2021

Turan problems in extremal graph theory and flexibility

Kyle Edinger Murphy
Iowa State University

Turán problems in extremal graph theory and graph flexibility

by

Kyle Edinger Murphy

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Mathematics

Program of Study Committee:
Bernard Lidický, Major Professor
Steve Butler
Claus Kadelka
Ryan Martin
Michael Young

Iowa State University

Ames, Iowa

2021

DEDICATION

I would like to dedicate this thesis to my family Jim, Gail and Sara, and my fiancée Virginia. The support from each of them provided me throughout graduate school.

TABLE OF CONTENTS

ACKNOWLEDGMENTS

I want to thank everyone who helped contribute to my graduate work at Iowa State. Thank you to Bernard Lidický for his guidance and friendship throughout my time here. I would also like to thank my committee members for their time and effort. Thank you to Tomáš Masařík, J.D. Nir, and Shira Zerbib for your collaborations and contributions. An additional thank you to Dr. Teresa Haynes and Dr. Robert Beeler for your guidance early in my career, and for inspiring me to get my Ph.D. Finally, I would like to especially thank my mom, who always encouraged and supported my interest in math. Without her help I would not have been able to do this.

ABSTRACT

In this work we will study two distinct areas of graph theory: generalized Turán problems and graph flexibility. In the first chapter, we will provide some basic definitions and motivation. Chapters 2 and 3 contain two submitted papers showing that two graphs, the cycle on five vertices and the path on four vertices, are maximized by the Turán graph when forbidding sufficiently large cliques. In chapter 4, we present new results on graph flexibility and weak flexibility for planar graphs.

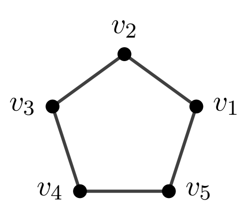
CHAPTER 1. GENERAL INTRODUCTION

This work focuses on two distinct areas of graph theory: extremal graph theory and graph coloring. In Chapters 2 and 3 we will focus on extremal graph theory, using the flag algebra method in order to calculate the generalized Turán numbers of two different graphs. In Chapter 4, we will turn our attention to graph coloring, presenting new results in the area of flexibility, which is a variant of list coloring. In this chapter, we will introduce some necessary definitions, along with background on each of the two subjects. Each of the chapters in this thesis contains a different paper which has been submitted for publication.

A *graph* G is a pair $G = (V(G), E(G))$ where $V(G)$ is a set whose elements are called *vertices*, and $E(G)$ is a set containing pairs of elements in $V(G)$, or *edges.* Two vertices connected by an edge are said to be *adjacent.* The *degree* of a vertex v, often denoted $d(v)$, is the number of vertices in $V(G)$ that are adjacent to v. The following well-known theorem is commonly referred to as the "Handshaking Lemma."

Theorem 1.0.1. *If G is a graph with m edges, then* $\sum_{v \in V(G)} d(v) = 2m.$

It is common to depict a graph using points drawn in the plane to represent vertices, and lines drawn between those points to represent edges.



$$V(G) = \{v_1, v_2, v_3, v_4, v_5\}$$

$$E(G) = \{v_1v_2, v_2v_3, v_3v_4, v_4v_5, v_5v_1\}$$

Figure 1.1 A graph $G = (V(G), E(G))$

A graph is said to be *simple* if it does not contain any multiedges or loops. A *multiedge* is a pair of vertices connected by more than one distinct edge, and a *loop* is an edge of the form vv for some vertex v. Moving forward, we will assume that all graphs being discussed are simple. In a graph G, two vertices u and v are called *connected* if there exists a sequence of edges of the following form $vx_1, x_1x_2, \ldots, x_{i-1}x_i, x_iu$. In other words, there exists a *path* from v to u. A graph G is said to be *connected* if every pair of vertices in G are connected. If G is not connected, then we say that it is disconnected. Unless specified otherwise, all graphs in this manuscript are connected.

Let G be a graph and let C be a set of colors. A *proper coloring* of G is a map $\phi : V(G) \to C$ such that for each edge $uv \in E(G)$, $\phi(u) \neq \phi(v)$. If $|C| = k$ and such a map exists, then we say that G is *k-colorable*. The smallest value of k for which G is k-colorable is denoted $\chi(G)$ and called the *chromatic number* of G. Note that $\chi(G)$ exists for each graph G, since one can always let $|C| = |V(G)|$ and simply let ϕ be any bijective map from $V(G)$ to C.

Two graphs $G = (V(G), E(G))$ and $H = (V(G), E(G))$ are said to be *isomorphic* if there exists a bijection $f : V(G) \to V(H)$ such that two vertices $u, v \in V(G)$ are adjacent if and only if $f(u)$ and $f(v)$ are adjacent in $V(H)$. A *subgraph* $F = (V(F), E(F))$ of a graph $G = (V(G), E(G))$ is a graph such that $V(F) \subseteq V(G)$ and $E(F) \subseteq E(G)$ so that for each edge $uv \in E(F)$, both u and v are contained in $V(F)$. A subgraph F is said to be *induced* if $E(F)$ contains all of the edges in $E(G)$ for which both vertices are contained in $V(F)$.

1.1 Extremal Graph Theory

Extremal graph theory focuses on maximizing or minimizing some function on a graph subject to some constraint. It has applications in computer science, and is useful for studying the properties of large graphs like the internet. In this thesis we will study *generalized Turán* problems, which deal with maximizing the frequency of a particular subgraph H in some host graph G while forbidding any occurrence of a different subgraph F. The forbidden graph F is most commonly a *complete graph* K_r, which is the graph on r vertices in which each pair of vertices is adjacent. The first result in extremal graph theory is Mantel's Theorem [7], concerning the graph K_3.

Theorem 1.1.1 (Mantel's Theorem). *The maximum number of edges among all K_3-free graphs on n vertices is* $\left\lfloor \frac{n^2}{4} \right\rfloor$.

