Results in Extremal Graph and Hypergraph Theory

Zelealem Belaineh Yilma

May 2011

Department of Mathematical Sciences
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Oleg Pikhurko, Chair
Tom Bohman
Po-Shen Loh
Asaf Shapira

Submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Keywords: Erdős-Ko-Rado, Turán graphs, Szemerédi's Regularity Lemma, Color-critical Graphs, Supersaturation

Abstract

In graph theory, as in many fields of mathematics, one is often interested in finding the maxima or minima of certain functions and identifying the points of optimality. We consider a variety of functions on graphs and hypegraphs and determine the structures that optimize them.

A central problem in extremal (hyper)graph theory is that of finding the maximum number of edges in a (hyper)graph that does not contain a specified forbidden substructure. Given an integer n, we consider hypergraphs on n vertices that do not contain a *strong simplex*, a structure closely related to and containing a *simplex*. We determine that, for n sufficiently large, the number of edges is maximized by a *star*.

We denote by $F(G, r, k)$ the number of edge r-colorings of a graph G that do not contain a monochromatic clique of size k. Given an integer n, we consider the problem of maximizing this function over all graphs on n vertices. We determine that, for large n, the optimal structures are $(k-1)^2$-partite *Turán graphs* when $r = 4$ and $k \in \{3, 4\}$ are fixed.

We call a graph F *color-critical* if it contains an edge whose deletion reduces the chromatic number of F and denote by $F(H)$ the number of copies of the specified color-critical graph F that a graph H contains. Given integers n and m, we consider the minimum of $F(H)$ over all graphs H on n vertices and m edges. The *Turán number* of F, denoted $\mathrm{ex}(n, F)$, is the largest m for which the minimum of $F(H)$ is zero. We determine that the optimal structures are supergraphs of Turán graphs when n is large and $\mathrm{ex}(n, F) \leq m \leq \mathrm{ex}(n, F) + cn$ for some $c > 0$.

Contents

Chapter 1

Introduction

In this chapter, we briefly summarize our results.

1.1 Strong Simplices

We start by introducing some terms.

A *hypergraph* (or *family*) $\mathcal{F}$ is a collection of subsets of a ground set V; we refer to the elements of V as the *vertices* and the members of $\mathcal{F}$ as *hyperedges* (or *edges*) of $\mathcal{F}$. If all edges in $\mathcal{F}$ are k-element subsets of V for some fixed integer $k \geq 2$, we call $\mathcal{F}$ a *k-uniform hypergraph* (or a *k-graph* for short). For $X \subseteq V$, we use the shorthand $\binom{X}{k}$ to refer to the set of all subsets of X with exactly k elements and denote by $[n]$ the set $\{1, 2, \ldots, n\}$. $\mathcal{F}$ is a *star* if there exists an element x that lies in all edges of $\mathcal{F}$.

A problem of immense interest in extremal graph theory is determining the maximum number of edges a hypergraph can contain if it does not contain a specified forbidden configuration (or a set of forbidden configurations). One of the most important results in extremal combinatorics is the Erdős-Ko-Rado Theorem [EKR61] which states that if the members of $\mathcal{F} \subset \binom{[n]}{k}$ are pairwise intersecting, then $\mathcal{F} \leq \binom{n-1}{k-1}$.

In Chapter 2, we consider those hypergraphs that do not contain *strong simplices*, a configuration introduced (with historical motivation) by Mubayi [Mub07] and closely related to *simplices*. To be precise, a *d-simplex* is a collection of $d+1$ sets such that every d of them have nonempty intersection and the intersection of all of them is empty. A *strong d-simplex* is a collection of $d+2$ sets $A, A_1, \ldots, A_{d+1}$ such that $\{A_1, \ldots, A_{d+1}\}$ is a d-simplex, while A contains an element of each d-wise intersection. Answering a conjecture of Mubayi and Ramadurai [MR09], we prove for $k \geq d+2 \geq 2$ and n large, that a k-graph on n vertices not containing a d-simplex has at most $\binom{n-1}{k-1}$ edges. Furthermore, equality holds only for a star.

This is joint work with Tao Jiang and Oleg Pikhurko [*SIAM J. Disc. Math.*, 24 (2010) 1038–1045].

1.2 Monochromatic Cliques

Let $G = (V, E)$ be a graph. An *edge r-coloring* of G is a mapping $\chi : E \to [r]$. Given G, we wish to count the number of edge r-colorings of G that do not contain a monochromatic copy of K_k, the complete graph on k vertices. We denote this quantity by $F(G, r, k)$. For example, if G itself contains no copy of K_k, one easily observes that $F(G, r, k) = r^{|E|}$.

We consider the related quantity $F(n, r, k)$, the maximum of $F(G, r, k)$ over all graphs G of order n, introduced by Erdős and Rothschild [Erd74, Erd92]. Recent work by Alon, Balogh, Keevash and Sudakov [ABKS04], shows that for all $k \geq 3$, $r \in \{2, 3\}$ and n large enough, the unique graph achieving this maximum is the Turán graph $T_{k-1}(n)$, the complete $(k-1)$-partite graph on n vertices with parts of size $\lfloor \frac{n}{k-1} \rfloor$ or $\lceil \frac{n}{k-1} \rceil$. In particular, as $T_{k-1}(n)$ contains no copy of K_k, they determine that $F(n, r, k) = r^{t_{k-1}(n)}$, where $t_{k-1}(n)$ is the number of edges in $T_{k-1}(n)$.

Surprisingly, they show that one can do significantly better for $r \geq 4$; that is, $F(n, r, k)$ is exponentially larger than $r^{t_{k-1}(n)}$ for n large enough. In Chapter 3, we build upon their result to show that $F(n, 4, 3)$ and $F(n, 4, 4)$ are attained by $T_4(n)$ and $T_9(n)$, respectively.

This is joint work with Oleg Pikhurko (submitted to *J. London Math. Soc.*).

1.3 Supersaturated Graphs

As mentioned above, the Turán graph $T_{k-1}(n)$ contains no copy of K_k. Furthermore, the celebrated result of Turán [Tur41] states that the largest size of a K_k-free graph on n vertices is $t_{k-1}(n)$, with $T_{k-1}(n)$ being the unique graph attaining this value. In light of this, the *generalized Turán number* of a graph F, denoted $\mathrm{ex}(n, F)$, is the maximum number of edges in a graph on n vertices that does not contain a (not necessarily induced) copy of F.

Interestingly, $T_{k-1}(n)$ is the extremal graph for a larger class of forbidden graphs, namely, the class of *color-critical* graphs. A graph F is called *r-critical* if its chromatic number is $r+1$, but it contains an edge whose deletion reduces the chromatic number. A result of Simonovits [Sim68] states that if F is r-critical, then $\mathrm{ex}(n, F) = t_r(n)$ for large n.

Call a graph G *supersaturated* with respect to F, if G has n vertices and at least $\mathrm{ex}(n, F) + 1$ edges. It readily follows that $F(G)$, the number of copies of F contained in G, is at least one. However, typically, $F(G)$ is much larger than 1, and given integers n and $m > \mathrm{ex}(n, F)$, an interesting question is to determine the minimum of $F(G)$ over graphs G on n vertices and m edges. The earliest result in this field, due to Rademacher (1941, unpublished), states that we have at least $\lfloor n/2 \rfloor$ copies of K_3 if $m \geq \lfloor n^2/4 \rfloor + 1$.

In Chapter 4, we address this problem for color-critical graphs F and $m = \mathrm{ex}(n, F) + q$, where $q = O(n)$. We show that there exists a limiting constant $c_1(F) > 0$, such that, for all $\epsilon > 0$ and $q < (c_1(F) - \epsilon)n$, the extremal graph(s) for $F(H)$ may be obtained by adding q edges to the Turán graph $T_r(n)$. We prove that our bound is tight for many graphs (and classes of graphs) and give an example for which $c_1(F)$ is irrational. In addition, we provide a threshold $c_2(F)$ for the asymptotic optimality of graphs obtained from $T_r(n)$ by adding extra edges.

This is joint work with Oleg Pikhurko (in preparation).

Chapter 2

Set Systems without a Strong Simplex

2.1 Forbidden Configurations

We begin by recalling the following result of Erdős, Ko and Rado — one of the most important results in extremal combinatorics.

Theorem 2.1.1 (Erdős, Ko, and Rado [EKR61]). *Let $n \geq 2k$ and let $\mathcal{F} \subseteq \binom{[n]}{k}$ be an intersecting family. Then $|\mathcal{F}| \leq \binom{n-1}{k-1}$. If $n > 2k$ and equality holds, then $\mathcal{F}$ is a star.*

The forbidden configuration in Theorem 2.1.1 consists of a pair of disjoint sets. A generalization of this configuration, with geometric motivation, is as follows.

Definition 2.1.2. *Fix $d \geq 1$. A family of sets is* d-wise-intersecting *if every d of its members have nonempty intersection. A collection of $d+1$ sets $A_1, A_2, \ldots, A_{d+1}$ is a* d-dimensional simplex *(or a* d-simplex*) if it is d-wise-intersecting but not $(d+1)$-wise-intersecting (that is, $\cap_{i=1}^{d+1} A_i = \emptyset$).*

Note that a 1-simplex is a pair of disjoint edges, and Theorem 2.1.1 states that if $\mathcal{F} \subseteq \binom{[n]}{k}$ with $n \geq 2k$ and $|\mathcal{F}| > \binom{n-1}{k-1}$, then $\mathcal{F}$ contains a 1-simplex. In general, it is conjectured that the same threshold for $\mathcal{F}$ guarantees a d-simplex for every d, $1 \leq d \leq k-1$. For $d = 2$, this was a question of Erdős [Erd71], while the following general conjecture was formulated by Chvátal.

Conjecture 2.1.3 (Chvátal [Chv75]). *Suppose that $k \geq d+1 \geq 2$ and $n \geq k(d+1)/d$. If $\mathcal{F} \subseteq \binom{[n]}{k}$ contains no d-simplex, then $|\mathcal{F}| \leq \binom{n-1}{k-1}$. Equality holds only if $\mathcal{F}$ is a star.*

Another motivation (see [Chv75, page 358]) is that when we formally let $d = k$, then we obtain the famous open problem of finding the Turán function of the hypergraph $\binom{[k+1]}{k}$, posed by Turán [Tur41] in 1941.

Various partial results on the case $d = 2$ of the conjecture were obtained in [BF77, Chv75, CK99, Fra76, Fra81] until this case was completely settled by Mubayi and Verstraëte [MV05]. Conjecture 2.1.3 has been proved by Frankl and Füredi [FF87] for every fixed k, d if n is sufficiently large. Keevash and Mubayi [KM10] have also proved the conjecture when k/n and $n/2 - k$ are both bounded away from zero.

Mubayi [Mub07] proved a stability result for the case $d = 2$ of Conjecture 2.1.3 and conjectured that a similar result holds for larger d.

Conjecture 2.1.4 (Mubayi [Mub07])**.** *Fix $k \geq d+1 \geq 3$. For every $\delta > 0$, there exist $\epsilon > 0$ and $n_0 = n_0(\epsilon, k)$ such that the following holds for all $n > n_0$. If $\mathcal{F} \subseteq \binom{[n]}{k}$ contains no d-simplex and $|\mathcal{F}| > (1-\epsilon)\binom{n-1}{k-1}$, then there exists a set $S \subseteq [n]$ with $|S| = n-1$ such that $|\mathcal{F} \cap \binom{S}{k}| < \delta\binom{n-1}{k-1}$.*

Subsequently, Mubayi and Ramadurai [MR09] proved Conjecture 2.1.4 in a stronger form except in the case $k = d+1$, as follows.

Definition 2.1.5. *Fix $d \geq 1$. A collection of $d+2$ sets $A, A_1, A_2, \ldots, A_{d+1}$ is a* strong d-simplex *if $\{A_1, A_2, \ldots, A_{d+1}\}$ is a d-simplex and A contains an element of $\cap_{j \neq i} A_j$ for each $i \in [d+1]$.*

Note that a strong 1-simplex is a collection of three sets A, B, C such that $A \cap B$ and $B \cap C$ are nonempty, and $A \cap C$ is empty. Note also that if a family $\mathcal{F}$ contains no d-simplex, then certainly it contains no strong d-simplex (but not vice versa). The main result of Mubayi and Ramadurai [MR09] can be formulated using asymptotic notation as follows, where $o(1) \to 0$ as $n \to \infty$.

Theorem 2.1.6 (Mubayi and Ramadurai [MR09])**.** *Fix $k \geq d+2 \geq 3$. Let $\mathcal{F} \subseteq \binom{[n]}{k}$ contain no strong d-simplex. If $|\mathcal{F}| \geq (1-o(1))\binom{n-1}{k-1}$, then there exists an element $x \in [n]$ such that the number of sets of $\mathcal{F}$ omitting x is $o(n^{k-1})$.*

Corollary 2.1.7 (Mubayi and Ramadurai [MR09])**.** *Fix $k \geq d+2 \geq 3$. Let $\mathcal{F} \subseteq \binom{[n]}{k}$ contain no strong d-simplex. Then $|\mathcal{F}| \leq (1+o(1))\binom{n-1}{k-1}$ as $n \to \infty$.*

In [KM10], a similar stability result was proved when k/n and $n/2-k$ are both bounded away from 0, and the result was used to settle Conjecture 2.1.3 in this range of n.

Let us describe our contribution. First, we observe that Theorem 2.1.6 does not hold when $k = d+1$.

Proposition 2.1.8. *Let $k = d + 1 \geq 2$. For every $\epsilon > 0$ there is n_0 such that for all $n \geq n_0$ there is a k-graph $\mathcal{F}$ with n vertices and at least $(1-\epsilon)\binom{n-1}{k-1}$ edges without a strong d-simplex such that every vertex contains at most ϵn^{k-1} edges of $\mathcal{F}$.*

The authors of [MR09] pointed out that that they were unable to use Theorem 2.1.6 to prove the corresponding exact result for large n (which would give a new proof of the result of Frankl and Füredi [FF87]). They subsequently made the following conjecture, which is a strengthening of Chvátal's conjecture.

Conjecture 2.1.9 (Mubayi and Ramadurai [MR09])**.** *Let $k \geq d+1 \geq 3$, $n > k(d+1)/d$, and $\mathcal{F} \subseteq \binom{[n]}{k}$ contain no strong d-simplex. Then $|\mathcal{F}| \leq \binom{n-1}{k-1}$ with equality only for a star.*

In section 2.4, we will prove Conjecture 2.1.9 for all fixed $k \geq d+2 \geq 3$ and large n.

Theorem 2.1.10. *Let $k \geq d+2 \geq 3$ and let n be sufficiently large. If $\mathcal{F} \subseteq \binom{[n]}{k}$ contains no strong d-simplex, then $|\mathcal{F}| \leq \binom{n-1}{k-1}$ with equality only for a star.*

The case $k = d+1$ behaves somewhat differently from the general case $k \geq d+2$ in that by Proposition 2.1.8 there are almost extremal configurations very different from a star. We were able to prove the case $k = d+1$ of Conjecture 2.1.9 for all $n \geq 5$ when $k = 3$. However, shortly after our result, Feng and Liu [FL10] solved the case $k = d+1$ for all k, using a weight counting method used by Frankl and Füredi in [FF87]. Independently, Füredi [Für] has obtained the same proof, which is short and follows readily from the counting method. For these reasons, we have chosen to not include this proof in our paper and in this thesis.

Independently of us, Füredi and Özkahya [FÖ09] have re-proved our main result, Theorem 2.1.10, in a stronger form (for $k \geq d+2$ and large n). Namely, they can additionally guarantee that (in the notation of Definition 2.1.5) the sets $A_1 \setminus A, \ldots, A_{d+1} \setminus A$ are pairwise disjoint, while the sets $A \setminus A_i, \ldots, A \setminus A_{d+1}$ partition A and have any specified nonzero sizes. Füredi and Özkahya's proof uses a sophisticated version of the delta system method that has been developed in earlier papers such as [FF87] and [F8̈3]. Their method is very different from ours.

The problem of forbidding a d-simplex where we put some extra restrictions on the sizes of certain Boolean combinations of edges has also been studied before, with one particularly interesting paper being that of Csákány and Kahn [CK99], which uses a homological approach.

Frankl and Füredi's proof [FF87] of Chvátal's conjecture for a d-simplex for large n is very complicated. Together with the stability result in [MR09], we obtained a new proof of a stronger result. One key factor seems be that having a special edge A in a strong d-simplex $\{A, A_1, \ldots, A_{d+1}\}$ that contains an element in every d-wise intersection in the d-simplex $\{A_1 \ldots, A_{d+1}\}$ facilitates induction arguments very nicely. This observation, already made in [MR09], further justifies the interest in strong d-simplices.

2.2 Shadows

Let $\mathcal{F} \subseteq \binom{[n]}{k}$. Recall that a strong d-simplex L in $\mathcal{F}$ consists of $d+2$ hyperedges $A, A_1, A_2, \ldots, A_{d+1}$ such that every d of $A_1, \ldots, A_{d+1}$ have nonempty intersection but $\cap_{i=1}^{d+1} A_i = \emptyset$. Furthermore, A contains an element of $\cap_{j \neq i} A_i$ for each $i \in [d+1]$. This means that we can find some $d+1$ elements $v_1, v_2, \ldots, v_{d+1}$ in A such that for each $i \in [d+1]$, $v_i \in \cap_{j \neq i} A_j$. Note that $v_1, v_2, \ldots, v_{d+1}$ are distinct because no element lies in all of $A_1, \ldots, A_{d+1}$. We call A the *special edge* for L and the set $\{v_1, v_2, \ldots, v_{d+1}\}$ a *special* $(d+1)$-*tuple* for L. (Note that there may be more than one choice of a special $(d+1)$-tuple.)

As usual, the *degree* $d_{\mathcal{F}}(x)$ (or simply $d(x)$) of a vertex x in $\mathcal{F}$ is the number of hyperedges that contain x. In addition, let

$$\begin{aligned} \mathcal{F} - x &= \{D : D \in \mathcal{F}, x \notin D\}, \\ \mathcal{F}_x &= \{D \setminus \{x\} : D \in \mathcal{F}, x \in D\}. \end{aligned}$$

For a positive integer p, the *p-shadow* of $\mathcal{F}$ is defined as

$$\Delta_p(\mathcal{F}) = \{S \subseteq [n] : |S| = p, S \subseteq D \text{ for some } D \in \mathcal{F}\}.$$

Also, we let

$$\mathcal{T}_{p+1}(\mathcal{F}) = \{T : T \text{ is a special } (p+1)\text{-tuple for some strong } p\text{-simplex in } \mathcal{F}\}.$$

For each $p \in [k-2]$, let

$$\partial_p^*(\mathcal{F}) = |\Delta_p(\mathcal{F})| + |\mathcal{T}_{p+1}(\mathcal{F})|.$$

2.3 Proof of Proposition 2.1.8

We have to show that if $k = d+1$, then there is no stability. Let $\epsilon > 0$ be given. Choose large m such that $\binom{m-1}{k-1} > (1 - \epsilon/2)\binom{m}{k-1}$. Let the complete star $\mathcal{H} \subseteq \binom{[m]}{k}$ consist of all

k-tuples containing 1. Clearly, $\mathcal{H}$ has no $(k-1)$-simplex. Let $n \to \infty$. A result of Rödl [Röd85] shows that we can find an m-graph $\mathcal{F} \subseteq \binom{[n]}{m}$ with at least $(1-\epsilon/2)\binom{n}{k-1}/\binom{m}{k-1}$ edges such that every two edges of $\mathcal{F}$ intersect in at most $k-2$ vertices. Replace every edge of $\mathcal{F}$ by a copy of the star $\mathcal{H}$. Since no k-subset of $[n]$ is contained in two edges of $\mathcal{F}$, the obtained k-graph $\mathcal{G}$ is well defined.

Next, we observe that $\mathcal{G}$ has no strong $(k-1)$-simplex S. Indeed the special k-set X of S intersects every other edge of S in $k-1$ vertices; thus if X belongs to some copy of the star $\mathcal{H}$, then every other edge of S belongs to the same copy, a contradiction.

The size of $\mathcal{G}$ is at least $(1-\epsilon/2)\binom{n}{k-1}/\binom{m}{k-1} \times (1-\epsilon/2)\binom{m}{k-1} > (1-\epsilon)\binom{n}{k-1}$. Also, by the packing property of $\mathcal{F}$, the number of edges of $\mathcal{G}$ containing any one vertex is at most $\binom{n-1}{k-2}/\binom{m-1}{k-2} \times \binom{m-1}{k-1} < \epsilon n^{k-1}$ when n is large. This establishes Proposition 2.1.8.

2.4 Proof of Theorem 2.1.10

In order to prove Theorem 2.1.10, we first establish a general lower bound on $\partial_p^*(\mathcal{F})$ in Theorem 2.4.5, which is of independent interest. Then we will use Theorem 2.4.5 to prove Theorem 2.1.10.

We need several auxiliary lemmas. The first follows readily from Corollary 2.1.7.

Lemma 2.4.1. *For each $k \geq d+2 \geq 3$, there exists an integer $n_{k,d}$ such that for all integers $n \geq n_{k,d}$ if $\mathcal{H} \subseteq \binom{[n]}{k}$ contains no strong d-simplex, then $|\mathcal{H}| \leq 2\binom{n-1}{k-1}$.*

Lemma 2.4.2. *For every $k \geq p+2 \geq 3$, there exists a positive constant $\beta_{k,p}$ such that the following holds.*

Let $n_{k,p}$ be defined as in Lemma 2.4.1. Let $\mathcal{H}$ be a k-uniform hypergraph with $n \geq n_{k,p}$ vertices and $m > 4\binom{n-1}{k-1}$ edges. Then $|\mathcal{T}_{p+1}(\mathcal{H})| \geq \beta_{k,p} m^{\frac{p}{k-1}}$.

Proof. From $m > 4\binom{n-1}{k-1}$, we get $n < \lambda_k m^{\frac{1}{k-1}}$ for some constant λ_k depending only on k. Since $m > 4\binom{n-1}{k-1}$ and $n \geq n_{k,p}$, by Lemma 2.4.1, $\mathcal{H}$ contains a strong p-simplex L with A being its special edge. Let us remove the edge A from $\mathcal{H}$. As long as $\mathcal{H}$ still has more than $m/2 > 2\binom{n-1}{k-1}$ edges left, we can find another strong p-simplex and remove its special edge from the hypergraph. We can repeat this at least $m/2$ times. This produces at least $m/2$ different special edges. Each special edge contains a special $(p+1)$-tuple. Each special $(p+1)$-tuple is clearly contained in at most $\binom{n-p-1}{k-p-1}$ special edges. So the number of distinct $(p+1)$-tuples in $\mathcal{T}_{p+1}(\mathcal{H})$ is at least $\frac{m}{2\binom{n-p-1}{k-p-1}}$. Using $n < \lambda_k m^{\frac{1}{k-1}}$, we get

$|\mathcal{T}_{p+1}(\mathcal{H})| \geq \beta_{k,p} m^{\frac{p}{k-1}}$ for some small positive constant $\beta_{k,p}$ depending on k and p only. $\square$

The next lemma provides a key step to our proof of Theorem 2.4.5. To some extent, it shows that the notions of strong simplices and special tuples facilitate induction very nicely.

Lemma 2.4.3. *Let $k \geq p+2 \geq 3$. Let $\mathcal{F}$ be a k-uniform hypergraph and x a vertex in $\mathcal{F}$. Suppose that $T \in \mathcal{T}_p(\mathcal{F}_x) \cap \Delta_p(\mathcal{F}-x)$. Then $T \cup \{x\} \in \mathcal{T}_{p+1}(\mathcal{F})$.*

Proof. Note that $\mathcal{F}_x$ is $(k-1)$-uniform. By our assumption, T is a special p-tuple for some strong $(p-1)$-simplex $L = \{A, A_1, \ldots, A_p\}$ in $\mathcal{F}_x$, where A is the special edge and $A \supseteq T$. By definition, $\{A_1, \ldots, A_p\}$ is $(p-1)$-wise-intersecting, but $\cap_{i=1}^p A_i = \emptyset$. Suppose that $T = \{v_1, \ldots, v_p\}$, where for each $i \in [p]$ we have $v_i \in \cap_{j \neq i} A_j$. Since $T \in \Delta_p(\mathcal{F}-x)$, there exists $D \in \mathcal{F}-x$ such that $T \subseteq D$.

For each $i \in [p]$, let $A'_i = A_i \cup \{x\}$. Let $A'_{p+1} = D$ and $A' = A \cup \{x\}$. Let $L' = \{A', A'_1, \ldots, A'_{p+1}\} \subseteq \mathcal{F}$. We claim that $\{A'_1, \ldots, A'_{p+1}\}$ is a p-simplex in $\mathcal{F}$. Indeed, $x \in \cap_{i=1}^p A'_i$. Also, for each $i \in [p]$, $v_i \in \cap_{j \in [p+1] \setminus \{i\}} A'_j$. So, $\{A'_1, \ldots, A'_{p+1}\}$ is p-wise-intersecting. Since $\cap_{i=1}^p A_i = \emptyset$, the only element in $\cap_{i=1}^p A'_i$ is x. But $x \notin D$ since $D \in \mathcal{F}-x$. So $\cap_{i=1}^{p+1} A'_i = \emptyset$. This shows that $\{A'_1, \ldots, A'_{p+1}\}$ is a p-simplex in $\mathcal{F}$.

Now, let $T' = T \cup \{x\} = \{x, v_1, \ldots, v_p\}$. Then A' contains T'. Let $v_{p+1} = x$. For all $i \in [p+1]$ we have $v_i \in \cap_{j \in [p+1] \setminus \{i\}} A'_j$. Since A' contains $v_1, \ldots, v_{p+1}$, L' is a strong p-simplex in $\mathcal{F}$ with T' being a special $(p+1)$-tuple. That is, $T' \in \mathcal{T}_{p+1}(\mathcal{F})$. $\square$

Lemma 2.4.4. *Let $k > j \geq 2$. Let $\mathcal{F}$ be a k-graph and let x be a vertex of $\mathcal{F}$. Then*

$$\partial_j^*(\mathcal{F}) \geq \partial_j^*(\mathcal{F}-x) + \partial_{j-1}^*(\mathcal{F}_x).$$

Proof. We want to prove that

$$|\Delta_j(\mathcal{F})| + |\mathcal{T}_{j+1}(\mathcal{F})| \geq |\Delta_j(\mathcal{F}-x)| + |\mathcal{T}_{j+1}(\mathcal{F}-x)| + |\Delta_{j-1}(\mathcal{F}_x)| + |\mathcal{T}_j(\mathcal{F}_x)|. \quad (2.1)$$

Let $T \in \mathcal{T}_j(\mathcal{F}_x)$; that is, T is a special j-tuple in $\mathcal{F}_x$. If $T \in \Delta_j(\mathcal{F}-x)$, we say that T is of Type 1. If $T \notin \Delta_j(\mathcal{F}-x)$, we say T is of Type 2. Suppose $\mathcal{T}_j(\mathcal{F}_x)$ contains a Type 1 special j-tuples and b Type 2 special j-tuples. Then $a + b = |\mathcal{T}_j(\mathcal{F}_x)|$.

For each Type 1 special j-tuple T of $\mathcal{F}_x$, by Lemma 2.4.3, $T \cup \{x\} \in \mathcal{T}_{j+1}(\mathcal{F})$. Furthermore, it is not in $\mathcal{T}_{j+1}(\mathcal{F}-x)$ since $T \cup \{x\}$ contains x. Hence

$$|\mathcal{T}_{j+1}(\mathcal{F})| \geq |\mathcal{T}_{j+1}(\mathcal{F}-x)| + a. \quad (2.2)$$

For each Type 2 special j-tuple T of $\mathcal{F}_x$, we have $T \in \Delta_j(\mathcal{F})$ since T is contained in some special edge in $\mathcal{F}_x$ which in turn is contained in some edge of $\mathcal{F}$. Also, by our definition of Type 2 special tuples, $T \notin \Delta_j(\mathcal{F} - x)$. Furthermore, T is not of the form $S \cup \{x\}$ since it does not contain x. Also, for each $S \in \Delta_{j-1}(\mathcal{F}_x)$, $S \cup \{x\}$ is an element in $\Delta_j(\mathcal{F})$ that is not in $\Delta_j(\mathcal{F} - x)$. Hence,

$$|\Delta_j(\mathcal{F})| \geq |\Delta_j(\mathcal{F} - x)| + |\Delta_{j-1}(\mathcal{F}_x)| + b. \tag{2.3}$$

When we add (2.2) and (2.3), we obtain the desired inequality (2.1) completing the proof of Lemma 2.4.4. $\square$

Theorem 2.4.5. *For all $k \geq p + 2 \geq 3$, there exists a positive constant $c_{k,p}$ such that the following holds: if $\mathcal{F}$ is a k-uniform hypergraph and $m = |\mathcal{F}|$, then $\partial_p^*(\mathcal{F}) \geq c_{k,p} m^{\frac{p}{k-1}}$.*

Proof. Let us remove all isolated vertices from $\mathcal{F}$. Let n denote the number of remaining (i.e., non-isolated) vertices of $\mathcal{F}$. Let $n_{k,p}$ be defined as in Lemma 2.4.1, which depends only on k and p. Suppose that $n < n_{k,p}$. Since clearly $m \leq \binom{n}{k} < \binom{n_{k,p}}{k}$, $m^{\frac{p}{k-1}}$ is upper bounded by some function of k and p. Hence, $\partial_p^*(\mathcal{F}) \geq \alpha_{k,p} m^{\frac{p}{k-1}}$ for some small enough constant $\alpha_{k,p}$. So, as long as we choose $c_{k,p}$ so that $c_{k,p} \leq \alpha_{k,p}$, the claim holds when $n \leq n_{k,p}$. To prove the general claim, we use induction on p. For each fixed p, we use induction on n noting that when $n \leq n_{k,p}$, the claim has already been verified.

For the basis step, let $p = 1$. Let $c_{k,1} = \min\{\alpha_{k,1}, \beta_{k,1}, 1/4\}$, where $\beta_{k,1}$ is defined in Lemma 2.4.2. First, suppose that $m \leq 4\binom{n-1}{k-1} < 4n^{k-1}$. Then $n > (m/4)^{\frac{1}{k-1}} > m^{\frac{1}{k-1}}/4$. We have $\partial_1^*(\mathcal{F}) \geq |\Delta_1(\mathcal{F})| = n \geq c_{k,1} m^{\frac{1}{k-1}}$.

Next, suppose that $m > 4\binom{n-1}{k-1}$. By Lemma 2.4.2, $\partial_1^*(\mathcal{F}) \geq |\mathcal{T}_2(\mathcal{F})| \geq \beta_{k,1} m^{\frac{1}{k-1}} \geq c_{k,1} m^{\frac{1}{k-1}}$. This completes the proof of the basis step.

For the induction step, let $2 \leq j \leq k-2$. Suppose the claim holds for $p < j$. We prove the claim for $p = j$. We use induction on n. Let

$$c_{k,j} = \min\left\{\alpha_{k,j},\, \beta_{k,j},\, \frac{1}{8k} c_{k-1,j-1}\right\}.$$

Suppose the claim has been verified for k-uniform hypergraphs on fewer than n vertices. Let $\mathcal{F}$ be a k-uniform on n vertices. Suppose $\mathcal{F}$ has m edges. Suppose first that $m > 4\binom{n-1}{k-1}$. By Lemma 2.4.2, $\partial_j^*(\mathcal{F}) \geq |\mathcal{T}_{j+1}(\mathcal{F})| \geq \beta_{k,j} m^{\frac{j}{k-1}} \geq c_{k,j} m^{\frac{j}{k-1}}$.

Next, suppose that $m \leq 4\binom{n-1}{k-1} < 4n^{k-1}$. Then $n > m^{\frac{1}{k-1}}/4$. Hence, the average degree of $\mathcal{F}$ is $km/n < 4km^{\frac{k-2}{k-1}}$. Let x be a vertex in $\mathcal{F}$ of minimum degree d. Then $d < 4km^{\frac{k-2}{k-1}}$.

Note that $\mathcal{F}_x$ is $(k-1)$-uniform with d edges. By the induction hypothesis, we have $\partial^*_{j-1}(\mathcal{F}_x) \geq c_{k-1,j-1} d^{\frac{j-1}{k-2}}$. Also, $\mathcal{F} - x$ is a k-uniform hypergraph on fewer than n vertices (and has $m-d$ edges). By the induction hypothesis, $\partial^*_j(\mathcal{F} - x) \geq c_{k,j}(m-d)^{\frac{j}{k-1}}$. Hence, by Lemma 2.4.4 we have

$$\partial^*_j(\mathcal{F}) \geq c_{k,j}(m-d)^{\frac{j}{k-1}} + c_{k-1,j-1} d^{\frac{j-1}{k-2}}. \tag{2.4}$$

Recall that $d \leq 4km^{\frac{k-2}{k-1}}$. Also, $d \leq km/n$. Since we assume that n is large (as a function of k), we may further assume that $d \leq m/2$.

Claim 2.4.6. *We have* $c_{k,j}(m-d)^{\frac{j}{k-1}} + c_{k-1,j-1} d^{\frac{j-1}{k-2}} \geq c_{k,j} m^{\frac{j}{k-1}}$.

Proof of Claim 2.4.6. By the mean value theorem, there exists $y \in (m-d, m) \subseteq (m/2, m)$ such that $c_{k,j} m^{\frac{j}{k-1}} - c_{k,j}(m-d)^{\frac{j}{k-1}} = c_{k,j} d \frac{j}{k-1} y^{\frac{j}{k-1}-1}$. It suffices to prove that $c_{k-1,j-1} d^{\frac{j-1}{k-2}} \geq c_{k,j} d \frac{j}{k-1} y^{\frac{j}{k-1}-1}$, which holds if $c_{k-1,j-1} y^{\frac{k-j-1}{k-1}} \geq c_{k,j} d^{\frac{k-j-1}{k-2}}$. Since $y \geq m/2$, $d \leq 4km^{\frac{k-2}{k-1}}$, and $c_{k,j} \leq \frac{1}{8k} c_{k-1,j-1}$, one can check that the last inequality holds. $\square$

By (2.4) and Claim 2.4.6, we have $\partial^*_j(\mathcal{F}) \geq c_{k,j} m^{\frac{j}{k-1}}$. This completes the proof.

Lemma 2.4.7. *Let $k \geq d+2 \geq 3$. Let $\mathcal{F} \subseteq \binom{[n]}{k}$ contain no strong d-simplex. Let $x \in [n]$. Let $C = \{u_1, \ldots, u_d\} \in \Delta_d(\mathcal{F} - x) \cap \Delta_d(\mathcal{F}_x)$. Let $A, B \in \mathcal{F}$ with $x \in A$, $x \notin B$ and $C \subseteq A \cap B$. Let $W \subseteq [n] \setminus (A \cup B)$ such that $|W| = k-d$. For each $i \in [d]$, let $E^i_W = (\{x\} \cup C \cup W) \setminus \{u_i\}$. Then for at least one $i \in [d]$, we have $E^i_W \notin \mathcal{F}$.*

Proof. Suppose on the contrary that, for all $i \in [d]$, $E^i_W \in \mathcal{F}$. Consider the collection $\{A, B, E^1_W, \ldots, E^d_W\}$. We have $\cap_{i=1}^d E^i_W = \{x\} \cup W$. For each $i \in [d]$, $\cap_{j \neq i} E^j_W = \{x, u_i\} \cup W$, and so $(\cap_{j \neq i} E^j_W) \cap B = \{u_i\}$. This also implies that $(\cap_{i=1}^d E^i_W) \cap B = \emptyset$. Hence $\{B, E^1_W, \ldots, E^d_W\}$ is d-wise-intersecting but not $(d+1)$-wise-intersecting. That is, it is a d-simplex. As A contains an element of each d-wise intersection among $\{B, E^1_W, \ldots, E^d_W\}$, these $d+2$ sets form a strong d-simplex in $\mathcal{F}$, a contradiction. $\square$

Now, we are ready to prove Theorem 2.1.10.

