G be a disjoint union of P_2 and P_2, and let H be a disjoint union of P_1 and P_3 (see Figure 4.2) where P_l is a path with l edges.

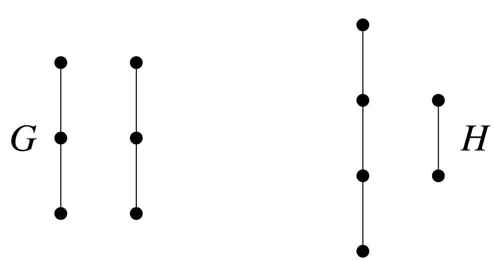

Figure 4.2: Counterexample to the converse of Lemma 4.2.1.

Then $f(G) = f(H) = 12$, but there is no bijection satisfying Proposition 4.2.4 part (ii) since, for example, G contains four vertices inducing no frustrated triangles while H does not. Therefore, $G \not\sim H$.

Let us move on and provide some formulae for $f(G)$ in terms of other graph variables, which also gives us some bounds on $f(G)$.

Proposition 4.2.5. *Let G be a graph with n vertices and e edges. Write p for the number of triangles in G, and q for the number of pairs of independent edges in G. Then*

(i) $f(G) = en - \sum_{v\in G} d_v^2 + 4p$

(ii) $f(G) = e(n-e-1) + 4p + 2q$

(iii) $f(G) \in [a_e, b_e]$

where d_v is the degree of vertex v.

Proof. (i) Given an edge xy, let $T_{xy} = \{vxy : v \neq x, y\}$ be the set of triples on xy. Let $T_x = \{vxy \in T_{xy} : vx \in E(G)\}$ and $T_y = \{vxy \in T_{xy} : vy \in E(G)\}$. Then, the number of triples on xy inducing exactly one edge is

$$\begin{aligned}|T_{xy} - (T_x \cup T_y)| &= |T_{xy}| - |T_x| - |T_y| + |T_x \cap T_y| \\ &= (n-2) - (d_x - 1) - (d_y - 1) + p_{xy} \\ &= n - d_x - d_y + p_{xy}\end{aligned}$$

where p_{xy} is the number of triangles containing xy. Hence,

$$\begin{aligned}f(G) &= p + \sum_{xy\in E(G)} (\#\text{triples on } xy \text{ inducing exactly one edge}) \\ &= p + \sum_{xy\in E(G)} (n - d_x - d_y + p_{xy}) \\ &= p + en - \sum_{xy\in E(G)} (d_x + d_y) + 3p \\ &= en - \sum_{v\in G} d_v^2 + 4p\end{aligned}$$

as required.

(ii) This follows immediately from (i). It is sufficient to show that $\sum_{v\in G} d_v^2 = e(e+1)-2q$. Since a pair of edges is either independent or dependent, we have

$$\binom{e}{2} = q + \sum_{v\in G} \binom{d_v}{2},$$

i.e.

$$e(e-1) = 2q + \sum_{v\in G} d_v^2 - 2e$$

as required.

(iii) The upperbound is obvious:

$$f(G) \leq (\text{\#triples containing at least one edge}) \leq e(n-2) = b_e,$$

and the lower bound follows from (ii):

$$f(G) = e(n-e-1) + 4p + 2q \geq e(n-e-1) = a_e$$

as required. □

The proof of part (iii) tells us that, amongst the graphs with n vertices and $e \leq n/2$ edges, the e-matching is the only graph with the maximum number of frustrated triangles and, amongst the graphs with n vertices and $e \leq n-1$ edges, the e-star is the only minimal graph.

Although part (iii) looks like what we would like for Theorem 4.1.1, it only gives us good bounds when e is small. For larger e, the lowerbound a_e becomes worse and can even be negative. However, we do not have to apply (iii) to G directly. By Lemma 4.2.1, we can apply (iii) to any graph H with $G \sim H$. Therefore, this motivates us to find a graph H with few edges such that $G \sim H$, which gives rise to the following crucial definition.

For a graph G, let $t_G = \min\{e(H) : G \sim H\}$ be the minimum number of edges we can have after some vertex flippings of G. Proposition 4.2.5 part (iii) implies that $f(G) \in [a_{t_G}, b_{t_G}]$. Therefore, to prove Theorem 4.1.1 part (i), it is enough to show that $t_G \leq t_{max}$ if $f(G) < a_{t_{max}+1}$. It is important to note that t_G is preserved under the flipping operation.

We now observe that t_G can also be viewed as a measure of how close G is to being a complete bipartite graph. We say that a pair of vertices uv is *odd* with respect to a bipartition $V(G) = X \cup X^c$ if

- $uv \notin E(G)$ and u, v are in different parts, or
- $uv \in E(G)$ and u, v are in the same part,

i.e. if uv is not what it should be in the complete bipartite graph between X, X^c. Therefore, flipping the odd pairs w.r.t. $V(G) = X \cup X^c$ would result in the complete bipartite graph between X, X^c. We have the following equivalent definition for t_G.

Proposition 4.2.6. *For any graph G,*

$$t_G = \min\{\textit{\#odd pairs w.r.t. } V(G) = X \cup X^c : X \subset V(G)\}.$$

Proof. For $X \subset V(G)$, we have

$$\begin{aligned}
&(\text{the set of odd pairs with respect to } V(G) = X \cup X^c) \\
&\quad = E(G[X]) \cup E(G[X^c]) \cup E^c(G[X, X^c]) \\
&\quad = (\text{the set of edges of the graph obtained from G by flipping each vertex in } X)
\end{aligned}$$

where $E^c(G[X, X^c])$ is the set of nonedges between X, X^c. □

The proof also gives us another necessary and sufficient condition for $G \sim H$.

Corollary 4.2.7. *Let G and H be graphs on n vertices. The following statements are equivalent.*

- *$G \sim H$.*

- *There are a bijection $\phi : V(G) \to V(H)$ and a subset $X \subset V(G)$ such that uv is an odd pair w.r.t. $V(G) = X \cup X^c$ in G iff $\phi(u)\phi(v)$ is an edge in H.* □

The definition of t_G allows us to state a generalization of Proposition 4.2.2,

$$f(G) \in [a_t, b_t] \text{ iff } t_G = t.$$

It is not hard to see that this is false for $t > t_{max}$ since $[a_t, b_t]$ and $[a_{t+1}, b_{t+1}]$ overlap. On the other hand, Theorem 4.1.1 part (ii) states that this generalization holds for $t \leq t_{max}$.

4.3 Proof of Theorem 4.1.1

We shall prove Theorem 4.1.1 in the following simple but stronger form.