Later, Turán [8] generalized this result to all K_r-free graphs for $r \geq 4$. In a moment we will state Turán's Theorem, but first we will require some definitions. A graph G is said to be *r-partite* if $V(G)$ can be partitioned into r sets $V_1, V_2, \ldots, V_r$ so that every edge of G joins two vertices in distinct sets. If $r = 2$, then we refer to G as *bipartite*. A r-partite graph G is said to be *complete r-partite* if uw is an edge of G if and only if $u \in V_i$ and $w \in V_j$ for $i \neq j$. The *Turán Graph* $T_r(n)$ is the complete r-partite graph on n vertices whose parts are of as equal size as possible. That is, the parts are all of size $\lfloor \frac{n}{r} \rfloor$ or $\lceil \frac{n}{r} \rceil$.

Theorem 1.1.2 (Turán's Theorem). *The maximum number of edges in an n-vertex graph with no K_{r+1} subgraph is exactly $|E(T_r(n))|$. Moreover, $T_r(n)$ is the unique graph attaining this maximum.*

Since $T_r(n)$ attains the maximum value in this case, it is called an *extremal graph*. For two graphs H and G, let $\nu(H, G)$ denote the (possibly non-induced) number of subgraphs of G which are isomorphic to H. If a graph G does not contain any subgraph isomorphic to F, then we say that G is *F-free*. Let $\text{ex}(n, H, F)$ denote the maximum value of $\nu(H, G)$ among all F-free graphs G on n vertices. The function $\text{ex}(n, H, F)$ has seen the most attention when $H = K_2$ is an edge. As such, we will use $\text{ex}(n, F)$ to denote $\text{ex}(n, K_2, F)$. Thus, one could restate Turán's Theorem in the following way:

$$\text{ex}(n, K_{r+1}) = \nu(K_2, T_r(n)).$$

In light of Turán's Theorem, it would be natural to ask for the value of $\text{ex}(n, F)$ where F is not a complete graph. The Erdős-Stone-Simonovits Theorem [5] determined that $\text{ex}(n, F)$ is asymptotically equal to $\text{ex}(n, K_r)$, where $r = \chi(F)$.

Theorem 1.1.3 (Erdős-Stone-Simonovits). *If F is a graph with chromatic number $\chi(F)$, then*

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

One of the most well-known results concerning $\text{ex}(n, H, F)$ when H is not an edge is Zykov's Theorem, stated below.

Theorem 1.1.4 (Zykov [9]). *Let k and t be integers such that $t < k+1$. Then for all n, the Turán graph $T_k(n)$ is the unique K_{k+1}-free graph on n vertices containing the maximum number of K_t subgraphs.*

Recently, there has been more attention on the function $\text{ex}(n, H, F)$ when H is not an edge, with the systematic study being introduced by Alon and Shikhelman [1]. Falling in line with many of the results discussed so far, it appears that for many graphs H, the value of the function $\text{ex}(n, H, K_r)$ is achieved by $T_r(n)$ for large enough n, and $r > \chi(H)$. As a result, Gerbner and Palmer [6] introduced the following definition:

Definition 1.1.5. *Fix an $(r+1)$-chromatic graph F and a graph T that does not contain F as a subgraph. We say that T is F-*Turán-good *if $ex(n, T, F) = \nu(T, T_r(n))$ for every n large enough.*

In this thesis, we will show that two graphs, C_5 and P_3, are K_r-Turán Good for $r \geq 4$. The *cycle* C_n is the graph with n vertices and n edges whose vertices can be labeled by $v_1, v_2, \ldots, v_n$ and whose edges are v_1v_n and v_iv_{i+1} for $i \in [n-1]$. The *path* P_n is the graph with $n+1$ vertices and n edges whose vertices can be labeled by $v_1, v_2, \ldots, v_n$ and whose edges are v_iv_{i+1} for $i \in [n-1]$. A note to the reader: in the literature, the path on n vertices is denoted by P_n (denoting the number of vertices) and P_{n-1} (denoting the number of edges), somewhat interchangeably. Since this thesis is a compilation of different papers written with different coauthors, we will use both definitions. In Chapter 3, we will use P_n to mean the path on $n+1$ vertices. In Chapters 2 and 4, we will use P_n to mean the path on n vertices.

The problem of showing that C_5 is K_r-Turán good for $r \geq 4$ was suggested by Cory Palmer at the "Mostly Manitoba, Michigan and Minnesota Combinatorics Graduate Students Workshop" at Iowa State University in 2018. The problem of showing that P_3 is K_r-Turán Good for $r \geq 4$ was suggested by Gerbner and Palmer in [6].

1.2 Graph Coloring

Graph coloring is one of the most widely studied areas of graph theory. It has a variety of wide ranging applications. One of the most well-known: scheduling problems. For example, when trying to schedule speakers for a conference, you might have a limited number of time slots, and certain speakers may need to present at different times. Let G be a graph whose vertices represent speakers, and let $uv \in E(G)$ if the speakers corresponding to u and v cannot be scheduled at the same time. Then a proper coloring of G gives an admissible schedule, and $\chi(G)$ gives the minimum number of necessary time slots.

A natural generalization of standard graph coloring is *list coloring*. A *list assignment* of G is a function L that assigns each vertex in $V(G)$ a list of colors $L(v)$. An *L-coloring* of G is a map $\phi : V(G) \to \cup_{v \in V(G)} L(v)$ such that

- $\phi(v) \in L(v)$ for each $v \in V(G)$, and
- $\phi(u) \neq \phi(v)$ if $uv \in E(G)$.

We say that a graph G is *k-choosable* if, for every list assignment L with $|L(v)| \geq k$ for all $v \in V(G)$, G is L-colorable. The *choosability* of a graph G, denoted $\mathrm{ch}(G)$ is the smallest k such that G is k-choosable. List coloring could be used to model a scheduling problem where each speaker at a conference is only available at a handful of times.

1.2.1 Flexibility

In many scheduling problems participants might have a specific request for a time slot from their list of available times. For example, when scheduling a conference, a speaker might be available at 8 AM, 9 AM, and 10 AM, but would prefer to present at 9 AM. In order to model this as a graph coloring problem, Dvořák, Norin, and Postle introduced the concept of ε-flexibility [4]. In this variant of the standard list coloring problem, a subset of the vertices have requests for colors from their lists. Since it is often the case that not all requests can be satisfied, the goal is to satisfy as many as possible. In this way, if an ε-proportion of the requests can be satisfied, we say that G is *ε-flexible* with lists of size L. More formal definitions are provided in Chapter 4.