Proof of Theorem 2.1.10. Given d and k, let n be large. Suppose on the contrary that $\mathcal{F} \subseteq \binom{[n]}{k}$ contains no strong d-simplex, $|\mathcal{F}| \geq \binom{n-1}{k-1}$, and $\mathcal{F}$ is not a star. We derive a contradiction. By Theorem 2.1.6, there exists an element $x \in [n]$ such that

$|\mathcal{F} - x| = o(n^{k-1})$ (that is, almost all edges of $\mathcal{F}$ contain x). Let

$$\begin{aligned}
\mathcal{B} &= \mathcal{F} - x, \\
\mathcal{M} &= \left\{ D \in \binom{[n]}{k} : x \in D, D \notin \mathcal{F} \right\}.
\end{aligned}$$

We call members of $\mathcal{B}$ *bad edges* and members of $\mathcal{M}$ *missing edges*. So, bad edges are those edges in $\mathcal{F}$ not containing x, and missing edges are those k-tuples containing x which are not in $\mathcal{F}$. Since $\binom{n-1}{k-1} \leq |\mathcal{F}| = \binom{n-1}{k-1} - |\mathcal{M}| + |\mathcal{B}|$, we have $|\mathcal{B}| \geq |\mathcal{M}|$. Let $b = |\mathcal{B}|$. By the definition of x,

$$b = o(n^{k-1}). \tag{2.5}$$

By Theorem 2.4.5, $|\Delta_d(\mathcal{B})| + |\mathcal{T}_{d+1}(\mathcal{B})| = \partial_d^*(\mathcal{B}) \geq c_{k,d} b^{\frac{d}{k-1}}$. Since $\mathcal{B} \subseteq \mathcal{F}$, $\mathcal{B}$ contains no strong d-simplex. So, $|\mathcal{T}_{d+1}(\mathcal{B})| = 0$. It follows that

$$|\Delta_d(\mathcal{B})| \geq c_{k,d} b^{\frac{d}{k-1}}.$$

Let

$$S_1 = \Delta_d(\mathcal{B}) \setminus \Delta_d(\mathcal{F}_x) \quad \text{and} \quad S_2 = \Delta_d(\mathcal{B}) \cap \Delta_d(\mathcal{F}_x).$$

We consider two cases.

Case 1. $|S_1| \geq |\Delta_d(\mathcal{B})|/2$.

For any $C \in S_1$ and a set $W \subseteq [n] \setminus (C \cup \{x\})$ of size $k-d-1$, the k-tuple $D = \{x\} \cup C \cup W$ does not belong to $\mathcal{F}$ because $C \notin \Delta_d(\mathcal{F}_x)$. So $D \in \mathcal{M}$. Doing this for each $C \in S_1$ yields a list of $\binom{n-d-1}{k-d-1}|S_1|$ edges (with multiplicity) in $\mathcal{M}$. An edge $D = \{x, y_1, \ldots, y_{k-1}\}$ may appear at most $\binom{k-1}{d}$ times in this list, as it is counted once for each d-subset of $\{y_1, \ldots, y_{k-1}\}$ that appears in S_1. Therefore,

$$b \geq |\mathcal{M}| \geq \frac{\binom{n-d-1}{k-d-1}|S_1|}{\binom{k-1}{d}} \geq c \cdot b^{\frac{d}{k-1}} n^{k-d-1} \tag{2.6}$$

for some properly chosen small positive constant c (depending on k only). Solving (2.6) for b, we get $b \geq c' \cdot n^{k-1}$ for some small positive constant c'. This contradicts (2.5) for sufficiently large n.

Case 2. $|S_2| \geq |\Delta_d(\mathcal{B})|/2$.

By Lemma 2.4.7, for every d-tuple $C \in S_2$ we may find two edges $A, B \in \mathcal{F}$ such that for every $(k-d)$-set $W \subseteq [n] \setminus (A \cup B)$ there exists $u \in C$ such that $(\{x\} \cup C \cup W) \setminus \{u\} \in \mathcal{M}$. So, we obtain a collection of at least $\binom{n-2k}{k-d}|S_2|$ members of $\mathcal{M}$. Pick an edge

$D = \{x, y_1, \ldots, y_{k-1}\}$ in $\mathcal{M}$ and consider its multiplicity in this collection. The edge D may appear each time a $(d-1)$-subset of $\{y_1, \ldots, y_{k-1}\}$ belongs to some d-tuple in S_2. There are $\binom{k-1}{d-1}$ such subsets, and each may be completed to form a d-tuple in at most $n - d + 1$ ways by picking the vertex u. Thus,

$$b \geq |\mathcal{M}| \geq \frac{\binom{n-2k}{k-d}|S_2|}{(n-d+1)\binom{k-1}{d-1}} \geq c'' \cdot b^{\frac{d}{k-1}} \cdot n^{k-d-1}$$

for some small positive constant c''. From this, we get $b \geq c''' \cdot n^{k-1}$ for some positive constant c''', which again contradicts (2.5) for sufficiently large n. This completes the proof of Theorem 2.1.10.

Chapter 3

The Number of K_3-free and K_4-free Edge 4-colorings

3.1 Introduction

Given a graph G and integers $k \geq 3$ and $r \geq 2$, let $F(G, r, k)$ denote the number of distinct edge r-colorings of G that are K_k*-free*, that is, do not contain a monochromatic copy of K_k, the complete graph on k vertices. Note that we do not require that these edge colorings are *proper* (that is, we do not require that adjacent edges get different colors). We consider the following extremal function:

$$F(n, r, k) = \max\{\, F(G, r, k) \,:\, G \text{ is a graph on } n \text{ vertices} \,\},$$

the maximum value of $F(G, r, k)$ over all graphs of order n.

One obvious choice for G is to take a maximum K_k-free graph of order n. The celebrated theorem of Turán [Tur41] states that $\mathrm{ex}(n, K_k)$, the maximum size of a K_k-free graph of order n, is attained by a unique (up to isomorphism) graph, namely the *Turán graph* $T_{k-1}(n)$ which is the complete $(k-1)$-partite graph on n vertices with parts of size $\lfloor \frac{n}{k-1} \rfloor$ or $\lceil \frac{n}{k-1} \rceil$. Thus

$$\mathrm{ex}(n, K_k) = t_{k-1}(n), \quad \text{for all } n, k \geq 2, \tag{3.1}$$

where $t_{k-1}(n)$ denotes the number of edges in $T_{k-1}(n)$. This gives the following trivial lower bound on our function:

$$F(n, r, k) \geq F(T_{k-1}(n), r, k) = r^{t_{k-1}(n)}. \tag{3.2}$$

Erdős and Rothschild (see [Erd74, Erd92]) conjectured that this is best possible when $r = 2$ and $k = 3$. Yuster [Yus96] proved that, indeed, $F(n, 2, 3) = 2^{t_2(n)} = 2^{\lfloor n^2/4 \rfloor}$ for large enough n. Both sets of authors further conjectured that this holds for all k when we have $r = 2$ colors. Alon, Balogh, Keevash, and Sudakov [ABKS04] not only settled this conjecture for large n but also showed that it holds for 3-colorings as well, i.e., we have equality in (3.2) when $r = 2, 3$, $k \geq 3$, and $n > n_0(k)$.

The generalization of the problem where one has to avoid a monochromatic copy of a general graph F was also studied in [ABKS04]. The papers [HKL09, LP, LPRS09, LPS] studied H-free edge colorings for general hypergraphs H. In particular, Lefmann, Person, and Schacht [LPS] proved that, for every k-uniform hypergraph F and $r \in \{2, 3\}$, the maximum number of F-free edge r-colorings over n-vertex hypergraphs is $r^{\mathrm{ex}(n,F)+o(n^k)}$. Interestingly, this result holds for every F even though the value of the Turán function $\mathrm{ex}(n, F)$ is known for very few hypergraphs F. Also, Balogh [Bal06] studied a version of the problem where a specific coloring of a graph F is forbidden. Alon and Yuster [AY06] considered this problem for directed graphs (where one counts admissible orientations instead of edge colorings).

Let us return to the original question. Surprisingly, Alon et al. [ABKS04] showed that one can do significantly better than (3.2) for larger values of r. In two particular cases, they were also able to obtain the best possible constant in the exponent. Namely they proved that

$$F(n, 4, 3) = 18^{n^2/8+o(n^2)}, \tag{3.3}$$

$$F(n, 4, 4) = 3^{4n^2/9+o(n^2)}. \tag{3.4}$$

Let us briefly show the lower bounds in (3.3) and (3.4), which are given by $F(T_4(n), 4, 3)$ and $F(T_9(n), 4, 4)$ respectively. Let $W_1, \ldots, W_k$ denote the parts of $T_k(n)$. Consider $T_4(n)$ first. Fix a function π that assigns to each pair $\{i, j\}$ of $\{1, \ldots, 4\}$ a list $\pi(\{i, j\})$ of two or three colors so that each color appears in exactly four lists with the corresponding four pairs forming a 4-cycle. Up to a symmetry, such an assignment is unique and we have two lists of size 2 and four lists of size 3. Generate an edge coloring of $T_4(n)$ by choosing for each edge $\{u, v\}$ with $u \in W_i$ and $v \in W_j$ an arbitrary color from $\pi(\{i, j\})$. Every obtained coloring is K_3-free and, if we assume that e.g. $n = 4m$, there are $3^{4m^2} \cdot 2^{2m^2} = 18^{n^2/8}$ such colorings. We proceed similarly for $T_9(n)$ except we fix the (unique up to a symmetry) assignment where each pair from $\{1, \ldots, 9\}$ gets a list of three colors while every color forms a copy of $T_3(9)$.

The goal of this chapter is to determine $F(n,4,3)$ and $F(n,4,4)$ exactly and describe all extremal graphs for large n. Specifically, we will show the following results.

Theorem 3.1.1. *There is N such that for all $n \geq N$, $F(n,4,3) = F(T_4(n),4,3)$ and $T_4(n)$ is the unique graph achieving the maximum.*

Theorem 3.1.2. *There is N such that for all $n \geq N$, $F(n,4,4) = F(T_9(n),4,4)$ and $T_9(n)$ is the unique graph achieving the maximum.*

Thus a new phenomenon occurs for $r \geq 4$: extremal graphs may have many copies of the forbidden monochromatic graph K_k. This makes the problem more interesting and difficult.

Similarly to [ABKS04], our general approach is to establish the stability property first: namely, that all graphs with the number of colorings close to the optimum have essentially the same structure. However, additionally to the approximate graph structure, we also have to describe how typical colorings look like. This task is harder and we do it in stages, getting more and more precise description of typical colorings (namely, the properties called *satisfactory*, *good*, and *perfect* in our proofs). We then proceed to show that the Turán graphs are, indeed, the unique graphs that attain the optimum. It is not surprising that our proofs are longer and more complicated than those in [ABKS04]. The case of $r \geq 4$ colors seems to be much harder than the case $r \leq 3$. It is not even clear if there is a simple closed formula for $F(T_4(n),4,3)$ and $F(T_9(n),4,4)$. Our proofs imply that

$$F(T_4(n),4,3) = (C+o(1)) \cdot 18^{t_4(n)/3}, \tag{3.5}$$

$$F(T_9(n),4,4) = (20160+o(1)) \cdot 3^{t_9(n)}, \tag{3.6}$$

where $C = (2^{14} \cdot 3)^{1/3}$ if $n \equiv 2 \pmod 4$ and $C = 36$ otherwise.

Unfortunately, we could not determine $F(n,r,k)$ for other pairs r,k, which seems to be an interesting and challenging problem. Hopefully, our methods may be helpful in obtaining further exact results. It is possible that for all large n, $n \geq n_0(k,r)$, all extremal graphs are complete partite (not necessarily balanced) but we could not prove nor disprove this.

This chapter is organized as follows. In Section 3.2 we state a version of Szemerédi's Regularity Lemma and some auxiliary definitions and results that we use in our arguments. Theorem 3.1.1 is proved in Section 3.3 and Theorem 3.1.2 is proved in Section 3.4.

3.2 Notation and Tools

For a set X and a non-negative integer k, let $\binom{X}{k}$ (resp. $\binom{X}{\le k}$) be the set of all subsets of X with exactly (resp. at most) k elements. Also, we denote $\binom{n}{\le k} = \sum_{i=0}^{k} \binom{n}{i}$ and $[k] = \{1, 2, \ldots, k\}$. We will often omit punctuation signs when writing unordered sets, abbreviating e.g. $\{i, j\}$ to ij.

As it is standard in graph theory, we use $V(G)$ and $E(G)$ to refer to the vertex and edge set, respectively, of a graph G. Also, $v(G) = |V(G)|$ and $e(G) = |E(G)|$ denote respectively the *order* and *size* of G. In addition, for disjoint $A, B \subseteq V(G)$, we use $G[A]$ to refer to the subgraph induced by A and $G[A, B]$ for the induced bipartite subgraph with parts A and B. Let

$$N_G(x) = \{y \in V(G) : xy \in E(G)\}$$

be the *neighborhood* of a vertex x in G. Let $K(V_1, \ldots, V_l)$ denote the complete l-partite graph with parts $V_1, \ldots, V_l$.

It will be often convenient to identify graphs with their edge sets. Thus, for example, $|G| = e(G)$ denotes the number of edges while $G \triangle H$ is the graph on $V(G) \cup V(H)$ whose edge set is the symmetric difference of $E(G)$ and $E(H)$.

As we make use of a multicolor version of Szemerédi's Regularity Lemma [Sze76], we remind the reader of the following definitions. Let G be a graph and A, B be two disjoint non-empty subsets of $V(G)$. The *edge density* between A and B is

$$d(A, B) = \frac{e(G[A, B])}{|A|\,|B|}.$$

For $\epsilon > 0$, the pair (A, B) is called *ϵ-regular* if for every $X \subseteq A$ and $Y \subseteq B$ satisfying $|X| > \epsilon|A|$ and $|Y| > \epsilon|B|$ we have

$$|d(X, Y) - d(A, B)| < \epsilon.$$

An *equitable partition* of a set V is a partition of V into pairwise disjoint parts $V_1, \ldots, V_m$ of almost equal size, i.e., $\big||V_i| - |V_j|\big| \le 1$ for all $i, j \in [m]$. An equitable partition of the set of vertices of G into parts $V_1, \ldots, V_m$ is called *ϵ-regular* if $|V_i| \le \epsilon|V|$ for every $i \in [m]$ and all but at most $\epsilon\binom{m}{2}$ of the pairs (V_i, V_j), $1 \le i < j \le m$, are ϵ-regular.

The following more general result can be deduced from the original Regularity Lemma of Szemerédi [Sze76] (cf. Theorems 1.8 and 1.18 in Komlós and Simonovits [KS96]).

Lemma 3.2.1 (Multicolor Regularity Lemma)**.** *For every $\epsilon > 0$ and integer $r \geq 1$, there is $M = M(\epsilon, r)$ such that for any graph G on $n > M$ vertices and any (not necessarily proper) edge r-coloring $\chi : E(G) \to [r]$, there is an equitable partition $V(G) = V_1 \cup \ldots \cup V_m$ with $1/\epsilon \leq m \leq M$, which is ϵ-regular simultaneously with respect to all graphs $(V(G), \chi^{-1}(i))$, $i \in [r]$.* □

Also, we will need the following special case of the Embedding Lemma (see e.g. [KS96, Theorem 2.1]).

Lemma 3.2.2 (Embedding Lemma)**.** *For every $\eta > 0$ and integer $k \geq 2$ there exists $\epsilon > 0$ such that the following holds for all large n. Suppose that G is a graph of order n with an equitable partition $V(G) = V_1 \cup \ldots \cup V_k$ such that every pair (V_i, V_j) for $1 \leq i < j \leq k$ is ϵ-regular of density at least η. Then G contains K_k.* □

While we have $t_k(n) = (1 - 1/k + o(1)) \binom{n}{2}$ for $n \to \infty$, the following easy bound holds for *all* $k, n \geq 1$:

$$\max\{e(G) : v(G) = n,\ G \text{ is } k\text{-partite}\} = t_k(n) \leq \left(1 - \frac{1}{k}\right) \frac{n^2}{2}. \tag{3.7}$$

We will also use the following stability result for the Turán function (3.1).

Lemma 3.2.3 (Erdős [Erd67] and Simonovits [Sim68])**.** *For every $\alpha > 0$ and integer $k \geq 1$, there exist $\beta > 0$ and n_0 such that, for all $n > n_0$, any K_{k+1}-free graph G on n vertices with at least $(1 - 1/k)n^2/2 - \beta n^2$ edges admits an equitable partition $V(G) = V_1 \cup \ldots \cup V_k$ with $|\, G \bigtriangleup K(V_1, \ldots, V_k)\,| < \alpha n^2$.* □

3.3 $F(n, 4, 3)$

In this section we prove Theorem 3.1.1. Here we have to overcome many new difficulties that are not present for 2 or 3 colors. So, unfortunately, the proof is long and complicated. In order to improve its readability we split it into a sequence of lemmas. Since we use the Regularity Lemma, the obtained value for N in Theorem 3.1.1 is very large and is of little practical value. Therefore we make no attempt to determine or optimize it.

First, let us state some important definitions that are extensively used in the whole proof. Fix positive constants

$$c_0 \gg c_1 \gg \ldots \gg c_{10},$$

each being sufficiently small depending on the previous ones. Let $M = 1/c_9$ and $n_0 = 1/c_{10}$.

Typically, the order of a graph under consideration will be denoted by n and will satisfy $n \geq n_0$. We will view n as tending to infinity with $c_0, \ldots, c_9$ being fixed and use the asymptotic terminology (such as, for example, the expression $O(1)$ or the phrase "almost every") accordingly.

Let $\mathcal{G}_n$ consist of graphs of order n that have many K_3-free edge 4-colorings. Specifically,

$$\mathcal{G}_n = \left\{ G : v(G) = n,\ F(G,4,3) \geq 18^{n^2/8} \cdot 2^{-c_8 n^2} \right\}.$$

Let $\mathcal{G} = \cup_{n \geq n_0} \mathcal{G}_n$. The lower bound in (3.3) (whose proof we sketched in the Introduction) shows that $\mathcal{G}_n$ is non-empty for each $n \geq n_0$.

Next, for an arbitrary graph G with $n \geq n_0$ vertices and a K_3-free 4-coloring χ of the edges of G, we will define the following objects and parameters. As the constants c_8 and M satisfy Lemma 3.2.1 (namely, we can assume that M is at least the function $M(c_8, 4)$ returned by Lemma 3.2.1), we can find a partition $V(G) = V_1 \cup \ldots \cup V_m$ with $1/c_8 \leq m \leq M$ that is c_8-regular with respect to each color. Next, we define the *cluster graphs* H_1, H_2, H_3, and H_4 on vertex set $[m]$, where H_ℓ consists of those pairs $ij \in \binom{[m]}{2}$ such that the pair (V_i, V_j) is c_8-regular and has edge density at least c_7 with respect to the ℓ-color subgraph $\chi^{-1}(\ell)$ of G. For $1 \leq s \leq 4$, let R_s be the graph on vertex set $[m]$ where $ab \in E(R_s)$ if and only if $ab \in E(H_\ell)$ for exactly s values of $\ell \in [4]$. Let $R = \cup_{s=1}^{4} R_s$ be the union of the graphs R_s. Let $r_s = 2e(R_s)/m^2$.

We view m, V_i, H_i, R_i, R, r_i as functions of the pair (G, χ). Although we may have some freedom when choosing the c_8-regular partition $V_1, \ldots, V_m$, we fix just one choice for each input (G, χ). We do not require any "continuity" property from these functions: for example, it may be possible that χ_1 and χ_2 are two colorings of the same graph G that differ on one edge only but $r_i(G, \chi_1)$ and $r_i(G, \chi_2)$ are quite far apart.

By Lemma 3.2.2, each cluster graph H_i is triangle-free and, by Turán's theorem (3.1), has at most $t_2(m)$ edges. By (3.7),

$$r_1 + 2r_2 + 3r_3 + 4r_4 = \frac{e(H_1) + e(H_2) + e(H_3) + e(H_4)}{m^2/2} \leq 2. \tag{3.8}$$

In addition, note that $R_3 \cup R_4$ is triangle-free because a triangle in $R_3 \cup R_4$ gives a triangle in some H_i. Therefore, by (3.1) and (3.7),

$$r_3 + r_4 \leq 1/2. \tag{3.9}$$

We also need the following "converse" procedure for generating all K_3-free edge 4-colorings of G. Our upper bounds on $F(G,4,3)$ and some structural information about typical colorings are obtained by estimating the possible number of outputs. Since the parameters $r_1,\dots,r_4$ play crucial role in these estimates, some guesses of the functions m, V_i, and H_i (and thus of R_i, R, and r_i) are also generated. The procedure is rather wasteful in the sense that it can generate a lot of "garbage". But the obtained inequalities (3.8) and (3.9) imply the crucial property that every K_3-free edge 4-coloring of G with the correct guess of m, V_i, and H_i is generated at least once provided $v(G) \ge n_0$.

The Coloring Procedure

1. Choose an arbitrary integer m' between $1/c_8$ and M.
2. Choose an arbitrary equitable partition $V(G) = V_1' \cup \cdots \cup V_{m'}'$.
3. Choose arbitrary graphs $H_1',\dots,H_4'$ with vertex set $[m']$ such that we have

$$r_1' + 2r_2' + 3r_3' + 4r_4' \le 2, \tag{3.10}$$

$$r_3' + r_4' \le 1/2, \tag{3.11}$$

   where R_i', and r_i' are defined by the direct analogy with R_i and r_i. (For example, for $i \in [4]$, R_i' is the graph on $[m']$ whose edges are those pairs of $\binom{[m']}{2}$ that are edges in exactly i graphs $H_1',\dots,H_4'$.)
4. Assign arbitrary colors to all edges of G that lie inside some part V_i'.
5. Select at most $4c_8\binom{m'}{2}$ elements of $\binom{[m']}{2}$ and, for each selected pair ij, assign colors to $G[V_i',V_j']$ arbitrarily.
6. For every color $l \in [4]$ and every $ij \in \binom{[m']}{2}$ color an arbitrary subset of edges of $G[V_i',V_j']$ of size at most $c_7|V_i'|\,|V_j'|$ by color l.
7. For every edge xy of G that is not colored yet, let us say $x \in V_i'$ and $y \in V_j'$, pick an arbitrary color from the set $C_{ij} = \{s \in [4] : ij \in H_s'\}$. If $C_{ij} = \emptyset$, then we color xy with color 1.

Lemma 3.3.1. *For every graph G of order $n \ge n_0$, the number of choices in Steps 1–6 of the Coloring Procedure is at most $2^{c_6 n^2}$.*

Proof. Clearly, the number of choices in Steps 1–3 is at most

$$M \cdot n^M \cdot \left(2^{\binom{M}{2}}\right)^4 = 2^{O(\log n)}. \tag{3.12}$$

Fix these choices. Since $m' \geq 1/c_8$, the number of edges that lie inside some part V_i' is at most $m'\binom{\lceil n/m' \rceil}{2} \leq c_6 n^2/8$; so the number of choices in Step 4 is at most $4^{c_6 n^2/8}$. In Step 5 we have at most $2^{\binom{m'}{2}} \cdot 4^{4c_8\binom{m'}{2}\lceil n/m' \rceil^2} < 2^{c_6 n^2/4}$ options. The number of choices in Step 6 is at most

$$\binom{\lceil n/m' \rceil^2}{\leq c_7 \lceil n/m' \rceil^2}^{4\binom{m'}{2}} < 2^{c_6 n^2/4}.$$

By multiplying these four bounds, we obtain the required. □

The number of options in Step 7 can be bounded from above by

$$\left(2^{e(R_2')} \cdot 3^{e(R_3')} \cdot 4^{e(R_4')}\right)^{\lceil n/m' \rceil^2} \leq \left(2^{r_2'} \cdot 3^{r_3'} \cdot 4^{r_4'}\right)^{n^2/2+O(n)} = 2^{Ln^2/2+O(n)}, \tag{3.13}$$

where $L = r_2' + \log_2(3)\, r_3' + 2r_4'$. One can easily show that the maximum of L given (3.10) and (3.11) (and the non-negativity of $r_1', \ldots, r_4'$) is $(\log_2 18)/4$, with the (unique) optimal assignment being $r_1' = r_4' = 0$, $r_2' = 1/4$, and $r_3' = 1/2$. When combined with Lemma 3.3.1, this allows one to conclude that, for example,

$$F(n, 4, 3) \leq 18^{n^2/8} \cdot 2^{2c_6 n^2}, \quad \text{for all } n \geq n_0. \tag{3.14}$$

This is essentially the argument from [ABKS04]. We need to take this argument further. As the first step, we derive some information about r_2 and r_3 for a typical coloring χ. We call a pair (G, χ) (or the coloring χ) *satisfactory* if

$$r_2 > 1/4 - c_5/2 \quad \text{and} \quad r_3 > 1/2 - c_5. \tag{3.15}$$

Otherwise, (G, χ) is *unsatisfactory*. Next, we establish some results about satisfactory colorings. Later, this will allow us to define two other important properties of colorings (namely, being *good* and being *perfect*).

Lemma 3.3.2. *For every graph G with $n \geq n_0$ vertices the number of unsatisfactory K_3-free edge 4-colorings is less than $18^{n^2/8} \cdot 2^{-c_6 n^2}$. In particular, if $G \in \mathcal{G}_n$ then almost every coloring is satisfactory.*

Proof. We use the Coloring Procedure and bound from above the number of outputs that give unsatisfactory colorings. By Lemma 3.3.1, the number of choices in Steps 1–6 is at most $2^{c_6 n^2}$. We use (3.13) to estimate the number of choices in Step 7.

The value of L under constraints (3.10), (3.11), and

$$r_3' \leq 1/2 - c_5, \tag{3.16}$$

(as well as the non-negativity of the variables r_i') is at most

$$L_{\max} = (1/4 + 3c_5/2) + (1/2 - c_5)\log_2 3 < (1/4 - c_5^2)\log_2 18.$$

This can be seen by multiplying (3.10), (3.11), and (3.16) by respectively $y_1 = 1/2$, $y_2 = 0$, and $y_3 = \log_2 3 - 3/2 > 0$, and adding these inequalities. The obtained inequality has $L_{\max}$ in the right-hand side while each coefficient of the left-hand is at least the corresponding coefficient of L, giving the required bound. (In fact, these reals y_i are the optimal dual variables for the linear program of maximizing L.)

Likewise, when we maximize L under constraints (3.10), (3.11), and

$$r_2' \leq 1/4 - c_5/2 \tag{3.17}$$

then we have the same upper bound $L_{\max}$ (with the optimal dual variables for (3.10), (3.11), and (3.17) being respectively $y_1 = 2 - \log_2 3 > 0$, $y_2 = 4\log_2 3 - 6 > 0$, and $y_3 = 2\log_2 3 - 3 > 0$). Since in Step 7 we have only two (possibly overlapping) cases depending on which of (3.17) or (3.16) holds, the total number of choices in Step 7 is by (3.13) at most

$$2 \cdot 2^{L_{\max} n^2/2 + O(n)} < 18^{(1/8 - c_5^2/3)\, n^2}.$$

By multiplying this by $2^{c_6 n^2}$, we obtain the required upper bound on the number of unsatisfactory colorings. $\square$

For each satisfactory coloring of $G \in \mathcal{G}$ we record the vector $\nu(\chi) = (m, V_i, H_i)$ of parameters. Call a vector (m, V_i, H_i) *popular* if

$$|\nu^{-1}((m, V_i, H_i))| \geq 18^{n^2/8} \cdot 2^{-3c_8 n^2},$$

that is, if it appears for at least $18^{n^2/8} \cdot 2^{-3c_8 n^2}$ satisfactory colorings, where $n = v(G)$. As the number of possible choices of vectors is bounded by (3.12), the number of satisfactory colorings for which the corresponding vector is not popular is at most

$$2^{O(\log n)} \cdot 18^{n^2/8} \cdot 2^{-3c_8 n^2} \leq 18^{n^2/8} \cdot 2^{-2c_8 n^2},$$

that is, $o(1)$-fraction of all colorings. Let $\mathrm{Pop}(G)$ be the set of all popular vectors and let

$$\mathcal{S}(G) = \nu^{-1}(\mathrm{Pop}(G)) \tag{3.18}$$

be the set of satisfactory K_3-free edge 4-colorings of G for which the corresponding vector is popular. By Lemma 3.3.2, $\mathcal{S}(G)$ is non-empty.

Our next goal is to exhibit a stability property, namely, that every graph $G \in \mathcal{G}$ is almost complete 4-partite. First we show that, for every popular vector $(m, V_i, H_i) \in \mathrm{Pop}(G)$, the cluster graph R is almost complete 4-partite. Then we extend this result to G.

Lemma 3.3.3. *Let $n \geq n_0$, $G \in \mathcal{G}_n$, and $(m, V_i, H_i) \in \mathrm{Pop}(G)$. Then there exist equitable partitions $[m] = A \cup B$, $A = U_1 \cup U_2$, and $B = U_3 \cup U_4$ such that*

$$\begin{aligned} \left| R_3 \triangle K(A,B) \right| &< c_4 m^2, &(3.19)\\ \left| R_2[A] \triangle K(U_1,U_2) \right| &< 2c_3 m^2, &(3.20)\\ \left| R_2[B] \triangle K(U_3,U_4) \right| &< 2c_3 m^2, &(3.21)\\ \left| R \triangle K(U_1,U_2,U_3,U_4) \right| &< 5c_3 m^2. &(3.22) \end{aligned}$$

Proof. We have already proved that R_3 is triangle-free. As (m, V_i, H_i) is associated with a satisfactory coloring, (3.15) is satisfied; in particular, $r_3 > 1/2 - c_5$. Therefore, $e(R_3) = r_3 m^2/2 > t_2(m) - c_5 m^2/2$. As $c_5 \ll c_4$, we can apply Lemma 3.2.3 to partition $V(R_3) = [m]$ into two sets A and B such that $|A| = \lfloor m/2 \rfloor$, $|B| = \lceil m/2 \rceil$, and (3.19) holds.

Since $R_2 \cap R_3 = \emptyset$, we have $|R_2 \cap K(A,B)| \leq \mid K(A,B) \setminus R_3 \mid < c_4 m^2$. This and (3.15) imply that

$$e(R_2[A]) + e(R_2[B]) > e(R_2) - c_4 m^2 = r_2 m^2/2 - c_4 m^2 > m^2/8 - 2c_4 m^2. \qquad (3.23)$$

What we show in the following sequence of claims is that $R_2[A]$ and $R_2[B]$ are both close to being triangle-free and have roughly $m^2/16$ edges each; then we can apply Lemma 3.2.3 to these graphs, obtaining the desired partitions of A and B.

For a vertex $a \in A$, let $B_a = N_{R_3}(a) \cap B$ be the set of R_3-neighbors of a that lie in B. Similarly, for a vertex $b \in B$, let $A_b = N_{R_3}(b) \cap A$.

Claim 3.3.4. *For every $a \in A$ the graph $R_2[B_a]$ is 4-partite.*

Proof of Claim. Each pair ab with $b \in B_a$ is contained in R_3 and, by definition, is labeled with a 3-element subset X_b of [4]. Color b by the unique element of $[4] \setminus X_b$. If two adjacent vertices b and b' of $R_2[B_a]$ receive identical color c, then the label of $bb' \in R_2$ (a 2-element subset of [4]) has a non-empty intersection with $[4] \setminus \{c\}$ which is the label of both $ab, ab' \in R_3$. This implies the existence of a triangle in some H_i, a contradiction. ∎

Claim 3.3.5. *If $a_1 a_2 \in E(R_2[A])$, then $K_3 \not\subseteq R_2[B_{a_1} \cap B_{a_2}]$.*

Proof of Claim. Suppose on the contrary that we have an edge $a_1 a_2$ in $R_2[A]$ and a triangle in $R_2[B_{a_1} \cap B_{a_2}]$ with vertices b_1, b_2, and b_3. Let S be the multiset produced by the union of the labels of the edges a_1a_2, a_ib_j, and b_ib_j. As each edge a_ib_j is labeled with a 3-element subset of [4] and the remaining four edges are labeled with a 2-element subset of [4], we have $|S| = 6 \cdot 3 + 4 \cdot 2 = 26$. By the Pigeonhole Principle, some member of [4] belongs to S with multiplicity at least 7. But this corresponds to some H_i having at least 7 edges among the 5 vertices a_1, a_2, b_1, b_2, b_3. By Turán's result (3.1), this implies that H_i has a triangle, a contradiction. ❙

Define

$$B' = \{b \in B : |A_b| > |A| - \sqrt{c_4}m\}.$$

As each vertex of $B \setminus B'$ contributes at least $\sqrt{c_4}m$ to $|K(A,B) \setminus R_3|$, there are less than $\sqrt{c_4}m$ such vertices by (3.19). Thus $|B'| > |B| - \sqrt{c_4}m \geq (1/2 - \sqrt{c_4})m$. Similarly, we can define A' to be the set of vertices $a \in A$ for which $|B_a| > |B| - \sqrt{c_4}m$ and note that $|A'| > |A| - \sqrt{c_4}m > 0$.

Claim 3.3.6. $K_3 \not\subseteq R_2[B']$.

Proof of Claim. Suppose on the contrary that b_1, b_2, b_3 form a K_3 in $R_2[B']$. Let $X = A_{b_1} \cap A_{b_2} \cap A_{b_3}$. By definition, $|A \setminus A_{b_i}| < \sqrt{c_4}m$. So, $|X| > |A| - 3\sqrt{c_4}m$. By Claim 3.3.5, there are no edges within X. So, $e(R_2[A]) \leq |A \setminus X| \cdot |A| < 3\sqrt{c_4}m^2$.

Let us estimate $e(R_2[B])$ from above. Consider B_a for some $a \in A'$. By definition, $|B_a| > |B| - \sqrt{c_4}m$ and, by Claim 3.3.4, B_a is 4-partite. By (3.7), the number of edges in $R_2[B]$ is at most $(3/4)\,|B_a|^2/2 + \sqrt{c_4}m|B|$.