Theorem 4.3.1. *Let G be a graph on $n \geq 3$ vertices and let $t \geq 0$.*

(i) If $f(G) < a_{t+1}$ then $t_G \leq t$.

(ii) If $f(G) > b_{t-1}$ then $t_G \geq t$.

Note that Theorem 4.3.1 part (i) covers a larger range than Theorem 4.1.1. Indeed, it describes the structure of graphs G with $f(G) \lesssim n^2$, since a_{t+1} can be as large as $\left(\frac{n-1}{2}\right)^2$.

Before proving Theorem 4.3.1, let us show that it immediately implies Theorem 4.1.1.

Proof of Theorem 4.1.1. (i) Suppose that $f(G) < a_{t_{max}+1}$. Then, by Theorem 4.3.1 part (i), $t_G \leq t_{max}$ and so we are done since $f(G) \in [a_{t_G}, b_{t_G}]$ by Proposition 4.2.5 part (iii). The uniqueness of t is trivial since $[a_0, b_0], [a_1, b_1,], \ldots, [a_{t_{max}}, b_{t_{max}}]$ are disjoint.

(ii) $(\Rightarrow)$ Suppose that $f(G) \in [a_t, b_t]$ for some $t \leq t_{max}$. By Proposition 4.2.6, it is sufficient to show that $t_G = t$. Since $t \leq t_{max}$, we have $f(G) \in [a_t, b_t] \subset (b_{t-1}, a_{t+1})$, and so $t_G = t$ by Theorem 4.3.1.

($\Leftarrow$) Suppose that G can be obtained from a complete bipartite graph by flipping exactly $t \leq t_{max}$ pairs of vertices. Equivalently, $G \sim H$ for some graph H with $e(H) = t$ by Corollary 4.2.7. Hence, by Proposition 4.2.5 part (iii),

$$f(G) = f(H) \in [a_{e(H)}, b_{e(H)}] = [a_t, b_t]$$

as required. □

Now let us prove Theorem 4.3.1. Note that part (ii) is easy, and the main content is in part (i).

Proof of Theorem 4.3.1. We start by proving part (ii), which is equivalent to the statement: if $t_G \leq t$ then $f(G) \leq b_t$. By Proposition 4.2.5 part (iii), $f(G) \leq b_{t_G} \leq b_t$ since b_t is increasing in t.

We now prove part (i) by induction on n. It is easy to check for $n = 3$. Let G be a graph on n vertices such that $f(G) < a_{t+1}$. Our aim is to find a bipartition of $V(G)$ with at most t odd pairs. We write a_s for a_s^n and a'_s for a_s^{n-1}. Since $f(G)$ and t_G are preserved under the flipping operation, we may assume that G has an isolated vertex v (as in the proof of Proposition 4.2.3). Let $G' = G - v$, and note that $f(G) = f(G') + e(G')$. Now observe that we are done if $e(G') \leq t$, since $t_G \leq e(G) = e(G')$ by the definition of t_G. So we may assume that $e(G') \geq t + 1$, and hence,

$$f(G') = f(G) - e(G') \leq f(G) - (t+1) < a_{t+1} - (t+1) = a'_{t+1}.$$

Therefore, we may now assume that $f(G') \in [a'_s, a'_{s+1})$ for some $s = 0, 1, \ldots, t$. Since $f(G') < a'_{s+1}$, we have, by the induction hypothesis, that G' has a bipartition $V(G') = X \cup Y$ with at most s odd pairs. Without loss of generality, X is the smaller part and we write x for $|X|$. We shall distinguish two cases.

Case 1: $x \leq t - s$

With respect to the bipartition $V(G) = X \cup (Y \cup \{v\})$, the number of odd pairs is at most $x+s \leq t$ as required.

Case 2: $x \geq t-s+1$

Since G' contains at most s odd pairs w.r.t. the biparition $V(G') = X \cup Y$ and $x \leq \frac{n-1}{2}$, we have

$$e(G') \geq x(n-1-x) - s = a_x - s \geq a_{t-s+1} - s,$$

and so

$$f(G) = f(G') + e(G') \geq a'_s + a_{t-s+1} - s = a_s + a_{t-s+1} - 2s.$$

It is sufficient to show that $a_s + a_{t-s+1} - 2s \geq a_{t+1}$ since it would contradict the fact that $f(G) < a_{t+1}$. The inequality is equivalent to

$$[s(n-1) - s^2] + [(t-s+1)(n-1) - (t-s+1)^2] - 2s \geq (t+1)(n-1) - (t+1)^2,$$

i.e.

$$(t+1)^2 - (t-s+1)^2 \geq s^2 + 2s.$$

That is, $2st \geq 2s^2$ which holds for $0 \leq s \leq t$ as required. □

Since F_n is symmetric, we automatically have, by taking complements, the corresponding result for $f(G) \in [\binom{n}{3} - n^{3/2}, \binom{n}{3}]$.

Corollary 4.3.2. *Let G be a graph on $n \geq 3$ vertices.*

(i) If $f(G) > \binom{n}{3} - a_{t_{max}+1}$ then $f(G) \in [\binom{n}{3} - b_t, \binom{n}{3} - a_t]$ for a unique $t \leq t_{max}$.

(ii) If $t \leq t_{max}$ then $f(G) \in [\binom{n}{3} - b_t, \binom{n}{3} - a_t]$ iff G can be obtained from a disjoint union of two cliques of orders summing to n by flipping exactly t pairs of vertices. □

4.4 Proof of Theorem 4.1.2

Before we prove Theorem 4.1.2, we shall observe that the number of frustrated triangles satisfies the following parity condition.

Proposition 4.4.1. *Let G be a graph on n vertices. Then*

- *$f(G)$ is even for n even.*
- *$f(G)$ has the same parity as $e(G)$ for n odd.*

Proof. Given an edge xy of G, let $V_1 = \{v \in G\backslash\{x,y\} : vxy \text{ is frustrated}\}$ be the set of vertices forming a frustrated triangle with xy, and let
$V_2 = \{v \in G\backslash\{x,y\} : vxy \text{ is not frustrated}\}$. Deleting the edge xy changes the parity of the number of edges induced by vxy. Therefore, $f(G-xy) = f(G) - |V_1| + |V_2|$. Since $|V_1|+|V_2| = n-2$, we have

$$f(G-xy) \equiv f(G)+n \mod 2.$$

If n is even, we see that $f(G) \equiv f(G-e_1) \equiv f(G-e_1-e_2) \equiv \cdots \equiv f(E_n) = 0 \mod 2$. If n is odd, we see that
$f(G) \equiv f(G-e_1)+1 \equiv f(G-e_1-e_2)+2 \equiv \cdots \equiv f(E_n)+e(G) = e(G) \mod 2.$ □

We also see from the proof that $|f(G-e)-f(G)| \leq n-2$; therefore, F_n can miss at most $n-3$ consecutive integers. The next corollary follows from Proposition 4.4.1 and Theorem 4.1.1.