1.2.2 The Discharging Method

Much of the current work in graph flexibility deals with planar graphs. This is to take advantage of the well-known *discharging method*. Here we will provide a short introduction and description. A more thorough review of the method can be found in [2]. A graph is said to be *planar* if it can be drawn in the plane with no edges crossing. Such a drawing is called a *planar embedding*. A graph G that has been drawn in this manner is called *plane*. If G is not planar, then we say it is *nonplanar*. A *face* in a plane graph is a region enclosed by the edges of a graph. The region surrounding the graph is considered the outer face. The *length* of a face f is the total length of the closed walks(s) bounding f.

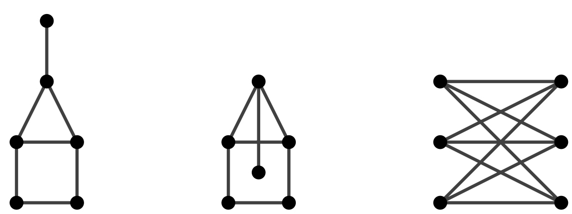


Figure 1.2 Left to Right: Planar (plane), Planar (not plane), Not Planar.

The following theorem is a well-known result on plane graphs, which we will need later on in this section. Let $F(G)$ denote the set of faces of a plane graph G, and let $\ell(f)$ denote the length of each face $f \in F(G)$.

Theorem 1.2.1. *Let G be a plane graph with m total edges. Then* $\sum_{f \in F(G)} \ell(f) = 2m$.

One of the most celebrated results in graph theory is Euler's Formula, which is stated below.

Theorem 1.2.2 (Euler's Formula)**.** *For every plane connected graph G with n vertices, m edges, and f faces,*

$$n - m + f = 2.$$

An outline of a discharging might proceed as follows. Suppose we were trying to prove some theorem on planar graphs. Let G be an arbitrary plane graph which is a minimum counterexample. One can often find a list of so called *reducible configurations* in G, which are small subgraphs that cannot appear in a minimum counterexample. By multiplying Euler's formula by an appropriately chosen constant, say -4, and invoking Theorems 1.0.1 and 1.2.1, one can show the following:

$$\sum_{v \in V(G)} (d(v) - 4) + \sum_{f \in F(G)} (\ell(f) - 4) = -8. \tag{1.1}$$

Given this fact, we can assign a charge of $d(v) - 4$ to each vertex $v \in V(G)$, and a charge of $\ell(f) - 4$ to each face $f \in F(G)$. We then move charge around the graph according to some predetermined set of discharging rules. The goal is to show that after redistributing charge, if G is a graph not containing any reducible configuration, then the sum of the charges in G is nonnegative. This however, would contradict 1.1, implying that such a graph G cannot exist.

In [3], the authors introduced tools to apply the discharging method to problems in flexibility. Our results in Chapter 4 apply these same methods.

1.3 References

[1] N. Alon and C. Shikhelman. Many T copies in H-free graphs. *J. Combin. Theory Ser. B*, 121:146–172, 2016.

[2] D. W. Cranston and D. B. West. An introduction to the discharging method via graph coloring. *Discrete Mathematics*, 340(4):766–793, 2017.

[3] Z. Dvořák, T. Masařík, J. Musílek, and O. Pangrác. Flexibility of triangle-free planar graphs, Feb. 2019.

[4] Z. Dvořák, S. Norin, and L. Postle. List coloring with requests. *Journal of Graph Theory*, 92(3):191–206, Jan. 2019.

[5] P. Erdős. On some problems in graph theory, combinatorial analysis and combinatorial number theory. *Graph theory and combinatorics (Cambridge, 1983), Academic Press, London*, pages 1–17, 1984.

[6] D. Gerbner and C. Palmer. Some exact results for generalized Turán problems. *arXiv e-prints*, page arXiv:2006.03756, June 2020.

[7] W. Mantel. Problem 28. *Winkundige Opgaven*, 10:60–61, 1907.

[8] P. Turán. Eine extremalaufgabe aus der graphentheorie. *Mat. Fiz. Lapok*, 48:436–452, 1941.

[9] A. A. Zykov. On some properties of linear complexes. *Matematicheskii sbornik*, 66:163–188, 1949.

CHAPTER 2. MAXIMIZING FIVE-CYCLES IN K_r-FREE GRAPHS

Bernard Lidický, Department of Mathematics, Iowa State University.

Kyle Murphy, Department of Mathematics, Iowa State University.

Modified from a manuscript currently under review in *The European Journal of Combinatorics*

2.1 Abstract

The Pentagon Problem of Erdős asks to find an n-vertex triangle-free graph that is maximizing the number of 5-cycles. The problem was solved using flag algebras by Grzesik and independently by Hatami, Hladký, Král', Norine, and Razborov. Recently, Palmer suggested a more general problem of maximizing the number of 5-cycles in K_{k+1}-free graphs. Using flag algebras, we show that every K_{k+1}-free graph of order n contains at most

$$\frac{1}{10k^4}(k^4 - 5k^3 + 10k^2 - 10k + 4)n^5 + o(n^5)$$

copies of C_5 for any $k \geq 3$, with the Turán graph being the extremal graph for large enough n.

2.2 Introduction

All graphs in this paper are simple. Let G, H, and F be graphs. We define $\nu(H, G)$ as the number of (possibly non-induced) subgraphs of G isomorphic to H. If G does not contain any subgraph isomorphic to F, then we say that G is *F-free*. Let $\text{ex}(n, H, F)$ denote the maximum value of $\nu(H, G)$ among all F-free graphs G on n vertices. The function $\text{ex}(n, H, F)$ is well-studied when H is an edge. As such, it is convention when $H = K_2$ to let $\text{ex}(n, F)$ denote $\text{ex}(n, K_2, F)$. The value of $\text{ex}(n, F)$ for any graph F is called the *Turán number* of F.