However, these upper bounds on $e(R_2[A])$ and $e(R_2[B])$ contradict (3.23). ❙

In particular, $R_2[B]$ may be made triangle-free by the removal of at most $|B \setminus B'| \cdot |B| < \sqrt{c_4}m^2$ edges. Hence, we have in fact that

$$e(R_2[B]) < (1 - \frac{1}{2})|B|^2/2 + \sqrt{c_4}m^2 \leq m^2/16 + 2\sqrt{c_4}m^2. \tag{3.24}$$

By (3.23) and (3.24), $e(R_2[A]) > m^2/16 - 2c_4m^2 - 2\sqrt{c_4}m^2$. As above, by removing at most $\sqrt{c_4}m^2$ edges, we can form a graph R_2' on vertex set A, which is triangle-free. We can now apply Lemma 3.2.3 to R_2', to find a partition $A = U_1 \cup U_2$ such that $|R_2' \triangle K(U_1, U_2)| < c_3m^2$. As R_2' and $R_2[A]$ differ in at most $\sqrt{c_4}m^2$ edges, we derive (3.20). The existence of an equitable partition $B = U_3 \cup U_4$ satisfying (3.21) is proved similarly.

By (3.19)–(3.21), we have $|(R_2 \cup R_3) \triangle K(U_1, U_2, U_3, U_4)| < 4c_3m^2 + c_4m^2$. Also, by (3.8) and (3.15), we have $r_1 + r_4 \leq 4c_5$ and $|R_1 \cup R_4| \leq 2c_5m^2$. Now (3.22) follows, finishing the proof of Lemma 3.3.3. □

For a graph $G \in \mathcal{G}$ and a popular vector $(m, V_i, H_i) \in \mathrm{Pop}(G)$, fix the sets $A, B, U_1, \ldots, U_4$ given by Lemma 3.3.3. For $i \in [4]$, let $\tilde{U_i} = \cup_{j \in U_i} V_j$ be the *blow-up* of U_i. Let $\tilde{F} = K(\tilde{U_1}, \tilde{U_2}, \tilde{U_3}, \tilde{U_4})$.

Lemma 3.3.7. *For every $n \geq n_0$, $G \in \mathcal{G}_n$, and $(m, V_i, H_i) \in \mathrm{Pop}(G)$, we have $|G \triangle \tilde{F}| < 12c_3n^2$.*

Proof. It routinely follows that the size of $G \setminus \tilde{F}$ is at most the sum of the following terms:

- $m\binom{\lceil n/m \rceil}{2}$, the number of edges of G inside parts V_i;
- $4c_8\binom{[m]}{2} \cdot \lceil n/m \rceil^2$, edges between parts which are not c_8-regular for at least one color graph;
- $4c_7\binom{n}{2}$, edges between parts of density at most c_7 for at least one color;
- $|R \setminus K(U_1, U_2, U_3, U_4)| \cdot \lceil n/m \rceil^2 \leq 5c_3m^2 \cdot \lceil n/m \rceil^2$, where we used (3.22).

Adding up, this gives less than $6c_3n^2$.

Next, we estimate $|\tilde{F} \setminus G|$ by bounding the number of satisfactory colorings of G that give our fixed vector (m, V_i, H_i). Again, we use the Coloring Procedure to generate all such colorings, where m, V_i, H_i are fixed in advance. By Lemma 3.3.1, we have at most $2^{c_6n^2}$ options in Steps 4–6. Once we have fixed the choices in these steps, the remaining uncolored edges of G are restricted to those between the parts while the graphs $R_1, \ldots, R_4$ specify how many choices of color each edge has. Thus the number of options in Step 7 is at most

$$\prod_{f=2}^{4} \prod_{ij \in R_f} f^{\lceil n/m \rceil^2 - |K(V_i,V_j) \setminus G|} \leq \left(2^{2c_6n^2} \cdot 18^{n^2/8}\right) \prod_{ij \in R_2 \cup R_3} 2^{-|K(V_i,V_j) \setminus G|}, \tag{3.25}$$

where we used the bound in (3.13) together with the maximization result mentioned immediately after (3.13). Let us look at the last factor in (3.25). If we replace the range of ij in the product by $K(U_1, U_2, U_3, U_4)$ instead of $R_2 \cup R_3$, this will affect at most $(c_4 + 4c_3)m^2$ pairs ij by (3.19)–(3.21) and we get an extra factor of at most $2^{5c_3n^2}$. Thus

$$\prod_{ij \in R_2 \cup R_3} 2^{-|K(V_i,V_j) \setminus G|} \leq 2^{-|\tilde{F} \setminus G|} \cdot 2^{5c_3n^2}.$$

Since the vector (m, V_i, H_i) is popular, we conclude that

$$|\tilde{F} \setminus G| \leq c_6n^2 + 5c_3n^2 + 2c_6n^2 + 3c_8n^2 \leq 6c_3n^2,$$

giving the required bound on $|G \triangle \tilde{F}|$. $\square$

Now, for every input graph G we fix a *max-cut* 4-partition $V(G) = W_1 \cup W_2 \cup W_3 \cup W_4$, that is, one that maximizes the number of edges of G across the parts.

Lemma 3.3.8 (Stability Property)**.** *Let $n \geq n_0$, $G \in \mathcal{G}_n$, and $W_1' \cup W_2' \cup W_3' \cup W_4'$ be a partition of $V(G)$ with*

$$\left| G \cap K(W_1', W_2', W_3', W_4') \right| \geq \left| G \cap K(W_1, W_2, W_3, W_4) \right| - c_3 n^2.$$

Then we have

$$\left| G \triangle K(W_1', W_2', W_3', W_4') \right| \leq 15 c_3 n^2 \tag{3.26}$$

and for every popular vector $(m, V_i, H_i) \in \mathrm{Pop}(G)$ there is a relabeling of $W_1', \ldots, W_4'$ such that for each $i \in [4]$,

$$\left| W_i' \triangle \tilde{U}_i \right| \leq 2000 c_3 n. \tag{3.27}$$

Also, we have $|\,|W_i| - n/4\,| \leq c_2 n$ for each $i \in [4]$ and $\left| G \triangle K(W_1, W_2, W_3, W_4) \right| \leq 15 c_3 n^2$.

Proof. Let $F' = K(W_1', W_2', W_3', W_4')$ and $F = K(W_1, W_2, W_3, W_4)$. As the max-cut partition $W_1 \cup \ldots \cup W_4$ maximizes the number of edges across parts, we have $|F' \cap G| + c_3 n^2 \geq |F \cap G| \geq |\tilde{F} \cap G|$. Since the partitions $[m] = U_1 \cup \cdots \cup U_4$ and $[n] = V_1 \cup \cdots \cup V_m$ are equitable, we have

$$|\,|\tilde{U}_i| - n/4\,| \leq m + n/m. \tag{3.28}$$

Thus we have $|\tilde{F}| \geq |F'| - c_7 n^2$ and, by Lemma 3.3.7,

$$\begin{aligned} |F' \triangle G| &= |F'| + |G| - 2\,|F' \cap G| \\ &\leq (|\tilde{F}| + c_7 n^2) + |G| - 2(\,|\tilde{F} \cap G| - c_3 n^2) \\ &= |\tilde{F} \triangle G| + c_7 n^2 + 2 c_3 n^2 \;\leq\; 15 c_3 n^2, \end{aligned} \tag{3.29}$$

proving the first part of the lemma.

We look for a relabeling of $W_1', \ldots, W_4'$ such that $|\tilde{U}_i \setminus W_i'| < 500 c_3 n$ for each $i \in [4]$. Suppose that no such relabeling exists. Then, since $c_3 \ll 1$ and e.g. each $|W_i'| \leq n/3$, there is $i \in [4]$ such that for every $j \in [4]$ we have that $|\tilde{U}_i \setminus W_j'| \geq 500 c_3 n$. Take $j \in [4]$ such that $|\tilde{U}_i \cap W_j'| \geq |\tilde{U}_i|/4$ and let $X = \tilde{U}_i \cap W_j'$ and $Y = \tilde{U}_i \setminus W_j'$. However, $X, Y \subseteq \tilde{U}_i$ and Lemma 3.3.7 imply that $e(G[X,Y]) < 12 c_3 n^2$ whereas $X \subseteq W_j'$, $Y \cap W_j' = \emptyset$, (3.28), and (3.29) imply that

$$e(G[X,Y]) \geq |X|\,|Y| - 15 c_3 n^2 \geq (n/16 - c_7 n) \cdot 500 c_3 n - 15 c_3 n^2 > 12 c_3 n^2,$$

a contradiction. So take the stated relabeling. Now, (3.27) follows from the observation that

$$W_i' \setminus \tilde{U}_i \subseteq \bigcup_{j\in[4]\setminus\{i\}} (\tilde{U}_j \setminus W_j').$$

Alternatively, one could use Lemma 3.3.7 that G is $12c_3n^2$-close to the complete 4-partite graph $\tilde{F}$ whose part sizes are close to $n/4$ by (3.28). One would get a weaker upper bound on $|W_i' \bigtriangleup \tilde{U}_i|$ (of order $\sqrt{c_3}n$) but which would also be sufficent for our proof.

Finally, the last two claims of Lemma 3.3.8 can be derived by taking $W_i' = W_i$ for $i \in [4]$ (and using (3.28)). $\square$

Define a *pattern* as an assignment $\pi : \binom{[4]}{2} \to \binom{[4]}{2} \cup \binom{[4]}{3}$ (to every edge of K_4 we assign a list of 2 or 3 colors) such that $\pi^{-1}(c)$ forms a 4-cycle for every color $c \in [4]$. Up to isomorphism (of colors and vertices) there is only one pattern. We say that an edge 4-coloring χ of $G \in \mathcal{G}_n$ *follows the pattern* π if for every $ij \in \binom{[4]}{2}$ we have

$$\big|\,\chi^{-1}([4] \setminus \pi(ij)) \cap G[W_i, W_j]\,\big| \le c_2n^2,$$

that is, at most c_2n^2 edges of $G[W_i, W_i]$ get a color not in $\pi(ij)$.

Recall that the set $\mathcal{S}(G)$ consists of all satisfactory colorings whose associated vector is popular.

Lemma 3.3.9. *For every graph $G \in \mathcal{G}_n$ with $n \ge n_0$, every coloring $\chi \in \mathcal{S}(G)$ follows a pattern.*

Proof. Take any $\chi \in \mathcal{S}(G)$. Recall that $A, B, U_1, \ldots, U_4$ are the sets given by Lemma 3.3.3. Let

$$R' = \big(R_3 \cap K(A,B)\big) \cup \big(R_2 \cap K(U_1,U_2)\big) \cup \big(R_2 \cap K(U_3,U_4)\big).$$

Let the *label* of an edge uv in R be $\hat{\chi}(uv) = \{i \in [4] : uv \in E(H_i)\}$. So, for all edges $u_iu_j \in R'$ across $U_i \times U_j$, we have

$$|\hat{\chi}(u_iu_j)| = \begin{cases} 2, & \text{if } \{i,j\} \in \{\{1,2\},\{3,4\}\}, \\ 3, & \text{otherwise.} \end{cases} \tag{3.30}$$

We show next that $\hat{\chi}$ has a very simple structure: with the exception of a small fraction of edges, $\hat{\chi}$ behaves as the blow up of some labeling on K_4. Furthermore, the latter labeling is isomorphic to some pattern π, as defined above.

Claim 3.3.10. *Let the sets $\{v_1, v_2, v_3, v_4\}$ and $\{w, v_2, v_3, v_4\}$ both span a K_4-subgraph in R', where $w \in U_1$ and each $v_i \in U_i$. Then $\hat{\chi}(v_1v_i) = \hat{\chi}(wv_i)$ for all $i \in \{2,3,4\}$.*

Proof of Claim. First consider the restriction of $\hat{\chi}$ to $X = \{v_1, v_2, v_3, v_4\}$. Let S be the multi-set produced by the union of $\hat{\chi}(v_iv_j)$, $1 \le i < j \le 4$. So, $|S| = 2 \cdot 2 + 4 \cdot 3 = 16$. As each $H_t[X]$ is triangle-free, it follows by the uniqueness of the Turán graph that $\hat{\chi}^{-1}(t)$ forms a 4-cycle on X for each $t \in [4]$. When taking (3.30) into consideration, we see that there is only one possible configuration (up to isomorphism). A nice property of this configuration is that $\hat{\chi}(v_iv_j) = \hat{\chi}(v_kv_\ell)$ whenever $\{i, j, k, \ell\} = [4]$, i.e., edges that form a matching on X receive identical labels. As $\{w, v_2, v_3, v_4\}$ also spans a copy of K_4, we have $\hat{\chi}(wv_j) = \hat{\chi}(v_kv_\ell) = \hat{\chi}(v_1v_\ell)$, where $\{j, k, \ell\} = \{2, 3, 4\}$, proving the claim. ▌

Now choose $X = \{v_1, v_2, v_3, v_4\}$, where $v_i \in U_i$, such that $R'[X] \cong K_4$ and, for each vertex $v_i \in X$, we have

$$|N_{R'}(v_i) \cap U_j| > |U_j| - 2\sqrt{c_3}m \quad \text{for all } j \in [4] \setminus \{i\}. \tag{3.31}$$

We may build such a set iteratively by picking $v_1 \in U_1$ satisfying (3.31), then $v_2 \in U_2 \cap N(v_1)$ satisfying (3.31), and so on. We are guaranteed the existence of such vertices as at most $2c_3m^2$ edges across a pair U_i, U_j are missing from R'. In fact, the number of vertices $u \in U_i$ that fail condition (3.31) is less than $3\sqrt{c_3}m$.

Let $A_i \subseteq U_i$ consist of those vertices that lie in $N_{R'}(v_j)$ for all $v_j \in X$ with $j \in [4] \setminus \{i\}$. As all vertices v_j satisfy (3.31), we have $|A_i| > |U_i| - 6\sqrt{c_3}m$. If $a_ia_j \in R'[A_i, A_j]$, then all three sets $X, \{a_i, v_j, v_k, v_\ell\}$, and $\{a_i, a_j, v_k, v_\ell\}$ form 4-cliques in R', where $\{i, j, k, \ell\} = [4]$. By Claim 3.3.10 we have that $\hat{\chi}(v_iv_j) = \hat{\chi}(a_iv_j) = \hat{\chi}(a_ia_j)$. Thus, the labeling on X determines the labeling on all edges of R' with the possible exception of at most $m \cdot 24\sqrt{c_3}m$ edges incident to vertices of $\bigcup_{i=1}^4 (U_i \setminus A_i)$. As $|R \setminus R'| < 5c_3m^2$, we have a pattern π such that $\hat{\chi}(u_iu_j) = \pi(ij)$ for all but at most $25\sqrt{c_3}m^2$ edges in R.

Now, Lemma 3.3.8 implies that for some relabeling of $W_1, \ldots, W_4$, we have

$$|K(W_1, W_2, W_3, W_4) \setminus K(\tilde{U_1}, \tilde{U_2}, \tilde{U_3}, \tilde{U_4})| < 4n \cdot 2000c_3n.$$

Then, including at most $5c_7n^2$ edges that disappear without a trace in any H_i during the application of the Regularity Lemma and at most $12c_3n^2$ edges lost in Lemma 3.3.7, we have that $\chi(w_iw_j) \in \pi(ij)$ for all but at most

$$5c_7n^2 + 12c_3n^2 + 25\sqrt{c_3}m^2 \cdot \lceil n/m \rceil^2 + 8000c_3n^2 < c_2n^2$$

edges w_iw_j in $G[W_i, W_j]$, proving the lemma. □

Since c_2 is small, Lemma 3.3.8 implies that the pattern π in Lemma 3.3.9 is unique. This allows us to make the following definition. A coloring $\chi \in \mathcal{S}(G)$ of a graph $G \in \mathcal{G}_n$ is *good* if for every $ij \in \binom{[4]}{2}$, all subsets $X_i \subseteq W_i$ and $X_j \subseteq W_j$ with $|X_i| \geq c_1 n$ and $|X_j| \geq c_1 n$, and every color $c \in \pi(ij)$ there is at least one edge xy in $G[X_i, X_j]$ with $\chi(xy) = c$, where π is the pattern of χ. Otherwise $\chi \in \mathcal{S}(G)$ is called *bad*.

Lemma 3.3.11. *The number of bad colorings of any $G \in \mathcal{G}_n$, $n \geq n_0$, is at most $18^{n^2/8} \cdot 2^{-c_1^2 n^2/8}$.*

Proof. The following procedure generates each bad coloring of G at least once.

1. Pick an arbitrary pattern π, a pair $ij \in \binom{[4]}{2}$, and a color $c \in \pi(ij)$.
2. Choose up to $6c_2n^2$ edges and color them arbitrarily.
3. Pick subsets $X_i \subseteq W_i$ and $X_j \subseteq W_j$ of size $\lceil c_1 n \rceil$ each.
4. Color edges inside a part W_i arbitrarily.
5. Color all edges in $X_i \times X_j$ using the colors from $\pi(ij) \setminus \{c\}$.
6. For each $k\ell \in \binom{[4]}{2}$ color all remaining edges of $G[W_k, W_\ell]$ using colors in $\pi(k\ell)$.

The number of choices in Steps 1–3 is bounded from above by

$$O(1)\binom{\binom{n}{2}}{\leq 6c_2n^2} 4^{6c_2n^2}\binom{|W_i|}{|X_i|}\binom{|W_j|}{|X_j|} < 2^{c_1^3 n^2}.$$

The number of choices at Step 4 is at most $4^{15c_3n^2}$ by Lemma 3.3.8. The number of choices in Steps 5–6 is at most

$$\left(\frac{|\pi(ij)|-1}{|\pi(ij)|}\right)^{|X_i|\,|X_j|} \prod_{k\ell\in\binom{[4]}{2}} |\pi(k\ell)|^{|W_k|\,|W_\ell|} \leq (2/3)^{c_1^2n^2}\,(2^2\,3^4)^{n^2/16+c_2n^2},$$

where we used Lemma 3.3.8. We obtain the required result by multiplying the above bounds. □

Call a good coloring χ of a graph $G \in \mathcal{G}$ *perfect* if $\chi(v_iv_j) \in \pi(ij)$ for every $ij \in \binom{[4]}{2}$ and every edge $v_iv_j \in G[W_i, W_j]$, where π is the pattern of χ. Let $\mathcal{P}(G)$ denote the set of perfect colorings of G.

The following lemma provides a key step of the whole proof.

Lemma 3.3.12. *Let G be a graph of order $n \geq n_0+2$ such that $F(G,4,3) \geq 18^{n^2/8} \cdot 2^{-c_9n^2}$*

and for every distinct $v, v' \in V(G)$ we have

$$\frac{F(G,4,3)}{F(G-v,4,3)} \geq (18-c_3)^{n/4}, \tag{3.32}$$

$$\frac{F(G,4,3)}{F(G-v-v',4,3)} \geq (18-c_3)^{(n+(n-1))/4}. \tag{3.33}$$

Then the following conclusions hold.

1. *G is 4-partite.*
2. *Almost every coloring of G is perfect; specifically,*

$$|\mathcal{P}(G)| \geq (1-2^{-c_9 n})\, F(G,4,3).$$

3. *If $G \ncong T_4(n)$, then there is a graph G' of order n with $F(G',4,3) > F(G,4,3)$.*

Proof. Since $F(G-v-v',4,3) > F(G-v,4,3)/4^n > F(G,4,3)/16^n$ for any $v, v' \in V(G)$, we have $G-v, G-v-v' \in \mathcal{G}$ and the notion of a good coloring with respect to $G-v$ or $G-v-v'$ is well-defined.

Claim 3.3.13. *For any distinct $v, v' \in V(G)$, there is a good coloring χ of $G-v$ (resp. of $G-v-v'$) such that the number of ways to extend it to the whole of G is at least $(18-c_2)^{n/4}$ (resp. at least $(18-c_2)^{n/2}$).*

Proof of Claim. By Lemma 3.3.11 the number of bad colorings of $G-v$ is at most $2^{-c_1^2 n^2/9} F(G,4,3)$. If the claim fails for all good colorings of $G-v$, then

$$F(G,4,3) \leq 4^n \cdot 2^{-c_1^2 n^2/9} F(G,4,3) + (18-c_2)^{n/4} F(G-v,4,3),$$

contradicting (3.32). The claim about $G-v-v'$ is proved in an analogous way. ▮

Claim 3.3.14. *For all $i \in [4]$ and $v \in W_i$, we have $|N(v) \cap W_i| < 8c_1 n$.*

Proof of Claim. Suppose on the contrary that some vertex v contradicts the claim. Take the good coloring χ of $G-v$ given by Claim 3.3.13.

For each class W_j (defined with respect to G), let $n_j = |N(v) \cap W_j|$. Note that

$$n_j \leq |W_j| \leq n/4 + c_2 n, \quad \text{for all } j \in [4], \tag{3.34}$$

by Lemma 3.3.8. Let $W_1' \cup W_2' \cup W_3' \cup W_4'$ be the selected max-cut partition of $G-v$. As

$$\big|\, G \cap K(W_1' \cup \{v\}, W_2', W_3', W_4') \,\big| > \big|\, G \cap K(W_1, W_2, W_3, W_4) \,\big| - n,$$

it follows again from Lemma 3.3.8 that, after a relabeling of $W_1', \ldots, W_4'$, we have

$$|W_i \triangle W_i'| \leq 4000 c_3 n + 1, \quad \text{for all } i \in [4]. \tag{3.35}$$

Also, let π be the pattern (with respect to $W_1', \ldots, W_4'$) associated with the good coloring χ of $G - v$.

For each extension $\bar{\chi}$ of χ to G, record the vector $\mathbf{x}$ whose i-th component is the number of colors c such that at least $2c_1 n$ edges of G between v and W_i get color c. Let $\mathbf{x} = (x_1, \ldots, x_4)$ be a vector that appears most frequently over all extensions $\bar{\chi}$. Fix some $\bar{\chi}$ that gives this $\mathbf{x}$. For a color c and a class W_j, let

$$Z_{j,c} = \{u \in W_j : \bar{\chi}(uv) = c\}.$$

(Thus x_j is the number of colors c with $|Z_{j,c}| \geq 2c_1 n$.) Analogously, for a color c, let y_c be the number of classes W_j for which $|Z_{j,c}| \geq 2c_1 n$. By (3.35), we have $|Z_{j,c} \cap W_j'| > c_1 n$ whenever $|Z_{j,c}| > 2c_1 n$.

Let us show that $y_c \leq 2$ for each $c \in [4]$. Indeed, if some $y_c \geq 3$, then among the three corresponding indices we can find two, say p and q, such that $c \in \pi(pq)$. Since χ is good, there is an edge $uw \in (Z_{p,c} \cap W_p') \times (Z_{q,c} \cap W_q')$ such that $\chi(uw) = c$, giving a $\bar{\chi}$-monochromatic triangle on $\{u, v, w\}$, a contradiction. In particular, we have

$$x_1 + x_2 + x_3 + x_4 = y_1 + y_2 + y_3 + y_4 \leq 8. \tag{3.36}$$

Since there are at most 5^4 choices of $(x_1, \ldots, x_4)$ and we fixed a most frequent vector, the total number of extensions of χ to G is at most

$$5^4 \prod_{j \in [4]} \binom{4}{x_j} \binom{n_j}{\leq 2c_1 n}^{4 - x_j} \max(x_j, 1)^{n_j} < 2^{c_0 n} \prod_{\substack{j \in [4] \\ x_j \neq 0}} x_j^{n_j}. \tag{3.37}$$

As $W_1 \cup W_2 \cup W_3 \cup W_4$ is a max-cut partition, we have $|N(v) \cap W_j| \geq 8c_1 n$ for all $j \in [4]$. By the pigeonhole principle, we have that $x_j \geq 1$ for all $j \in [4]$. This and (3.36) imply that $x_1 x_2 x_3 x_4 \leq 16$. By (3.34) and (3.37), the total number of extensions of χ is at most

$$2^{c_0 n} \cdot (x_1 x_2 x_3 x_4)^{n/4} \cdot 4^{4c_2 n} < 2^{2c_0 n} \cdot 16^{n/4} < (18 - c_2)^{n/4},$$

contradicting the choice of χ. ∎

We will now strengthen Claim 3.3.14 and prove Part 1 of the lemma.

Claim 3.3.15. *For all $i \in [4]$ and distinct $v, v' \in W_i$, we have $vv' \notin E(G)$.*

Proof of Claim. Suppose on the contrary that the claim fails for some v and v'. Assume without loss of generality that $v, v' \in W_1$.

Let χ be the good coloring of $G - v' - v \in \mathcal{G}_{n-2}$ with at least $(18 - c_2)^{n/2}$ extensions to G given by Claim 3.3.13. Let us recycle the definitions of Claim 3.3.14 that formally remain unchanged even though χ is undefined on edges incident to v'. On top of them, we define a few more parameters.

Specifically, we look at all extensions $\bar{\chi}$ that give rise to the fixed most frequent vector $\mathbf{x}$. For each such $\bar{\chi}$, we define $Z'_{j,c} = \{u \in W_j : \bar{\chi}(uv') = c\}$ and let x'_j be the number of colors c such that $|Z'_{j,c}| \geq 2c_1 n$. Then we fix a most popular vector $\mathbf{x}' = (x'_1, \ldots, x'_4)$ and take any extension $\bar{\chi}$ that gives both $\mathbf{x}$ and $\mathbf{x}'$ and, conditioned on this, such that the color $\bar{\chi}(vv')$ assumes its most frequent value, which we denote by s. We define y_c as before and let y'_c be the number of $j \in [4]$ such that $|Z'_{c,j}| \geq 2c_1 n$. This is consistent with the definitions of Claim 3.3.14 because there we did not have any restriction on $\bar{\chi}$ except that it gives the vector $\mathbf{x}$.

Claim 3.3.14, the upper bounds on n_i and $n'_i = |N(v') \cap W_i|$ of Lemma 3.3.8, and the argument leading to (3.37) show that the total number of extensions of χ to G is at most

$$(5^4)^2 \cdot 4 \cdot 2^{c_0 n} \cdot (4^{8c_1 n + 3c_2 n})^2 \cdot \prod_{i=2}^{4} \big(\max(x_i, 1) \cdot \max(x'_i, 1)\big)^{n/4}. \tag{3.38}$$

If some $|Z_{j,c}| \geq 2c_1 n$ but $c \notin \pi(\{1, j\})$, say $j = 3$, then the 4-cycle formed by color c visits indices $1, 2, 3, 4$ in this order and, since χ is good, we have $|Z_{2,c}| < 2c_1 n$ and $|Z_{4,c}| < 2c_1 n$ (otherwise $\bar{\chi}$ contains a color-c triangle via v). Thus y_c contributes at most 1 to $x_2 + x_3 + x_4$. Since each $y_i \leq 2$, we have that $x_2 + x_3 + x_4 \leq 7$. It follows that $\prod_{i=2}^{4} \max(x_i, 1) \leq 12$. Since $x'_2 + x'_3 + x'_4 \leq 8$, we have $\prod_{i=2}^{4} \max(x'_i, 1) \leq 18$. Thus the expression in (3.38) is at most $2^{2c_0 n} \cdot (12 \cdot 18)^{n/4}$, contradicting the choice of χ.

Thus $x_i \leq |\pi(\{1, i\})|$ for each $i \in \{2, 3, 4\}$ and all these inequalities are in fact equalities (otherwise $\prod_{i=2}^{4} \max(x_i, 1) \leq 12$, giving a contradiction as before). We conclude for $j \in \{2, 3, 4\}$ that $|Z_{j,c}| \geq 2c_1 n$ if and only if $c \in \pi(\{1, j\})$. The same applies to the parameters x'_i and $Z'_{j,c}$.

Let the special color $s = \bar{\chi}(vv')$ appear in, say $\pi(\{1, 2\})$ with $|\pi(\{1, 2\})| \geq 3$. Then for all $w \in W_2 \cap N(v) \cap N(v')$ there are at most $x_2 x'_2 - 1$ choices for the colors of vw and vw' when extending χ to G because s cannot occur on both edges. Also, if $w \in W_2 \setminus (N(v) \cap N(v'))$ then trivially there are at most 4 choices for this vertex w. This allows us to reduce the bound in (3.38) by factor $(8/9)^{n/4}$, giving the desired contradiction. ∎

Thus we have proved Part 1 of the lemma. Next, we prove Part 2. If it is false, then by Lemma 3.3.11 there are there are at least $(1/2) \cdot 2^{-c_9 n} \cdot F(G, 4, 3)$ colorings of G that are good but not perfect. For each such coloring there is a *wrong* edge vv' whose color does not conform to the pattern. Pick an edge vv' that appears most frequently this way, say $v \in W_1$ and $v' \in W_4$, and then a most frequent wrong color s of vv'.

By a version of (3.35), it is not hard to show that the number of good colorings χ of $G-v-v'$ for which there is an extension $\bar{\chi}$ which is a good coloring of G but with a *different* pattern than that of χ is at most, for example, $2^{-c_1^2 n^2/9} \cdot F(G,4,3)$. Indeed, the pattern π may change only when some color $c \in \pi(ij)$ appears very infrequently in $G[W_i, W_j]$ and the proportion of such (degenerate) colorings can be bounded by e.g. $(2/3 + c_0)^{(n/4)^2}$.

It follows that there is a good coloring χ of $G - v - v'$ that has at least $(18 - c_2)^{n/2}$ pattern-preserving extensions to G with vv' getting the wrong color s. Indeed, if this is false, then we would get a contradiction to (3.33) by an argument of Claim 3.3.13:

$$\frac{(1/2) \cdot 2^{-c_9 n} \cdot F(G,4,3)}{4 \cdot \binom{n}{2}} \le 2 \cdot 16^n \cdot 2^{-c_1^2 n^2/9} \cdot F(G,4,3) + (18 - c_2)^{n/2} F(G - v - v', 4, 3),$$

Here, we use use the trivial upper bound 16^n on the number of extensions for colorings of $G - v - v'$ that are bad or are good but admit an extension with a different pattern. On the other hand, a good but not perfect coloring of G may be overcounted as there may be more than one choice of vv' and s; we use the trivial upper bound of $4 \cdot \binom{n}{2}$ here.

Defining $\pi, x_i, x_i', Z_{j,c}, Z_{j,c}', y_i, y_i'$ as in Claim 3.3.15, one can argue similarly to (3.38) that the number of pattern-preserving extensions of χ is at most

$$2^{c_0 n} \left(\prod_{j=2}^{4} \max(x_j, 1) \cdot \prod_{j=1}^{3} \max(x_j', 1) \right)^{n/4}, \tag{3.39}$$

where all smaller terms are swallowed by $2^{c_0 n}$. Moreover, as before, $|Z_{j,c}| \ge 2c_1 n$ if and only if $c \in \pi(\{1,j\})$ while $|Z_{j,c}'| \ge 2c_1 n$ if and only if $c \in \pi(\{j,4\})$.

Since $s \notin \pi(\{1,4\})$, we have $s \in \pi(\{1,3\}) \cap \pi(\{3,4\})$. But then the number of choices per $w \in W_3 \cap N(v) \cap N(v')$ is at most $x_3 x_3' - 1$ (instead of $x_3 x_3'$) because we cannot assign color s to both vw and vw'. Also, if vw or vw' is not an edge, then we have at most 4 choices per w (while $|\pi(\{1,3\})| \cdot |\pi(\{3,4\}| \ge 6)$. This allows us to improve (3.39) by factor $(8/9)^{n/4}$. This contradicts the choice of χ and proves Part 2 of Lemma 3.3.12.

Let $H = K(W_1, \ldots, W_4)$. Suppose first that $G \not\cong H$, that is, G is not complete 4-partite. We know that almost every coloring χ of G is perfect. Moreover, if we start with a perfect coloring χ of G and color all remaining edges in $E(H) \setminus E(G)$ according to the pattern of χ then we get at least $2^{|H \setminus G|} \ge 2$ extensions to H none containing a monochromatic K_3. Thus $|\mathcal{P}(H)| \ge 2|\mathcal{P}(G)| > F(G,4,3)$ and we can take $G' = H$.

Finally, suppose that $G = H$ but $G \not\cong T_4(n)$. Let $d_i = |W_i|$ for $i \in [4]$. Assume, without loss of generality, that $d_1 \geq d_2 \geq d_3 \geq d_4$ with $d_1 \geq d_4 + 2$. Let G' be the complete 4-partite graph with parts of size $d_1 - 1, d_2, d_3, d_4 + 1$. We already know that almost every coloring of G is perfect. Thus, in order to finish the proof it is enough to show that, for example, $|\mathcal{P}(G')| > 1.1\,|\mathcal{P}(G)|$.

The number of perfect colorings of G is given by the following expression:

$$|\mathcal{P}(G)| = (12 + o(1))\,(S_1 + S_2 + S_3), \tag{3.40}$$

where

$$\begin{aligned} S_1 &= 2^{d_1d_2+d_3d_4}3^{d_1d_3+d_1d_4+d_2d_3+d_2d_4}, \\ S_2 &= 2^{d_1d_3+d_2d_4}3^{d_1d_2+d_1d_4+d_2d_3+d_3d_4}, \\ S_3 &= 2^{d_1d_4+d_2d_3}3^{d_1d_2+d_1d_3+d_2d_4+d_3d_4}. \end{aligned}$$

Note that we have an error term in (3.40) because some (degenerate) colorings are overcounted in the right-hand side. Also,

$$\begin{aligned} |\mathcal{P}(G')| &= (12 + o(1))\left(2^{-d_2+d_3}3^{d_1-d_4-1+d_2-d_3} \cdot S_1\right. \\ &+ \left. 2^{d_2-d_3}3^{d_1-d_4-1-d_2+d_3} \cdot S_2 + 2^{d_1-d_4-1} \cdot S_3\right). \end{aligned}$$

But, as $d_1 - d_4 \geq \max\{2, d_2 - d_3\}$, the coefficient in front of each S_i is at least $4/3$. Therefore $|\mathcal{P}(G')| > 1.1\,|\mathcal{P}(G)|$, finishing the proof of Lemma 3.3.12. □

Routine calculations (omitted) show that

$$|\mathcal{P}(T_4(n))| = (C + o(1)) \cdot 18^{t_4(n)/3}, \tag{3.41}$$

where $C = (2^{14} \cdot 3)^{1/3}$ if $n \equiv 2 \pmod 4$ and $C = 36$ otherwise.

Proof of Theorem 3.1.1. Let e.g. $N = n_0^2$. Let G be an extremal graph on $n \geq N$ vertices. Suppose on the contrary that $G \not\cong T_4(n)$. Let $G_n = G$.

We iteratively apply the following procedure. Given a current graph G_m on $m \geq n_0 + 2$ vertices with $F(G_m, 4, 3) \geq 18^{m^2/8} \cdot 2^{-c_9 m^2}$ we apply Lemma 3.3.12. If (3.32) fails for some vertex $v \in V(G_m)$, we let $G_{m-1} = G_m - v$, decrease m by 1, and repeat. Note that

$$F(G_{m-1}, 4, 3) \geq F(G_m, 4, 3)/(18 - c_3)^{m/4} \geq 18^{(m-1)^2/8} \cdot 2^{-c_9(m-1)^2}.$$

Likewise, if (3.33) fails for some distinct $v, v' \in V(G_m)$, we let $G_{m-2} = G - v - v'$, decrease m by 2, and repeat. If both (3.32) and (3.33) hold and $G_m \not\cong T_4(m)$, replace G_m by the graph G' returned by Lemma 3.3.12 and repeat the step (without decreasing m).