Corollary 4.4.2. *For $t \leq t_{max}$, we have*

$$F_n \cap [a_t, b_t] \subset \{a_t, a_t+2, \ldots, b_t-2, b_t\}.$$

Proof. If n is even, this follows immediately from Proposition 4.4.1 since a_t, b_t are even. Let n be odd and let G be a graph on n vertices with $f(G) \in [a_t, b_t]$ for some $t \leq t_{max}$. By

Theorem 4.1.1 part (ii), G can be obtained from a complete bipartite graph by flipping exactly t pairs of vertices. Equivalently, $G \sim H$ for some graph H with $e(H) = t$. Hence, by Proposition 4.4.1, we have

$$f(G) = f(H) \equiv t \equiv a_t \equiv b_t \mod 2$$

as required. □

We shall now prove Theorem 4.1.2 part (i). The proof is by construction of graphs consisting of four parts. By modifying the first part of the graphs, we obtain a sequence of even numbers belonging to F_n in the required interval with gaps at most $n-2$. Next the modification of the second part refines the partition of such interval such that the gaps are now at most $2(\sqrt{n}-1)$. Then the third part reduces the gaps to at most $2(\sqrt[4]{4n}-1)$. Finally, we modify the fourth part to obtain all even numbers in the interval.

Proof of Theorem 4.1.2 part (i). Let n be even. First, we shall note that this is best possible up to the second order term. Let s be the maximum t such that $b_t + 2 < a_{t+1}$ and so $s \sim \sqrt{n}$. Since $b_s + 2 \in (b_s, a_{s+1})$, we have $b_s + 2 = s(n-2) + 2 \sim n^{3/2}$ is even but is not a member of F_n by Theorem 4.3.1.

Now, for each even number $m \in [n^{3/2} + 2\sqrt{2}n^{5/4}, \binom{n}{3} - (n^{3/2} + 2\sqrt{2}n^{5/4})]$, our aim is to construct a graph G on n vertices with $f(G) = m$. Since F_n is symmetric about $\frac{1}{2}\binom{n}{3}$, it is sufficient to do so for each even number $m \in [n^{3/2} + 2\sqrt{2}n^{5/4}, \frac{1}{2}\binom{n}{3}]$. Let G be a graph on n vertices containing $\lceil \sqrt{n} \rceil + 2\lceil \sqrt[4]{4n} \rceil$ independent edges. For simplicity, we shall write $s_1 = \lceil \sqrt{n} \rceil$ and $s_2 = \lceil \sqrt[4]{4n} \rceil$. Let H be a graph obtained from G by adding all the edges between the isolated vertices of G, i.e. H is a disjoint union of a complete graph of order $r = n - 2s_1 - 4s_2$ and a matching of size $s_1 + 2s_2$. Then $f(G) = (s_1 + 2s_2)(n-2)$ and $f(H) = \binom{r}{3} + \binom{r}{2}(n-r) + (s_1 + 2s_2)(n-2)$. It is sufficient to show that every even number between $f(G)$ and $f(H)$ belongs to F_n since $f(G) \leq n^{3/2} + 2\sqrt{2}n^{5/4}$ and $f(H) \geq \binom{n-2\sqrt{n}-4\sqrt[4]{4n}}{3} \geq \frac{1}{2}\binom{n}{3}$ for sufficiently large n.

We shall break G into four parts and modify each part separately to obtain new graphs. Let $V(G) = V_1 \sqcup V_2 \sqcup V_3 \sqcup V_4$ where

- $G[V_1]$ is empty with $|V_1| = r$,
- $G[V_2]$ is a matching of size $\sqrt{n}$ with $|V_2| = 2\sqrt{n}$, and
- $G[V_3]$ and $G[V_4]$ are matchings of size $\sqrt[4]{4n}$ with $|V_3| = |V_4| = 2\sqrt[4]{4n}$, and
- $G[V_i, V_j]$ is empty for all $i \neq j$.

Now we add an edge one by one inside $G[V_1]$ until we obtain H. Each time we add an edge, f can change by at most $n-2$ by the proof of Proposition 4.4.1. Hence, F_n contains a sequence of even numbers $f(G) = f(G_1), f(G_2), \ldots, f(G_{\binom{r}{2}}) = f(H)$ with $|f(G_i) - f(G_{i+1})| \leq n-2$ for all i. Therefore, it is sufficient to show that F_n contains every even number between $f(G_i)$ and $f(G_i) - (n-2)$ for all i.

Let us fix i. By construction, $V(G_i) = V_1 \sqcup V_2 \sqcup V_3 \sqcup V_4$ where V_1 induces i edges, V_2 induces a matching of size $\sqrt{n}$ and each of V_3, V_4 induces a matching of size $\sqrt[4]{4n}$. We shall modify $G_i[V_2], G_i[V_3]$ and $G_i[V_4]$ to obtain new graphs. Let $\{x_1y_1, \ldots, x_{\sqrt{n}}y_{\sqrt{n}}\}$ be the matching inside V_2. First, we delete x_2y_2 and replace it with x_1y_2. This decreases f by 2. Next, we delete x_3y_3 and replace it with x_1y_3 which decreases f by 4 more. We continue similarly (see Figure 4.3). When we delete x_jy_j and replace it with x_1y_j, it decreases f by $2(j-1)$.

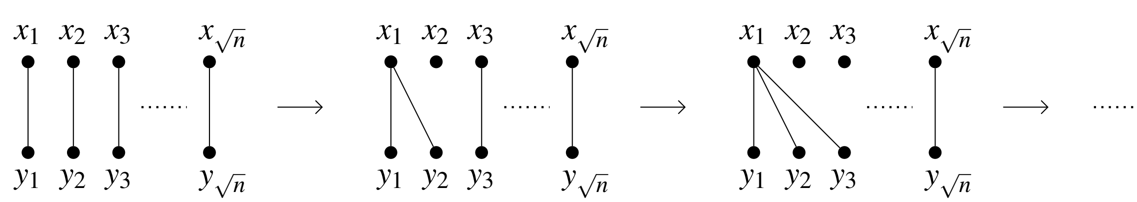

Figure 4.3: Star accumulation of V_2.