One of the first results in extremal graph theory was Mantel's Theorem [27] which states that for all $n \geq 3$, $\text{ex}(n, K_3) \leq \lfloor \frac{n^2}{4} \rfloor$. When $k \geq 3$, the value of $\text{ex}(n, K_{k+1})$ was determined by Turán.

Theorem 2.2.1 (Turán's Theorem [34])**.** *For all $k \geq 3$, and all n,*

$$ex(n, K_{k+1}) \leq \left(1 - \frac{1}{k}\right)\frac{n^2}{2}.$$

Moreover, the Turán graph $T_k(n)$, which is the complete balanced $(k-1)$-partite graph on n vertices, is the unique K_{k+1}-free graph on n vertices which contains the maximum possible number of edges.

The Erdős-Stone-Simonovits Theorem [11] determined the asymptotic value of $\mathrm{ex}(n, F)$ when F is not a complete graph. Let $\chi(F)$ denote the chromatic number of F. Then for all F for which $\chi(F) \geq 2$,

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

The systematic study of the function $\mathrm{ex}(n, H, F)$ was initiated by Alon and Shikhelman [2], although there were some prior results. When $t < k + 1$, Zykov [35] showed that the Turán graph $T_k(n)$ is also the unique graph with the maximum number of K_t subgraphs among all K_{k+1}-free graphs.

Theorem 2.2.2 (Zykov [35])**.** *Let k and t be integers such that $t < k + 1$. Then for all n, the Turán graph $T_k(n)$ is the unique K_{k+1}-free graph on n vertices containing the maximum number of K_t subgraphs.*

We will need the following corollary in our calculations.

Corollary 2.2.3. *Let G be a K_{k+1}-free graph on n vertices. Then*

$$\nu(K_5, G) \leq \frac{k^4 - 10k^3 + 35k^2 - 50k + 24}{k^4}n^5 + o(n^5).$$

Alon and Shikhelman [2] proved the following analogue of the Kővári-Sós-Turán Theorem:

$$\mathrm{ex}(n, K_3, K_{s,t}) = O(n^{3-3/s}).$$

They also proved that for fixed integers $t < k$, if F is a k-chromatic graph:

$$\mathrm{ex}(n, K_t, F) = \binom{k-1}{t}\left(\frac{n}{k-1}\right)^t + o(n^t).$$

In [23], Győri, Pach, and Simonovits studied a handful of cases where $F = K_r$. The order of magnitude of $\mathrm{ex}(n, C_k, C_\ell)$ is known for all $\ell \geq 3$ and $k \geq 3$, see Gishboliner and Shapira [19]. The asymptotic value of $\mathrm{ex}(n, C_k, C_4)$ was determined by Gerbner, Győri, Methuku, and Vizer [16]. They proved a variety of results on $\mathrm{ex}(n, F, H)$ when F and H were both cycles. This includes showing that $\mathrm{ex}(n, C_{2\ell}, C_{2k}) = \Theta(n^\ell)$ for $k, \ell \geq 2$ and

$$\mathrm{ex}(n, C_4, C_{2k}) = (1 + o(1))\frac{(k-1)(k-2)}{4}n^2$$

for $k \geq 2$. In [17], Gerbner and Palmer provided more general bounds on $\mathrm{ex}(n, H, F)$. In particular, they showed that if H and F are graphs and $\chi(F) = k$, then

$$\mathrm{ex}(n, H, F) \leq \mathrm{ex}(n, H, K_k) + o\left(n^{|H|}\right).$$

Additionally, they extended the result of Gishboliner and Shapira to show that for all k and t,

$$\mathrm{ex}(n, C_k, K_{2,t}) = \left(\frac{1}{2k} + o(1)\right)(t-1)^{k/2}n^{k/2},$$

and

$$\mathrm{ex}(n, P_k, K_{2,t}) = \left(\frac{1}{2} + o(1)\right)(t-1)^{(k-1)/2}n^{(k+1)/2},$$

where P_k is a path on k vertices.

In [10], Cutler, Nir, and Radcliffe determined the asymptotic value of $\mathrm{ex}(n, S_t, K_{k+1})$, where S_t is the star with t leaves. In particular, they showed that while the extremal graph must be complete multi-partite, it is not always isomorphic to the Turán graph $T_k(n)$. The study of the function $\mathrm{ex}(n, K_3, H)$ has seen recent attention as well. In particular, the function $\mathrm{ex}(n, K_3, C_5)$ was studied in [2, 8, 13]. In [29], Mubayi and Mukherjee studied the function $\mathrm{ex}(n, K_3, H)$ for a handful of other 3-chromatic graphs H.

In [18], Gerbner and Palmer found a handful of cases where the value of $\mathrm{ex}(n, H, F)$ is achieved by the Turán graph and in [15], Gerbner studied the function $\mathrm{ex}(n, H, F)$ when H and F each have at most 4 vertices. Recently, the authors of [25] studied the problem of maximizing the number of copies of a graph H in some graph G embedded in a particular surface.

In 1984, Erdős conjectured that the balanced blow-up of C_5 on n vertices maximizes the number of five-cycles among all triangle-free graphs of order n. If G is a graph on m vertices, then the *balanced blow-up* of G on n vertices is the graph $G(n)$ obtained from G by replacing each vertex of G with an independent set of size $\lfloor \frac{n}{m} \rfloor$ or $\lceil \frac{n}{m} \rceil$, and replacing each edge in G with a complete bipartite graph on the corresponding sets. The problem of determining $\text{ex}(n, C_5, K_3)$ was known as the Pentagon Problem of Erdős. In a sense, a graph with $\text{ex}(n, C_5, K_3)$ five-cycles is the "least bipartite" triangle-free graph on n vertices when measured by the number of 5-cycles. In posing this question, Erdős also proposed the following two measures of "non-bipartiteness" [12].

1. The minimal possible number of edges in a subgraph spanned by half the vertices.

2. The minimal possible number of edges that have to be removed to make the graph bipartite, which is equivalent to the problem of max cut.