Note that for every m for which G_m is defined we have

$$F(G_m, 4, 3) \geq F(G, 4, 3) \cdot (18 - c_3)^{-(n+(n-1)+\ldots+(m+1))/4}. \tag{3.42}$$

It follows that we never reach $m < n_0 + 2$ for otherwise, when this happens for the first time, we get the contradiction

$$F(G_m, 4, 3) \geq \frac{18^{n^2/8} \cdot 2^{-c_9 n^2}}{(18 - c_3)^{(\binom{n}{2} - \binom{m}{2})/4}} > 4^{\binom{m}{2}}.$$

Thus we stop for some $m \geq n_0 + 2$, having $G_m \cong T_4(m)$. We cannot have $m = n$, for otherwise $T_4(n)$ strictly beats G. By Lemma 3.3.12, almost every coloring of $G_m \cong T_4(m)$ is perfect. Thus, by (3.42),

$$2 \cdot |\mathcal{P}(T_4(m))| > F(T_4(m), 4, 3) \geq F(G, 4, 3) \cdot (18 - c_3)^{-(n+(n-1)+\cdots+(m+1))/4}. \tag{3.43}$$

Also, note that $t_4(\ell) - t_4(\ell - 1) = \lfloor 3\ell/4 \rfloor$. Thus, (3.41) implies that, for example, $|\mathcal{P}(T_4(\ell))| \geq 18^{\ell/4-1} |\mathcal{P}(T_4(\ell - 1))|$ for all $\ell \geq n_0$. By the extremality of G, we conclude that

$$F(G, 4, 3) \geq F(T_4(n), 4, 3) \geq |\mathcal{P}(T_4(n))| \geq \frac{18^{(n+\cdots+(m+1))/4}}{18^{n-m}} |\mathcal{P}(T_4(m))|. \tag{3.44}$$

But (3.43) and (3.44) give a contradiction to $n > m$, proving Theorem 3.1.1. $\square$

Remark. If we set $G = T_4(n)$ with $n \geq N$ in the above argument, then we conclude that $m = n$ (otherwise we get a contradiction as before). Thus we do not perform any iterations at all, which implies that (3.32) and (3.33) hold for $T_4(n)$. By Part 2 of Lemma 3.3.12 almost every coloring of $T_4(n)$ is perfect. Thus the estimate (3.5) that was claimed in the Introduction follows from (3.41).

3.4 $F(n, 4, 4)$

In this section we prove Theorem 3.1.2. Some parts of the proof closely follow those of Theorem 3.1.1. We omit many details that have already been presented or are obvious modifications of those in Section 3.3. We start by fixing positive constants

$$c_0 \gg c_1 \gg \ldots \gg c_{10}.$$

Let $M = 1/c_9$ and $n_0 = 1/c_{10}$. Define

$$\mathcal{F}_n = \left\{ G : v(G) = n,\ F(G,4,4) \geq 3^{4n^2/9} \cdot 2^{-c_8 n^2} \right\}.$$

and let $\mathcal{F} = \bigcup_{n \geq n_0} \mathcal{F}_n$. The lower bound in (3.4) shows that $\mathcal{F}_n$ is non-empty for every $n \geq n_0$.

Using the obvious analogs of the previous definitions, we define the parameters $(m, V_i, H_i, R_i, R, r_i)$ arising from an arbitrary graph G and a K_4-free 4-coloring χ of the edges of G and fix one such vector for each pair (G, χ).

By Lemma 3.2.2, each cluster graph H_i is K_4-free and, by Turán's theorem (3.1), has at most $t_3(m)$ edges. Thus by (3.7)

$$r_1 + 2r_2 + 3r_3 + 4r_4 = \frac{e(H_1) + e(H_2) + e(H_3) + e(H_4)}{m^2/2} \leq \frac{8}{3}. \tag{3.45}$$

We also have a procedure for generating all K_4-free edge 4-colorings of G at least once. This procedure is identical to the Coloring Procedure provided in Section 3.3 with the only difference being that in Step 3 the parameters r_i (where we omit primes for convenience) now satisfy (3.45) instead of (3.10) and (3.11). So, Lemma 3.3.1 that bounds the number of choices in Steps 1–6 still holds.

The number of options in Step 7 is again bounded by (3.13), i.e., the expression $2^{Ln^2/2+O(n)}$, where $L = r_2 + \log_2(3)\, r_3 + 2r_4$. Under the constraint (3.45) and the non-negativity of the r_i's, the maximum of L is $(8/9)\log_2 3$ with the (unique) optimal assignment being $r_1 = r_2 = r_4 = 0$ and $r_3 = 8/9$. We conclude that

$$F(n,4,4) \leq 3^{4n^2/9} \cdot 2^{2c_6 n^2},$$

as it was also shown in [ABKS04].

We will now obtain structural information about the cluster graphs (and, indirectly, about G). We call a pair (G, χ) (or the coloring χ) *unsatisfactory* if

$$r_3 \leq 8/9 - c_4. \tag{3.46}$$

Otherwise, (G, χ) is *satisfactory.*

Lemma 3.4.1. *For every graph G with $n \geq n_0$ vertices the number of unsatisfactory K_4-free edge 4-colorings is less than $3^{4n^2/9} \cdot 2^{-c_6 n^2}$.*

Proof. The maximum of L under constraints (3.45) and (3.46) (and the non-negativity of r_i's) is

$$L_{\max} = (8/9 - c_4)\log_2(3) + 3c_4/2 < (8/9)\log_2(3) - c_5,$$

with the optimal dual variables for (3.45) and (3.46) being $y_1 = 1/2$ and $y_2 = \log_2(3) - 3/2 > 0$ respectively. Therefore, the total number of choices is at most $2^{c_6 n^2} \cdot 2^{L_{\max} n^2/2 + O(n)}$, giving the required upper bound on the number of unsatisfactory colorings. □

Call a vector (m, V_i, H_i) *popular* if it appears for at least $3^{4n^2/9} \cdot 2^{-3c_8 n^2}$ satisfactory K_4-free edge 4-colorings of G. As before, (3.12) guarantees that the number of colorings for which the associated vector is not popular is at most $3^{4n^2/9} \cdot 2^{-2c_8 n^2}$. Let $\mathrm{Pop}(G)$ be the set of all popular vectors and let $\mathcal{S}(G)$ consist of all satisfactory colorings for which the associated vector is popular.

Lemma 3.4.2. *For any $n \geq n_0$, a graph $G \in \mathcal{F}_n$, and a popular vector $(m, V_i, H_i) \in \mathrm{Pop}(G)$, there exists an equitable partition $[m] = U_1 \cup \ldots \cup U_9$ such that*

$$\left| R_3 \triangle K(U_1, \ldots, U_9) \right| < c_3 m^2, \tag{3.47}$$

$$\left| R \triangle K(U_1, \ldots, U_9) \right| < 2c_3 m^2. \tag{3.48}$$

Proof. Suppose that some $Y \subseteq [m]$ induces a clique of order 10 in R_3. Then $R_3[Y]$ contains $\binom{10}{2} = 45$ edges, each of which, by definition, belongs to exactly 3 cluster graphs H_i. Each H_i is K_4-free so, by Turán's Theorem (3.1), $H_i[Y]$ has at most $t_3(10) = 33$ edges. But $4 \cdot 33 < 3 \cdot 45$, a contradiction.

Thus $K_{10} \not\subseteq R_3$. Since $e(R_3) \geq (8/9 - c_4)m^2/2$, Lemma 3.2.3 gives an equitable partition $[m] = U_1 \cup \ldots \cup U_9$ satisfying (3.47). This partition also satisfies (3.48) because $r_1 + r_2 + r_4 \leq 3c_4$ by (3.45) and the negation of (3.46). □

For a graph G and a popular vector $(m, V_i, H_i) \in \mathrm{Pop}(G)$, fix the equitable 9-partition $[m] = U_1 \cup \cdots \cup U_9$ given by Lemma 3.4.2. For $i \in [9]$, let $\tilde{U}_i = \cup_{j \in U_i} V_j$ be the *blow-up* of U_i. Let $\tilde{F} = K(\tilde{U}_1, \ldots, \tilde{U}_9)$.

Lemma 3.4.3. *For any $n \geq n_0$, $G \in \mathcal{F}_n$, and $(m, V_i, H_i) \in \mathrm{Pop}(G)$, we have $|G \triangle \tilde{F}| < 6c_3 n^2$.*

Proof. First consider $G \setminus \tilde{F}$. Up to $5c_7 n^2$ edges may be lost by application of the Regularity Lemma. In addition, at most $|R \setminus K(U_1, \ldots, U_9)| \cdot \lceil n/m \rceil^2$ edges are missing in $\tilde{F}$. Overall, $|G \setminus \tilde{F}| < 3c_3 n^2$.

On the other hand, we may estimate $|\tilde{F} \setminus G|$ by bounding the number of colorings of G associated with our vector (m, V_i, H_i). We revert to the Coloring Procedure and compute the number of options in Step 7:

$$\begin{aligned} \prod_{f=2}^{4} \prod_{ij \in R_f} f^{\lceil n/m \rceil^2 - |K(V_i,V_j) \setminus G|} &\leq \left(3^{4n^2/9} \cdot 2^{2c_6 n^2}\right) \prod_{ij \in R_3} 2^{-|K(V_i,V_j)\setminus G|} \\ &\leq \left(3^{4n^2/9} \cdot 2^{2c_6 n^2}\right) \cdot 2^{-|\tilde{F} \setminus G| + 2c_3 n^2 + O(n)}. \end{aligned}$$

Since the vector (m, V_i, H_i) is popular, we have

$$|\, \tilde{F} \setminus G \,| \leq c_6 n^2 + 2c_3 n^2 + 2c_6 n^2 + 3c_8 n^2 + O(n) \leq 3c_3 n^2,$$

as required. □

For each graph G fix a max-cut partition $V(G) = W_1 \cup \cdots \cup W_9$.

Lemma 3.4.4 (Stability Property)**.** *Let $n \geq n_0$, $G \in \mathcal{F}_n$, and $V(G) = W_1' \cup \ldots \cup W_9'$ be a partition with*

$$\big|\, G \cap K(W_1', \ldots, W_9') \,\big| \geq \big|\, G \cap K(W_1, \ldots, W_9) \,\big| - c_3 n^2.$$

Then $|\, G \triangle K(W_1', \ldots, W_9') \,| \leq 9c_3 n^2$ and, for any $(m, V_i, H_i) \in \mathrm{Pop}(G)$, there is a relabeling of $W_1', \ldots, W_9'$ such that

$$\big|\, W_i' \triangle \tilde{U}_i \,\big| \leq 12000\, c_3 n, \quad \textit{for each } i \in [9]. \tag{3.49}$$

Also we have $|\, |W_i| - n/9 \,| \leq c_2 n$ for each $i \in [9]$ and $|\, G \triangle K(W_1, \ldots, W_9) \,| \leq 9c_3 n^2$.

Proof. Let $F' = K(W_1', \ldots, W_9')$ and $F = K(W_1, \ldots, W_9)$. As $W_1 \cup \ldots \cup W_9$ is a max-cut partition, we have $|F' \cap G| + c_3 n^2 \geq |F \cap G| \geq |\tilde{F} \cap G|$. In addition, both $[m] = U_1 \cup \cdots \cup U_9$ and $[n] = V_1 \cup \cdots \cup V_m$ are equitable partitions, so $|\, |\tilde{U}_i| - n/9 \,| < m + n/m$. It follows that $|\tilde{F}| \geq |\, F'| - c_7 n^2$, and

$$|\, F' \triangle G \,| \leq |\, \tilde{F} \triangle G \,| + c_7 n^2 + 2c_3 n^2 \leq 9c_3 n^2, \tag{3.50}$$

where we used Lemma 3.4.3. This proves the first part of the lemma.

To prove the next part, we look for a relabeling of $W_1', \ldots, W_9'$ such that $|\tilde{U}_i \setminus W_i'| < 1250 c_3 n$ for each $i \in [9]$. If no such relabeling exists, we have some $i \in [9]$ such that $|\tilde{U}_i \setminus W_j'| \geq 1250 c_3 n$ for all $j \in [9]$. However, for some j, $|\tilde{U}_i \cap W_j'| \geq |\tilde{U}_i|/9$. Let $X = \tilde{U}_i \cap W_j'$ and $Y = \tilde{U}_i \setminus W_j'$. Then, by Lemma 3.4.3, we have $e(G[X, Y]) < 6c_3 n^2$

while $X \subseteq W'_j$, $Y \cap W'_j = \emptyset$ and (3.50) imply that $e(G[X,Y]) > |X||Y| - 9c_3n^2 > 6c_3n^2$, a contradiction.

The desired estimate (3.49) follows from the observation that

$$W'_i \setminus \tilde{U}_i \subseteq \bigcup_{j\in[9]\setminus\{i\}} (\tilde{U}_j \setminus W'_j).$$

The last two claims of the lemma follow by taking $W'_i = W_i$. $\square$

A *pattern* is an assignment $\pi : \binom{[9]}{2} \to \binom{[4]}{3}$ (to every edge of K_9 we assign a list of 3 colors) such that $\pi^{-1}(i)$ is isomorphic to $T_3(9)$ for each $i \in [4]$. It is easy to check that up to isomorphism (of colors and vertices) there is only one pattern. It can be explicitly described as follows. Identify the 9-point vertex set with $(\mathbb{F}_3)^2$, the 2-dimensional vector space over the 3-element finite field $\mathbb{F}_3$. There are 4 possible directions of 1-dimensional subspaces. Let the color $i \in [4]$ be present in the pattern in those pairs whose difference is not parallel to the i-th direction.

We say that an edge 4-coloring χ of $G \in \mathcal{F}_n$ *follows the pattern* π if for every $ij \in \binom{[9]}{2}$ we have

$$\left|\, \chi^{-1}([4] \setminus \pi(ij)) \cap G[W_i, W_j] \,\right| \le c_2n^2.$$

Lemma 3.4.5. *Let $n \ge n_0$ and $G \in \mathcal{F}_n$. Then every coloring $\chi \in \mathcal{S}(G)$ follows a pattern.*

Proof. Let $\chi \in \mathcal{S}(G)$ and (m, V_i, H_i) be the associated popular vector. Let $[m] = U_1 \cup \ldots \cup U_9$, be the partition given by Lemma 3.4.2.

Let the *label* of an edge $uv \in R_3$ be $\hat{\chi}(uv) = \{i \in [4] : uv \in E(H_i)\}$. So, $|\hat{\chi}(uv)| = 3$ for all edges $uv \in R_3$.

Claim 3.4.6. *Let $Y = \{v_1, \ldots, v_9\}$ be a subset of $[m]$ such that $R_3[Y] \cong K_9$ and $v_i \in U_i$ for each $i \in [9]$. Let $v'_j \in U_j$ be such that $Y' = (Y \setminus \{v_j\}) \cup \{v'_j\}$ also spans K_9 in R_3. Then $\hat{\chi}(v_jv_i) = \hat{\chi}(v'_jv_i)$ for all $i \in [9] \setminus \{j\}$.*

Proof of Claim. The identity $3 \cdot \binom{9}{2} = 4 \cdot t_3(9)$ and Turán's theorem imply that each K_4-free graph $H_i[Y]$ has exactly $t_3(9)$ vertices and thus is isomorphic to the Turán graph $T_3(9)$. Let $Y_{i,1}$, $Y_{i,2}$, and $Y_{i,3}$ be the parts of $H_i[Y]$. The family of 3-sets $\{Y_{i,j} : i \in [4],\ j \in [3]\}$ forms a Steiner triple system on Y, that is, every pair is covered exactly once. Thus if we delete a vertex from Y, then the four triples that contain it are uniquely reconstructible. It follows that if we know $H_i[Y] - v_j$ for each $i \in [4]$, then the labels of the eight pairs containing v_j are uniquely determined. This and the analogous statement for Y' imply the claim. ▮

We can iteratively build a set $Y = \{v_1, \ldots, v_9\}$ such that $R_3[Y] \cong K_9$ and for all $i \in [9]$ we have $v_i \in U_i$ and

$$|N_{R_3}(v_i) \cap U_j| > |U_j| - \sqrt{c_3}m \quad \text{for all } j \in [9] \setminus \{i\}. \tag{3.51}$$

Let $A_i \subseteq U_i$ consist of those vertices that lie in $N_{R_3}(v_j)$ for all $j \in [9] \setminus \{i\}$. As all $v_1, \ldots, v_9$ satisfy (3.51), we have $|A_i| > |U_i| - 8\sqrt{c_3}m$. Now, if $a_i a_j \in R_3[A_i, A_j]$ (without loss of generality assume that $(i,j) = (1,2)$), then all three sets $\{v_1, v_2, \ldots, v_9\}$, $\{a_1, v_2, \ldots, v_9\}$, and $\{a_1, a_2, v_3, \ldots, v_9\}$ form 9-cliques. By Claim 3.4.6 we have $\hat{\chi}(v_i v_j) = \hat{\chi}(a_i v_j) = \hat{\chi}(a_i a_j)$. Therefore the labeling on Y determines the labeling on all edges of R_3 with the possible exception of at most $72\sqrt{c_3}m^2$ edges incident to vertices of $\bigcup_{i=1}^{9}(U_i \setminus A_i)$. We, therefore, have a pattern π such that $\hat{\chi}(u_i u_j) = \pi(ij)$ for all but at most $73\sqrt{c_3}m^2$ edges in R.

By applying (3.49) to $W_i' = W_i$ and arguing as in the proof of Lemma 3.3.9, one can show that χ follows the pattern π. $\square$

A coloring $\chi \in \mathcal{S}(G)$ is called *good* if for every distinct $i, j, k \in [9]$, all sets $X_i \subseteq W_i, X_j \subseteq W_j, X_k \subseteq W_k$ each of size at least $c_1 n$, and a color $c \in \pi(ij) \cap \pi(ik) \cap \pi(jk)$, we can find a monochromatic triangle in color c with one vertex in each of X_i, X_j, X_k. Otherwise, call χ *bad*.

We make use of the following result [ABKS04, Lemma 3.1] that is proved by the standard embedding argument, see e.g. [SS91, Theorem 5].

Lemma 3.4.7. *Let G be a graph and let $V_1, \ldots, V_k$ be subsets of vertices of G such that, for every $i \neq j$ and every pair of subsets $X_i \subseteq V_i$ and $X_j \subseteq V_j$ with $|X_i| \geq 10^{-k}|V_i|$ and $|X_j| \geq 10^{-k}|V_j|$, there are at least $\frac{1}{10}|X_i||X_j|$ edges between X_i and X_j in G. Then G contains a copy of K_k with one vertex in each set V_i.* $\square$

As a consequence of this lemma, a coloring fails to be good only if there are c, i, j such that $c \in \pi(ij)$ but for some sets $X_i \subseteq W_i$ and $X_j \subseteq W_j$ with $|X_i|, |X_j| \geq c_1 n/1000$, $\chi^{-1}(c)$ has at most $|X_i||X_j|/10$ edges between X_i and X_j. The proof of Lemma 3.3.11 with obvious modifications gives the following.

Lemma 3.4.8. *The number of bad colorings is at most $3^{4n^2/9} \cdot 2^{-c_1^2 n^2/10^7}$.* $\square$

A good coloring χ of G is *perfect* if $\chi(v_i v_j) \in \pi(ij)$ for every pair $ij \in \binom{[9]}{2}$ and every edge $v_i v_j \in G[W_i, W_j]$. Let $\mathcal{P}(G)$ consist of all perfect colorings of G.

Lemma 3.4.9. *Let $G \in \mathcal{F}_n$ be a graph of order $n \geq n_0 + 2$ such that $F(G, 4, 4) \geq$*

$3^{4n^2/9} \cdot 2^{-c_9 n^2}$ *and for every distinct* $v, v' \in V(G)$ *we have*

$$\frac{F(G,4,4)}{F(G-v,4,4)} \geq (3-c_3)^{8n/9}, \tag{3.52}$$
$$\frac{F(G,4,4)}{F(G-v-v',4,4)} \geq (3-c_3)^{(8/9)(n+(n-1))}. \tag{3.53}$$

Then the following conclusions hold.

1. G *is* 9*-partite.*
2. $|\mathcal{P}(G)| \geq (1-2^{-c_9 n})\, F(G,4,4)$.
3. *If* $G \not\cong T_9(n)$*, then there is a graph* G' *with* $v(G') = n$ *and* $F(G',4,4) > F(G,4,4)$.

Proof. As in the proof of Lemma 3.3.12, the notion of a good coloring is well-defined for $G - X$ provided $|X| \leq 2$.

Claim 3.4.10. *For each* $i \in [9]$ *and every* $v \in W_i$, $|N(v) \cap W_i| < 8c_1 n$.

Proof of Claim. Suppose that a vertex v violates the claim. Let $W_1' \cup \dots \cup W_9'$ be the selected max-cut partition of $G - v$. Similarly to Claim 3.3.13 there is a good coloring χ of $G - v$ with at least $(3-c_2)^{8n/9}$ extensions to G. Let π be the pattern of χ (with respect to $W_1', \dots . W_9'$) and $n_i = |N(v) \cap W_i|$ for $i \in [9]$. As in the proof of Lemma 3.3.12, we take an extension $\bar{\chi}$ of χ that gives a most frequent vector $\mathbf{x} = (x_1, \dots, x_9)$, where x_i is the number of colors c such that $Z_{i,c} = \{u \in W_i : \bar{\chi}(uv) = c\}$ has at least $2c_1 n$ elements. Also, let y_c be the number of $j \in [9]$ such that $|Z_{j,c}| \geq 2c_1 n$. We have

$$x_1 + x_2 + \ldots + x_9 = y_1 + y_2 + y_3 + y_4. \tag{3.54}$$

By the max-cut property, each $x_i \geq 1$. The argument of (3.37) shows that the number of extensions of χ to G is at most $2^{c_0 n} \prod_{i=1}^{9} x_i^{n_i}$.

Suppose that $y_c \geq 7$ for some color c. Any 7 vertices of the color-c graph that is isomorphic to $T_3(9)$ span a triangle. The three c-neighborhoods of v in the corresponding parts W_i' have at least $|Z_{i,c}| - 24000c_3 n > c_1 n$ vertices each by (3.49). Since χ is good, this gives a copy of K_4 of color c in $\bar{\chi}$, a contradiction.

Thus $y_c \leq 6$ for every $c \in [4]$ and the sum of x_i's is at most 24. Since each x_i is a positive integer, their product is at most $2^3 3^6$ (it is clearly maximized when the factors are nearly equal). Also, each $n_i \leq n/9 + c_2 n$ by Lemma 3.4.4. Thus the number of extensions of χ is at most $2^{2c_0 n}(2^3 3^6)^{n/9} < (3-c_2)^{8n/9}$, a contradiction that proves the claim. ∎

Claim 3.4.11. *If* $x_1, \dots, x_8$ *are non-negative integers with sum* 24 *then* $\prod_{i=1}^{8} \max(x_i, 1) \leq 3^8$ *with equality if and only if each* x_i *equals* 3.

Proof of Claim. Clearly, the product of two positive integers k and l given their sum is maximum when $|k-l| \leq 1$. Thus if we fix t, the number of non-zero x_i's, then their product is maximized when each positive x_i is $\lfloor 24/t \rfloor$ or $\lceil 24/t \rceil$. Thus, for $t = 8, 7, \ldots, 1$ the maximum of the product is respectively $3^8 = 6561$, $3^4 \cdot 4^3 = 5184$, $4^6 = 4096$, $4 \cdot 5^4 = 2500$, $6^4 = 1296$, $8^3 = 512$, $12^2 = 144$, and 24. Here, 3^8 is the largest entry. ∎

Claim 3.4.12. *For all $i \in [9]$ and all $v, v' \in W_i$, we have $vv' \notin E(G)$.*

Proof of Claim. Assume for a contradiction that $vv' \in E(G)$, where without loss of generality $v, v' \in W_9$. As in Claim 3.3.13, one can find a good coloring χ of $G - v - v' \in \mathcal{F}_{n-2}$ with at least $(3-c_2)^{16n/9}$ extensions to G. Define the parameters $\pi, n_i, Z_{i,c}, x_i, y_i, n_i', Z_{i,c}', x_i', y_i', \bar{\chi}$ and a most frequent color s of vv', as it was done in Claim 3.3.15. Then a version of (3.38) states that the total number of extensions of χ is at most

$$(5^9)^2 \cdot 4 \cdot 2^{c_0 n} \cdot (4^{8c_1 n + 8c_2 n})^2 \cdot \prod_{i=1}^{8} (\max(x_i, 1) \cdot \max(x_i', 1))^{n/9}. \tag{3.55}$$

Since each $y_c \leq 6$, we have $\sum_{i=1}^{8} x_i \leq 24$. By Claim 3.4.11 we have that $x_i = x_i' = 3$ for each $i \in [8]$, for otherwise the bound in (3.55) is strictly less than $(3-c_2)^{16n/9}$, a contradiction to the choice of χ.

Assume that the parts of $H_s \cong T_3(9)$ are $A_1 = \{1, 2, 3\}$, $A_2 = \{4, 5, 6\}$, and $A_3 = \{7, 8, 9\}$.

Suppose first that there is $j \in [8]$ such that $|Z_{j,s}| \geq 2c_1 n$ but $s \notin \pi(\{j, 9\})$, say $j = 8$. By (3.49), we have $|Z_{8,s} \cap W_8'| \geq c_1 n$. Since χ is good, in order to avoid a color-c K_4 in $\bar{\chi}$ we must have $|Z_{i,s}| < 2c_1 n$ for all $i \in A_1$ or for all $i \in A_2$. Thus y_s contributes at most 5 to $\sum_{i=1}^{8} x_i$ and (since any other y_t is at most 6) this sum is at most 23, giving a contradiction by Claim 3.4.11 and (3.55).

Also, this implies that $|Z_{j,s}| \geq 2c_1 n$ for all $j \in [6]$ (for otherwise $y_s \leq 5$). The same claim applies to $|Z_{j,s}'|$. Let $y_1 z_1, \ldots, y_m z_m$ be a maximal matching formed by color-s edges between W_1 and W_4. Since χ is good, we have that

$$m \geq \min(|W_1|, |W_4|) - c_1 n \geq n/9 - 2c_1 n.$$

When we extend the coloring χ to G, the number of choices to color the edges of $G[vv', y_i z_i]$ is at most $3^4 - 1$ for every $i \in [m]$ because, if all 4 pairs are present in G, then we are not allowed to color all of them with color c while otherwise we have at most $4^3 < 3^4$

choices. This allows us to improve the bound in (3.55) by factor $(80/81)^{n/10}$, giving the desired contradiction. ∎

Thus we have proved Part 1 of the lemma.

Suppose on the contrary that the conclusion of Part 2 does not hold. As in the proof of Lemma 3.3.8, we can find an edge $vv' \in G$, say with $v \in W_1$ and $v' \in W_9$, a color s, and a good coloring χ of $G - v - v'$ such that there are at least $(3 - c_2)^{16n/9}$ good extensions of χ to G that preserve the pattern π of χ and assign the "wrong" color s to vv'. Defining $x_i, x'_i, Z_{j,c}, Z'_{j,c}, y_i, y'_i$ by the direct analogy with the definitions of Claim 3.3.15, one can argue similarly to (3.38) that the total number of extensions of χ is at most

$$2^{c_0 n} \cdot \left(\prod_{i=2}^{9} \max(x_i, 1) \cdot \prod_{i=1}^{8} \max(x'_i, 1) \right)^{n/9}. \tag{3.56}$$

By Claim 3.4.11, we have $x_i = 3$ for each $2 \le i \le 9$ and $x'_i = 3$ for each $i \in [8]$. Thus each y_i and each y'_i is equal to 6. It follows that, for any $2 \le j \le 9$ and $c \in [4]$, we have $|Z_{j,c}| \ge 2c_1 n$ if and only if $c \in \pi(\{1, j\})$. Also, the analogous claim holds for $|Z'_{j,c}|$. Since $s \notin \pi(\{1, 9\})$, we can find distinct $i, j \in \{2, \ldots, 8\}$ such that s belongs to $\pi(ij)$ as well as to the label of each pair in $\{1, 9\} \times \{i, j\}$. As before, by considering a maximal color-s matching in $G[W_i, W_j]$, we can improve (3.56) by a factor $(80/81)^{n/10}$, getting a contradiction and proving Part 2 of the lemma.

Let us prove Part 3. If G is not complete 9-partite, then by Part 2 we can take $G' = K(W_1, \ldots, W_9)$: indeed, $|\mathcal{P}(G')| \ge 3|\mathcal{P}(G)| > F(G, 4, 4)$. So suppose that G is complete 9-partite.

Let us determine the number of possible patterns (with distinguishable colors and vertices). For the color-1 graph we have $\binom{8}{2} \cdot \binom{5}{2}$ choices (there are $\binom{8}{2}$ choices for the part $A_1 \in \binom{[9]}{3}$ containing 1, then $\binom{5}{2}$ choices for the part A_2 containing the smallest element of $[9] \setminus A_1$.) Then we have $9 \cdot 4$ choices for color 2, then 2 choices for color 3, and one choice for color 4. Thus the total number of patterns is $20160 = 9!/18$. The same answer can be obtained by noting that, when we permute $[9]$, then we have a transitive action on patterns and every pattern is fixed by 18 permutations.

It follows that G has $(9!/18 + o(1))\, 3^{e(G)}$ perfect colorings in total, since every edge of G has exactly 3 choices for a given pattern. Since $G \not\cong T_9(n)$, we have $|\mathcal{P}(T_9(n))| \ge (3+o(1))|\mathcal{P}(G)|$ and we can take $G' = T_9(n)$. This completes the proof of Lemma 3.4.9. □

Now, Theorem 3.1.2 can be deduced from Lemma 3.4.9 in the same way (modulo some obvious modifications) as Theorem 3.1.1 was deduced from Lemma 3.3.12.

3.5 Remarks

In the preceding sections we extended the results obtained by Alon et al. [ABKS04] on $F(n,2,k)$ and $F(n,3,k)$ to provide the extremal graphs for $F(n,4,3)$ and $F(n,4,4)$. As the general problem of evaluating $F(n,r,k)$ remains difficult, one may wonder for which values of r and k it seems to be tractable. A natural extension may be fixing $r=4$ and evaluating $F(n,4,k)$, possibly starting with $F(n,4,5)$. On the other hand, one may study monochromatic-triangle-free colorings by fixing $k=3$ and looking at $F(n,r,3)$ in general and $F(n,5,3)$, in particular. Another interesting question happens to be that of $F(n,q+1,q+1)$ where q is a prime or a prime power where, as seen in the extremal structure for $F(n,4,4)$, the affine plane $\mathbb{F}_q^2$ suggests that $T_{q^2}(n)$ may be the extremal graph. We would like to say a few words on each.

In tackling $F(n,r,k)$, our method, as in [ABKS04], will consist of applying the regularity lemma to a given graph G and a coloring χ to obtain cluster graphs $H_1,\dots,H_r$ on vertex set $[m]$. The cluster graphs allow us to construct the graphs R_s, where $ab\in R_s$ if and only if $ab\in H_i$ for exactly s cluster graphs H_i. We also have the quantities e_s and r_s where e_s is the number of edges in R_s and $r_s=2e_s/m^2$.

Now, loosely speaking, the number of colorings χ which give rise to a particular vector (m,V_i,H_i,R_i,e_i,p_i) is on the order of $\left(\prod_{s=1}^r s^{r_s}\right)^{n^2/2}$. So we try to optimize this function, or rather

$$\sum_{s=1}^{r} r_s\log s, \tag{3.57}$$

under some constraints on the quantities r_s. One such quantity is obtained by noting that each graph H_i is K_k-free. Hence,

$$\sum_{i=1}^{r} i\cdot r_i \le r\frac{k-2}{k-1}. \tag{3.58}$$

We may also obtain inequalities of the form

$$\sum_{i=s}^{r} r_i \le \frac{t-2}{t-1} \tag{3.59}$$

by observing that $K_t\not\subset R_s\subset\bigcup_{i\ge s}R_i$. For example, $r_3\le 1/2$ was used in Section 3.3 and

$r_3 \leq 8/9$ may have been utilized in computing this value for $F(n,4,4)$ (in that case, (3.58) was a dominating inequality).

3.5.1 $F(n,4,k)$

For $r = 4$ and $k \geq 5$, the optimal value for (3.57) under constraint (3.58), is obtained when $r_3 = 1$ and $r_2 = \frac{k-2}{2(k-1)}$. Unfortunately, this value is unattainable and, in order to obtain tighter bounds, we will have to find inequalities resembling (3.59). Clearly, $r_4 \leq \frac{k-2}{k-1}$ is dominated by (3.58) and does not play a role, so the key question becomes finding a bound for r_3.

However, this becomes a Ramsey-type problem in which edges receive multiple colors. To state it more generally, let

$$N(r,s,k) = \min\{n \mid \forall \chi : E(K_n) \to \binom{[r]}{s} \text{ there exists } i \in [r] \text{ with } \chi^{-1}(i) \supset K_k\}.$$

Then, the Ramsey number $R(k,k)$ is simply $N(2,1,k)$, and, as seen earlier, $N(4,3,4) = 10$. In this particular case, we are interested in finding $N(4,3,k)$ where $k \geq 5$. It is not at all clear what this value should be even for $k = 5$, although it is bounded above by $R(5,5)$.

Assuming we know the value of $N = N(4,3,k)$, we may start exploring an extremal structure for $F(n,4,k)$. One obvious choice is $T_{N-1}(n)$. This graph would be optimal unless $r_2 > 0$, in which case one would suspect that each part of $T_{N-1}(n)$ will be partitioned further to furnish $T_{c(N-1)}(n)$.

3.5.2 $F(n,r,3)$

This case also requires the same Ramsey-type constraints and does not seem easier to handle for general r. So, we restrict ourselves to the case when $r = 5$. It is simple to obtain the inequalities $r_4, r_5 \leq 1/2$ by showing that $N(5,5,3) = N(5,4,3) = 3$. However, we will show below that $N(5,3,3) = 5$, giving us the inequality $r_3 \leq 3/4$. With the inclusion of this inequality, the optimal point for (3.57) is $r_3 = 3/4, r_2 = 1/8$ giving us the bound $F(n,5,3) \leq (2^{1/8}3^{3/4})^{\binom{n}{2}+o(n^2)}$. This would suggest $T_8(n)$ as the extremal graph, but unfortunately, no labeling $\pi : E(T_8(n)) \to \binom{[5]}{3} \cup \binom{[5]}{2}$ satisfies the above.

Claim 3.5.1. *$K_5 \not\subseteq R_3$.*

Proof of Claim. Assume $W = \{w_1, \ldots, w_5\}$ form a K_5 in R_3. As a triangle-free graph on 5 vertices may have at most 6 edges, each color is used on at most 6 edges. In addition, since there are $\binom{5}{2} = 10$ edges and each edge receives 3 colors, each color must appear exactly 6 times, i.e., each $H_i[W] \cong K_{2,3}$. For each i, let $A_i \cup B_i = W$ be the bipartition, where $|A_i| = 2$. So, an edge $w_j w_k$ receives color i iff exactly one of w_j, w_k lies in A_i. As each edge receives 3 colors, without loss of generality, w_j belongs to A_i, for some i, more than w_k does. They appear together the rest of the time. However, this implies that the sum of degrees of w_j is greater than that of w_k, whereas the sum of degrees at each vertex needs to be exactly $4 \cdot 3 = 12$. ∎

3.5.3 $F(n, q+1, q+1)$

In this case, where q is prime or a prime power, we have the inequality $p_q \leq \frac{q^2-1}{q^2}$ as $N(q+1, q, q+1) = q^2 + 1$. We may construct an edge k-coloring χ of the complete graph on q^2 vertices by using the affine plane $\mathbb{F}_q^2$ as follows.