Therefore, F_n contains a decreasing sequence of even numbers $f(G_i) = f(G_{i,1}), f(G_{i,2}), \ldots, f(G_{i,\sqrt{n}})$ with $|f(G_{i,j}) - f(G_{i,j+1})| \leq 2(\sqrt{n}-1)$ for all j.

Moreover, $f(G_{i,\sqrt{n}})$ is larger than $f(G_i)-(n-2)$ by at most $2(\sqrt{n}-1)$. Indeed,

$$
\begin{aligned}
f(G_{i,\lceil\sqrt{n}\rceil}) &= f(G_i)-(2+4+\cdots+2(\lceil\sqrt{n}\rceil-1)) \\
&= f(G_i)-(\lceil\sqrt{n}\rceil-1)\lceil\sqrt{n}\rceil \\
&\le f(G_i)-(n-2)+2(\lceil\sqrt{n}\rceil-1).
\end{aligned}
$$

Hence, it is sufficient to show that F_n contains every even number between $f(G_{i,j})$ and $f(G_{i,j})-2(\sqrt{n}-1)$ for all j. Let us fix j and we shall modify the third part of the graph $G_{i,j}$ as follows. Let $\{x'_1y'_1,\ldots,x'_{\sqrt[4]{4n}}y'_{\sqrt[4]{4n}}\}$ be the matching inside V_3. We perform the same 'star' procedure as the second part. First, we delete $x'_2y'_2$ and replace it with $x'_1y'_2$. This decreases f by 2. Next, we delete $x'_3y'_3$ and replace it with $x'_1y'_3$ which decreases f by 4 more. We continue similarly. When we delete $x'_ky'_k$ and replace it with $x'_1y'_k$, it decreases f by $2(k-1)$.

Therefore, F_n contains a decreasing sequence of even numbers $f(G_{i,j})=f(G_{i,j,1}),f(G_{i,j,2}),\ldots,f(G_{i,j,\sqrt[4]{4n}})$ with $|f(G_{i,j,k})-f(G_{i,j,k+1})|\le 2(\sqrt[4]{4n}-1)$ for all k. Moreover, $f(G_{i,j,\sqrt[4]{4n}})$ is larger than $f(G_{i,j})-2(\sqrt{n}-1)$ by at most $2(\sqrt[4]{4n}-1)$. Indeed,

$$
\begin{aligned}
f(G_{i,j,\lceil\sqrt[4]{4n}\rceil}) &= f(G_{i,j})-(2+4+\cdots+2(\lceil\sqrt[4]{4n}\rceil-1)) \\
&= f(G_{i,j})-(\lceil\sqrt[4]{4n}\rceil-1)\lceil\sqrt[4]{4n}\rceil \\
&\le f(G_{i,j})-2(\lceil\sqrt{n}\rceil-1)+2(\lceil\sqrt[4]{4n}\rceil-1).
\end{aligned}
$$

Hence, it is sufficient to show that F_n contains every even number between $f(G_{i,j,k})$ and $f(G_{i,j,k})-2(\sqrt[4]{4n}-1)$ for all k. To prove this, let us fix k and we shall modify the fourth part of the graph $G_{i,j,k}$ as follows. Let $\{z_1w_1,\ldots,z_{\sqrt[4]{4n}}w_{\sqrt[4]{4n}}\}$ be the matching inside V_4. First, we delete z_2w_2 and replace it with w_1w_2. This decreases f by 2. Next, we delete z_3w_3 and replace it with w_2w_3 which decreases f by 2 more. We continue similarly (see

Figure 4.4). When we delete $z_l w_l$ and replace it with $w_{l-1}w_l$, it decreases f by 2.

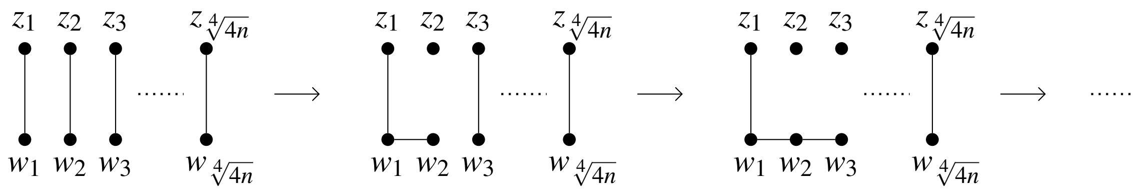

Figure 4.4: Path accumulation of V_4.

Since the modification has $\sqrt[4]{4n}-1$ steps, we conclude that F_n contains every even number between $f(G_{i,j,k})$ and $f(G_{i,j,k})-2(\sqrt[4]{4n}-1)$ as required. □

The proof for part (ii) is by the same method but with more care. Observe that, for n odd, the sequence obtained by adding edges to the first part alternates between odd and even numbers. Since the odd and even subsequences have larger gaps than before, of size at most $2(n-2)$, we shall follow the same proof for each subsequence by taking larger matchings.

Proof of Theorem 4.1.2 part (ii). Let n be odd. First, we shall note that this is best possible up to the second order term. Let s be the maximum t such that $b_{t-1}+2 < a_{t+1}$ and so $s \sim \sqrt{2n}$. We claim that at least one of $b_{s-1}+1$ and $b_{s-1}+2$ is not a member of F_n. Suppose for contradiction that they are both in F_n. Since they lie in (b_{s-1}, b_{s+1}), it follows from Theorem 4.3.1 that there are two graphs with n vertices and s edges containing $b_{s-1}+1$ and $b_{s-1}+2$ frustrated triangles respectively contradicting Proposition 4.4.1. We are then done since $b_{s-1} = (s-1)(n-2) \sim \sqrt{2}n^{3/2}$.

Now, for each number $m \in [\sqrt{2}n^{3/2}+2\sqrt[4]{8}n^{5/4}, \binom{n}{3}-(\sqrt{2}n^{3/2}+2\sqrt[4]{8}n^{5/4})]$, our aim is to construct a graph G on n vertices with $f(G)=m$. Since F_n is symmetric about $\frac{1}{2}\binom{n}{3}$, it is sufficient to do so for each $m \in [\sqrt{2}n^{3/2}+2\sqrt[4]{8}n^{5/4}, \frac{1}{2}\binom{n}{3}]$. Let G be a graph on n vertices containing $\lceil\sqrt{2n}\rceil+2\lceil\sqrt[4]{8n}\rceil$ independent edges. For simplicity, we shall write $s_1 = \lceil\sqrt{2n}\rceil$ and $s_2 = \lceil\sqrt[4]{8n}\rceil$. Let H be a graph obtained from G by adding all the edges between the isolated vertices of G, i.e. H is a disjoint union of a complete graph of order

$r = n - 2s_1 - 4s_2$ and a matching of size $s_1 + 2s_2$. Then $f(G) = (s_1 + 2s_2)(n-2)$ and $f(H) = \binom{r}{3} + \binom{r}{2}(n-r) + (s_1+2s_2)(n-2)$. It is sufficient to show that every number between $f(G)$ and $f(H) - (n-2)$ belongs to F_n since $f(G) \leq \sqrt{2}n^{3/2} + 2\sqrt[4]{8}n^{5/4}$ and $f(H) - (n-2) \geq \binom{n-2\sqrt{2n}-4\sqrt[4]{8n}}{3} \geq \frac{1}{2}\binom{n}{3}$ for sufficiently large n.