In 1989, Győri [22] showed that a triangle-free graph on n vertices contains at most $1.03 \left(\frac{n}{5}\right)^5$ five-cycles. In 2012, Grzesik [21] and independently in 2013, Hatami Hladký, Král', Norine, and Razborov [24] showed that a triangle-free graph on n vertices contains at most $\left(\frac{n}{5}\right)^5 + o(n^5)$ five cycles. Moreover, a matching lower bound is given by the balanced blow-up of C_5 when n is divisible by 5. The authors of [24] also proved that for large enough n, the balanced blow-up of a C_5 on n vertices is the unique extremal graph. In 2018, Lidický and Pfender [26] proved that a balanced C_5 blow-up is the unique extremal construction for all n, with the exception of $n = 8$. This observation was made by Michael [28], who showed that the Möbius ladder on 8 vertices contains the same number of five cycles as the balanced C_5 blow-up.

Palmer [30] suggested a generalization to the Pentagon Problem of Erdős: maximizing the number of five-cycles in K_{k+1}-free graphs for $k \geq 3$. Observe that in the more general case, the problem of maximizing the number of non-induced C_5 subgraphs is different from maximizing the number of induced C_5 subgraphs.

In this paper, we will discuss the non-induced case. Let H and G be graphs on n_1 and n_2 vertices, respectively. The *density* $d(H,G)$ of H in G is given by

$$d(H,G) = \nu(H,G)\binom{n_2}{n_1}^{-1}.$$

Normally, $n_2^{-n_1}$ would be used as the scaling factor for defining the density of H in G. We will use $\binom{n_2}{n_1}^{-1}$, since this is more natural in proofs involving the flag algebra method. Let

$$\mathrm{OPT}_k(C_5) = \lim_{n\to\infty} \max_{G_n \in \mathcal{F}_n^k} d(C_5, G_n), \tag{2.1}$$

where $\mathcal{F}_n^k$ is the set of all K_{k+1}-free graphs on n vertices. Note that since $d(C_5, G)$ measures the density of non-induced C_5 subgraphs in a graph G, this parameter will often have a value greater than one. For example, $d(C_5, K_\ell) = 12$ for all $\ell \geq 5$. Our main goal is to prove the following theorem.

Theorem 2.2.4. *Let $k \geq 3$ be an integer. Then*

(i) $\mathrm{OPT}_k(C_5) = \frac{1}{k^4}(12k^4 - 60k^3 + 120k^2 - 120k + 48)$.

(ii) If n is sufficiently large, then $T_k(n)$ is the unique K_{k+1}-free graph on n vertices for which $\nu(C_5, T_k(n)) = ex(n, C_5, K_{k+1})$.

Since our result forbids $(k+1)$-cliques, Turán's Theorem implies that the number of edges in an extremal graph cannot be more than in $T_k(n)$. Interestingly, the authors of [6] proved that if G is a graph with at least $\frac{k-1}{k}\binom{n}{2}$ edges, then the Turán graph provides a lower bound on the number of five-cycles contained in G.

The proof of Theorem 2.2.4(i) uses flag algebras to calculate the upper bound for $\mathrm{OPT}_k(C_5)$. The second part is done by stability and exact structure arguments. Unlike typical applications of the flag algebra method, our result does not need computer assistance for the calculations involving flag algebras. However, it is still convenient to use a computer for the purpose of multiplying and expanding rational functions.

In the next section, we will give a brief overview of the flag algebra method. Section 2 contains the proof of Theorem 2.2.4(i). Then we prove a stability lemma in Section 3, and use it to prove Theorem 2.2.4(ii) in Section 4. We will end with some concluding remarks and conjectures concerning the general behavior of the function $\mathrm{ex}(n, H, F)$.

2.2.1 The Flag Algebra Method

Introduced by Razborov [32], the flag algebra method provides a framework for computationally solving problems in extremal combinatorics. Flag algebras have been used to solve problems on hypergraphs [4, 14, 20, 31], permutations [5], graph decomposition problems [7], and oriented graphs [9] among many other applications. Here we will give a brief introduction and description of the notation and theory we will need for our result. We will not prove any claims since they have already been proven by Razborov [32]. Another overview of flag algebras can be found in [33].

Let H and G be graphs on n_1 and n_2 vertices, respectively, such that $n_1 \leq n_2$. If $X \subseteq V(G)$, we will denote the induced subgraph of G on the vertices of X by $G[X]$. Let a subset X be selected uniformly at random from $V(G)$ such that $|X| = n_1$. Then $P(H, G)$ is the probability that $G[X]$ is isomorphic to H.

A sequence of graphs $(G_n)_{n\geq 1}$ of increasing order is said to be *convergent* if for every finite graph H, the following limit converges:

$$\lim_{n\to\infty} P(H, G_n).$$

Let $\mathcal{F}$ denote the set of all graphs up to isomorphism, and let $\mathcal{F}_\ell$ denote the set of all graphs on ℓ vertices up to isomorphism. Let $\mathbb{R}\mathcal{F}$ denote the set of all formal finite linear combinations of graphs in $\mathcal{F}$. A *type of size* k is a graph σ on k labelled vertices labeled by $[k] = \{1, \ldots, k\}$. If σ is a type of size k and F is a graph on at least k vertices, then an *embedding* of σ into F is an injective function $\theta : [k] \to V(F)$, such that θ gives an isomorphism between σ and $F[\mathrm{im}(\theta)]$. A σ-flag is a pair (F, θ) where F is a graph and θ is an injective function from $[k]$ to $V(F)$ that defines a graph isomorphism of $F[\mathrm{im}(\theta)]$ and σ. In this way, σ can be thought of as a labelled subgraph of F. Two σ-flags F and G are isomorphic if there exists a graph isomorphism between F and G that preserves the labelled subgraph σ.

Let $\mathcal{F}^\sigma$ denote the set of all σ-flags and $\mathcal{F}^\sigma_\ell$ denote the set of all σ-flags on ℓ vertices. Observe that if σ is the empty graph, then $\mathcal{F}^\sigma = \mathcal{F}$. For two σ-flags F and G with $|V(F)| \leq |V(G)|$, let $P(F, G)$ denote the probability that an injective map from $V(F)$ to $V(G)$ that fixes the labeled graph σ induces a copy of F in G. Razborov showed that there exists an algebra $\mathcal{A}^\sigma$ after some factorization of $\mathbb{R}\mathcal{F}^\sigma$. In doing so, he defined addition and multiplication on the elements of $\mathbb{R}\mathcal{F}^\sigma$. Addition can be defined in the natural way, by simply adding coefficients of the elements in $\mathbb{R}\mathcal{F}^\sigma$. We will now describe how to define multiplication of elements in $\mathcal{A}^\sigma$.