Let the vertices of K_{q_2} be of the form $v = (\alpha, \beta)$ where $\alpha, \beta \in \mathbb{F}_q$. For each $\rho, \sigma \in \mathbb{F}_q$, we have the linear subsets $L(\rho, \sigma) = \{(\alpha, \beta) : \beta = \rho\alpha + \sigma\}$. Then, $L(\rho) = \{L(\rho, \sigma) : \sigma \in \mathbb{F}_q\}$ partitions $V(K_{q^2})$ into q sets of size q. Define some mapping $\theta : [r] \to \mathbb{F}_q$ and for each $i \in [r]$, let $i \in \chi(uv)$ if and only if $\{u, v\} \notin L(\theta(i), \sigma)$ for all $\sigma \in \mathbb{F}_q$.

By blowing up this coloring, we obtain $F(n, q+1, q+1) \geq q^{\frac{q^2-1}{q^2}\frac{n^2}{2}}$, which we believe is tight up to a constant factor. However, the optimal point for the linear optimization (3.57) gives an upper bound that does not match the constructive lower bound.

Chapter 4

Counting Color-Critical Graphs

4.1 Supersaturation

In 1907, Mantel [Man07] proved that a triangle-free graph on n vertices contains at most $\lfloor n^2/4 \rfloor$ edges. The celebrated result of Turán generalizes this result as follows: a graph containing no copy of K_{r+1} has at most $t_r(n)$ edges, where $t_r(n)$ is the number of edges in the Turán graph $T_r(n)$, the complete r-partite graph with parts of size $\lfloor n/r \rfloor$ or $\lceil n/r \rceil$.

Stated in the contrapositive, this implies that a graph with $t_r(n) + 1$ edges contains at least one copy of K_{r+1}. A result of Rademacher (1941, unpublished), which is perhaps the first result in the so-called "theory of supersaturated graphs," shows that a graph on $\lfloor n^2/4 \rfloor + 1$ edges contains not just one but at least $\lfloor n/2 \rfloor$ copies of a triangle. Subsequently, Erdős [Erd62a, Erd62b] proved that there exists some small constant $c_r > 0$ such that $t_r(n) + q$ edges (where $0 < q < c_r n$) guarantee at least as many copies of K_{r+1} as contained in the graph obtained from $T_r(n)$ by adding q edges to one of its maximal classes. Lovász and Simonovits [LS75, LS83] later proved that the same statement holds with $c_r = 1/r$. In fact, in [LS83], they determine the extremal graphs for $q = o(n^2)$. Erdős [Erd69] also considered the supersaturation problem for odd cycles C_{2k+1} and proved that a graph on $2n$ vertices and $t_2(2n) + 1$ edges contains at least $n(n-1) \cdots (n-k+1)(n-2) \cdots (n-k)$ copies of C_{2k+1}.

We study the supersaturation problem for the broader class of color-critical graphs, that is, graphs containing an edge whose deletion reduces the chromatic number.

Definition 4.1.1. *Call a graph F r-critical if $\chi(F) = r + 1$ but it contains an edge e such that $\chi(F - e) = r$.*

Simonovits [Sim68] proved that for an r-critical graph F, the Turán number $\mathrm{ex}(n,F)$ coincides with that of K_{r+1} for n large enough. That is, a graph containing no copy of F has at most $t_r(n)$ edges. Mubayi [Mub10] proved the following supersaturation result for color-critical graphs:

Definition 4.1.2. *Fix $r \geq 2$ and let F be an r-critical graph. Let $c(n,F)$ be the minimum number of copies of F in the graph obtained from $T_r(n)$ by adding one edge.*

Here, a copy of F in H is a subset of $f = |V(F)|$ vertices and $|F| = |E(F)|$ edges of H such that the subgraph formed by this set of vertices and edges is isomorphic to F.

Theorem 4.1.3 (Mubayi [Mub10])**.** *For an r-critical graph F, there exists a constant $c_0 = c_0(F) > 0$ such that for all sufficiently large n and $1 \leq q < c_0 n$, every n vertex graph with $t_r(n) + q$ edges contains at least $qc(n,F)$ copies of F.*

Mubayi proves that Theorem 4.1.3 is sharp for some graphs including odd cycles C_{2k+1} and $K_4 - e$, the graph obtained from K_4 by deleting an edge. However, the value of c_0 obtained in the proof is small.

In this chapter, we extend these results in various ways. Our main result involves the characterization of extremal graphs for certain values of q. Let $\mathcal{H}_F(n,q)$ be the set of graphs on n vertices and $t_r(n) + q$ edges which contain the fewest number of copies of F. Let $\mathcal{T}_r^q(n)$ be the set of graphs obtained from the Turán graph $T_r(n)$ by adding q edges. Then, with

$$\begin{aligned} q_F(n) &= \max\{q : \mathcal{H}_F(n,q') \subseteq \mathcal{T}_r^{q'}(n) \text{ for all } q' \leq q\}, \\ c_1(F) &= \liminf_{n\to\infty} \left\{\frac{q_F(n)}{n}\right\} \end{aligned}$$

we show

Theorem 4.1.4. *For every connected r-critical graph F, $c_1 = c_1(F) > 0$.*

In fact, our proof proceeds by showing a lower bound on $c_1(F)$. Roughly speaking, we prove that $\mathcal{T}_r^q(n)$ fails to be optimal if one of two phenomena occur: the number of copies of F may be decreased either by using a non-equitable partition of the vertex set or by rearranging the neighborhood of some vertex of large degree. Interestingly, in the first scenario, the congruence class of n modulo r affects the value of $q_F(n)$. Hence, we also define the following r constants

$$c_{1,i}(F) = \liminf_{\substack{n\to\infty \\ n\equiv i \bmod r}} \left\{\frac{q_F(n)}{n}\right\}.$$

For the special case where F is an odd cycle, we obtain the exact result $c_{1,1}(C_{2k+1}) = 1$, whereas $c_1(C_{2k+1}) = c_{1,0}(C_{2k+1}) = 1/2$ (see Lemma 4.4.4 and Theorem 4.5.2). We also compute the exact value of $c_1(F)$ for graphs obtained by deleting an edge from K_{r+2} where $r \geq 2$ (Theorem 4.5.6).

We also examine a threshold for the asymptotic optimality of $\mathcal{T}_r^q(n)$. Formally, with $\#F(H)$ being the number of copies of F in H,

$$c_2(F) = \sup\left\{c : \forall \epsilon > 0\ \exists n_0 \text{ s.t.} \forall n \geq n_0, q \leq cn\,(H \in \mathcal{H}_F(n,q)) \Rightarrow \left(\frac{\#F(H)}{qc(n,F)} \geq 1-\epsilon\right)\right\}.$$

That is, if $q = cn$, $G_n \in \mathcal{T}_r^q(n)$ minimizes $F(G)$ over all $G \in \mathcal{T}_r^q(n)$, and $H_n \in \mathcal{H}_F(n,q)$, then $\lim_{n\to\infty} \#F(H_n)/F(G_n) = 1$ if $c \leq c_2(F)$. In Section 4.2, we introduce a graph parameter that allows one to calculate $c_2(F)$ exactly (see Theorem 4.3.1). As $c_2(F) \geq c_1(F)$, the above parameter provides an upper bound on $c_1(F)$.

While the focus of this chapter is on the supersaturation problem with $q = o(n^2)$, and in particular $q = O(n)$, it is worth mentioning that, for cliques, the case $q = \Omega(n^2)$ has been actively studied and proved notoriously difficult. Only recently, Razborov [Raz08], using techniques developed in [Raz07], proved asymptotically sharp bounds on the number of triangles in a graph with given edge density $\rho \geq 1/2$. The case for K_4 was answered by Nikiforov [Nik11], but the question remains open for larger cliques.

The rest of the chapter is organized as follows. In the next section we introduce the functions and parameters with which we work. In Section 4.3, we set up the general framework and prove results for $q = O(n^{2-\eta})$ for some $\eta > 0$. In particular, we use this setup to prove our bound on $c_2(F)$. Theorem 4.1.4 is proved in Section 4.4. We use the last section to prove an upper bound on $c_1(F)$ for a special class of graphs.

4.2 Parameters

We will first reproduce part of Mubayi's argument in order to define some recurring constants and structures. We will then extend his arguments to prove our results in subsequent sections.

In the arguments and definitions to follow, F will be an r-critical graph and we let $f = |V(F)|$ be the number of vertices of F. We write $x = y \pm z$ to mean $|x - y| \leq z$. We begin with an expression for $c(n, F)$.

Lemma 4.2.1. *Let F be an r-critical graph on f vertices. There is a positive constant α_F such that*

$$c(n, F) = \alpha_F n^{f-2} \pm O(n^{f-3}).$$

This is proved by Mubayi [Mub10] by providing an explicit formula for $c(n, F)$. If F is an r-critical graph, we call an edge e (resp., a vertex v) a *critical edge* (resp., a *critical vertex*) if $\chi(F - e) = r$ (resp., $\chi(F - v) = r$).

Given disjoint sets $V_1, \ldots, V_r$, let $K(V_1, \ldots, V_r)$ be formed by connecting all vertices $v_i \in V_i, v_j \in V_j$ with $i \neq j$, i.e., $K(V_1, \ldots, V_r)$ is the complete r-partite graph on vertex classes $V_1, \ldots, V_r$. Let H be obtained from $K(V_1, \ldots, V_r)$ by adding one edge xy in the first part and let $c(n_1, \ldots, n_r; F)$, where $n_i = |V_i|$, denote the number of copies of F contained in H. Let $uv \in F$ be a critical edge and let χ_{uv} be a proper r-coloring of $F - uv$ where $\chi_{uv}(u) = \chi_{uv}(v) = 1$. Let x^i_{uv} be the number of vertices of F excluding u, v that receive color i. An edge preserving injection of F into H is obtained by picking a critical edge uv of F, mapping it to xy, then mapping the remaining vertices of F to H such that no two adjacent vertices get mapped to the same part of H. Such a mapping corresponds to some coloring χ_{uv}. So, with $Aut(F)$ denoting the number of automorphisms of F, we obtain

$$c(n_1, \ldots, n_r; F) = \frac{1}{Aut(F)} \sum_{uv \text{ critical}} \sum_{\chi_{uv}} 2(n_1 - 2)_{x^1_{uv}} \prod_{i=2}^{r} (n_i)_{x^i_{uv}}.$$

We obtain a formula for $c(n, F)$ by picking $H \in \mathcal{T}^1_r(n)$. In particular, if $n_1 \leq n_2 \leq \cdots \leq n_r$ and $n_r - n_1 \leq 1$,

$$c(n, F) = \min\{c(n_1, \ldots, n_r; F), c(n_r, \ldots, n_1; F)\}. \tag{4.1}$$

If $r \,|\, n$, we get a polynomial expression in n of degree $f - 2$ and α_F is taken to be the leading coefficient.

A recurring argument in our proof involves the deletion of one edge and its replacement by another. Here, we define quantities that allow us to compare and bound the number of copies gained and lost by this swapping process.

Definition 4.2.2. *Let F be an r-critical graph. Given a graph $H = K_{n_1,\ldots,n_r} + uv$, denote by $m(H, F)$ the maximum number of copies of F to which an edge $uw \neq uv$ belongs.*

Definition 4.2.3. *Let $m(n, F) = \max\{m(H, F) : H \in \mathcal{T}^1_r(n)\}$.*

It is easy to see that $\lim_{n\to\infty} \frac{m(n,F)}{n^{f-3}}$ exists; we denote it by μ_F.

Another operation involves moving vertices or edges from one class to another, potentially changing the partition of n. To this end, we compare the values of $c(n_1, \ldots, n_r; F)$. In [Mub10], Mubayi proves that there exists some γ_F such that

$$c(n_1, \ldots, n_r; F) \geq c(n, F) - \gamma_F a n^{f-3}$$

for all partitions $n_1 + \ldots + n_r = n$ where $\lfloor n/r \rfloor - a \leq n_i \leq \lceil n/r \rceil + a$ for every $i \in [r]$. We need the following, more precise estimate:

Lemma 4.2.4. *There exists a constant ζ_F such that the following holds for all $\delta > 0$ and $n > n_0(\delta, F)$. Let $c(n, F) = c(n'_1, \ldots, n'_r; F)$ as in (4.1). Let $n_1 + \ldots + n_r = n$, $a_i = n_i - n'_i$ and $M = \max\{|a_i| \ : \ i \in [r]\}$. If $M < \delta n$, then*

$$c(n, F) - c(n_1, \ldots, n_r; F) = \zeta_F a_1 n^{f-3} \pm O(M^2 n^{f-4}).$$

Proof. We bound $c(n_1, \ldots, n_r; F)$ using the Taylor expansion about $(n'_1, \ldots, n'_r)$. We first note that $c(n_1, \ldots, n_r; F)$ is symmetric in the variables $n_2, \ldots, n_r$. Hence,

$$\frac{\partial c}{\partial n_i}(n/r, \ldots, n/r) = \frac{\partial c}{\partial n_j}(n/r, \ldots, n/r)$$

for all $2 \leq i, j \leq r$. Furthermore, as $|n'_i - n/r| \leq 1$ for all $1 \leq i \leq r$, there is some constant C_1 such that

$$\frac{\partial c}{\partial n_i}(n'_1, \ldots, n'_r) - \frac{\partial c}{\partial n_i}(n/r, \ldots, n/r) < C_1 n^{f-4}.$$

Then, we have the inequality

$$\begin{aligned} c(n_1, \ldots, n_r; F) - c(n'_1, \ldots, n'_r; F) &\leq \sum_{j=1}^{r} a_j \frac{\partial c}{\partial n_j}(\frac{n}{r}, \ldots, \frac{n}{r}) + C_2 M^2 n^{f-4} \\ &= a_1 \frac{\partial c}{\partial n_1}(\frac{n}{r}, \ldots, \frac{n}{r}) + \frac{\partial c}{\partial n_2}(\frac{n}{r}, \ldots, \frac{n}{r}) \sum_{j=2}^{r} a_i + C_2 M^2 n^{f-4}. \end{aligned}$$

Similarly, we have

$$c(n_1, \ldots, n_r; F) - c(n'_1, \ldots, n'_r; F) \geq a_1 \frac{\partial c}{\partial n_1}(\frac{n}{r}, \ldots, \frac{n}{r}) + \frac{\partial c}{\partial n_2}(\frac{n}{r}, \ldots, \frac{n}{r}) \sum_{j=2}^{r} a_i - C_3 M^2 n^{f-4}.$$

As $\sum_{i=1}^{r} a_i = 0$, the lemma follows with ζ_F being the coefficient of n^{f-3} in $\frac{\partial c}{\partial n_2}(n/r, \ldots, n/r) - \frac{\partial c}{\partial n_1}(n/r, \ldots, n/r)$. $\square$

Definition 4.2.5. *For an r-critical graph F, let* $\pi_F = \begin{cases} \frac{\alpha_F}{|\zeta_F|} & \text{if } \zeta_F \neq 0 \\ \infty & \zeta_F = 0 \end{cases}.$

To give a brief foretaste of the arguments to come, we compare the number of copies of a 2-critical graph F in some $H \in \mathcal{T}_2^q(n)$ and a graph H' with $K(V_1, V_2) \subseteq H'$ where $n = 2\ell$ is even, $|V_1| = \ell+1$, and $|V_2| = \ell-1$. While H contains q 'extra' edges, $(\ell+1)(\ell-1) = \ell^2-1$ implies that the number of 'extra' edges in H' is $q+1$. Ignoring, for now, the copies of F that use more than one 'extra' edge, we compare the quantities $qc(n, F) \approx q\alpha_F n^{f-2}$ and $(q+1)(\alpha_F n^{f-2} - \zeta_F n^{f-3})$. It becomes clear that the ratio α_F/ζ_F will play a significant role in bounding the value $c_1(F)$.

Another phenomenon of interest is the existence of a vertex with large degree. Let $\mathbf{d} = (d_1, \ldots, d_r)$ and let $F(n_1, \ldots, n_r; \mathbf{d})$ be the number of copies of F in the graph $H = K(V_1, \ldots, V_r) + z$ where $|V_i| = n_i$ and z has a neighborhood of size d_i in V_i. Let $F(n, \mathbf{d})$ correspond to the case when $n_1 + \ldots + n_r = n - 1$ are almost equal and $n_1 \geq \ldots \geq n_r$.

We have the following formula for $F(n_1, \ldots, n_r; \mathbf{d})$. An edge preserving injection from F to H is obtained by choosing a critical vertex u, mapping it to z, then mapping the remaining vertices of F to H so that neighbors of u get mapped to neighbors of z and no two adjacent vertices get mapped to the same part. Such a mapping is given by an r-coloring χ_u of $F - u$ and the number of mappings associated with χ_u is the number of ways the vertices colored i can get mapped to the ith part of H (with neighbors of u colored i being mapped to neighbors of z in the ith part). This gives

$$F(n_1, \ldots, n_r; \mathbf{d}) = \frac{1}{Aut(F)} \sum_{u \text{ critical}} \sum_{\chi_u} \prod_{i=1}^{r} (n_i - y_i)_{x_i} (d_i)_{y_i}$$

where y_i is the number of neighbors of u that receive color i and x_i is the number of non-neighbors that receive color i. In particular, when $H \supseteq T_r(n-1)$ and $r|(n-1)$, we have

$$F(n, \mathbf{d}) = \frac{1}{Aut(F)} \sum_{u \text{ critical}} \sum_{\chi_u} \prod_{i=1}^{r} \left(\frac{n-1}{r} - y_i \right)_{x_i} (d_i)_{y_i}.$$

We find it convenient to write $d_i = \xi_i n$ and work instead with the following polynomial. For $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_r) \in \mathbb{R}^r$, let

$$P_F(\boldsymbol{\xi}) = \frac{1}{Aut(F)} \sum_{u \text{ critical}} \sum_{\chi_u} \prod_{i=1}^{r} \frac{1}{r^{x_i}} \xi_i^{y_i}.$$

As a first exercise, let us characterize all connected graphs for which $\deg(P_F) = r$ (we will later treat such graphs separately).

Lemma 4.2.6. *If F is a connected r-critical graph and* $\deg(P_F) = r$*, then $F = K_{r+1}$ or $r = 2$ and $F = C_{2k+1}$ is an odd cycle.*

Proof. The degree of P_F is determined by the critical vertex with the largest degree. To be precise, as $\deg(P_F) = \max_{u \text{ critical}} \{\deg(u) = \sum_i y_i(u)\}$, it follows that $\deg(u) \leq r$ for each critical vertex $u \in F$. However, any r-coloring χ_u of $F - u$ must assign all r colors to the neighbors of u. Thus, $\deg(u) = r$ and $y_i = 1$ for each critical vertex u and all $i \in [r]$. Therefore, every edge incident to u is a critical edge and, by extension, every neighbor of u is a critical vertex. As F is connected, it follows that every vertex is critical and has degree r. By Brooks' Theorem [Bro41], F is either the complete graph K_{r+1} or $r = 2$ and F is an odd cycle C_{2k+1} for some $k \geq 1$. $\square$

The case $F = K_{r+1}$ has been solved in a stronger sense by Lovász and Simonovits [LS83], whose results imply that $c_{1,1}(K_{r+1}) = 2/r$ while $c_{1,t}(K_{r+1}) = 1/r$ for $t \not\equiv 1 \pmod r$.

We now discuss some properties of the polynomial $P_F(\boldsymbol{\xi})$.

Lemma 4.2.7. *$P_F(\boldsymbol{\xi})$ is a symmetric polynomial with nonnegative coefficients.*

Proof. Let σ be a permutation of $[r]$. For a critical vertex u and an r-coloring χ_u of $F - u$, let $\sigma(\chi_u)$ be the coloring obtained by permuting the color classes of χ_u. It follows that $\sigma(\chi_u)$ is an r-coloring of $F - u$. Then,

$$\begin{aligned} P_F(\xi_{\sigma^{-1}(1)}, \ldots, \xi_{\sigma^{-1}(r)}) &= \frac{1}{Aut(F)} \sum_{u \text{ critical}} \sum_{\chi_u} \prod_{i=1}^{r} \frac{1}{r^{x_i}} (\xi_{\sigma^{-1}(i)})^{y_i} \\ &= \frac{1}{Aut(F)} \sum_{u \text{ critical}} \sum_{\sigma(\chi_u)} \prod_{i=1}^{r} \frac{1}{r^{x_{\sigma(i)}}} \xi_i^{y_{\sigma(i)}} \\ &= P_F(\boldsymbol{\xi}). \end{aligned}$$

In addition, as $x_i \geq 0$ and $y_i \geq 1$, each term in the product has a positive coefficient. $\square$

We now restrict the domain of P_F to those $\boldsymbol{\xi}$ which may arise as the density vector of some vertex. As $\xi_i = d_i/n$, it follows that $\xi_i \geq 0$ for all $i \in [r]$. Furthermore, $\sum_i d_i \leq n - 1$ implies that $\sum_i \xi_i \leq 1$. However, as we mostly encounter equitable partitions, we use the more restrictive set

$$\mathcal{S} = \{\boldsymbol{\xi} \in \mathbb{R}^r : 0 \leq \xi_i \leq 1/r \ \forall i \in [r]\}.$$

Lemma 4.2.8. *There exists a constant C_F such that for any $\boldsymbol{\xi}$ and $\boldsymbol{\xi}'$ with $\boldsymbol{\xi}, \boldsymbol{\xi}' \geq 0$ and* $\max(\|\boldsymbol{\xi}\|_1, \|\boldsymbol{\xi}'\|_1) \leq 1$*, $|P_F(\boldsymbol{\xi}) - P_F(\boldsymbol{\xi}')| < C_F \|\boldsymbol{\xi} - \boldsymbol{\xi}'\|_1$.*

Proof. Take C_F to be the maximum gradient of P_F on the set $\{\boldsymbol{\xi} \geq 0 \ : \ \|\boldsymbol{\xi}\|_1 \leq 1\}$. $\square$

Most of the arguments that follow involve minimizing P_F, usually over some subset of $\mathcal{S}$. One such subset is $\mathcal{S}_\rho = \{\boldsymbol{\xi} \in \mathcal{S} : \sum_i \xi_i = \rho\}$ where $\rho \in [0,1]$. Let $p(\rho) = \min\{P_F(\boldsymbol{\xi}) : \boldsymbol{\xi} \in \mathcal{S}_\rho\}$ and let $\mathcal{D}_\rho$ be the set of minimizers of $P_F(\boldsymbol{\xi})$ over the elements of $\mathcal{S}_\rho$. Let $\mathcal{D} = \bigcup_{\rho \in [0,1]} \mathcal{D}_\rho$. In many instances, the minimizers belong to the set

$$\mathcal{S}^* = \{\boldsymbol{\xi} \in \mathcal{S} : \exists j \in [r] \text{ with } \xi_j = 0 \text{ or } \sum_{i \neq j} \xi_i = 1 - 1/r\},$$

those density vectors which may be obtained from the vertices of (a subgraph of) some graph $H \in \mathcal{T}_r^q(n)$. This leads us to our next definition:

Definition 4.2.9. *Let* $\rho_F^* = \inf\{\rho \in [0,1] : \mathcal{D}_\rho \not\subseteq \mathcal{S}^*\}$*. If* $\mathcal{D}_\rho \subseteq \mathcal{S}^*$ *for all* $\rho \in [0,1]$*, then* $\rho_F^* = \infty$*.*

Lemma 4.2.10. $\rho_F^* > 1 - 1/r$*.*

Proof. First observe that $P_F(\boldsymbol{\xi}) > 0$ unless $\xi_i = 0$ for some $i \in [r]$. Thus, $\mathcal{D}_\rho \subseteq \mathcal{S}^*$ for $\rho \in [0, 1 - 1/r]$ and $\rho_F^* \geq 1 - 1/r$.

We now show that there exists some $\epsilon > 0$ such that for all $\rho = 1 - 1/r + \epsilon'$ with $\epsilon' < \epsilon$ and all $\boldsymbol{\xi} \in \mathcal{S}_\rho$, $p(\rho) = P_F(\boldsymbol{\xi})$ if and only if $\xi \in \mathcal{S}^*$.

Let $\boldsymbol{\xi} \in \mathcal{S}_\rho$ where $\rho = 1 - 1/r + \epsilon'$. As P_F is symmetric, we may assume that $\xi_1 \leq \xi_i$ for all $i \in [r]$. We will first identify a threshold δ and provide lower bounds on $P_F(\boldsymbol{\xi})$ for the cases $\xi_1 < \delta$ and $\xi_1 \geq \delta$. We then pick ϵ accordingly.

If $\xi_1 < \delta$, we bound $P_F(\boldsymbol{\xi})$ using the Taylor expansion about $(0, 1/r, \ldots, 1/r)$. That is,

$$P_F(\boldsymbol{\xi}) = \sum_{a_1=0}^{f-1} \cdots \sum_{a_r=0}^{f-1} \frac{\delta_1^{a_1}(-\delta_2)^{a_2} \cdots (-\delta_r)^{a_r}}{a_1! \cdots a_r!} \frac{\partial^{a_1+\ldots+a_r} P_F}{\partial \xi_1^{a_1} \cdots \partial \xi_r^{a_r}}(0, 1/r, \ldots, 1/r),$$

where $\delta_1 = \xi_1$ and $\delta_i = 1/r - \xi_i$ for $2 \leq i \leq r$. Observe that $\delta_1 = \sum_{i=2}^r \delta_i + \epsilon'$ and $\deg(P_F) \leq f - 1$.

However, as $\delta_1 < \delta$ is small (that is, by choice of δ), we have

$$\begin{aligned}
P_F(\boldsymbol{\xi}) &= \left(\sum_{i=1}^{f-1} \frac{\delta_1^i}{i!} \frac{\partial P_F}{\partial \xi_1^i} + \sum_{\substack{\mathbf{a}\in\mathbb{Z}_+^r \\ a_2+\ldots+a_r>0}} \frac{\delta_1^{a_1}(-\delta_2)^{a_2}\cdots(-\delta_r)^{a_r}}{a_1!\cdots a_r!} \frac{\partial^{a_1+\ldots+a_r} P_F}{\partial \xi_1^{a_1}\cdots\partial\xi_r^{a_r}} \right)(0,1/r,\ldots,1/r) \\
&\geq \left(\sum_{i=1}^{f-1} \frac{\delta_1^i}{i!} \frac{\partial P_F}{\partial \xi_1^i} - 2\delta_1 \left(\sum_{i=2}^{r} \delta_i\right) \frac{\partial^2 P_F}{\partial\xi_1\partial\xi_2} \right)(0,1/r,\ldots,1/r) \\
&\geq \left(\sum_{i=1}^{f-1} \frac{\delta_1^i}{i!} \frac{\partial P_F}{\partial \xi_1^i} - \frac{1}{2} \left(\sum_{i=2}^{r} \delta_i\right) \frac{\partial P_F}{\partial\xi_1} \right)(0,1/r,\ldots,1/r) \\
&\geq \left(\sum_{i=2}^{f-1} \frac{\delta_1^i}{i!} \frac{\partial P_F}{\partial \xi_1^i}(0,1/r,\ldots,1/r) + \left(\epsilon' + \frac{1}{2} \left(\sum_{i=2}^{r} \delta_i\right)\right) \frac{\partial P_F}{\partial\xi_1} \right)(0,1/r,\ldots,1/r) \\
&\geq P_F(\epsilon', 1/r,\ldots,1/r) + \frac{1}{2}\left(\sum_{i=2}^{r} \delta_i\right) \frac{\partial P_F}{\partial\xi_1}(0,1/r,\ldots,1/r)
\end{aligned}$$

where the first inequality follows using a second order approximation and the fact that

$$\frac{\partial^{a_1+\ldots+a_r} P_F}{\partial \xi_1^{a_1}\cdots\partial\xi_r^{a_r}}(0,1/r,\ldots,1/r) = 0$$

whenever $a_1 = 0$. Hence, if $\xi_1 < \delta$, then $P_F(\boldsymbol{\xi}) > P_F(\epsilon', 1/r, \ldots, 1/r)$ unless $\sum_{i=2}^r \delta_i = 0$, that is, $\boldsymbol{\xi} = (\epsilon', 1/r, \ldots, 1/r)$.

On the other hand, as all coefficients in P_F are non-negative, it follows that $P_F(\boldsymbol{\xi}) \geq P_F(\xi_1, \ldots, \xi_1)$. Hence, if $\xi_1 \geq \delta$, we have the lower bound $P_F(\boldsymbol{\xi}) \geq P_F(\delta, \ldots, \delta)$. As δ is a fixed constant, we may pick ϵ such that $P_F(\delta, \ldots, \delta) > P_F(\epsilon, 1/r, \ldots, 1/r) \geq P_F(\epsilon', 1/r, \ldots, 1/r)$, to complete the proof. □

We now illustrate a relationship between α_F and vectors in $\mathcal{S}^*$ by observing that

$$\frac{\partial P_F}{\partial \xi_1}(0, 1/r, \ldots, 1/r) = \alpha_F.$$

To be precise, if $\boldsymbol{\xi} = (\xi_1, 1/r, \ldots, 1/r) \in \mathcal{S}_\rho \cap \mathcal{S}^*$, we have $P_F(\boldsymbol{\xi}) = \sum_{i=1}^{f-1} \frac{\xi_1^i}{i!} \frac{\partial P_F}{\partial \xi_1^i}(0, 1/r, \ldots, 1/r)$. So, $P_F(\boldsymbol{\xi}) = \alpha_F(\rho - (1 - 1/r))$ precisely when $\deg(P_F) = r$. It also implies that if $\deg(P_F) \geq r + 1$ and $\rho \in (\frac{r-1}{r}, \rho_F^*)$, then $p(\rho) > \alpha_F(\rho - \frac{r-1}{r})$. We are interested in the values of ρ for which the last inequality holds. We now define the following two, closely related parameters.

Definition 4.2.11. *Let* $\rho_F = \inf\left\{\rho \in [\rho^*, 1] : p(\rho) \leq \alpha_F(\rho - \frac{r-1}{r})\right\}$. *If* $\rho_F^* = \infty$ *or* $p(\rho) > \alpha_F(\rho - (r-1)/r)$ *for all* $\rho \in [\rho_F^*, 1]$, *then* $\rho_F = \infty$.

Definition 4.2.12. *Let* $\hat{\rho}_F = \inf\left\{\rho \in [\rho^*, 1] : p(\rho) < \alpha_F(\rho - \frac{r-1}{r})\right\}$. *If* $\rho^*_F = \infty$ *or* $p(\rho) \geq \alpha_F(\rho - (r-1)/r)$ *for all* $\rho \in [\rho^*_F, 1]$, *then* $\hat{\rho}_F = \infty$.

As a consequence of Theorem 4.1.3 (Mubayi [Mub10]), we observe that $\hat{\rho}_F - (1-1/r) \geq c_0(F) > 0$. In fact, we will prove in Section 4.3 that $c_2(F) = \hat{\rho}_F - (1-1/r)$. On the other hand, ρ_F appears as one of our bounds on $c_1(F)$.

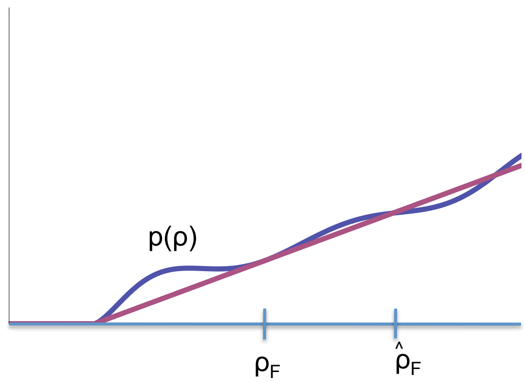

Figure 4.1: ρ_F and $\hat{\rho}_F$.

To give a better picture of proceedings, let us recall some previous parameters. First, consider starting with the Turán graph and 'growing' the graph by adding extra edges. Loosely speaking, the number of copies of F grows 'linearly' with q with a slope of α_F. On the other hand, if we start with a slight perturbation of the partition sizes, we have a slope slightly smaller than α_F (but a higher intercept). The ratio π_F gives the intersection of these two curves. Alternatively, we may start with a Turán graph on one fewer vertices and grow the graph by introducing a vertex of appropriate degree. The number of copies then grows according to $p(\rho)$. In this scenario, ρ_F and $\hat{\rho}_F$ identify, respectively, the first time this curve intersects and crosses the line of slope α_F. In a sense, the values ρ_F and $\hat{\rho}_F$ signify critical densities when comparing $H \in \mathcal{T}_r^q(n)$ with those graphs obtained by altering the neighborhoods of certain vertices.

For generality, the values ρ_F and $\hat{\rho}_F$ in Figure 4.1 do not coincide. However, $\rho_F = \hat{\rho}_F$ for all graphs we have thus far encountered, and we believe equality holds for all graphs. In many instances, this would imply that $c_1(F) = c_2(F)$.

Lemma 4.2.13. *If* $\rho_F < \infty$, *then* $\rho_F < 1$.

Proof. First, note that if $\deg(P_F) = r$, then $P_F(\boldsymbol{\xi}) = C\prod_{i=1}^r \xi_i$, where C is some positive constant. As $\prod_{i=1}^r \xi_i$ is minimized by maximizing the variance of $\xi_1, \ldots, \xi_r$, it follows that

$\rho_F^* = \infty$ and, by definition, $\rho_F = \infty$.

On the other hand, if $\deg(P_F) \geq r+1$, then

$$\begin{aligned} p(1) = P_F(1/r,\ldots,1/r) &= \sum_{i=1}^{f-1}(1/r)^i \frac{\partial^i P_F}{\partial \xi_1^i}(0,1/r,\ldots,1/r) \\ &> \alpha_F/r = \alpha_F(1-(1-1/r)) \end{aligned}$$

as $\frac{\partial P_F}{\partial \xi_1}(0,1/r,\ldots,1/r) = \alpha_F$ and $\sum_{i=2}^{f-1} \frac{\partial^i P_F}{\partial \xi_1^i}(0,1/r,\ldots,1/r) > 0$. Hence, if $\rho_F < \infty$, then $\rho_F < 1$. □

We now relate $P_F(\boldsymbol{\xi})$ and $F(n,\mathbf{d})$.

Lemma 4.2.14. *For all $\delta > 0$ and n large enough, if $\xi_i = d_i/n$, then $|F(n,\mathbf{d}) - n^{f-1}P_F(\boldsymbol{\xi})| < \delta n^{f-1}$.*

Proof. As $n^t \geq (n)_t$ for all n and t, it follows that

$$\begin{aligned} F(n,\mathbf{d}) &= \frac{1}{Aut(F)} \sum_{v \text{ critical}} \sum_{\chi_v} \prod_{i=1}^{r} (n_i - y_i)_{x_i} (\xi_i n)_{y_i} \\ &\leq \frac{1}{Aut(F)} \sum_{v \text{ critical}} \sum_{\chi_v} \prod_{i=1}^{r} \left(\frac{n}{r}\right)^{x_i} \xi_i^{y_i} n^{y_i} \\ &= n^{f-1} \frac{1}{Aut(F)} \sum_{v \text{ critical}} \sum_{\chi_v} \prod_{i=1}^{r} \frac{1}{r^{x_i}} \xi_i^{y_i} \\ &= n^{f-1}\ F(\boldsymbol{\xi}) \end{aligned}$$

Note that $P_F(\boldsymbol{\xi}) = 0$ if $\xi_i = 0$ for some $i \in [r]$. By Lemma 4.2.8, we may pick $\delta' \ll \delta$ such that $P_F(\boldsymbol{\xi}) < \delta$ whenever $\xi_i < \delta'$ for some i. Let N be such that $(n)_t \geq n^t(1-\delta'/2)$ for all $n \geq N$ and all $t \leq r$. Let $n_0 = N/\delta'$.