We shall break G into four parts and modify each part separately to obtain new graphs. Let $V(G) = V_1 \sqcup V_2 \sqcup V_3 \sqcup V_4$ where

- $G[V_1]$ is empty with $|V_1| = r$,
- $G[V_2]$ is matching of size $\sqrt{2n}$ with $|V_2| = 2\sqrt{2n}$, and
- $G[V_3]$ and $G[V_4]$ are matchings of size $\sqrt[4]{8n}$ with $|V_3| = |V_4| = 2\sqrt[4]{8n}$, and
- $G[V_i, V_j]$ is empty for all $i \neq j$.

Now we add an edge one by one inside $G[V_1]$ until we obtain H. Each time we add an edge, f can change by at most $n-2$ by the proof of Proposition 4.4.1. Hence, F_n contains a sequence $f(G) = f(G_1), f(G_2), \ldots, f(G_{\binom{r}{2}}) = f(H)$ with $|f(G_i) - f(G_{i+1})| \leq n-2$ for all i. By Proposition 4.4.1, this sequence alternates between odd and even numbers. We claim that it is sufficient to show that F_n contains

$$f(G_i), f(G_i) - 2, f(G_i) - 4, \ldots, f(G_i) - 2(n-2)$$

for all i. Indeed, let $m \in [f(G), f(H) - (n-2)]$. Then there is an i such that $f(G_i)$ has the same parity as m and $0 \leq f(G_i) - m \leq 2(n-2)$ since $|f(G_j) - f(G_{j+2})| \leq 2(n-2)$ and $f(G_j), f(G_{j+2})$ have the same parity for all j. Hence, $m \in \{f(G_i), f(G_i) - 2, f(G_i) - 4, \ldots, f(G_i) - 2(n-2)\} \subset F_n$ as required. The rest of the proof is similar to the previous proof.

Let us fix i. By construction, $V(G_i) = V_1 \sqcup V_2 \sqcup V_3 \sqcup V_4$ where V_1 induces i edges, V_2 induces a matching of size $\sqrt{2n}$ and each of V_3, V_4 induces a matching of size $\sqrt[4]{8n}$. We shall modify $G_i[V_2], G_i[V_3]$ and $G_i[V_4]$ to obtain new graphs. Let $\{x_1y_1, \ldots, x_{\sqrt{2n}}y_{\sqrt{2n}}\}$ be the matching inside V_2. First, we delete x_2y_2 and replace it with x_1y_2. This decreases f by 2. Next, we delete x_3y_3 and replace it with x_1y_3 which decreases f by 4 more. We continue similarly. When we delete x_jy_j and replace it with x_1y_j, it decreases f by $2(j-1)$.

Therefore, F_n contains a decreasing sequence $f(G_i) = f(G_{i,1}), f(G_{i,2}), \ldots, f(G_{i,\sqrt{2n}})$ with $|f(G_{i,j}) - f(G_{i,j+1})| \leq 2(\sqrt{2n}-1)$ for all j. Moreover, $f(G_{i,\sqrt{2n}})$ is larger than $f(G_i) - 2(n-2)$ by at most $2(\sqrt{2n}-1)$. Indeed,

$$
\begin{aligned}
f(G_{i,\lceil\sqrt{2n}\rceil}) &= f(G_i) - (2+4+\cdots+2(\lceil\sqrt{2n}\rceil-1)) \\
&= f(G_i) - (\lceil\sqrt{2n}\rceil-1)\lceil\sqrt{2n}\rceil \\
&\leq f(G_i) - 2(n-2) + 2(\lceil\sqrt{2n}\rceil-1).
\end{aligned}
$$

Hence, it is sufficient to show that F_n contains

$$
f(G_{i,j}), f(G_{i,j})-2, f(G_{i,j})-4, \ldots, f(G_{i,j})-2(\sqrt{2n}-1)
$$

for all j. Let us fix j and we shall modify the third part of the graph $G_{i,j}$ as follows. Let $\{x'_1y'_1, \ldots, x'_{\sqrt[4]{8n}}y'_{\sqrt[4]{8n}}\}$ be the matching inside V_3. First, we delete $x'_2y'_2$ and replace it with $x'_1y'_2$. This decreases f by 2. Next, we delete $x'_3y'_3$ and replace it with $x'_1y'_3$ which decreases f by 4 more. We continue similarly. When we delete $x'_ky'_k$ and replace it with $x'_1y'_k$, it decreases f by $2(k-1)$.

Therefore, F_n contains a decreasing sequence $f(G_{i,j}) = f(G_{i,j,1}), f(G_{i,j,2}), \ldots, f(G_{i,j,\sqrt[4]{8n}})$ with $|f(G_{i,j,k}) - f(G_{i,j,k+1})| \leq 2(\sqrt[4]{8n}-1)$ for all k. Moreover, $f(G_{i,j,\sqrt[4]{8n}})$ is larger than $f(G_{i,j}) - 2(\sqrt{2n}-1)$ by at most $2(\sqrt[4]{8n}-1)$. Indeed,

$$
\begin{aligned}
f(G_{i,j,\lceil\sqrt[4]{8n}\rceil}) &= f(G_{i,j}) - (2+4+\cdots+2(\lceil\sqrt[4]{8n}\rceil-1)) \\
&= f(G_{i,j}) - (\lceil\sqrt[4]{8n}\rceil-1)\lceil\sqrt[4]{8n}\rceil
\end{aligned}
$$

$$\leq f(G_{i,j}) - 2(\lceil\sqrt{2n}\rceil - 1) + 2(\lceil\sqrt[4]{8n}\rceil - 1).$$

Hence, it is sufficient to show that F_n contains

$$f(G_{i,j,k}), f(G_{i,j,k}) - 2, f(G_{i,j,k}) - 4, \ldots, f(G_{i,j,k}) - 2(\sqrt[4]{8n} - 1)$$

for all k. To prove this, let us fix k and we shall modify the fourth part of the graph $G_{i,j,k}$ as follows. Let $\{z_1w_1, \ldots, z_{\sqrt[4]{8n}}w_{\sqrt[4]{8n}}\}$ be the matching inside V_4. First, we delete z_2w_2 and replace it with w_1w_2. This decreases f by 2. Next, we delete z_3w_3 and replace it with w_2w_3 which decreases f by 2 more. We continue similarly. When we delete z_lw_l and replace it with $w_{l-1}w_l$, it decreases f by 2. We are done since the modification has $\sqrt[4]{8n} - 1$ steps. □

4.5 Proof of Theorem 4.1.3

We have seen a special case of Theorem 4.1.3 from the proof of Proposition 4.2.5 part (iii) that, amongst the graphs with n vertices and $e \leq n-1$ edges, the e-star is the only graph with the minimum number of frustrated triangles. We deduce Theorem 4.1.3 using Theorem 4.3.1 together with this fact.