Let $(G, \theta) \in \mathcal{F}^\sigma$ be a σ-flag on n vertices. Let $(F_1, \theta_1), (F_2, \theta_2) \in \mathcal{F}^\sigma$ be two σ-flags for which $|V(F_1)| + |V(F_2)| \leq n + |V(\sigma)|$. Let X_1 and X_2 be two disjoint sets of sizes $|V(F_1)| - |V(\sigma)|$ and $|V(F_2)| - |V(\sigma)|$ respectively, selected uniformly at random from $V(G) \setminus \mathrm{im}(\theta)$. We will define the *density of* F_1 *and* F_2 *in* G, denoted $P(F_1, F_2; G)$ as the probability that $(G[X_1 \cup \mathrm{im}(\theta)], \theta)$ is isomorphic to (F_1, θ_1) and $(G[X_2 \cup \mathrm{im}(\theta)], \theta)$ is isomorphic to (F_2, θ_2).

It can be shown that as n grows, then the density of F_1 and F_2 is approximately equal to the product of their individual densities:

$$|P(F_1, F_2; G) - P(F_1, G)P(F_2, G)| \leq O(n^{-1}). \tag{2.2}$$

Given this fact, if $|V(F_1)| + |V(F_2)| = \ell - |V(\sigma)|$ we could ideally define multiplication in $\mathcal{A}^\sigma$ by

$$F_1 \cdot F_2 = \sum_{F \in \mathcal{F}^\sigma_\ell} P(F_1; F_2; F)F. \tag{2.3}$$

The issue with this, however, is that the product $F_1 \cdot F_2$ could be also written as a linear combination of elements in $\mathcal{F}^\sigma_{\ell'}$ for any $\ell' > \ell$. Hence, before defining $\mathcal{A}^\sigma$ we factor out all expressions of the form

$$F - \sum_{F' \in \mathcal{F}^\sigma_\ell} P(F, F')F' \tag{2.4}$$

from $\mathbb{R}\mathcal{F}^\sigma$. Note that (2.4) corresponds to the law of total probability and hence it should behave as 0 when added to another linear combination. Let $\mathcal{K}^\sigma$ be the linear subspace of $\mathbb{R}\mathcal{F}^\sigma$ containing all expressions of the form (2.4). We define $\mathcal{A}^\sigma$ to be $\mathbb{R}\mathcal{F}^\sigma$ factorized by $\mathcal{K}^\sigma$, and we define multiplication in $\mathcal{A}^\sigma$ by naturally extending (2.3).

Returning to the idea of convergent sequences of graphs, let $\mathrm{Hom}^+(\mathcal{A}^\sigma, \mathbb{R})$ be the set of all homomorphisms from $\mathcal{A}^\sigma$ to $\mathbb{R}$ such that for each $\phi \in \mathrm{Hom}^+(\mathcal{A}^\sigma, \mathbb{R})$ and $H \in \mathcal{F}^\sigma$, $\phi(H) \geq 0$. If σ has order 0, we omit it in the notation. Razborov showed that each homomorphism in $\mathrm{Hom}^+(\mathcal{A}, \mathbb{R})$ corresponds to some convergent graph sequence and vice versa [32].

In any fixed graph G, we can express $d(C_5, G)$ as the sum of induced densities in the following way:

$$d(C_5, G) = \sum_{F_i \in \mathcal{F}_5} c_{F_i}^{C_5} P(F_i, G), \tag{2.5}$$

where $c_{F_i}^{C_5} = \nu(C_5, F_i)$. Hence, for any sequence of unlabelled graphs $(G_n)_{n \geq 1}$ and its corresponding homomorphism $\phi \in \mathrm{Hom}^+(\mathcal{A}, \mathbb{R})$,

$$\lim_{n \to \infty} d(C_5, G_n) = \sum_{F_i \in \mathcal{F}_5} c_{F_i}^{C_5} \phi(F_i). \tag{2.6}$$

Quite often in our computations to simplify notation, we will drop the function notation and simply write F_i or draw the graph F_i in place of $\phi(F_i)$. Under this notation equation (2.6) would be

$$\lim_{n \to \infty} d(C_5, G_n) = \sum_{F_i \in \mathcal{F}_5} c_{F_i}^{C_5} F_i.$$

Finally, while we will often work with σ-flags where σ is not empty, flag algebras are often applied to questions concerning unlabelled graphs. In order to translate information from $\mathrm{Hom}^+(\mathcal{A}^\sigma, \mathbb{R})$ to $\mathrm{Hom}^+(\mathcal{A}, \mathbb{R})$ Razborov defined the *unlabelling operator* which is a linear operator $[\![\cdot]\!]_\sigma$ such that

$$[\![\cdot]\!]_\sigma : \mathbb{R}\mathcal{F}^\sigma \to \mathbb{R}\mathcal{F},$$

where for any σ-flag $F = (H, \theta)$, $[\![F]\!]_\sigma = q_\sigma(F)H$, where $q_\sigma(F)$ is equal to the probability (H, θ') is isomorphic to F where θ' is a randomly chosen injective mapping θ' from $[k]$ to $V(H)$. It can be shown that for any $a \in \mathcal{A}^\sigma$ and any $\phi \in \mathrm{Hom}^+(\mathcal{A}, \mathbb{R})$,

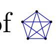
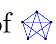
$$\phi\left([\![a \cdot a]\!]_\sigma\right) \geq 0. \tag{2.7}$$

We will frequently make use of this fact in our computations.

If a flag algebra calculation has a constant number of terms and operations, then it can be interpreted as a calculation in a graph of order n with an error term $O(n^{-1})$ coming from (2.2).

2.3 Proof of Theorem 2.2.4(i)

In this section we will prove Theorem 2.2.4(i). First we will provide a lower bound by counting the number of five cycles in the Turán graph. Next, using the flag algebra method, we will provide a matching upper bound. The proof of the upper bound when $k = 3$ is slightly different than the proof when $k \geq 4$.