Now, if $d_i = \xi_i n \geq N$ for all i, where $n \geq n_0$, then

$$\begin{aligned} F(n,\mathbf{d}) &= \frac{1}{Aut(F)} \sum_{v \text{ critical}} \sum_{\chi_v} \prod_{i=1}^{r} (n_i - y_i)_{x_i} (\xi_i n)_{y_i} \\ &\geq \frac{1}{Aut(F)} \sum_{v \text{ critical}} \sum_{\chi_v} \prod_{i=1}^{r} (1-\delta') \left(\frac{n}{r}\right)^{x_i} \xi_i^{y_i} n^{y_i} \\ &= (1-\delta')^r n^{f-1} \frac{1}{Aut(F)} \sum_{v \text{ critical}} \sum_{\chi_v} \prod_{i=1}^{r} \frac{1}{r^{x_i}} \xi_i^{y_i} \\ &\geq (1-\delta')^r n^{f-1}\ F(\boldsymbol{\xi}) \\ &\geq n^{f-1}\ F(\boldsymbol{\xi}) - \delta n^{f-1}. \end{aligned}$$

On the other hand, if $\xi_j n < N$ for some j, then $\xi_j < \delta'$ and $n^{f-1}P_F(\boldsymbol{\xi}) < \delta n^{f-1}$. $\square$

We now allow for inequitable partitions.

Lemma 4.2.15. *For every $\epsilon > 0$, there exists $\delta > 0$ satisfying the following: if $n = \sum_i n_i$ and $|n_i - n/r| \leq \delta n$ for all $i \in [r]$, then for all $\mathbf{d} = (d_1, \ldots, d_r)$ with $d_i \leq n_i$, there exists $\boldsymbol{\xi}' \in \mathcal{S}$ such that $|\xi_i' - d_i/n| \leq \delta$ and $|F(n_1, \ldots, n_r; \mathbf{d}) - n^{f-1}P_F(\boldsymbol{\xi}')| < \epsilon n^{f-1}$.*

Proof. Let $H = K(V_1, \ldots, V_r) + u$ where $|V_i| = n_i$ and u has d_i neighbors in each V_i. Then $\#F(H) = F(n_1, \ldots, n_r; \mathbf{d})$. Now let $H' \subseteq H \subseteq H''$, where $H' = K(V_1', \ldots, V_r') + u'$ and $H'' = K(V_1'', \ldots, V_r'') + u''$ with $|V_i'| = \lfloor (1/r - \delta)n \rfloor$ and $|V_i''| = \lceil (1/r + \delta)n \rceil$. Let $\boldsymbol{\xi}'$ be the density vector for both u' and u'', with $\xi_i' = \min(\xi_i, 1/r)$, where $\xi = d_i/n$. So, $\|\boldsymbol{\xi} - \boldsymbol{\xi}'\|_\infty \leq \delta$ and $\|\boldsymbol{\xi} - \boldsymbol{\xi}'\|_1 \leq \delta r$. Therefore

$$\begin{aligned} F(n_1, \ldots, n_r; \mathbf{d}) &\geq F((1 - \delta r)n, \boldsymbol{\xi}') \\ &\geq (1 - \delta r)^{f-1} n^{f-1} P_F(\boldsymbol{\xi}') - \epsilon n^{f-1}/2 \\ &\geq n^{f-1} P_F(\boldsymbol{\xi}') - \epsilon n^{f-1}. \end{aligned}$$

Similarly, $F(n_1, \ldots, n_r; \mathbf{d}) \leq n^{f-1}P_F(\boldsymbol{\xi}') + \epsilon n^{f-1}$.

We note here that $\boldsymbol{\xi}' \leq \boldsymbol{\xi}$ and thus $\sum_i \xi_i \leq \sum_i d_i/n$. $\square$

4.3 Asymptotic Optimality of $\mathcal{T}_r^q(n)$

In this section we prove the following result.

Theorem 4.3.1. *Let F be an r-critical graph. Then, $c_2(F) = \hat{\rho}_F - (1 - 1/r)$.*

We use the proof of Theorem 4.3.1, to build the framework that allows us to prove Theorem 4.1.4. In a sense, the proof of Theorem 4.3.1 provides a strong stability condition that we exploit in the sections to come.

As was the case in Mubayi's result [Mub10], the graph removal lemma (see [KS96, Theorem 2.9]) and the Erdős-Simonovits Stability Theorem are key components of our proof.

Theorem 4.3.2 (Graph Removal Lemma)**.** *Let F be a graph with f vertices. Suppose that an n-vertex graph H has at most $o(n^f)$ copies of F. Then there is a set of edges of H of size $o(n^2)$ whose removal from H results in a graph with no copies of F.*

For hypergraph versions of the Graph Removal Lemma, see [Gow07, NRS06, RS06, Tao06].

Theorem 4.3.3 (Erdős [Erd67] and Simonovits [Sim68])**.** *Let $r \geq 2$ and F be a graph with chromatic number $r+1$. Let H be a graph with n vertices and $t_r(n) - o(n^2)$ edges that contains no copy of F. Then there is a partition of the vertex set of H into r parts so that the number of edges contained within a part is at most $o(n^2)$. In other words, H can be obtained from $T_r(n)$ by adding and deleting a set of $o(n^2)$ edges.*

Proof of Theorem 4.3.1 Let us first show that $c_2(F) \leq \hat{\rho}_F - (1 - 1/r)$. We may assume that $\hat{\rho}_F$ is finite (and, as a consequence of Lemma 4.2.13, less than 1). Given arbitrary $c > \hat{\rho}_F - (1 - 1/r)$, we produce an infinite sequence of graphs, H_n on $n+1$ vertices and $t_r(n+1) + q$ edges (where $q < cn$) for which $\#F(H_n) < q(1-\epsilon)c(n+1, F)$ for some $\epsilon > 0$ independent of n.

As $p(\rho)$ is continuous, we may pick $\lambda, \lambda' \ll c - \hat{\rho}_F + (1-1/r)$ such that $p(\rho) < \alpha_F(\rho - \frac{r-1}{r})$ whenever $\rho - \hat{\rho}_F \in (\lambda, \lambda')$. Then, for ϵ small enough, there exists $\delta > 0$ (with $\delta < (\lambda' - \lambda)/2$) such that $p(\rho) \leq (1 - 8\epsilon)\alpha_F(\rho - \frac{r-1}{r})$ for $\rho - \hat{\rho}_F \in [\lambda + \delta, \lambda' - \delta]$. We pick $\boldsymbol{\xi} \in \mathbb{Q}^r$, with $\sum_i \xi_i = \rho$ and $\rho - \hat{\rho}_F \in [\lambda + \delta, \lambda' - \delta]$ for which $P_F(\boldsymbol{\xi}) < (1 - 4\epsilon)\alpha_F(\rho - \frac{r-1}{r})$. Now, for n large enough, such that $\boldsymbol{\xi} n \in r\mathbb{Z}^r$ (there are infinitely many such n), we construct the graph $H_n = T_r(n) + u$ where u has exactly $\xi_i n$ neighbors in each part V_i. That is, H_n has $n+1$ vertices and $t_r(n+1) + q$ edges, where $q = (\rho - \frac{r-1}{r})n$. Let $F(u)$ be the number of copies of F that use the vertex u. Then,

$$
\begin{aligned}
\#F(H_n) &= F(u) \\
&\leq (1 - 2\epsilon)\alpha_F(\rho - \frac{r-1}{r})n^{f-1} \\
&= (1 - 2\epsilon)q\alpha_F n^{f-2} \\
&< (1 - \epsilon)qc(n+1, F),
\end{aligned}
$$

proving that $c_2(F) \leq \hat{\rho}_F - (1 - 1/r)$.

Let us now prove the converse. For more generality, we initially assume only that $q \leq n^{2-\eta}$ for arbitrary $\eta > 0$. Given arbitrary $\epsilon > 0$, we define some constants satisfying the following hierarchy:

$$1/n_0 \ll \delta_1 \ll \delta_2 \ll \delta_3 \ll \delta_4 \ll \delta_5 \ll \delta_6 \ll \delta_7 \ll \epsilon.$$

Let $n \geq n_0$ and H be a graph on $[n]$ with $t_r(n) + q$ edges, where $q \leq n^{2-\eta}$, containing the fewest number of copies of F.

For comparison, we first consider $\#F(H^*)$ for some graph $H^* \in \mathcal{T}_r^q(n)$. In particular, we pick H^* where all q 'extra' edges belong to one part, say V_1, and form a regular (or almost regular) bipartite graph (so, each vertex $v \in V_1$ has about $2rq/n \leq 2n^{1-\eta}$ neighbors in V_1). Now, each of the q edges produces $c(n, F)$ copies of F. On the other hand, the number of copies that contain more than one extra edge may be bounded by

$$q^2 f^4 2^{f^2} n^{f-4} + q \cdot 2(2rq/n) f^4 2^{f^2} n^{f-3} = O(n^{f-2\eta}),$$

where the first term counts the number of copies using disjoint extra edges and the second term counts those copies that use adjacent extra edges. This gives us an upper bound of $qc(n, F) + o(n^{f-\eta}) = O(n^{f-\eta})$ on the minimum number of copies of F contained in a graph with $t_r(n) + q$ edges.

As $\#F(H) \leq \#F(H^*)$, it follows that $\#F(H) < n^{f-\eta/2}$. This allows us to apply the Removal Lemma to obtain a subgraph H' containing $t_r(n) + q - \delta_1 n^2$ edges and no copies of F. We then apply the Erdős-Simonovits Stability Theorem and obtain an r-partite subgraph $H'' \subseteq H'$ with at least $t_r(n) + q - \delta_2 n^2$ edges.

Let $V(H) = V_1 \cup \ldots \cup V_r$ be maximum cut r-partition of $V(H)$. We call the edges of H that intersect two parts *good* and those that lie within one part we call *bad*. We denote the sets of good and bad edges by G and B, respectively. Let $M = K(V_1, \ldots, V_r) \setminus G$ be the set of *missing* edges. We observe here that $n_i = |V_i| = n/r \pm \delta_3 n$ and $|B| \leq \delta_2 n^2$. We also assume, without loss of generality, that $c(n_1, n_2, \ldots, n_r; F) \leq c(n_{\sigma(1)}, \ldots, n_{\sigma(r)}; F)$ for all permutations σ of $[r]$.

Let $F(e)$ be the number of copies of F in H containing the edge $e \in B$ but no other bad edge. For $uv \in M$, we denote by $F'(uv)$ the number of potential copies of F using uv, that is, $F'(uv) = \#F(H') - \#F(H)$, where $H' = H + uv$. Recall that $F(u)$ is the number of copies of F that contain the vertex u.

If $M = \emptyset$, $\#F(H) \geq |B| \cdot c(n_1, \ldots, n_r; F)$. However, $|B| \geq q$ and by Lemma 4.2.4, $c(n_1, \ldots, n_r; F) \geq c(n, F) - 2\zeta_F(\delta_3 n) n^{f-3} \geq (1-\epsilon) c(n, F)$, giving the desired result.

Now assume $M \neq \emptyset$ and let $uv \in M$. For each edge $xw \in B$ where $x \in \{u, v\}$, there are at most $m(n, F) + \delta_4 n^{f-3}$ potential copies of F that use xw and uv (where $m(n, F)$ is the quantity defined in Definition 4.2.2). In addition, bad edges not incident to u or v may contribute up to $|B| f^4 2^{f^2} n^{f-4} < \delta_4 n^{f-2}$ potential copies via uv. So,

$$F'(uv) \leq \mu_F n^{f-3}(d_B(u) + d_B(v)) + 3\delta_4 n^{f-2}. \tag{4.2}$$

On the other hand, for $u'v' \in B$

$$F(u'v') \geq \alpha n^{f-2} - \mu_F n^{f-3}(d_M(u') + d_M(v')) - 3\delta_4 n^{f-2}. \tag{4.3}$$

However, $H \in \mathcal{H}_F(n,q)$ implies that $F'(uv) \geq F(u'v')$ for any pair $uv \in M$ and any edge $u'v' \in B$, as otherwise, one may delete the edge $u'v'$ and replace it with uv to obtain a graph with fewer copies of F. Therefore,

$$\begin{aligned} 0 &\leq F'(uv) - F(u'v') \\ &\leq \mu_F n^{f-3}(d_B(u) + d_B(v) + d_M(u') + d_M(v')) - \alpha_F n^{f-2} + 6\delta_4 n^{f-2} \end{aligned}$$

and

$$d_B(u) + d_B(v) + d_M(u') + d_M(v') \geq \left(\frac{\alpha_F}{\mu_F} - 4\delta_5\right) n. \tag{4.4}$$

We now use the following identity

$$\sum_{uv \in M} (d_B(u) + d_B(v)) = \sum_{u \in V} d_M(u) d_B(u) = \sum_{u'v' \in B} (d_M(u') + d_M(v'))$$

and the fact that $|B| \geq |M|$, to obtain a pair $uv \in M$ and an edge $u'v' \in B$ such that

$$d_B(u) + d_B(v) \geq d_M(u') + d_M(v'). \tag{4.5}$$

Then, (4.4) and (4.5) imply the existence of a vertex u such that

$$d_B(u) \geq \left(\frac{\alpha_F}{4\mu_F} - \delta_5\right) n. \tag{4.6}$$

Consider the set $X = \{x \in V(H) : d_B(x) > \delta_7 n\}$. Note that (4.6) implies that $X \neq \emptyset$. Let us first analyze the case where $\deg(P_F) = r$. We know, by Lemma 4.2.6, that $F = K_{r+1}$ or $F = C_{2k+1}$.

Claim 4.3.4. *Let F be an r-critical graph and let $\deg(P_F) = r$. Let $H \in \mathcal{H}_F(n,q)$ and $u \in V_i$ with associated density vector $\boldsymbol{\xi}$. If $\xi_i > \delta_7$, then $\xi_j > 1/r - 2\delta_6$ for all $j \neq i$.*

Proof of Claim. Assume there is a vertex u violating the above condition. Say, without loss of generality, that $u \in V_1$, $\xi_1 > \delta_7$ and $\xi_2 \leq 1/r - 2\delta_6$. Consider replacing u with a vertex u' with corresponding density vector $\boldsymbol{\xi}'$, such that $\xi'_1 = \xi_1 - \delta_6$, $\xi'_2 = \xi_2 + \delta_6$ and $\xi'_i = \xi_i$ for $i \in [r]$, $i \neq 1, 2$. As $V_1 \cup \ldots \cup V_r$ is a max-cut partition, we may assume that $\xi_i \geq \xi_1 > \delta_7$ for all $i \in [r]$. Then, for some positive constant $C = C(F)$,

$$\begin{aligned} P_F(\boldsymbol{\xi}') &= C \prod_{i \in [r]} \xi'_i \\ &= C(\xi_1 - \delta_6)(\xi_2 + \delta_6) \prod_{i=3}^{r} \xi_i \\ &\leq C \prod_{i \in [r]} \xi_i - C\delta_6^2 \prod_{i=3}^{r} \xi_i \\ &\leq P_F(\boldsymbol{\xi}) - 3\delta_5. \end{aligned}$$

Lemma 4.2.15 implies that $F(u) \geq n^{f-1}P_F(\xi) - \delta_5 n^{f-1}$ and $F(u') \leq n^{f-1}P_F(\xi') + \delta_5 n^{f-1}$. Hence, we reduce the number of copies of F by making this alteration, contradicting the optimality of H. ▍

As $F(u) \geq n^{f-1}(P_F(\boldsymbol{\xi}) - \delta_5)$, it follows that $F(x) = \Theta(n^{f-1})$ for all $x \in X$. Furthermore, as $\#F(H) < n^{f-\eta/2}$, it must be that $|X| = O(n^{1-\eta/2}) < \delta_5 n$. Let $M(H-X)$ be the set of missing edges in the graph $H - X$.

Claim 4.3.5. $M(H-X) = \emptyset$.

Proof of Claim. Assume there exists $uv \in M$ with $u, v \notin X$. Then, by (4.2), $F'(uv) \leq 3\delta_7 \mu n^{f-2}$. On the other hand, consider a vertex $x \in X$. There is a bad edge xw such that $d_M(w) < \delta_6 n$ (otherwise, $|B| \geq |M| > \delta_2 n^2$). Then, for this edge, $F(xw) \geq \alpha_F n^{f-2} - 2r\delta_6 n^{f-2} > F'(uv)$, resulting in a contradiction. ▍

Claims 4.3.4 and 4.3.5 imply that for any vertex u,

$$d_M(u) \leq \max(|X|, (r-1)(2\delta_6 + \delta_3)) < 2r\delta_6 n. \tag{4.7}$$

It follows that for any bad edge $u'v'$, $d_M(u') + d_M(v') < 4r\delta_6 n$. That is, by (4.3)

$$F(u'v') \geq (1-\epsilon)c(n, F), \tag{4.8}$$

implying that $\#F(H) \geq (1-\epsilon)qc(n, F)$ for all H whenever $\deg(P_F) = r$.

On the other hand, if $\deg(P_F) \geq r+1$, we have the following claim:

Claim 4.3.6. *If $\deg(P_F) \geq r+1$ and $H \in \mathcal{H}_F(n, q)$, then for all $u \in V_i$ with associated density vector $\boldsymbol{\xi} \in \mathcal{S}_\rho$, $\rho \notin [\frac{r-1}{r} + \delta_6, \rho_F - \delta_6]$.*

Proof of Claim. Consider $p(\rho)$. By the definition of ρ_F, $p(\rho) > \alpha_F(\rho - \frac{r-1}{r})$ for $\rho \in (\frac{r-1}{r}, \rho_F)$. So, $p(\rho) - \alpha_F(\rho - \frac{r-1}{r})$ has a positive lower bound over any compact subset of $(\frac{r-1}{r}, \rho_F)$. In particular, by choice of δ_5 and δ_6, we have $p(\rho) - \alpha_F(\rho - \frac{r-1}{r}) \geq 5\delta_5$ whenever $\rho \in [\frac{r-1}{r} + \delta_6, L]$, (where $L = 1$ if $\rho_F = \infty$ and $L = \rho_F - \delta_6$ otherwise).

Therefore, if u (in V_1, say) has density vector $\boldsymbol{\xi} \in \mathcal{S}_\rho$ with $\rho \in [\frac{r-1}{r} + \delta_6, L]$, then $F(u) \geq (\alpha_F(\rho - \frac{r-1}{r}) + 4\delta_5)n^{f-1}$. On the other hand, we may replace u with a vertex u' whose associated density vector $\boldsymbol{\xi}' \in \mathcal{S}_{\rho'}$ has $\xi'_1 = 0$ and $\xi'_i = |V_i|/n$ for $i = 2, \ldots, r$. We note here that $|\rho' - \frac{r-1}{r}| < \delta_4$ and it follows that $F(u') \leq \delta_5 n^{f-1}$. Next, we distribute the remaining $(\rho - \rho')n$ edges evenly amongst vertices in V_1 with bad degree at most $\delta_4 n$. As these create at most $(\rho - \frac{r-1}{r} + \delta_4)nc(n, F) + \delta_5 n^{f-1}$ new copies of F, we lower the number of copies and contradict the optimality of H. ▍

It follows immediately that

$$d(x) > (\rho_F - \delta_6)n \quad \text{for any } x \in X. \tag{4.9}$$

Otherwise, $F(x) \geq n^{f-1}(P_F(\delta_7, \ldots, \delta_7) - \delta_5)$ will exceed the $\delta_6 C_F n^{f-1}$ upper bound for a vector $\boldsymbol{\xi} \in \mathcal{S}^*$ with $\sum_i \xi_i \leq \frac{r-1}{r} + \delta_6$. As a result, (4.6) implies that

$$(\rho_F = \infty) \Rightarrow (X = \emptyset) \Rightarrow (M = \emptyset) \tag{4.10}$$

and we are done. We have the following claim if ρ_F is finite (that is, $\rho_F < 1$).

Claim 4.3.7. *If $\rho_F < 1$, $0 < c < \hat{\rho}_F - (1-1/r)$, and $q \leq cn$, then $\#F(H) \geq q(1-\epsilon)c(n, F)$.*

Proof of Claim. By assumption, $M \neq \emptyset$ and we have a vertex u satisfying (4.6). If $d(u) \geq \hat{\rho}_F n$, then

$$F(u) \geq n^{f-1}(p(\hat{\rho}_F) - \delta_4) \geq n^{f-1}(\alpha_F(\hat{\rho}_F - (1 - 1/r)) - \delta_4) > q(1 - \epsilon)c(n, F).$$

Therefore, $(\rho_F - \delta_6)n \leq d(u) \leq \hat{\rho}_F n$. As $p(\rho) \geq \alpha_F(\rho - \frac{r-1}{r})$ for $\rho \leq \hat{\rho}_F$, it follows that

$$F(u) \geq n^{f-1}(\alpha_F(d(u)/n - (r-1)/r) - \delta_4) \geq (d(u) - (r-1)n/r - \delta_5)\, c(n, F),$$

which implies the result if $q < (\rho_F - (1-1/r) - \delta_5)n$. So, assume $q \geq (\rho_F - (1-1/r) - \delta_5)n$ and let $H_1 = H - u = H_0 - u_0$. If $H_1 \notin \mathcal{H}(n-1, q')$ (for appropriate q'), we pick $H_1' \in \mathcal{H}(n-1, q')$ and iteratively create sequences H_i' and u_i until we reach H_k' where $M(H_k') = \emptyset$ or $|H_k'| \leq t_r(n-k)$. Now,

$$\#F(H) = \#F(H_k') + \sum_{i<k} F(u_i) \geq \#F(H_k') + c(n, F)(1 - \delta_6) \sum_{i<k} (d(u_i) - (r-1)n/r).$$

As $d(u_i) \geq (\rho_F - \delta_6)n$ for all $i < k$, it follows that $k < 2c/\rho_F$. If $|H_k'| \leq t_r(n-k)$, then $\sum_{i<k}(d(u_i) - (r-1)n/r) \geq q - k > (1 - \delta_6)q$. On the other hand, if $M(H_k') = \emptyset$, then for all $vw \in B(H_k')$, $F(vw) \geq c(n_1', \ldots, n_r'; F) > (1 - \delta_6)c(n, F)$, where $n_1' + \cdots + n_r' = n - k$. That is, $\#F(H_k') \geq |B(H_k')|(1 - \delta_6)c(n, F)$. As $|B(H_k')| + \sum_{i<k}(d(u) - (r-1)n/r) \geq q - k$, it once again follows that $\#F(H) \geq q(1 - \epsilon)c(n, F)$. ▮

Therefore, $c_2(F) \geq \hat{\rho}_F - (1 - 1/r)$. This completes the proof of Theorem 4.3.1. □

4.4 Optimality of $\mathcal{T}_r^q(n)$

We prove a stronger result that implies Theorem 4.1.4. Recall that the sign of $x \in \mathbb{R}$, denoted $\mathrm{sgn}(x)$, is 0 if x is 0 and $x/|x|$, otherwise. Let $\theta_F = \rho_F - (1 - 1/r)$.

We reuse many of the results and structures introduced in the previous section, including the sequence of constants $1/n_0 \ll \delta_1 \ll \cdots \ll \delta_7 \ll \epsilon$.

Theorem 4.4.1. *Let F be a connected r-critical graph. Then, $c_{1,i}(F) \geq c_1(F) \geq \min(\pi_F, \theta_F)$ for all $i \in [r]$. Furthermore, if $t = \mathrm{sgn}(\zeta_F)$, then $c_{1,t}(F) \geq \min(2\pi_F, \theta_F)$.*

Proof of Theorem 4.4.1 Let $c(n, F) = c(n'_1, \ldots, n'_r; F)$. Pick some $c > 0$ and consider $H \in \mathcal{H}_F(n, q)$ where $q < cn$. Let G, B and M be the sets of good, bad and missing edges, respectively, of a max-cut partition $V(H) = V_1 \cup \ldots \cup V_r$, with $|V_1| \geq |V_2| \geq \ldots \geq |V_r|$. Let $a = \max(|V_1| - \lceil n/r \rceil, \lfloor n/r \rfloor - |V_r|)$. It follows that $|B| \geq q + m + \frac{a^2 r}{2(r-1)}$, where $m = |M|$.

Let us first consider the case $m = 0$. Then, by Lemma 4.2.4 we have,

$$\begin{aligned}
\#F(H) &\geq (q + \frac{a^2 r}{2(r-1)})(c(n, F) - |\zeta_F| a n^{f-3} - O(a^2 n^{f-4})) \\
&\geq qc(n, F) + \frac{a^2 r}{2(r-1)}(\alpha_F n^{f-2} - O(n^{f-3})) - c\,|\zeta_F| a n^{f-2} - o(a^2 n^{f-2}) \\
&\geq qc(n, F) + a n^{f-2} \left(\alpha_F \frac{ar}{2(r-1)} - c\,|\zeta_F| - \delta_4 a \right)
\end{aligned}$$

So, if $a \geq 2(c|\zeta_F| + 1)/\alpha_F$ we have $\#F(H) > qc(n, F) + a n^{f-2}$. On the other hand, as seen in the proof of Theorem 4.3.1, there is a graph in $\mathcal{H}_F(n, q)$ containing at most $qc(n, F) + c^2 f^4 2^{f^2} n^{f-2}$ copies of F. Therefore, as H is optimal, $a \leq \max(c^2 f^4 2^{f^2},\ 2(c|\zeta_F| + 1)/\alpha_F) = O(1)$.

We now refine the argument to show that if $c < \pi_F - \epsilon$, then all optimal graphs are contained in the set $\mathcal{T}_r^q(n)$. In other words, if $|H| = t_r(n) + q$ and H contains the complete r-partite graph on parts of size $n_1, \ldots, n_r$ where $n = n_1 + \ldots + n_r$ and $n_1 \geq n_r + 2$, then H is not optimal unless $q > (\pi_F - \epsilon)n$.

For $i \in [r]$, let $B_i = B[V_i]$ be the set of bad edges contained in V_i.

Claim 4.4.2. *If $|V_j| = |V_k| + s$, where $s > 1$, then*

$$(|B_j| - |B_k|)\zeta_F \geq (s - 1)(1 - \delta_4)\alpha_F n.$$

Proof of Claim. Assume otherwise. Consider H' obtained from H by moving one vertex from V_j to V_k. Let $v \in V_j$ with $d_B(v) \leq d_B(u)$ for all $u \in V_j$. We replace v with a vertex v' such that $uv' \in H'$ for all $u \notin V_k$. Next, we pick $d_B(v)$ vertices in V_k with the lowest bad degrees as neighbors of v'. However, as $(|V_j| - 1)(|V_k| + 1) = |V_j||V_k| + s - 1$, we remove $s - 1$ bad edges chosen arbitrarily.

Let us now consider the net effect of this change. As $0 \leq |B_j|, |B_k| < q + a^2$, it follows that $0 \leq d_B(v) < 3rc$. In addition, we may assume that $d_B(u) < 6rc$ for all u such that

$uv' \in B(H')$. The copies of F that use more than one bad edge that we introduce due to this alteration are at most $f^4 2^{f^2}\left(6rc\, d_B(v)n^{f-3} + |B|d_B(v)n^{f-4} + |B|^2 n^{f-5}\right) = O(n^{f-3})$. For the copies that contain exactly one bad edge, we use Lemma 4.2.4. First, the $d_B(v)$ edges incident to v in H but to v' in H' may add $(2a)(3rc)|\zeta_F|n^{f-3} = O(n^{f-3})$ copies of F. All other bad edges either remain in the same part B_i or are deleted. Let $F_H(e)$ and $F_{H'}(e)$ denote the number of copies of F that use the bad edge e in H (resp., H'). As $F_H(e) = c(n,F) - C_i\zeta_F n^{f-3} \pm O(n^{f-4})$, where $C_i \in \{|V_i| - \lfloor n/r \rfloor, |V_i| - \lceil n/r \rceil\}$,

$$F_H(xy) - F_{H'}(xy) = \begin{cases} c(n,F) \pm O(n^{f-3}) & \text{if } xy \text{ was deleted} \\ -\zeta_F n^{f-3} \pm O(n^{f-4}) & \text{if } xy \in B_j \\ \zeta_F n^{f-3} \pm O(n^{f-4}) & \text{if } xy \in B_k \\ \pm O(n^{f-4}) & \text{otherwise.} \end{cases}$$

Here, we obtain the first expression as $F_{H'}(e) = 0$ for each deleted edge. Furthermore, as $|V_i|$ remains unchanged for $i \notin \{j,k\}$, $F_{H'}(xy)$ differs from $F_H(xy)$ only in the error term (which is $O(n^{f-4})$). The change in the cardinality of V_j and V_k, however, contributes a difference of order n^{f-3}.

As the number of deleted edges is $|B(H)| - |B(H')| = s-1$, we have

$$\begin{aligned} 0 &\geq \#F(H) - \#F(H') \\ &\geq \sum_{e \in B}(F_H(e) - F_{H'}(e)) - O(n^{f-3}) \\ &\geq (s-1)c(n,F) - (|B_j| - |B_k|)\zeta_F n^{f-3} - O(n^{f-3}). \end{aligned}$$

Thus, $(|B_j| - |B_k|)\zeta_F \geq (s-1)\alpha_F n - \delta_3 n$, proving the claim. ∎

Note that if $t = 0$, then $\zeta_F = 0$ and Claim 4.4.2 cannot be satisfied. Therefore, $|V_j| - |V_k| \leq 1$ for all $1 \leq j,k \leq r$, that is, $H \in \mathcal{T}_r^q(n)$.

Suppose that $t \in \{-1,1\}$ and $H \notin \mathcal{T}_r^q(n)$. Then there exist j,k such that $|V_j| - |V_k| \geq 2$. Hence, $|B| \geq \max(|B_j|,|B_k|) \geq (1-\delta_4)\pi_F n$, and $q \geq |B| - a^2 \geq (1-\epsilon)\pi_F n$ as required. So, if $n \equiv t \pmod r$, there exist j,k,l with $|V_j| - |V_k| \geq 3$ or $j \neq k$ and $|V_j| - |V_l| = |V_k| - |V_l| = 2t$. In the first case, we apply Claim 4.4.2 directly to obtain $q \geq 2(1-\epsilon)\pi_F$. On the other hand, if the second case holds with $t \cdot \zeta_F > 0$, we have, by applying Claim 4.4.2 twice, that $|B_j|, |B_k| \geq (1-\delta_4)\pi_F$, again implying that $q \geq 2(1-\epsilon)\pi_F$.

Now assume $M \neq \emptyset$. We will handle the two case $\deg(P_F) = r$ and $\deg(P_F) \geq r+1$ separately. Let us first consider the case $\deg(P_F) \geq r+1$.

If $\deg(P_F) \geq r+1$ and $\theta_F = \infty$, (4.10) implies $M = \emptyset$. Therefore, we may assume $\theta_F \in (0, 1/r)$. Once again, if $M \neq \emptyset$, the set $X = \{x \in H : d_B(x) \geq \delta_7 n\}$ is nonempty. Then (4.9) implies that for any $x \in X$,

$$F(x) \geq n^{f-1}(p(\rho_F - \delta_6) - \delta_5) > n^{f-1}(\alpha_F \theta_F - \delta_7) > (\theta_F - \epsilon) n c(n, F) + \frac{\epsilon}{2} n^{f-1}.$$

Thus, if $c \leq \theta_F - \epsilon$, the number of copies of F at some vertex $x \in X$ exceeds $qc(n,F) + c^2 2^{f^2} n^{f-2}$, contradicting our assumption that $H \in \mathcal{H}_F(n,q)$.

This completes the proof of Theorem 4.4.1 for r-critical graphs with $\deg(P_F) \geq r+1$.

We now consider the case when $\deg(P_F) = r$, that is, $F = K_{r+1}$ or $r = 2$ and $F = C_{2k+1}$. It was already shown in [Mub10] that $c_1(C_{2k+1}) > 0$. In what follows, we improve this to show that $c_1(C_{2k+1}) \geq 1/2$. Let us first compute the values of α_F and ζ_F.

Claim 4.4.3. *If F is r-critical and $\deg(P_F) = r$, then $\pi_F = \frac{1}{r}$.*

Proof of Claim. We observe that

$$c(n_1, \ldots, n_r; K_{r+1}) = \prod_{i=2}^{r} n_i,$$

with $c(n,F)$ attained when $n_1 = \lceil n/r \rceil$. Therefore, $\alpha_{K_{r+1}} = r^{-r+1}$. Furthermore,

$$\frac{\partial c}{\partial n_1}(n/r, \ldots, n/r) = 0 \quad \text{and} \quad \frac{\partial c}{\partial n_2}(n/r, \ldots, n/r) = (n/r)^{r-2}.$$

It follows that $\zeta_{K_{r+1}} = r^{-r+2}$ and $\pi_{K_{r+1}} = 1/r$.

On the other hand,

$$c(n_1, n_2; C_{2k+1}) = n_2\,(n_2 - 1) \cdots (n_2 - k + 1)\,(n_1 - 2) \cdots (n_1 - k)$$

and

$$\frac{\partial c}{\partial n_i}(n/2, n/2) = (k + i - 2)(n/2)^{2k-2} + O(n^{2k-3}).$$

So $\alpha_{C_{2k+1}} = 2^{-2k+1}$, $\zeta_{C_{2k+1}} = 2^{-2k+2}$ and $\pi_{C_{2k+1}} = 1/2 = 1/r$.

We also note that $\frac{\alpha_F}{\mu_F} = 1/2$ as

$$m(n_1, n_2; C_{2k+1}) = (n_2 - 1) \cdots (n_2 - k + 1)\,(n_1 - 2) \cdots (n_1 - k) = \frac{c(n_1, n_2; C_{2k+1})}{n_2}. \quad \blacksquare$$

As we noted earlier, if $F = K_{r+1}$, Theorem 4.4.1 follows from a result of Lovász and Simonovits [LS83]. Therefore, we focus on the case where $r = 2$, $k \geq 2$ and $F = C_{2k+1}$. While proving $c_1(C_{2k+1}) \geq 1/2$ is relatively straightforward, showing that $c_{1,1}(C_{2k+1}) \geq 1$ is quite involved. We will prove these lower bounds in a stronger form.

Lemma 4.4.4. *Let $F = C_{2k+1}$. There exists $n_0 = n_0(k)$ such that $\mathcal{H}_F(n,q) \subseteq \mathcal{T}_2^q(n)$ for all $n \geq n_0$ and $q \leq n/2 - 2$. Furthermore, if n is odd, $\mathcal{H}_F(n,q) \subseteq \mathcal{T}_2^q(n)$ for $q \leq n - 10k^2$.*

Proof of Lemma Our proof resembles that of [LS75] in that an approximate structure is refined repeatedly until, finally, it is seen that all optimal graphs belong to the special set $\mathcal{T}_2^q(n)$. We use results from Section 4.3 to identify our initial approximate structure. However, subsequent computations become more complicated as we have to account for copies of C_{2k+1} that may appear in various configurations.