Proof of Theorem 4.1.3. By Proposition 4.2.6, we see that G can be obtained from a complete bipartite graph by flipping t_G pairs of vertices. Therefore, $e \in [x(n-x) - t_G, x(n-x) + t_G]$ for some $0 \leq x \leq \frac{n}{2}$, and hence $g(e) \leq t_G$ by the definition of g. Now, by Theorem 4.3.1 part (i), we have $f(G) \geq a_{g(e)}$ as required.

Next, we show that there is a graph G with n vertices and $e \leq \lfloor \frac{n^2}{4} \rfloor + \lfloor \frac{n-1}{2} \rfloor - 1$ edges such that $f(G) = a_{g(e)}$. We see that $g(e)$, the distance from e to the sequence (c_x), is at most $\lfloor \frac{n-1}{2} \rfloor$. By the definition of g, there is $0 \leq x \leq \frac{n}{2}$ such that $g(e) = |e - c_x|$. We shall distinguish two cases.

If $e = c_x - g(e)$ then let G be the graph obtained from the complete bipartite graph

$K_{x,n-x}$ by deleting a $g(e)$-star with center in the smaller side and leaves in the larger side of $K_{x,n-x}$. This is possible since $g(e) \leq \lfloor \frac{n-1}{2} \rfloor \leq n-x$, and if $x = 0$ then $g(e) = 0$. By Corollary 4.2.7, $G \sim g(e)$-star, and hence $f(G) = a_{g(e)}$.

On the other hand, if $e = c_x + g(e)$ then let G be the graph obtained from the complete bipartite graph $K_{x,n-x}$ by adding a $g(e)$-star inside the larger side of $K_{x,n-x}$. This is possible since $g(e)+1 \leq \lfloor \frac{n-1}{2} \rfloor + 1 \leq n-x$. By Corollary 4.2.7, $G \sim g(e)$-star, and hence $f(G) = a_{g(e)}$.

We shall now show that these are the only extremal graphs. Suppose that G is a graph with n vertices and $e \leq \lfloor \frac{n^2}{4} \rfloor + \lfloor \frac{n-1}{2} \rfloor - 1$ edges such that $f(G) = a_{g(e)}$. We know that $g(e) \leq \lfloor \frac{n-1}{2} \rfloor$. First, we consider the case $g(e) = \lfloor \frac{n-1}{2} \rfloor$. This can only happen when $e = \lfloor \frac{n-1}{2} \rfloor$ or $\lceil \frac{n-1}{2} \rceil$. If $e = \lfloor \frac{n-1}{2} \rfloor$ then $e \leq n-1$ and so G is the e-star. Since $g(e) = \lfloor \frac{n-1}{2} \rfloor = e$, we see that G is obtained from $K_{0,n}$ by adding a $g(e)$-star. Similarly, if $e = \lceil \frac{n-1}{2} \rceil$ then $e \leq n-1$ and so G is the e-star. Since $g(e) = \lfloor \frac{n-1}{2} \rfloor = (n-1)-e$, we see that G is obtained from $K_{1,n}$ by deleting a $g(e)$-star.

Now we may assume that $g(e) \leq \lfloor \frac{n-1}{2} \rfloor - 1$. Therefore, $f(G) = a_{g(e)} < a_{g(e)+1}$ since a_t is increasing when $t \leq \frac{n-1}{2}$. By Theorem 4.3.1 part (i), we conclude that $t_G \leq g(e)$. Recall from the beginning of this proof that $t_G \geq g(e)$ and so we must have $t_G = g(e)$. By the definition of t_G, $G \sim H$ for some graph H with $g(e)$ edges. Since $f(H) = f(G) = f(g(e)\text{-star})$ and $g(e) \leq n-1$, we conclude that H must be the $g(e)$-star. By Corollary 4.2.7, we have that G can be obtained from a complete bipartite graph by flipping $g(e)$ pairs of vertices forming a star. Since $g(e)$ is the distance from e to the sequence of number of edges of a complete bipartite graph, the $g(e)$ pairs of vertices that we flip must be all edges or all nonedges. □

Let us remark that there are at most two extremal graphs for each $e \leq \lfloor \frac{n^2}{4} \rfloor + \lfloor \frac{n-1}{2} \rfloor - 1$. Indeed, there are at most two c_x's such that $g(e) = |e - c_x|$. The size of the complete bipartite graph is determined by this c_x while the choice of deleting or adding edges is determined by the sign of $e - c_x$.

Since F_n is symmetric, we have, by taking complements, the corresponding result for maximising the number of frustrated triangles amongst the graphs with a fixed number of edges.

Corollary 4.5.1. *If G is a graph on n vertices with e edges then* $f(G) \le \binom{n}{3} - a_{g\left(\binom{n}{3}-e\right)}$. *Moreover, the bound can be achieved when* $e \ge \binom{n}{3} - \lfloor \frac{n^2}{4} \rfloor - \lfloor \frac{n-1}{2} \rfloor + 1$. *In this case, the extremal graphs are obtained from a disjoint union of two complete graphs of orders summing to n by deleting or adding* $g\left(\binom{n}{3} - e\right)$ *edges forming a star.* □

4.6 Open problems

We conclude by mentioning questions and conjectures that would merit further study. Let $f(n)$ be the maximum nonmember of F_n which is less than $\frac{1}{2}\binom{n}{3}$. Theorem 4.1.2 shows that $f(n) = n^{3/2} + O(n^{5/4})$ for n even and $f(n) = \sqrt{2}n^{3/2} + O(n^{5/4})$ for n odd. A careful modification of the construction could solve the following problem.