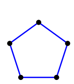
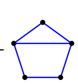
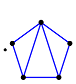
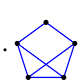
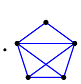
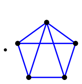
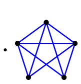
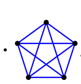
Proof of Theorem 2.2.4(i). First we will count the number of five-cycles in $T_k(n)$, which will give an asymptotic lower bound. Observe that the only induced subgraphs of $T_k(n)$ on five vertices containing a five-cycle are , , and . There are $\binom{k}{5}\left(\frac{n}{k}\right)^5$ copies of  in $T_k(n)$, with every such graph containing 12 distinct C_5 subgraphs. There are $4\binom{k}{4}\binom{n/k}{2}\left(\frac{n}{k}\right)^3$ copies of  in $T_k(n)$, with every such graph containing 6 distinct C_5 subgraphs. Finally, there are $3\binom{k}{3}\binom{n/k}{2}^2\frac{n}{k}$ copies of  in $T_k(n)$, with every such graph containing 4 distinct C_5 subgraphs. Thus,

$$\nu(C_5, T_k(n)) = 12\binom{k}{5}\left(\frac{n}{k}\right)^5 + 24\binom{k}{4}\binom{n/k}{2}\left(\frac{n}{k}\right)^3 + 12\binom{k}{3}\binom{n/k}{2}^2\frac{n}{k} + o(n^5),$$

where the error term $o(n^5)$ accounts for the cases where n is not divisible by k. This implies that for all n,

$$d(C_5, T_k(n)) \geq \nu(C_5, T_k(n))\binom{n}{5}^{-1} = \frac{1}{k^4}(12k^4 - 60k^3 + 120k^2 - 120k + 48) + o(1).$$

Now we will calculate an asymptotic upper bound. Unless it is stated otherwise, assume that $k \geq 3$. Let $\mathcal{F}_5 = \{F_0, \ldots, F_{33} = K_5\}$ be the set of unlabeled graphs (up to isomorphism) on five vertices. Each of these graphs is pictured in Table 2.8 in the Appendix. After removing each $c_{F_i}^{C_5}$ for which $c_{F_i}^{C_5} = 0$ from (2.5) we get

$\lim_{n\to\infty} d(C_5, G_n) =$  $+$  $+$  $+ 2\cdot$  $+ 2\cdot$  $+ 4\cdot$  $+ 6\cdot$  $+ 12\cdot$ . (2.8)

Since $\mathcal{F}_5$ contains all graphs on five vertices (up to isomorphism) and $\sum_{i=0}^{33} F_i = 1$,

$$\lim_{n\to\infty} d(C_5, G_n) \leq \max\{c_{F_i}^{C_5} : F_i \in \mathcal{F}_5\}.$$

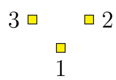
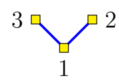
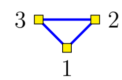
Therefore,

$$\mathrm{OPT}_k(C_5) \leq \max\{c_{F_i}^{C_5} : F_i \in \mathcal{F}_5\}.$$

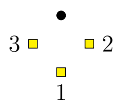
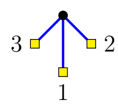
Given this fact, our goal is to find appropriate constants c_{F_i} so that

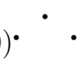
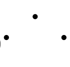
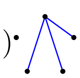
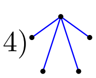
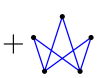
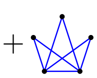
$$\lim_{n\to\infty} d(C_5, G_n) \leq \sum_{F_i \in \mathcal{F}_5} c_{F_i} F_i,$$

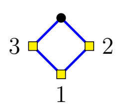
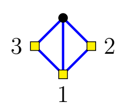
and $\max\{c_{F_i} : F_i \in \mathcal{F}_5\}$ is as small as possible. To do so, we can take advantage of properties that we know must be true of all $\phi \in \mathrm{Hom}^+(\mathcal{A}, \mathbb{R})$ which correspond to K_{k+1}-free convergent sequences of graphs. Additionally, using labeled flags, we can derive nonnegative expressions of unlabeled graphs in $\mathcal{F}_5$. We define

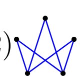
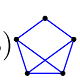
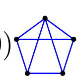
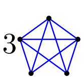
$\sigma_1 =$  $\sigma_2 =$  $\sigma_3 =$ 

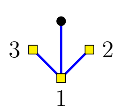
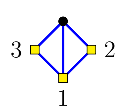
so that $\mathcal{F}_4^{\sigma_1}$, $\mathcal{F}_4^{\sigma_2}$, and $\mathcal{F}_4^{\sigma_3}$ denote three sets of labeled flags on four vertices. By (2.7), the following expressions are nonnegative for all $k \geq 3$.

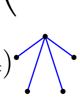
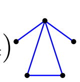
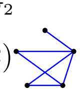
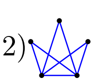
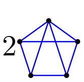
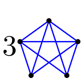
1. $P_1(k) = 10 \cdot [\![\big((k-1)$  $-$  $\big)^2]\!]_{\sigma_1} =$

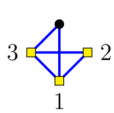
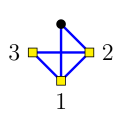
   $(10k^2 - 20k + 10)$  $+ (k^2 - 2k + 1)$  $+ (-k + 1)$  $+ (-4k + 4)$  $+$  $+$ 

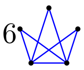
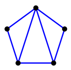
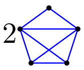
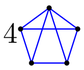
2. $P_2(k) = 30 \cdot [\![\big((k-2)$  $-$  $\big)^2]\!]_{\sigma_2} =$

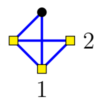
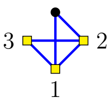
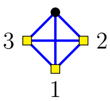
   $(3k^2 - 12k + 12)$  $+ (k^2 - 6k + 8)$  $+ (-4k + 10)$  $+ 3$ 

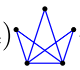
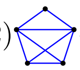
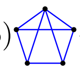
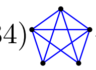
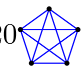
3. $P_3(k) = 30 \cdot [\![\big((k-2)$  $-$  $\big)^2]\!]_{\sigma_2} =$