As in previous arguments, elements of this proof will involve alterations that contradict the optimality of a given graph H. In addition, we present two graphs to serve as reference points, i.e., these graphs provide an upper bound on $\#F(H)$. We will show that one of these graphs is an optimal configuration; we make no such claims for the other.

If $q \leq n/2 - 2$, we propose as an optimal graph the graph $H^*(n,q) \in \mathcal{T}_2^q(n)$ constructed as follows: $V(H^*) = U_1 \cup U_2$, $|U_1| = \lceil n/2 \rceil$, $|U_2| = \lfloor n/2 \rfloor$ and $E(H^*) = K(U_1, U_2) \cup K(\{u^*\}, W)$, where $u^* \in U_1$, $W \subseteq U_1 \setminus \{u^*\}$ and $|W| = q$. That is, $H^*(n,q)$ is obtained from $T_2(n)$ by adding (the edges of) a star of size q in U_1. For larger values of q and a partition $n = n_1 + n_2$, we provide the graph $H''(n_1, n_2, q)$ on parts U_1'' and U_2'' with $|U_i''| = n_i$ and $E(H'') \supseteq K(U_1'', U_2'')$. The remaining $|B(H'')| = q + \lfloor n^2/4 \rfloor - n_1 n_2$ edges form a regular (or almost regular) bipartite graph in U_1''.

Let us now bound $\#C_{2k+1}(H^*)$ and $\#C_{2k+1}(H'')$. As $H^* - u^*$ is bipartite, u^* is contained in every copy of C_{2k+1}. Furthermore, each copy of C_{2k+1} uses an even number of edges across the (U_1, U_2) cut and must use an odd number of bad edges. It follows that every C_{2k+1} uses exactly one bad edge $u^* w$ and $\#C_{2k+1}(H^*) = qc(n, C_{2k+1})$.

To bound $\#C_{2k+1}(H'')$, it becomes important to consider the number of copies containing more than one bad edge. Let us first introduce a few configurations (each using three bad edges) that will appear repeatedly in what follows.

For a graph H on partitions of size $|V_1| = n_1$ and $|V_2| = n_2$, and $K(V_1, V_2) \subseteq H$, we have the following types of copies of F that use 3 bad edges:

(i) If $x_0x_1, x_1x_2, x_2x_3 \in H[V_1]$ form a path of length 3, we have $\frac{c(n_1,n_2;F)(n_1-k-1)}{(n_1-2)(n_1-3)(n_2-k+1)}$ copies of F that contain this path.

(ii) If $x_0x_1, x_1x_2, x_3x_4 \in H[V_1]$, there are $(2k-4)\frac{c(n_1,n_2;F)(n_1-k-1)}{(n_1-2)(n_1-3)(n_1-4)(n_2-k+1)}$ copies of F that use these edges.

(iii) If $x_0x_1, x_1x_2 \in H[V_1]$, $y_1y_2 \in H[V_2]$, we have $(2k-2)\frac{c(n_2,n_1;F)(n_1-k)}{n_1(n_1-1)(n_1-2)(n_2-k)}$ copies of F

that use these edges.

(iv) If $x_1x_2, y_1y_2, z_1z_2 \in H[V_1]$ are disjoint edges, there are at most $8\binom{k-3}{2}\frac{c(n_1,n_2;F)(n_1-k-1)}{(n_1-2)(n_1-3)(n_1-4)(n_1-5)(n_2-k+1)}$ copies of F containing this triple.

(v) If $x_1x_2, y_1y_2 \in H[V_1]$ and $z_1z_2 \in H[V_2]$ are disjoint edges, there are at most $8\binom{k-2}{2}\frac{c(n_1,n_2;F)}{n_2(n_2-1)(n_1-2)(n_1-3)}$ copies of F containing this triple.

These values are obtained simply by contracting two edges and counting the number of $C_{2(k-1)+1}$ that contain these 'clusters' and the remaining bad edge. For example, to count copies of Type (ii), we replace the path $x_0x_1x_2$ by a vertex y and form a copy of $C_{2(k-1)+1}$ by picking $(k-1)$ vertices in V_2, $(k-3)$ vertices in $V_1 \setminus \{y, x_3, x_4\}$, picking one of $(k-2)$ positions for y and one of two orientations for the path $x_0x_1x_2$.

We note that these are not the only ways to form copies of F using more than one bad edge. In fact, there may be copies of F using $2j+1$ bad edges for any $j \leq k$. However, if $j \geq 2$, we have stronger upper bounds on the number of such copies and their contribution will be minimal.

H'' contains at most

- $|B(H'')|\left(\frac{2|B(H'')|}{n_1}\right)^2$ paths of length 3,
- $\frac{1}{2}|B(H'')|^2 \cdot 2\left(\frac{2|B(H'')|}{n_1}\right)$ triples of the form in Type (ii) copies,
- $\binom{|B(H'')|}{3}$ disjoint edges as in Type (iv).

Therefore, we may bound $\#C_{2k+1}(H'')$ from above by

$$c(n_1,n_2;F)\left(|B(H'')| + \frac{2k(k+1)}{3}\frac{|B(H'')|^3(1+o(1))}{n_1^4}\right) + \sum_{t=5}^{2k+1}\binom{2k+1}{t}|B(H'')|^t n^{2k+1-2t}$$

As $|B(H'')| \leq 2n_1$, we note that the term $|B(H'')|c(n_1,n_2;F)$ dominates the above sum. However, in most of the arguments that follow, our candidate graph H will contain around $|B(H)|c(n_1,n_2;F)$ copies of C_{2k+1}, where $|B(H)| \approx |B(H'')|$. Therefore, the lower order terms (at most $17k(k+1)c(n_1,n_2;F)/3n_1$ for H'') become more significant when comparing the two values $\#C_{2k+1}(H)$ and $\#C_{2k+1}(H'')$.

Now fix $k \geq 2$ and let $F = C_{2k+1}$. We fix $\epsilon = \epsilon(k) > 0$ small enough and let n be large enough to satisfy (4.8). Let $H \in \mathcal{H}_F(n,q)$, where $q < n$, and let G, B and M be the good, bad and missing edges, respectively, of a max-cut partition $V(H) = V_1 \cup V_2$ with $|V_i| = n_i$ and $|V_1| \geq |V_2|$. Our goal is to show that $M = \emptyset$. The result then follows after applying a stronger version of Claim 4.4.2.

Claim 4.4.5. *There do not exist three disjoint edges in M.*

Proof of Claim. We first observe that

$$\#F(H''(\lceil n/2\rceil, \lfloor n/2\rfloor, q)) \le c(n,F)\left(q + (1+\epsilon)\frac{2k(k+1)q^3}{3\lceil n/2\rceil^4}\right) \le (1+\epsilon)qc(n,F).$$

Therefore, if $H \in \mathcal{H}_F(n,q)$, it follows that $\#F(H) \le (1+\epsilon)qc(n,F)$. However, by (4.8), $F(u'v') \ge (1-\epsilon)c(n,F)$ for all $u'v' \in B = B(H)$ and we have the inequality

$$(1-\epsilon)|B|c(n,F) \le \#F(H) \le (1+\epsilon)qc(n,F)$$

which implies that

$$|B| \le (1+3\epsilon)q < (1+3\epsilon)n. \tag{4.11}$$

However, any missing pair $uv \in M$ must satisfy the inequality $F'(uv) \ge F(u'v')$. So, by (4.2)

$$d_B(u) + d_B(v) > \left(\frac{\alpha_F}{\mu_F} - 2\epsilon\right)n = (1/2 - 2\epsilon)n. \tag{4.12}$$

Therefore, if there exist three disjoint pairs $u_1v_1, u_2v_2, u_3v_3 \in M$,

$$|B| \ge \sum_{i=1}^{3}(d_B(u_i) + d_B(v_i)) - 6 \ge (3/2 - 7\epsilon)n,$$

violating (4.11). ▮

As $|B| = q + \lfloor n^2/4\rfloor - n_1n_2 + |M| = q + \lfloor (n_1-n_2)^2/4\rfloor + |M|$, it follows from (4.11) that $n_1 - n_2 < 4\sqrt{\epsilon n}$ and $|M| \le 3\epsilon n$. We also observe, using the definitions of $c(n_1,n_2;F)$ and $c(n_2,n_1;F)$ that

$$\frac{c(n_1,n_2;F)}{n_2(n_2-1)(n_1-k)} = \frac{c(n_2,n_1;F)}{n_1(n_1-1)(n_2-k)}. \tag{4.13}$$

Claim 4.4.6. *There do not exist a pair of disjoint edges in M.*

Proof of Claim. Assume $w_1w_3, w_2w_4 \in M$ are disjoint with $w_1, w_2 \in V_1$. Let $W = \{w_1, w_2, w_3, w_4\}$ and let $B_W = \{e \in B : e \cap W \ne \emptyset\}$. By (4.11) and (4.12), we have for $(i,j) \in \{(1,3),(2,4)\}$ that

$$(1/2 - 2\epsilon)n \le d_B(w_i) + d_B(w_j) \le (1/2 + 5\epsilon)n.$$

It follows that $(1-5\epsilon)n \le |B_W| \le q$ and $|B \setminus B_W| \le 5\epsilon n$.

Each bad edge $uv \in B[V_1] \setminus B_W$ is adjacent to at most 4 missing edges and is, therefore, contained in at least

$$c(n_1, n_2; F) - \frac{4c(n_1, n_2; F)}{n_2} - 2k|M|\frac{c(n_1, n_2; F)}{n_2(n_2 - 2)} \tag{4.14}$$

copies of F. Similarly, if $uv \in B[V_2] \setminus B_W$, the number of copies of F containing uv is at least

$$c(n_2, n_1; F) - \frac{4c(n_2, n_1; F)}{n_1} - 2k|M|\frac{c(n_2, n_1; F)}{n_1(n_1 - 2)}. \tag{4.15}$$

We note, as a consequence of (4.13), that the quantity in (4.15) is at least as large as the quantity in (4.14). Thus, the number of copies of F using edges in $B \setminus B_W$ is at least

$$c(n_1, n_2; F)(|B \setminus B_W| - (5\epsilon n)(4 + \epsilon)/n_2) > c(n_1, n_2; F)(|B \setminus B_W| - 41\epsilon). \tag{4.16}$$

For $w \in W$, the copies of F using exactly one bad edge incident to w amount to at least

$$d_B(w)c(n_1, n_2; F)\left(1 - \frac{1}{n_2}d_M(w) - \frac{2k}{n_2(n_2 - 2)}(|M| - d_M(w))\right). \tag{4.17}$$

We obtain (4.17) using $w \in V_1$, but, once again, as a consequence of (4.13), we may use the same bound for $w \in V_2$.

Now let $x_i = 2d_B(w_i)/n$. The number of copies of F of Types (i), (ii) and (iii) that use the vertices of W is at least $(1 - 10\epsilon)c(n_1, n_2; F)F_3(x_1, x_2, x_3, x_4)$, where

$$\begin{aligned} F_3(x_1, x_2, x_3, x_4) \;&=\; (x_1 + x_2)\max(x_1 + x_2 - 1, 0) + (x_3 + x_4)\max(x_3 + x_4 - 1, 0) \\ &+\; (k - 2)x_1x_2(x_1 + x_2) + (k - 2)x_3x_4(x_3 + x_4) \\ &+\; (k - 1)(x_1^2 + x_2^2)(x_3 + x_4) + (k - 1)(x_1 + x_2)(x_3^2 + x_4^2). \end{aligned}$$

The number of copies of F using edges in B_W is then at least

$$c(n_1, n_2; F)\left(|B_W| + (1 - 10\epsilon)F_3(x_1, x_2, x_3, x_4) - \sum_{w \in W}\frac{d_B(w)}{n_2}(d_M(w) + 13\epsilon k)\right). \tag{4.18}$$

Observe that, as $V_1 \cup V_2$ is a max-cut partition, $d_B(w) < n_2$ for all $w \in W$. Therefore, adding (4.16) and (4.18), we have

$$\frac{\#F(H)}{c(n_1, n_2; F)} \;\geq\; |B| + (1 - 10\epsilon)F_3(x_1, x_2, x_3, x_4) - 100k\epsilon - \frac{1}{n_2}\sum_{w \in W} d_B(w)d_M(w).$$

We focus on the last term of the above expression and write

$$\begin{aligned}
\frac{1}{n_2}\sum_{w\in W} d_B(w)d_M(w) &= \frac{1}{n_2}\sum_{w\in W}\sum_{wy\in M} d_B(w)\\
&= \sum_{\substack{wy\in M\\ w\in W,\ y\notin W}} \frac{1}{n_2}d_B(w) \ + \sum_{w_iw_j\in M[W]} \frac{1}{n_2}(d_B(w_i)+d_B(w_j))\\
&\leq |M| -| M[W]| + 4\epsilon + \sum_{w_iw_j\in M[W]} (x_i+x_j).
\end{aligned}$$

Optimizing the expression

$$F_3(x_1,x_2,x_3,x_4) + |M[W]| - \sum_{w_iw_j\in M[W]} (x_i+x_j)$$

subject to the constraints $x_1+x_3 = x_2+x_4 = 1$ and $|M[W]| \in \{2,3,4\}$, we obtain a minimum of 3/4 when $|M[W]| = 3, x_1 = x_4 = 3/4$ and $x_2 = x_3 = 1/4$. Therefore,

$$\#F(H) \geq c(n_1,n_2;F)\left(|B| -| M| + 3/4 - 200k\epsilon\right).$$

On the other hand, $\#F(H''(n_1,n_2,q)) \leq c(n_1,n_2;F)(|B(H'')|+\epsilon)$. However, as $|B(H'')| = |B(H)| -| M|$, we have $\#F(H'') < \#F(H)$, contradicting the optimality of H. ▮

It follows that if $M \neq \emptyset$ there exists some vertex u such that $u \in e$ for all $e \in M$.

Claim 4.4.7. *There exists u covering all missing edges with $d_B(u) \geq (1/6-4\epsilon)n$.*

Proof of Claim. Let u cover M and assume $d_B(u) < (1/6-4\epsilon)n$. Then, $d_B(u') \geq (1/3+2\epsilon)n$ for all u' with $uu' \in M$. If $d_M(u) = 1$, u' satisfies the claim. If $d_M(u) \geq 3$, then $d_B(u) + 3(1/3+2\epsilon)n > (1+3\epsilon)n$, violating (4.11). It follows that $|M| = d_M(u) = 2$.

Let $\{v_1,v_2\} = N_M(u)$. If $w \in N_B(v_1)\cap N_B(v_2)$ and $xy \in B\setminus\{v_1w, v_2w, v_1v_2\}$ with $\{x,y\}\cap\{v_1,v_2\} \neq \emptyset$, we may form a path of length 3 using the edges xy, v_1w and v_2w. So, we have at least

$$(d_B(v_1)+d_B(v_2)-3)|N_B(v_1)\cap N_B(v_2)| \geq (d_B(v_1)+d_B(v_2)-3)(d_B(v_1)+d_B(v_2)-n_1) \quad (4.19)$$

paths of length 3 using v_1 and v_2. The number of copies of F lost due to missing edges is at most

$$(2d_B(u)+d_B(v_1)+d_B(v_2))\frac{c(n_1,n_2;F)}{n_2}. \quad (4.20)$$

Here, we use the fact that $\frac{c(n_1,n_2;F)}{n_2} \geq \frac{c(n_2,n_1;F)}{n_1}$, which follows from (4.13), to simplify the sum and ignore the fact that u and v_1, v_2 belong to opposite classes.

We now let $(x, y, z) = 2(d_B(u), d_B(v_1), d_B(v_2))/n$ and maximize the difference of (4.20) and the number of copies of F of Type (i) formed using the paths counted in (4.19). Subject to the constraints $0 \leq x \leq 1/3$ and $x + y, x + z \geq 1$, we obtain a maximum of 14/9 for the expression

$$(2x + y + z) - (y + z)(y + z - 1).$$

As , $|B(H'')| = |B(H)| -| M| = |B| - 2$,

$$\#F(H) \geq (|B| - 15/9)c(n_1, n_2; F) > (|B| - 2 + \epsilon)c(n_1, n_2; F) \geq \#F(H''),$$

contradicting the optimality of H. ❙

Claim 4.4.8. *$d_B(u) \geq (1/2 - 10\epsilon)n$.*

Proof of Claim. Assume $d_B(u) < (1/2 - 10\epsilon)n$. We wish to show that, under this assumption, $\#F(H) > c(n_1, n_2; F)(|B| -| M| + \epsilon)$, exceeding the bound provided by $\#F(H'')$.

Let us first consider the number of potential copies (that is, missing copies) of F that use a pair $uv \in M$. Each bad edge incident to u or v is involved in at most $c(n_1, n_2; F)/n_2$ potential copies. On the other hand, if $e \in B$ and $e \cap \{u, v\} = \emptyset$, there are at most $\frac{2k\, c(n_1,n_2;F)}{n_2(n_2-2)}$ potential copies of F using e. So, we may write

$$\frac{\#F(H)}{c(n_1, n_2; F)} \geq |B| - \sum_{v \in N_M(u)} \left(\frac{d_B(u) + d_B(v)}{n_2} + \frac{2k(|B| - d_B(u) - d_B(v))}{n_2(n_2 - 2)} \right). \tag{4.21}$$

However, the above sum ignores the contribution of copies that use more than one bad edge. For example, if $vw \in B$ where both $v, w \in N_M(u)$ we have at least $(d_B(v)-1)(d_B(w)-2)$ paths of length 3 of the form xv, vw, wy. It follows from (4.12) that $d_B(v), d_B(u) > 8\epsilon n$, so these result in at least $60\epsilon^2 c(n_1, n_2; F)$ copies of F of Type (i).

On the other hand, for each $v \in N_M(u)$, let $B'(v) = N_B(v) \setminus N_M(u)$ and let $d'_B(v) = |B'(v)|$. We associate with the missing edge uv, the Type (iii) copies of F that use bad edges incident to u and v, i.e., copies using triples of the form uw_1, uw_2, vz_1 or uw_1, vz_1, vz_2 where $w_i \in N_B(u)$ and $z_i \in B'(v)$. The number of such copies is at least

$$(2k - 2)\frac{c(n_1, n_2; F)}{n_2^3}\Big(d_B(u)(d_B(u) - 1)d'_B(v) + d_B(u)d'_B(v)(d'_B(v) - 1)\Big).$$

We now rewrite (4.21) by including the extra copies of F counted above. As $60\epsilon^2 > 2/n_2$, we ignore bad edges with both endpoints in $N_M(u)$. Let $x = d_B(u)/n_2$ and $y_v =$

$d'_B(v)/n_2$. Then,

$$\begin{aligned}\frac{\#F(H)}{c(n_1,n_2;F)} &\geq |B| - \sum_{v\in N_M(u)} ((x+y_v+\epsilon)-(2k-2)xy_v(x+y_v)) \\ &\geq |B| -| M|(1-5\epsilon) \\ &> |B| -| M| + 5\epsilon,\end{aligned}$$

as required. Here, the first inequality follows by noting that $2k-2 \geq 2$ and maximizing the expression $(x+y)(1-xy)$ subject to the constraints $x+y \geq 1$, $x \geq 1/3-\epsilon$ and $y \geq 8\epsilon$. ∎

We now analyze the class of graphs that satisfy the conditions proved thus far. Let u be the vertex covering all missing edges and let $d_B(u) = n_2 - t$ where $t < 10\epsilon n$. We separately look at the cases where $u \in V_1$ and $u \in V_2$, paying particular attention to the difference between $c(n_1, n_2; F)$ and $c(n_2, n_1; F)$. Also of importance will be copies of F of Type (i), (ii) and (iii) using the vertex u and we count such copies carefully.

The remainder of the proof involves detailed computation as we account for the various copies of F that contain a specific edge. Loosely speaking, each bad edge contributes $c(n_1, n_2; F)$ copies of F. We then add or subtract fractions of $c(n_1, n_2; F)$ as necessary. For example, if $uv \in M$, each bad edge vw has about $\frac{1}{n_2}c(n_1, n_2; F)$ missing copies. Similarly, if $vw \in B[V_2]$, it is contained in an extra $c(n_2, n_1; F) - c(n_1, n; F) \approx \frac{n_1-n_2}{n_2}c(n_1, n_2; F)$ copies of F.

Once again, we will use the graph $H''(n_1, n_2, q)$ as a reference point. However, the slightly loose bound $\#F(H'') \leq c(n_1, n_2; F)(|B(H'')| + \epsilon)$ is no longer sufficient. We use the stronger $\#F(H'') \leq c(n_1, n_2; F)(|B(H'')| + 17k^2/3n_1)$, instead.

Claim 4.4.9. *If $n_1 > n_2$ or $m > 0$, then $u \notin V_2$.*

Proof of Claim. Assume $u \in V_2$. We first classify the bad edges of H. Let $B(u)$ be bad edges incident to u, and for $0 \leq \ell \leq 2$, let $B_1^\ell = \{vw \in B[V_1] : |\{v, w\} \cap N_M(u)| = \ell\}$ and $B_2^\ell = \{vw \in B[V_2] : |\{v, w\} \cap N_B(u)| = \ell\}$.

Let us count the number of copies of F that contain each type of bad edge.

1. If $uv \in B(u)$, each missing edge uz contributes at most $c(n_2, n_1; F)/n_1$ potential copies of F. Therefore, $F(uv) \geq c(n_2, n_1; F)(1 - \frac{m}{n_1})$.
2. If $vw \in B_1^\ell$, we may form copies of F of Type (iii) using vw and $ux, uy \in B(u)$. We subtract the missing copies using vw and the ℓ missing edges adjacent to vw (on the order of $c(n_1, n_2; F)/n_2$) as well as those using the $m - \ell$ missing edges not adjacent

to vw (each contributing on the order of $c(n_1, n_2; F)/n_2^2$). Thus, $F(vw)$ is at least

$$c(n_1, n_2; F)\left(1 - \frac{\ell(n_1 - m)}{n_2(n_1 - 2)} - \frac{2k(m - \ell)}{n_2(n_1 - 2)} + \frac{(k-1)(n_2 - t)(n_2 - t - 1)(1 - \epsilon)}{n_2(n_2 - 1)(n_2 - 2)}\right). \tag{4.22}$$

3. If $vw \in B_2^\ell$, we may form copies of F of Types (i) and (ii). On the other hand, as vw is not adjacent to any missing edges, the number of potential copies of F containing both vw and a missing edge uz is at most $2kc(n_2, n_1; F)/n_1(n_2 - 2)$. We have the following lower bound on $F(vw)$:

$$c(n_2, n_1; F)\left(1 - \frac{2km}{n_1(n_2 - 2)} + \frac{(k-2)(n_2 - t)(n_2 - t - 1)(1 - \epsilon)}{(n_2 - 2)(n_2 - 3)(n_2 - 4)} + \frac{\ell(n_2 - t - \ell)(1 - \epsilon)}{(n_2 - 2)(n_2 - 3)}\right). \tag{4.23}$$

As $m \le 3\epsilon n$ and $n_1 \ge n/2$, we observe that $2km/n_1(n_2 - 2) < 13k\epsilon/n_2$ (a similar bound holds also for $2km/n_2(n_1 - 2)$). So, if $vw \in B_2$, it follows from (4.23) that $F(vw) > c(n_2, n_1; F)(1 - 13k\epsilon/n_2)$. Also note that $(n_1 - m) < (1 + \epsilon)(n_1 - 2)$ and $(n_2 - t)(n_2 - t - 1) > (1 - 50\epsilon)n_2^2$. Hence, we simplify (4.22) to obtain

$$F(vw) \ge c(n_1, n_2; F)(1 + (k - 1 - \ell)/n_2 - 65k\epsilon/n_2)$$

for $vw \in B_1^\ell$. That is, $F(vw) \ge c(n_1, n_2; F)(1 - 65k\epsilon/n_2)$ unless $k = \ell = 2$. Furthermore, as a consequence of (4.13), we have

$$\frac{c(n_2, n_1; F)}{c(n_1, n_2; F)} > \left(1 + \frac{n_1 - n_2}{n_2} - \frac{k(n_1 - n_2)}{n_2(n_2 - 1)}\right) = \left(\frac{n_1}{n_2} - \frac{k(n_1 - n_2)}{n_2(n_2 - 1)}\right). \tag{4.24}$$

Therefore,

$$\begin{aligned}
\frac{\#F(H)}{c(n_1, n_2; F)} &\ge |B \setminus B(u)|(1 - 65k\epsilon/n_2) - |B_1^2|/n_2 + \frac{c(n_2, n_1; F)}{c(n_1, n_2; F)}|B(u)|(1 - m/n_1) \\
&> |B| - 100k\epsilon - \binom{m}{2}\frac{1}{n_2} + \left(\frac{n_1 - n_2}{n_2} - \frac{m}{n_2}\right)(n_2 - t) \\
&> |B| + (1 - 10\epsilon)(n_1 - n_2) - m^2/2n_2 - m(n_2 - t)/n_2 - 100k\epsilon
\end{aligned}$$

On the other hand, $\#F(H'') \le c(n_1, n_2; F)(|B| - m + 17k(k+1)/n_1)$. So,

$$\begin{aligned}
\frac{\#F(H) - \#F(H'')}{c(n_1, n_2; F)} &> (1 - 10\epsilon)(n_1 - n_2) + m\frac{(n_2 - m/2 - (n_2 - t))}{n_2} - 200k\epsilon \\
&= (1 - 10\epsilon)(n_1 - n_2) + \frac{m(t - m/2)}{n_2} - 200k\epsilon.
\end{aligned}$$

As $t \ge m$, the above difference is positive unless $n_1 = n_2$ and $m = 0$, thereby proving the claim. ▮

Let us now consider the case $u \in V_1$. We modify the definition of B_i^ℓ to reflect this fact and perform analogous computations.

1. If $uv \in B(u)$, then $F(uv) \geq c(n_1, n_2; F)(1 - \frac{m}{n_2})$.
2. If $vw \in B_1^\ell$, it is not adjacent to any missing edges and we may form copies of F of Types (i) and (ii). Thus, $F(vw)$ is at least
$$c(n_1, n_2; F)\left(1 - \frac{2km}{n_2(n_1-2)} + \frac{(k-2)(n_2-t)(n_2-t-1)}{(n_1-2)(n_1-3)(n_1-4)} + \frac{\ell(n_2-t-\ell)}{(n_1-2)(n_1-3)}\right). \tag{4.25}$$
3. If $vw \in B_2^\ell$, we have potential copies due to missing edges adjacent to vw as well as those not adjacent to vw. We also form copies of F of Type (iii). So, $F(vw)$ is at least
$$\begin{aligned} & c(n_2, n_1; F)\left(1 - \frac{\ell(n_2-m)}{n_1(n_2-2)} - \frac{2k(m-\ell)}{n_1(n_2-2)} + \frac{(k-1)(n_2-t)(n_2-t-1)}{n_1(n_1-1)(n_1-2)}\right) \\ \geq \quad & c(n_1, n_2; F)\left(1 + \frac{n_1-n_2}{n_2} - \frac{\ell(n_2-m)}{n_2(n_2-2)} - \frac{2k(m-\ell)}{n_2(n_2-2)} + \frac{(k-1)(1-50\epsilon)}{n_2}\right) \end{aligned} \tag{4.26}$$

where we use (4.13) and (4.24).

Putting these together, we have the following lower bound on $\#F(H)/c(n_1, n_2; F)$:
$$|B| - \frac{1}{n_2}\left(m(n_2-t) + \frac{2km|B-B(u)|}{n_2-2} + \sum_{\ell=0}^{2}|B_2^\ell|(n_1-n_2-\ell+(k-1)(1-50\epsilon))\right).$$
In particular, if $q \leq n/2-2$, we can show $\#F(H) > c(n_1, n_2; F)(|B|-|M|)$ by noting that $|B - B(u)| < 13\epsilon n$ and $|B_2^2| \leq m(m-1)/2$. That is, as $k \geq 2$,
$$\frac{\#F(H)}{c(n_1, n_2; F)} - (|B|-m) \geq \frac{1}{n_2}\Big(mn_2 - m(n_2-t) - 26\epsilon km - (1+50\epsilon k)m(m-1)/2\Big) > 0.$$
It follows that $\#F(H) > (q + \lfloor (n_1-n_2)^2/4 \rfloor)c(n_1, n_2; F)$. However, $\#F(H^*) = qc(n, F)$ and for $n_1 > n_2$, $(q + \lfloor (n_1-n_2)^2/4 \rfloor)c(n_1, n_2; F) > qc(n, F)$, thereby proving the first part of the lemma.

So we may assume $n_1 + n_2$ is odd and $n/2 - 2 < q \leq n - 5k$. We first prove that m, t and $n_1 - n_2$ are all bounded by constants.

Claim 4.4.10. *$m \leq 10k$ and $t \leq 20k^2$.*

Proof of Claim. As a result of (4.25), edges in B_1 are contained in at least
$$|B_1|(c(n_1, n_2; F) - 2km/n_2(n_1-2)) > |B_1|c(n_1, n_2; F) - 3km/n_2$$

copies of F.

On the other hand, if $vw \in B_2^\ell$, by (4.26) and the fact that $n_1 \geq n_2 + 1$ we have $F(vw) > c(n_1, n_2; F)(1 + (1-\ell)/n_2)$. In particular, $F(vw) > c(n_1, n_2; F)$ if $vw \in B_2^0 \cup B_2^1$, and $F(vw) > c(n_1, n_2; F)(1 - 1/n_2)$ if $vw \in B_2^2$. Note that $|B_2^2| \leq \binom{m}{2}$.

We now compare $\#F(H)$ and $\#F(H'')$. Then,

$$\begin{aligned}\frac{\#F(H) - \#F(H'')}{c(n_1, n_2; F)} &\geq \left(|B| - \frac{1}{n_2}\left(m(n_2 - t) - 3km - \binom{m}{2}\right)\right) - \left(|B| - m + 10k^2/n_2\right) \\ &\geq \frac{1}{n_2}\left(mt - m^2/2 - 3km - 10k^2\right),\end{aligned}$$

and H fails to be optimal unless $m \leq 10k$ and $t \leq 20k^2$. ❙

It follows that $|B_2^2| < 50k^2$ and $|B_1^0| \leq \binom{t+n_1-n_2}{2} < 10\epsilon n$. We now strengthen the bounds obtained from (4.25) and (4.26) and show that $|B \setminus B(u)| < 12\epsilon n$. First, as $t \leq 20k^2$, for $vw \in B_1^1 \cup B_1^2$, we may strengthen $F(vw)$ to $c(n_1, n_2; F)(1 + (1-\epsilon)(\ell + k - 2)/n_2)$. Similarly, if $vw \in B_2^0 \cup B_2^1$, we have $F(vw) \geq c(n_1, n_2; F)(1 + (1-\epsilon)(k-1)/n_2)$. Therefore, if $|B_1^1 \cup B_1^2 \cup B_2^0 \cup B_2^1| > 200k$, we will gain an extra $(1-\epsilon)200k(k-1)/n_2$ copies of F, thereby surpassing the deficit of $(10k^2 + 30k^2 + 50k^2)/n_2$ obtained when $m = 10k$. It follows that, $|B \setminus B(u)| < 12\epsilon n$ and $|B| < (1/2 + 15\epsilon)n$.

Claim 4.4.11. $n_1 = n_2 + 1$.

Proof of Claim. Assume $n_1 > n_2 + 1$. As n is odd, we have $n_1 \geq n_2 + 3$. We show that $\#F(H)$ may be decreased by moving a vertex from V_1 to V_2. In particular, let $v \in \widetilde{V}_1 = V_1 \setminus (N_B(u) \cup \{u\})$. We delete v and replace it by v' where v' is connected to all $w \in V_1 \setminus \{v\}$ but not to any vertex in V_2. As $d(v') - d(v) = (n_1 - 1) - (n_2 + d_B(v))$, if $d_B(v) \geq n_1 - n_2 - 1$, we distribute the remaining $d_B(v) - (n_1 - n_2 - 1)$ edges among vertices in V_2. Otherwise, if $d_B(v) < (n_1 - n_2 - 1)$, we delete $(n_1 - n_2 - 1) - d_B(v)$ bad edges. This operation reduces $|B|$ by $n_1 - n_2 - 1$. As $F(vw) \geq c(n_1, n_2; F)(1 - 10k/n_2)$ for all $vw \in B$, we reduce $\#F(H)$ by $(n_1 - n_2 - 1)c(n_1, n_2; F)(1 - 10k/n_2)$.

On the other hand, each remaining bad edge gains at most

$$c(n_1 - 1, n_2 + 1; F) - c(n_1, n_2; F) \ll c(n_1, n_2; F)/n_2$$

copies of F that use v'. Furthermore, each of the $d_B(v) - (n_1 - n_2 - 1)$ edges placed in V_2 gain $c(n_2 + 1, n_1 - 1; F) - c(n_1, n_2; F) \leq c(n_1, n_2; F)(1 + n_1 - n_2)/n_2$ copies. These edges also gain at most $(k-1)c(n_1, n_2; F)/n_2$ copies of Types (iii). However, as $|B_1^1| < 200k$, we have $d_B(v) - (n_1 - n_2 - 1) < t + 200k \leq 120k^2$. The net gain is, therefore, at most

$$\begin{aligned}
& \frac{c(n_1, n_2; F)}{n_2}(|B| + 120k^2(k + n_1 - n_2) - (n_1 - n_2 - 1)(n_2 - 10k)) \\
< \; & \frac{c(n_1, n_2; F)}{n_2}((1/2 + 15\epsilon)n + 120k^2(k + n_1 - n_2) - 2(n_2 - 10k)) \\
< \; & \frac{c(n_1, n_2; F)}{n_2}(120k^2(k + n_1 - n_2) - n_2/2) \\
< \; & -\frac{1}{4}c(n_1, n_2; F),
\end{aligned}$$

contradicting the optimality of H. ∎

Claim 4.4.12. $|B_2^1| \geq m$.

Proof of Claim. Assume $B_2^1 < m$. Then, there exists a vertex $v \in N_M(u)$ with $d_B(v) \leq m-1$. As $q > n_2-2$, we have $|B \backslash B(u)| \neq 0$. We alter H by removing an edge $xy \in B \backslash B(u)$ and adding the edge uv. This procedure deletes at least $F(xy) \geq c(n_1, n_2; F)(1 - \epsilon/n_2)$ copies of F that use the edge xy. On the other hand, for each edge $uw \in N_B(u)$ and $vx' \in B_2$, we obtain at most $c(n_1, n_2; F)/n_2$ extra copies of F using the edge uv. In addition, bad edges disjoint from uv gain at most $2kc(n_1, n_2; F)/n_2(n_2 - 2)$ copies of F. As $t \geq m$, these amount to at most

$$\frac{c(n_1, n_2; F)}{n_2}\left(n_2 - t + (m-1) + \frac{2k|B \setminus B(u)|}{n_2 - 2}\right) < c(n_1, n_2; F)(1 - (1 - 50\epsilon k)/n_2).$$

Therefore, the alteration reduces the number of copies of F, contradicting the optimality of H. ∎

Claim 4.4.13. $t \geq m + k$.