Problem 4.6.1. *Determine the second order term of* $f(n)$ *in both cases.*

Let $t \le t_{max}$. Combining Theorem 4.1.1 part (ii) and Corollary 4.2.7, we have

$$F_n \cap [a_t, b_t] = \{f(G) : e(G) = t\}.$$

Corollary 4.4.2 tells us that $F_n \cap [a_t, b_t] \subset \{a_t, a_t + 2, \ldots, b_t - 2, b_t\}$. However, it is not true that every number in $\{a_t, a_t + 2, \ldots, b_t - 2, b_t\}$ appears in F_n. For example, by considering all graphs with 4 edges, we see that
$\{f(G) : e(G) = 4\} = \{a_4, a_4 + 2, \ldots, b_4 - 2, b_4\} - \{a_4 + 2\}$. Therefore, we ask the following question.

Problem 4.6.2. *Determine the set* $[a_t, b_t] - \{f(G) : e(G) = t\}$ *for each* $t \le t_{max}$.

We have seen from the proof of Proposition 4.2.5 part (iii) that, amongst the graphs with n vertices and $e \le n/2$ edges, the e-matching is the only graph with the maximum number of frustrated triangles. Furthermore, Theorem 4.1.3 and Corollary 4.5.1 partially answer the following question.

Problem 4.6.3. *Given the number of vertices and the number of edges, which graphs on n vertices maximise/minimise the number of frustrated triangles?*

Note that extremal graphs with minimum number of triangles of Rozborov [57] provide an asymptotic answer to this question.

There are several ways one could generalize the definition of frustrated triangles. For instance, we can replace triangle with another subgraph. The most natural generalization is to cycles. For $k \geq 3$ and vertices $v_1, v_2, \ldots, v_k$ of a graph G, we say that a cyclic ordering $v_1 v_2 \ldots v_k$ is a *frustrated k-cycle* if

$$\left|\{v_1v_2, v_2v_3, \ldots, v_{k-1}v_k, v_kv_1\} \cap E(G)\right|$$

is odd. Let $f_k(G)$ be the number of frustrated k-cycles in a graph G. We conjecture the following generalization of Theorem 4.3.1.

Conjecture 4.6.4. *For every $k \geq 3$, $t \geq 0$, and all sufficiently large n, the following holds. If $f_k(G) < f_k\big((t+1)$-star on n vertices$\big)$, then either G or $\overline{G}$ can be obtained from a complete bipartite graph by flipping at most t edges/non-edges.*

Note that when n is odd, this should hold for just G instead of 'G or $\overline{G}$'.

We could also try to define frustrated triangles in hypergraphs and study the analogue set. One thing to note is that if we wish to proceed along the same method then our new definition should at least give us Lemma 4.2.1.

REFERENCES

[1] S. S. Ahuja, S. Ramasubramanian, and M. M. Krunz. Single-link failure detection in all-optical networks using monitoring cycles and paths. *IEEE/ACM Trans. Netw.*, 17:1080–1093, 2009.

[2] N. Alon and B. Bollobás. Graphs with a small number of distinct induced subgraphs. *Discrete Math.*, 75:23–30, 1989.

[3] N. Alon and Joel H. Spencer. *The probabilistic method*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., Hoboken, NJ, 3$^{\text{rd}}$ edition, 2008.

[4] M. Axenovich, R. R. Martin, and T. Ueckerdt. Twins in graphs. *European J. Combin.*, 39:188–197, 2014.

[5] J. Balogh, P. Hu, B. Lidický, F. Pfender, J. Volec, and M. Young. Rainbow triangles in three-colored graphs. *preprint*, 2014. arXiv: 1408.5296.

[6] Zs. Baranyai. On the factorization of the complete uniform hypergraph. In *Infinite and finite sets (Colloq., Keszthely, 1973; dedicated to P. Erdős on his 60th birthday)*, volume 10 of *Colloq. Math. Soc. János Bolyai*, pages 91–108. North-Holland, Amsterdam, 1975.

[7] I. Ben-Eliezer and M. Krivelevich. Perfectly balanced partitions of smoothed graphs. *Electron. J. Combin.*, 16:Note 14, 2009.

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

[9] B. Bollobás. *Random graphs*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2$^{\text{nd}}$ edition, 2001.

[10] B. Bollobás, T. Kittipassorn, B. P. Narayanan, and A. D. Scott. Disjoint induced subgraphs of the same order and size. *European J. Combin.*, to appear, 2014.

[11] B. Bollobás and A. D. Scott. On separating systems. *European J. Combin.*, 28:1068–1071, 2007.

[12] B. Bollobás and A. D. Scott. Separating systems and oriented graphs of diameter two. *J. Combin. Theory Ser. B*, 97:193–203, 2007.

[13] B. Bollobás and A. Thomason. Graphs which contain all small graphs. *European J. Combin.*, 2:13–15, 1981.

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

[15] J. A. Bondy. Perfect path double covers of graphs. *J. Graph Theory*, 14:259–272, 1990.

[16] M. Cai. On separating systems of graphs. *Discrete Math.*, 49:15–20, 1984.

[17] Y. Caro and R. Yuster. Large disjoint subgraphs with the same order and size. *European J. Combin.*, 30:813–821, 2009.

[18] M. Chudnovsky and S. Safra. The Erdős-Hajnal conjecture for bull-free graphs. *J. Combin. Theory Ser. B*, 98:1301–1310, 2008.

[19] F. R. K. Chung, R. L. Graham, and R. M. Wilson. Quasi-random graphs. *Combinatorica*, 9:345–362, 1989.

[20] J. Cummings, D. Král', F. Pfender, K. Sperfeld, A. Treglown, and M. Young. Monochromatic triangles in three-coloured graphs. *J. Combin. Theory Ser. B*, 103(4):489–503, 2013.

[21] P. Erdős and R. Rado. A combinatorial theorem. *J. London Math. Soc.*, 25:249–255, 1950.

[22] P. Erdős. On a theorem of Rademacher-Turán. *Illinois J. Math.*, 6:122–127, 1962.

[23] P. Erdős and A. Hajnal. On the number of distinct induced subgraphs of a graph. *Discrete Math.*, 75:145–154, 1989.

[24] P. Erdős and A. Hajnal. Ramsey-type theorems. *Discrete Appl. Math.*, 25:37–52, 1989.

[25] P. Erdős, J. Pach, and L. Pyber. Isomorphic subgraphs in a graph. In *Combinatorics (Eger, 1987)*, volume 52 of *Colloq. Math. Soc. János Bolyai*, pages 553–556. North-Holland, Amsterdam, 1988.

[26] M. Erickson. A conjecture concerning Ramsey's theorem. *Discrete Math.*, 126:395–398, 1994.