Proof of Claim. As $|\widetilde{V}_1| = n_1 - d_B(u) - 1 = t$ and $vw \in B_1^0$ implies $v, w \in \widetilde{V}_1$, if $t \leq m+k-1$, then $|B_1^0 \cap B(v)| \leq m+k-2$ for all $v \in \widetilde{V}_1$. Let $xy \in B_2^1$, $v \in \widetilde{V}_1$, and $w \in N_M(u)$. Consider removing the edge xy and all edges $vz \in B$ and replace them with the edge uv and $d_B(v)$ bad edges incident to w. As

$$\begin{aligned}
F(xy) - F(uv) \; &\geq \; c(n_1, n_2; F)(1 + (k - 1 - \epsilon)/n_2) - c(n_1, n_2; F)(1 - (m - \epsilon)/n_2) \text{ and} \\
F(vz) - F(wz') \; &\geq \; c(n_1, n_2; F)(1 + (k - 2 - \ell - \epsilon)/n_2) - c(n_1, n_2; F)(1 + (k - 1 + \epsilon)/n_2),
\end{aligned}$$

for $vz \in B_1^\ell$, the net loss is at least

$$\frac{c(n_1, n_2; F)}{n_2}(k - 1 + m - |B_1^0 \cap B(v)| - 2\epsilon d_B(v)) > 0,$$

contradicting the optimality of H. ∎

Claim 4.4.14. $H \notin \mathcal{H}_F(n,q)$.

Proof of Claim. Pick $uv \in B(u)$ with $d_B(v) = 1$. Consider the following alterations:

1. If $|B_2^1| \geq t-1$, remove the edge uv as well as $t-1$ edges in B_2^1 and include the edge vw for each $w \in \widetilde{V}_1$.
2. If $|B_2^1| = p < t-1$, remove the edge uv and all edges in B_2^1 and add p edges vw with $w \in \widetilde{V}_1$ as well as an edge ux for some $x \in N_M(u)$.

In the first instance, the t edges we created in B_1^0 each give at most $c(n_1,n_2;F)(1 + (k-2+\epsilon)/n_2)$ copies of F. On the other hand, as $F(uv) \geq c(n_1,n_2;F)(1-m/n_2)$ and $F(xy) \geq c(n_1,n_2;F)(1+(k-1-\epsilon)/n_2)$ for the $t-1$ edges removed from B_2^1, we lose at least

$$\frac{c(n_1,n_2;F)}{n_2}\Big(m-(t-1)(k-1-\epsilon)+t(k-2+\epsilon)\Big) > \frac{c(n_1,n_2;F)}{n_2}\Big(t-m-k+1-2t\epsilon\Big) > 0$$

copies.

In the second case, we delete $c(n_1,n_2;F)((p+1)+(p(k-1)-m)/n_2)$ copies of F. The number of copies added is at most

$$pc(n_1,n_2;F) + (d_B(u) - 1 + d'_B(x) + p(k-2))c(n_1,n_2;F)/n_2,$$

where $d'_B(x)$ counts the number of remaining bad edges of x. However, as $|B_2^1| = 0$ after the alteration, $d'_B(x) \leq m-1$. Furthermore, by Claim 4.4.12, $p \geq m$. So, the net loss in the number of copies of F is at least

$$\frac{c(n_1,n_2;F)}{n_2}(n_2+p-m-(n_2-t-1)-(m-1)) \geq \frac{c(n_1,n_2;F)}{n_2}(t+p+2-2m) > 0,$$

and we reduce the number of copies of F.

Therefore, $M(H) \neq \emptyset$ implies that H is not optimal. ∎

The result follows by comparing the various ways to partition n. In particular, it is readily checked that $c(\lceil n/2\rceil, \lfloor n/2\rfloor; F)(q + 2(17k^2/3)/n)$ is less than $(q+2)c(\lceil n/2\rceil + 1, \lfloor n/2\rfloor - 1; F)$ for $q < n - 10k^2$.

□

4.5 Special Graphs

In this section we obtain upper bounds on $c_{1,i}(F)$ for a class graphs and compute the exact value for some special instances. We also give an example of a graph with $c_1(F)$ strictly greater than $\min(\pi_F, \theta_F)$.

4.5.1 Pair-free graphs

Definition 4.5.1. *Let F be an r-critical graph. Say F is* pair-free *if for any two (different, but not necessarily disjoint) edges u_1v_1, u_2v_2, there is no proper r-coloring χ of $F - u_1v_1 - u_2v_2$ where $\chi(u_1) = \chi(u_2) = \chi(v_1) = \chi(v_2)$.*

Alternatively, let $H \in \mathcal{T}_r^2(n)$ be formed by including two edges in one part. Then, the class of pair-free graphs consists of r-critical graphs whose copies in H use exactly one of the two bad edges. Many interesting graphs belong to this class. For instance, it is easily seen that complete graphs K_t are pair-free and, as argued in the previous section, no copy of the odd cycle C_{2k+1} uses exactly two bad edges. In addition, graphs obtained from the complete bipartite graph $K_{s,t}$ by adding an edge to the part of size s are pair-free if $t \geq 3$.

Theorem 4.5.2. *Let F be pair-free and let $t = \operatorname{sgn}(\zeta_F)$. Then $c_{1,t}(F) \leq 2\pi_F$ and $c_{1,i}(F) \leq \pi_F$ for $i \not\equiv t \pmod r$.*

Proof. Let n be large and $q = (\pi_F + \epsilon)n$ for some $\epsilon > 0$. We prove the case $n \not\equiv t \pmod r$; the other case follows in a similar manner. Write $n = n_1 + \ldots + n_r$, where $c(n, F) = c(n_1, \ldots, n_r; F)$ and consider the partition $n = n_1' + n_2' + \ldots + n_r'$ where $n_1' = n_1 + t$, $n_2' = n_2 - t$ and $n_i' = n_i$ for $i = 3, \ldots r$. Construct H' as follows: $H' \supseteq K(V_1', \ldots, V_r')$ with $|V_i'| = n_i'$. Next place $q + 1$ bad edges in V_1' to form a regular (or almost regular) bipartite graph. We claim that $\#F(H') < \#F(H)$ for any $H \in \mathcal{T}_r^q(n)$.

First of all, each bad edge in H' is contained in at most $c(n, F) - |\zeta_F| n^{f-3} + O(n^{f-4})$ copies of F that contain only one bad edge. As F is pair-free, no copy of F contains exactly two bad edges. In addition, we may bound the number of copies of F using at least three bad edges by $Cq^3 n^{f-6} = C' n^{f-3}$ for some constants C, C'.

On the other hand, $\#F(H) \geq qc(n, F)$. Therefore,

$$\begin{aligned}
\#F(H') - \#F(H) &\leq (q+1)\big(c(n, F) - |\zeta_F| n^{f-3} + O(n^{f-4})\big) + C' n^{f-3} - qc(n, F) \\
&< \alpha_F n^{f-2} - (\pi_F + \epsilon) n |\zeta_F| n^{f-3} - C'' n^{f-3} \\
&< -\epsilon |\zeta_F| n^{f-2}/2,
\end{aligned}$$

proving the theorem. □

In particular, as Theorem 4.3.1 implies that $c_{1,i}(F) \leq \hat{\rho}_F - (1 - 1/r)$, if F is pair-free and $\rho_F = \hat{\rho}_F$, we have the exact value of $c_{1,i}(F)$. This is the case for $F = K_{s,t}^+$.

Lemma 4.5.3. *Let $s, t \geq 2$ and $F = K_{s,t}^{+}$ be obtained from the complete bipartite graph $K_{s,t}$ by adding an edge to the part of size s. Then $\theta_F = \hat{\rho}_F - (1 - 1/r)$.*

Proof. Clearly, $F = K_{s,t}^{+}$ is 2-critical and

$$c(n_1, n_2; K_{s,t}^{+}) = \binom{n_2}{t}\binom{n_1 - 2}{s - 2}.$$

It readily follows that

$$\alpha_F = \frac{2^{-(t+s-2)}}{t!(s-2)!}, \quad \zeta_F = (t - s + 2)\frac{2^{-(t+s-3)}}{t!(s-2)!}, \text{ and}$$

$$\pi_F = \begin{cases} \infty & \text{if } t = s - 2 \\ (2(t - s + 2))^{-1} & \text{otherwise.} \end{cases}$$

On the other hand,

$$P_F(\boldsymbol{\xi}) = \frac{2^{-s+2}}{t!(s-2)!}(\xi_1\xi_2^t + \xi_1^t\xi_2).$$

As $P_F(c \cdot \boldsymbol{\xi}) = c^{t+1}P_F(\boldsymbol{\xi})$ for all $\boldsymbol{\xi} \in \mathbb{R}_+^2$ and $c \in \mathbb{R}_+$, we consider the behavior of $P_F(\boldsymbol{\xi})$ where $\xi_1 + \xi_2 = 1$. Let $\xi_1 = 1/2 + y$, $\xi_2 = 1/2 - y$ and express

$$P_F(\boldsymbol{\xi}) = p_t(y) = \frac{2^{-s+2}}{t!(s-2)!}\Big((1/2 + y)(1/2 - y)^t + (1/2 + y)^t(1/2 - y)\Big).$$

We observe that $p_t(y)$ is an even function with $p_t(1/2) = p_t(-1/2) = 0$ and $p_t(0) = \alpha_F$. Now consider the coefficient s_k of y^k in

$$t!(s-2)!2^{s-2}p_t'(y) = (1/2-y)^t - t(1/2+y)(1/2-y)^{t-1} + t(1/2+y)^{t-1}(1/2-y) - (1/2+y)^t.$$

We obtain the following values from each term above:

1. $(-1)^k\binom{t}{k}2^{-t+k}$
2. $-t/2 \cdot (-1)^k\binom{t-1}{k}2^{-t+k+1} - t \cdot (-1)^{k-1}\binom{t-1}{k-1}2^{-t+k}$
3. $t/2 \cdot \binom{t-1}{k}2^{-t+k+1} - t \cdot \binom{t-1}{k-1}2^{-t+k}$
4. $-\binom{t}{k}2^{-t+k}$.

Combining the above, we have

$$\begin{aligned} s_k &= 2^{-t+k}\left[(-1)^k - 1\right]\left[\binom{t}{k} - t\binom{t-1}{k} + t\binom{t-1}{k-1}\right] \\ &= 2^{-t+k}\left[(-1)^k - 1\right]\left[(k+1)\binom{t}{k} - t\binom{t-1}{k}\right] \\ &= 2^{-t+k}\left[(-1)^k - 1\right](2k + 1 - t)\binom{t}{k} \end{aligned}$$

It follows that $s_k \in \begin{cases} \mathbb{Z}_- & \text{if } k > (t-1)/2 \text{ and } k \text{ is odd} \\ \{0\} & \text{if } k = (t-1)/2 \text{ or } k \text{ is even} \\ \mathbb{Z}_+ & \text{if } k < (t-1)/2 \text{ and } k \text{ is odd.} \end{cases}$

That is, for $t \geq 4$, the coefficients of $p'(t)$ change sign exactly once. By Descartes' Rule of Signs [Des37], $p_t'(y)$ has exactly one positive root and, by symmetry, exactly one negative root. As $p_t''(0) > 0$ for $t \geq 4$, it follows that $(0, \alpha_F)$ is the unique local minimum for p_t with the two roots of p_t' providing local maxima.

In addition, if $t = 2, 3$, $p_t'(y)$ is a decreasing odd polynomial. So, $(0, \alpha)$ is the unique maximum point of $p_t(y)$ and no other local maxima or minima exist. It follows that $\theta_F^* = \theta_F = \infty$ in these two cases.

If $t \geq 4$, we may solve for ρ_F^* and ρ_F as the roots of certain polynomial equations. On one hand, if $\boldsymbol{\xi} \in \mathcal{S}_\rho$ with $\xi_1 = \xi_2 = \rho/2$, we have $P_F(\boldsymbol{\xi}) = \alpha_F \rho^{t+1}$. On the other hand, if $\boldsymbol{\xi} \in \mathcal{S}^* \cap \mathcal{S}_\rho$ with $\xi_1 = \rho - 1/2$ and $\xi_2 = 1/2$, $P_F(\boldsymbol{\xi}) = \alpha_F(\rho - 1/2) + \alpha_F 2^{t-1}(\rho - 1/2)^t$. Hence, ρ_F^* is the smallest root (that is at least 1/2) of $\rho^{t+1} = (\rho - 1/2) + 2^{t-1}(\rho - 1/2)^t$. Similarly, ρ_F is the smallest root of $\rho^{t+1} = \rho - 1/2$. We observe that $2^{-t-1} < \theta_F < 2^{-t}$ for $t \geq 4$.

Now, if $\theta_F \neq \hat{\rho}_F - (1 - 1/r)$, then the two curves $\alpha_F \rho^{t+1}$ and $\alpha_F(\rho - (1 - 1/2))$ must be tangent at ρ_F. Therefore, ρ_F is not only a root of $g_1(\rho) = \rho^{t+1} - \rho + 1/2$, but also of its derivative $g_1'(\rho) = (t+1)\rho^t - 1$. However, as $(t+1) < (5/3)^t$ and $2^{-t} < 1/10$ for $t \geq 4$, $\rho_F = (t+1)^{-1/t} > .6 > (.5 + 2^{-t})$, resulting in a contradiction. Hence, $\theta_F = \hat{\rho}_F - (1 - 1/r)$. $\square$

As both π_F and θ_F have been determined for $F = K_{s,t}^+$, Theorems 4.4.1 and 4.5.2 give us the exact value of $c_{1,i}(F)$.

Theorem 4.5.4. *Let $s, t \geq 2$ and $F = K_{s,t}^+$. Then $c_1(F) = c_{1,0}(F) = \min(\pi_F, \theta_F)$ and $c_{1,1}(F) = \min(2\pi_F, \theta_F)$.*

4.5.2 Non-tightness of Theorem 4.4.1

We now exhibit a graph for which $c_1(F) > \min(\pi_F, \theta_F)$. Let F be the graph in Figure 4.2. Interestingly, for this graph, $\rho_F = \hat{\rho}_F = \infty$, so we only show that $c_1(F) > \pi_F$. This non-tightness occurs as a consequence of Claim 4.4.2. Specifically, if $q < (\pi_F - \epsilon)n$ and the max-cut partition of a graph H has parts of size $n = n_1 + n_2$ where $n_1 \geq n_2 + 2$, Claim 4.4.2 implies that all bad edges lie in the same part. However, if $|n_1 - n_2| \leq 1$, we reduce the number of copies of F by distributing the bad edges among the two parts.

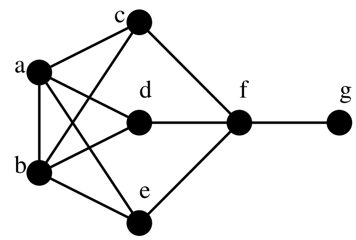

Figure 4.2: Example for non-tightness of Theorem 4.4.1.

Note that F is 2-critical and ab is the unique critical edge. There is a unique (up to isomorphism) 2-coloring χ of $F - ab$ with $\chi^{-1}(1) = \{a, b, f\}$ and $\chi^{-1}(2) = \{c, d, e, g\}$. It readily follows that

$$c(n_1, n_2; F) = \binom{n_2}{3}(n_1 - 2)(n_2 - 3)$$

and $\alpha_F = (3! \cdot 2^5)^{-1}$. Taking derivatives, we observe that $\zeta_F = 2^{-5}$ and $\pi_F = 1/6$.

We also have

$$P_F(\boldsymbol{\xi}) = \frac{1}{4 \cdot 3!}(\xi_1\xi_2^3 + \xi_1^3\xi_2),$$

which, if we fix $\xi_1 + \xi_2$, is minimized by maximizing the difference. Hence, $\theta_F = \infty$.

However, note that F is not pair-free as there exists a 2-coloring χ^* of $F - ab - fg$ with $\chi^*(a) = \chi^*(b) = \chi^*(f) = \chi^*(g) = 1$. In fact, if u_1v_1, u_2v_2 are two distinct edges in F, there is no 2-coloring χ'' of $F - u_1v_1 - u_2v_2$ with $\chi''(u_1) = \chi''(v_1)$ and $\chi''(u_2) = \chi''(v_2)$ unless $\{u_1v_1, u_2v_2\} = \{ab, fg\}$ and χ'' is isomorphic to χ^*. That is, for any $H \in \mathcal{T}_r^q(n)$, the only copies of F in H that use exactly two bad edges correspond to χ^*.

We recall the set $X = X(H)$ of vertices with large bad degree. As $\theta_F = \infty$, (4.10) implies that both $X(H)$ and $M(H)$ are empty for any $H \in \mathcal{H}_F(n, q)$. That is, if $\epsilon > 0$ is fixed and if n is large enough, then $d_B(x) < \epsilon n$ for any $x \in V(H)$. In addition, by Claim 4.4.2, if $H \supseteq K(V_1, V_2)$ with $|V_1| \geq| V_2|$, then

$$|B_1| \geq |B_1| -| B_2| \geq (1 - \epsilon)(|V_1| -| V_2| - 1)\pi_F n.$$

In particular, if n is even and $|V_1| \geq |V_2| + 2$, we have at least $(\pi_F n)^2(1 - 2\epsilon)/2$ disjoint pairs of edges in B_1, each of which form $4\binom{|V_2|}{3}$ copies of F. Therefore,

$$\#F(H) \geq (q + 1)(c(n, F) - n^4/2^5) + \frac{1}{864}n^5 - \epsilon n^5. \qquad (4.27)$$

On the other hand, if $H^* \in \mathcal{T}_2^q(n)$, we may place $q/2$ edges in each of B_1 and B_2, thereby forming at most $q^2/4$ pairs of bad edges that lie in the same part. Thus,

$$\#F(H^*) \le qc(n,F) + q^2\frac{n^3}{48} + \epsilon n^5. \tag{4.28}$$

Comparing the above quantities, and solving the resulting quadratic inequality, we see that $c_{1,0}(F) \ge \frac{5\sqrt{5}-9}{12} > 1/6 = \pi_F$. In fact, a careful analysis will show

Theorem 4.5.5. $c_{1,0}(F) = \frac{3-\sqrt{5}}{4}$.

Proof. We prove the lower bound by showing that if $|V_1| \ge |V_2| + 2$ and $q < n/5$, then $B_2 = \emptyset$. As a result of Claim 4.4.2 we may initially assume that $|B_1| \ge (1-\epsilon)n/6$. Now, if $B_2 \ne \emptyset$, an edge $uv \in B_2$ is contained in at least $c(n_2, n_1; F) > c(n_1, n_2; F) + 2(\zeta_F - \epsilon)n^4$ copies of F. However, if we remove uv and replace it with an edge xy where $x, y \in V_1$ have $d_B(x), d_B(y) < 3$, we form at most $c(n_1, n_2; F) + qn^3/12 + \epsilon n^4$ copies of F. As

$$2\zeta_F n^4 = 2^{-4}n^4 > n^4/60 \ge qn^3/12,$$

this alteration reduces the number of copies of F. So, $\#F(H)$ is minimized by making $B_2 = \emptyset$. Therefore, we have $q^2(1-\epsilon)/2$ pairs of disjoint bad edges in B and we improve (4.27) to

$$\#F(H) \ge (q+1)(c(n,F) - n^4/2^5) + q^2\frac{n^3}{24} - \epsilon n^5. \tag{4.29}$$

The root of the resulting quadratic is now $\frac{3-\sqrt{5}}{4}$.

The upper bound follows by noting that inequalities (4.28) and (4.29) may be changed to equations by replacing the last term with $\pm\epsilon n^5$. $\square$

4.5.3 $K_{r+2} - e$.

Let $r \ge 2$ and let $F = K_{r+2} - e$ be obtained from the complete graph K_{r+2} by deleting one edge. We obtain the following exact result for $c_1(F)$.

Theorem 4.5.6. *If $r \ge 2$ and $F = K_{r+2} - e$, then $c_1(F) = \pi_F = \frac{r-1}{r^2}$.*

By definition, $\chi(F) = r+1$. In addition, if uv is the edge removed from K_{r+2}, we may further reduce the chromatic number by removing an edge xy where $\{x, y\} \cap \{u, v\} = \emptyset$. It follows that F is r-critical and

$$c(n_1, \ldots, n_r; F) = \sum_{i=2}^{r} \binom{n_i}{2} \prod_{\substack{2\le j\le r \\ j\ne i}} n_j = \frac{(n - n_1 - r + 1)}{2} \prod_{i=2}^{r} n_i.$$

Therefore, $\alpha_F = \frac{r-1}{2r^r}$, $\zeta_F = \frac{1}{2r^{r-2}}$, and $\pi_F = \frac{r-1}{r^2}$.

On the other hand,

$$P_F(\boldsymbol{\xi}) = \frac{1}{2}\sum_{i=1}^{r} \xi_i^2 \prod_{\substack{1\le j\le r\\ j\neq i}} \xi_j = \frac{1}{2}\left(\sum_{i=1}^{r}\xi_i\right)\prod_{i=1}^{r}\xi_i.$$

Therefore, if $\sum_i \xi_i = \rho$ is fixed, by convexity, $P_F(\boldsymbol{\xi})$ is minimized by picking $\boldsymbol{\xi} \in \mathcal{S}^*$, implying that $\theta_F = \infty$.

Theorem 4.4.1 now implies that $c_1(F) \geq \pi_F$, so we only prove the upper bound $c_1(F) \leq \pi_F$.

As $M(H) = \emptyset$ for $H \in \mathcal{H}_F(n, q)$, we compare graphs obtained from a complete r-partite graph by adding extra edges. First, given $H \in \mathcal{T}_r^q(n)$, let us compute the number of copies of F formed by pairs of bad edges. Let $u_1v_1 \in B_i$ and $u_2v_2 \in B_j$, where $r \mid n$ and $|V_k| = n/r$ for all $k \in [r]$.

1. If $i = j$ and $\{u_1, v_1\} \cap \{u_2, v_2\} = \emptyset$, no copy of F contains both bad edges as any 4 vertices in F span at least 5 edges and only 2 edges are present among $\{u_1, u_2, v_1, v_2\}$.
2. If $i = j$ and $|\{u_1, v_1\} \cap \{u_2, v_2\}| = 1$, we may create a copy of F by picking one vertex each from V_k where $k \neq i$. Therefore, we have $(n/r)^{r-1}$ copies of F containing both bad edges.
3. If $i \neq j$, we form a copy of K_{r+2} by picking a vertex from each of the parts V_k where $k \notin \{i, j\}$. We may then choose any of the $\binom{r+2}{2} - 2$ edges (except for u_1v_1 and u_2v_2) to be the one missing in F. In addition, for any choice of $k_1, k_2 \notin \{i, j\}$, we may pick 2 vertices from V_{k_1}, no vertices from V_{k_2} and one vertex each from V_l where $l \notin \{i, j, k_1, k_2\}$, to form a copy of F. So, the number of copies of F containing both edges is
$$\left(\binom{r+2}{2} - 2\right)(n/r)^{r-2} + (r-2)\binom{n/r}{2}(r-3)(n/r)^{r-4}.$$

For $q \leq n/r$, let $H^* \in \mathcal{T}_r^q(n)$ be formed by placing all q bad edges in V_1. In particular, consider enumerating the bad edges as $e_1, e_2, \ldots, e_q$ and the vertices in V_1 as $v_1, v_2, \ldots, v_{n/r}$. Then, construct H^* such that $e_i = v_{2i-1}v_{2i}$ for $i \leq n/(2r)$, and $e_{n/2r+j} = v_{2j}v_{2j+1}$ for $1 \leq j \leq q - n/(2r)$.

Claim 4.5.7. *If $q \leq n/r$, $\#F(H^*) \leq \#F(H)$ for all $H \in \mathcal{T}_r^q(n)$.*

Proof of Claim. If $q \leq n/(2r)$, $\#F(H^*) = qc(n, F)$, which is a trivial lower bound for all $H \in \mathcal{T}_r^q(n)$. So we consider the case $n/(2r) \leq q \leq n/r$.

Let H be the minimizer of $\#F(H)$ over the set $\mathcal{T}_r^q(n)$. Assume, without loss of generality, that $|B_1| \geq| B_i|$ for al $i \in [r]$. If $|B_1| < n/(2r)$, then $B \setminus B_1 \neq \emptyset$. Say $B_2 \neq \emptyset$, and consider removing an edge in B_2 and replacing it by adding an extra edge to B_1. We may place this edge so that no new vertices of degree 2 are created. Then, the number of copies of F is reduced by at least $(|B_1| -| B_2| + 1)(n/r)^{r-2}(\binom{r+2}{2} - 2)$, contradicting optimality of H.

On the other hand, if $|B_1| \geq n/(2r)$, then every edge $uv \in B \setminus B_1$ forms at least

$$c(n, F) + \left(\binom{r+2}{2} - 2\right)(n/r)^{r-2}|B_1| \geq c(n, F) + 2(n/r)^{r-1}$$

copies of F.

So we may assume that $B = B_1$. Now, by convexity, $\sum_{v\in V_1}\binom{d_B(v)}{2}$ is minimized when there are exactly $2|B_1| - n/r$ vertices of degree 2 and all remaining vertices have degree 1. However, each vertex of degree 2 gives $(n/r)^{r-1}$ copies of F that use both edges incident to it. It follows that

$$\begin{aligned}
\#F(H) &\geq qc(n, F) + 2(n/r)^{r-1}\left(|B \setminus B_1| + |B_1| - n/(2r)\right) \\
&= qc(n, F) + (2q - n/r)(n/r)^{r-1} \\
&= \#F(H^*). \ \blacksquare
\end{aligned}$$

Now consider a graph H on partition $n = n_1 + n_2 + \ldots + n_r$ where $n_1 = n/r + 1$, $n_2 = n/r - 1$ and $n_i = n/r$ for $i \geq 3$ with $K(V_1, \ldots, V_r) \subseteq H$ and all $q + 1$ bad edges contained in V_1 as in H^*. Then

$$\#F(H) \leq (q + 1)(c(n, F) - \zeta_F n^{r-1}) + (2q + 2 - n/r)(n/r)^{r-1} + \epsilon n^{r-1}.$$

In particular, if $q \geq (\pi_F + \epsilon)n$, then

$$\#F(H) - \#F(H^*) \leq \alpha_F n^r - (\pi_F + \epsilon)\zeta_F n^r + n^{r-1/2} < -\epsilon\zeta_F n^{r-1}/2,$$

thus proving the upper bound $c_{1,0}(F) \leq \pi_F$.

Bibliography

[ABKS04] N. Alon, J. Balogh, P. Keevash, and B. Sudakov, *The number of edge colorings with no monochromatic cliques*, J. London Math. Soc. **70** (2004), no. 2, 273–288. 1.2, 3.1, 3.1, 3.3, 3.4, 3.4, 3.5

[AY06] N. Alon and R. Yuster, *The number of orientations having no fixed tournament*, Combinatorica **26** (2006), 1–16. 3.1

[Bal06] J. Balogh, *A remark on the number of edge colorings of graphs*, Europ. J. Combin. **27** (2006), 565–573. 3.1

[BF77] J. C. Bermond and P. Frankl, *On a conjecture of Chvátal on m-intersecting hypergraphs*, Bull. London Math. Soc. **9** (1977), 309–312. 2.1

[Bro41] R. L. Brooks, *On colouring the nodes of a network*, Mathematical Proceedings of the Cambridge Philosophical Society **37** (1941), no. 2, 194–197. 4.2

[Chv75] V. Chvátal, *An extremal set-intersection theorem*, J. London Math. Soc. **9** (1974/75), 355–359. 2.1.3, 2.1

[CK99] R. Csákány and J. Kahn, *A homological approach to two problems on finite sets*, J. Algebraic Comb. **9** (1999), 141–149. 2.1, 2.1

[Des37] R. Descartes, *La géométrie*, p. 57, Ian Maire, Leiden, 1637. 4.5.1

[EKR61] P. Erdős, C. Ko, and R. Rado, *Intersection theorem for systems of finite sets*, Quart. J. Math. Oxford Set. **12** (1961), 313–320. 1.1, 2.1.1

[Erd62a] P. Erdős, *On a theorem of Rademacher-Turán*, Illinois Journal of Math **6** (1962), 122–127. 4.1

[Erd62b] ———, *On the number of complete subgraphs contained in certain graphs*, Magy. Tud. Acad. Mat. Kut. Int. K ozl. **7** (1962), no. 5, 459–474. 4.1

[Erd67] ———, *Some recent results on extremal problems in graph theory. Results*, Theory of Graphs (Internat. Sympos., Rome, 1966), Gordon and Breach, New York, 1967, pp. 117–123 (English); pp. 124–130 (French). 3.2.3, 4.3.3

[Erd69] ______, *On the number of complete subgraphs and circuits contained in graphs*, Casopis Pest. Mat. **94** (1969), 290–296. 4.1

[Erd71] ______, *Topics in combinatorial analysis*, Proceedings of the Second Louisiana Conference on Combinatorics, Graph Theory, and Computing (1971), 2–20. 2.1

[Erd74] ______, *Some new applications of probability methods to combinatorial analysis and graph theory*, Congres. Numer. **10** (1974), 39–51. 1.2, 3.1

[Erd92] ______, *Some of my favorite problems in various branches of combinatorics*, Matematiche (Catania) **47** (1992), 231–240. 1.2, 3.1

[F8̈3] Z. Füredi, *On finite set-systems whose every intersection is a kernel of a star*, Discrete Mathematics **47** (1983), 129–132. 2.1

[FF87] P. Frankl and Z. Füredi, *Exact solution of some turán-type problems*, J. Combinatorial Theory Series A **45** (1987), 226–262. 2.1, 2.1, 2.1

[FL10] M. Feng and X.J. Liu, *Note on set systems without a strong simplex*, Discrete Mathematics **310** (2010), 1645–1647. 2.1

[FÖ09] Z. Füredi and L. Özkahya, *Unavoidable subhypergraphs: a-clusters*, Electronic Notes in Disc. Math. **34** (2009), 63–67. 2.1

[Fra76] P. Frankl, *On Sperner families satisfying an additional condition*, J. Combin. Theory (A) **20** (1976), 1–11. 2.1

[Fra81] ______, *On a problem of Chvátal and Erdős on hypergraphs containing no special simplex*, J. Combin. Theory (A) **30** (1981), 169–182. 2.1

[Für] Z. Füredi, personal communications. 2.1

[Gow07] W. T. Gowers, *Hypergraph regularity and the multidimensional Szemerédi theorem*, Ann. of Math. **(2) 166** (2007), no. 3, 897–946. 4.3

[HKL09] C. Hoppen, Y. Kohayakawa, and H. Lefmann, *Kneser colorings of uniform hypergraphs*, Electronic NOtes in Disc. Math. **34** (2009), 219–223. 3.1

[JPY10] T. Jiang, O. Pikhurko, and Z. Yilma, *Set systems without a strong simplex*, SIAM J. Disc. Math. **24** (2010), no. 3, 1038–1045.

[KM10] P. Keevash and D. Mubayi, *Set systems without a simplex or a cluster*, Combinatorica **30** (2010), no. 2, 175–200. 2.1, 2.1

[KS96] J. Komlós and M. Simonovits, *Szemerédi's regularity lemma and its application in graph theory*, Paul Erdős is Eighty (D. Miklós, V. Sós, and T. Szőnyi, eds.), vol. 2, Bolyai Mathematical Society, 1996, pp. 295–352. 3.2, 3.2, 4.3

[LP] H. Lefmann and Y. Person, *The number of hyperedge colorings for certain classes of hypergraphs*, Manuscript. 3.1

[LPRS09] H. Lefmann, Y. Person, V. Rödl, and M. Schacht, *On colorings of hypergraphs without monochromatic Fano planes*, Combinatorics, Probability and Computing **18** (2009), 803–818. 3.1

[LPS] H. Lefmann, Y. Person, and M. Schacht, *A structural result for hypergraphs with many restricted edge colorings*, Submitted, 2010. 3.1

[LS75] L. Lovász and M. Simonovits, *On the number of complete subgraphs of a graph*, Proc. of Fifth British Comb. Conf. (Aberdeen), 1975, pp. 431–442. 4.1, 4.4

[LS83] ______, *On the number of complete subgraphs of a graph II*, Studies in pure mathematics, Birkhuser, Basle, 1983, pp. 459–495. 4.1, 4.2, 4.4

[Man07] W. Mantel, *Problem 28*, Wiskundige Opgaven **10** (1907), 60–61. 4.1

[MR09] D. Mubayi and R. Ramadurai, *Simplex stability*, Combinatorics, Probability and Computing **18** (2009), 441–454. 1.1, 2.1, 2.1, 2.1.6, 2.1.7, 2.1, 2.1.9, 2.1

[Mub07] D. Mubayi, *Structure and stability of triangle-free systems*, Trans. Amer. Math. Soc. **359** (2007), 275–291. 1.1, 2.1, 2.1.4

[Mub10] ______, *Counting substructures I : Color critical graphs*, Advances in Mathematics **225** (2010), no. 5, 2731–2740. 4.1, 4.1.3, 4.2, 4.2, 4.2, 4.3, 4.4

[MV05] D. Mubayi and J. Verstraëte, *Proof of a conjecture of Erdős on triangles in set systems*, Combinatorica **25** (2005), 599–614. 2.1

[Nik11] V. Nikiforov, *The number of cliques in graphs of given order and size*, Trans. Amer. Math. Soc. **363** (2011), no. 2, 1599–1618. 4.1

[NRS06] B. Nagle, V. Rödl, and M. Schacht, *The counting lemma for regular k-uniform hypergraphs*, Random Structures Algorithms **28** (2006), no. 2, 113–179. 4.3

[PY] O. Pikhurko and Z. Yilma, *The maximum number of K_3-free and K_4-free edge 4-colorings*, Submitted.

[Raz07] A. Razborov, *Flag algebras*, J. Symbolic Logic **72** (2007), 1239–1282. 4.1

[Raz08] ______, *On the minimal density of triangles in graphs*, Combinatorics, Probability and Computing **17** (2008), no. 4, 603–618. 4.1

[Röd85] V. Rödl, *On a packing and covering problem*, European J. Combin. **5** (1985), 69–78. 2.3

[RS06] V. Rödl and J. Skokan, *Application of the regularity lemma for uniform hypergraphs*, Random Structures Algorithms **28** (2006), no. 2, 180–194. 4.3

[Sim68] M. Simonovits, *A method for solving extremal problems in graph theory, stability problems*, Theory of Graphs (Proc. Colloq., Tihany, 1966), Academic Press, 1968, pp. 279–319. 1.3, 3.2.3, 4.1, 4.3.3

[SS91] M. Simonovits and V. T. Sós, *Szemerédi's partition and quasirandomness*, Random Structures Algorithms **2** (1991), 1–10. 3.4

[Sze76] E. Szemerédi, *Regular partitions of graphs*, Proc. Colloq. Int. CNRS (Paris), 1976, pp. 309–401. 3.2

[Tao06] T. Tao, *A variant ofthe hypergraph removal lemma*, J. Combin. Theory Ser. A **113** (2006), no. 7, 1257–1280. 4.3

[Tur41] P. Turán, *On an extremal problem in graph theory (in hungarian)*, Mat. Fiz. Lapok **48** (1941), 436–452. 1.3, 2.1, 3.1

[Yus96] R. Yuster, *The number of edge colorings with no monochromatic triangle*, J. Graph Theory **21** (1996), 441–452. 3.1