[27] V. Falgas-Ravry, T. Kittipassorn, D. Korándi, S. Letzter, and B. P. Narayanan. Separating path systems. *J. Comb.*, 5(3):335–354, 2014.

[28] C.W. Fay. *Statistical Mechanics of Vertex Cover*. Michigan State University. Department of Physics and Astronomy, 2007.

[29] F. Foucaud and M. Kovše. Identifying path covers in graphs. *J. Discrete Algorithms*, 2013. To appear.

[30] J. Fox and B. Sudakov. Induced Ramsey-type theorems. *Adv. Math.*, 219:1771–1800, 2008.

[31] A. W. Goodman. On sets of acquaintances and strangers at any party. *Amer. Math. Monthly*, 66:778–783, 1959.

[32] A. W. Goodman. Triangles in a complete chromatic graph. *J. Austral. Math. Soc. Ser. A*, 39(1):86–93, 1985.

[33] R. L. Graham, B. L. Rothschild, and J. H. Spencer. *Ramsey theory*. Wiley-Interscience Series in Discrete Mathematics and Optimization. John Wiley & Sons, Inc., New York, 2nd edition, 1990.

[34] A. Gyárfás. Transitive edge coloring of graphs and dimension of lattices. *Combinatorica*, 22:479–496, 2002.

[35] G. Hansel. Nombre minimal de contacts de fermeture nécessaires pour réaliser une fonction booléenne symétrique de n variables. *C. R. Acad. Sci. Paris*, 258:6037–6040, 1964.

[36] N. J. A. Harvey, M. Patrascu, Yonggang W., S. Yekhanin, and V. W. S. Chan. Non-adaptive fault diagnosis for all-optical networks via combinatorial group testing on graphs. In *26th IEEE International Conference on Computer Communications*, pages 697–705, 2007.

[37] D. Hefetz and M. Tyomkyn. Universality of graphs with few triangles and anti-triangles. *preprint*, 2014. arXiv:1401.5735.

[38] N. Hindman, I. Leader, and D. Strauss. Forbidden distances in the rationals and the reals. *J. London Math. Soc.*, 73:273–286, 2006.

[39] G. O. H. Katona. On separating systems of a finite set. *J. Combinatorial Theory*, 1:174–194, 1966.

[40] T. Kittipassorn and G. Mészáros. Frustrated triangles. *Submitted*, 2015. arXiv: 1411.1749.

[41] T. Kittipassorn and B. P. Narayanan. Approximations to m-coloured complete infinite hypergraphs. *J. Graph Theory*, 2014. to appear.

[42] T. Kittipassorn and B. P. Narayanan. A canonical Ramsey theorem for exactly m-coloured complete subgraphs. *Combin. Probab. Comput.*, 23(1):102–115, 2014.

[43] A. Kündgen, D. Mubayi, and P. Tetali. Minimal completely separating systems of k-sets. *J. Combin. Theory Ser. A*, 93:192–198, 2001.

[44] C. Laflamme, N. W. Sauer, and V. Vuksanovic. Canonical partitions of universal structures. *Combinatorica*, 26:183–205, 2006.

[45] J. A. Larson. Counting canonical partitions in the random graph. *Combinatorica*, 28:659–678, 2008.

[46] C. Lee, P. Loh, and B. Sudakov. Self-similarity of graphs. *SIAM J. Discrete Math.*, 27:959–972, 2013.

[47] H. Li. Perfect path double covers in every simple graph. *J. Graph Theory*, 14:645–650, 1990.

[48] N. Linial and A. Morgenstern. Graphs with few 3-cliques and 3-anticliques are 3-universal. *J. Graph Theory*, 2014. to appear.

[49] L. Lovász. On covering of graphs. In *Theory of Graphs (Proc. Colloq., Tihany, 1966)*, pages 231–236. Academic Press, New York, 1968.

[50] L. Lovász and M. Simonovits. On the number of complete subgraphs of a graph. II. In *Studies in pure mathematics*, pages 459–495. Birkhäuser, Basel, 1983.

[51] Bhargav P. Narayanan. Exactly m-coloured complete infinite subgraphs. *J. Combin. Theory Ser. B*, 106:163–173, 2014.

[52] E. A. Nordhaus and B. M. Stewart. Triangles in an ordinary graph. *Canad. J. Math.*, 15:33–41, 1963.

[53] D. Olpp. A conjecture of Goodman and the multiplicities of graphs. *Australas. J. Combin.*, 14:267–282, 1996.

[54] H. J. Prömel and V. Rödl. Non-Ramsey graphs are $c\log n$-universal. *J. Combin. Theory Ser. A*, 88:379–384, 1999.

[55] C. Ramsay and I. T. Roberts. Minimal completely separating systems of sets. *Australas. J. Combin.*, 13:129–150, 1996.

[56] F. P. Ramsey. On a problem of formal logic. *Proc. London Math. Soc.*, 30:264–286, 1930.

[57] A. Razborov. On the minimal density of triangles in graphs. *Combin. Probab. Comput.*, 17(4):603–618, 2008.

[58] A. Rényi. On random generating elements of a finite Boolean algebra. *Acta Sci. Math. Szeged*, 22:75–81, 1961.

[59] J. H. Spencer. Minimal completely separating systems. *J. Combinatorial Theory*, 8:446–447, 1970.

[60] A. Stacey and P. Weidl. The existence of exactly m-coloured complete subgraphs. *J. Combin. Theory Ser. B*, 75:1–18, 1999.

[61] J. Tapolcai, L. Rónyai, and P. Ho. Link fault localization using bi-directional M-trails in all-optical mesh networks. *IEEE Transactions on Communications*, 61:291–300, 2013.

[62] R. Thomas and P. Wollan. An improved linear edge bound for graph linkages. *European J. Combin.*, 26:309–324, 2005.

[63] A. Thomason. Pseudorandom graphs. In *Random graphs ’85 (Poznań, 1985)*, volume 144 of *North-Holland Math. Stud.*, pages 307–331. North-Holland, Amsterdam, 1987.

[64] G. Toulouse. The frustration model. In *Modern Trends in the Theory of Condensed Matter*, volume 115 of *Lecture Notes in Physics*, pages 195–203. Springer Berlin Heidelberg, 1980.

[65] J Vannimenus and G Toulouse. Theory of the frustration effect. ii. ising spins on a square lattice. *Journal of Physics C: Solid State Physics*, 10(18):L537, 1977.

[66] I. Wegener. On separating systems whose elements are sets of at most *k* elements. *Discrete Math.*, 28:219–222, 1979.