   $(6k^2 - 24k + 24)$  $+ (k^2 - 4k + 4)$  $+ (-k + 2)$  $+ (-6k + 12)$  $+ 2$  $+ 3$ 

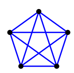
4. $P_4(k) = 30 \cdot [\![\big($  $-$  $\big)^2]\!]_{\sigma_3} =$

6  $-$  $+\,2$  $-\,4$ 

5. $P_5(k) = 30 \cdot \Big[\!\Big[\Big((k-3)$  $+ (k-3)$  $- 2$  $\Big)^2 \Big]\!\Big]_{\sigma_3} =$

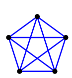
$(6k^2-36k+54)$  $+(2k^2-20k+42)$  $+(4k^2-24k+36)$  $+(-24k+84)$  $+120$ 

Additionally, we can apply Corollary 2.2.3, which states that for any K_{k+1}-free convergent sequence of graphs,

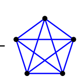
 $$\leq \frac{k^4 - 10k^3 + 35k^2 - 50k + 24}{k^4}.$$

At this point we will split the proof into the two cases where $k \geq 4$ and $k = 3$. To gain some intuition as to why this is necessary, we can consider the previous inequality. When $k = 3$ or $k = 4$, the previous bound implies that  $= 0$. The issue is that it does not give any information about the density of K_4, which is also equal to zero when $k = 3$. Thus, two slightly different proofs are required for $k = 3$ and $k \geq 4$.

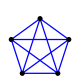
Case 1: Suppose that $k \geq 4$. Since $\sum_{i=0}^{33} F_i = 1$,

 $$\leq \sum_{i=0}^{33} F_i \left(\frac{k^4 - 10k^3 + 35k^2 - 50k + 24}{k^4} \right). \tag{2.9}$$

After rearranging the terms from (2.9) we obtain the following constraint on the elements of $\mathcal{F}_5$:

$$0 \leq \sum_{i=0}^{32} F_i \cdot \frac{k^4 - 10k^3 + 35k^2 - 50k + 24}{k^4} +$$  $$\cdot \frac{-10k^3 + 35k^2 - 50k + 24}{k^4}. \tag{2.10}$$

Let

$$Z(k) = \sum_{i=0}^{32} F_i \cdot \frac{k^4 - 10k^3 + 35k^2 - 50k + 24}{k^4} +$$  $$\cdot \frac{-10k^3 + 35k^2 - 50k + 24}{k^4}.$$

It is straightforward to verify that the following rational functions are nonnegative for all $k \geq 4$.

$$p_1(k) = \tfrac{3(k^5-8k^4+22k^3-24k^2+8k)}{5k^7-35k^5+75k^4-48k^3} \qquad p_2(k) = \tfrac{10k^5-60k^4+109k^3-76k^2+18k}{5k^7-35k^5+75k^4-48k^3}$$

$$p_3(k) = \tfrac{5k^5-28k^4+45k^3-28k^2+6k}{5k^7-35k^5+75k^4-48k^3} \qquad p_4(k) = \tfrac{5k^7-30k^6+53k^5-52k^4+94k^3-96k^2+24k}{4(5k^7-35k^5+75k^4-48k^3)}$$

$$p_5(k) = \tfrac{15k^5-60k^4+78k^3-40k^2+8k}{4(5k^7-35k^5+75k^4-48k^3)} \qquad z(k) = \tfrac{120(5k^3-20k^2+30k-16)}{5k^3-35k^2+75k-48}$$

Thus, for any convergent sequence $(G_n)_{n\geq 1}$ of K_{k+1}-free graphs with $k \geq 4$,

$$\lim_{n\to\infty} d(C_5, G_n) \leq \sum_{i=0}^{33} c_{F_i}^{C_5} F_i + z(k)Z(k) + \sum_{j=1}^{5} p_j(k)P_j(k) = \sum_{i=0}^{33} c_{F_i} F_i, \tag{2.11}$$

where c_{F_i} is the coefficient of F_i after all expansions. Then

$$\mathrm{OPT}_k(C_5) \leq \max\{c_{F_i} : F_i \in \mathcal{F}_5\}.$$

The values of c_{F_i} for each $F_i \in \mathcal{F}_5$ are listed below.

- $C_1(k) = c_{\cdot} = c_{\cdot} = c_{\cdot} = c_{\cdot} = c_{\cdot} = c_{\cdot} = c_{\cdot} = \frac{60k^7-720k^6+3600k^5-9876k^4+16320k^3-16440k^2+9360k-2304}{5k^7-35k^5+75k^4-48k^3}.$

- $C_2(k) = c_{\cdot} = \frac{33k^7-450k^6+2547k^5-7824k^4+14214k^3-15360k^2+9144k-2304}{5k^7-35k^5+75k^4-48k^3}.$

- $C_3(k) = c_{\cdot} = c_{\cdot} = c_{\cdot} = c_{\cdot} = c_{\cdot} = c_{\cdot} = c_{\cdot} = c_{\cdot} = c_{\cdot} = c_{\cdot} = c_{\cdot} = c_{\cdot} = c_{\cdot} = c_{\cdot} = c_{\cdot} = c_{\cdot} = c_{\cdot} = c_{\cdot} = \frac{30k^7-420k^6+2430k^5-7596k^4+13980k^3-15240k^2+9120k-2304}{5k^7-35k^5+75k^4-48k^3}.$

- $C_4(k) = c_{\cdot} = \frac{30k^7-423k^6+2457k^5-7686k^4+14118k^3-15336k^2+9144k-2304}{5k^7-35k^5+75k^4-48k^3}.$

- $C_5(k) = c_{\cdot} = \frac{35k^7-468k^6+2607k^5-7916k^4+14278k^3-15367k^2+9144k-2304}{5k^7-35k^5+75k^4-48k^3}.$

- $C_6(k) = c_{\cdot} = \frac{30k^7-425k^6+2468k^5-7697k^4+14098k^3-15302k^2+9132k-2304}{5k^7-35k^5+75k^4-48k^3}.$

- $C_7(k) = c_{\cdot} = c_{\cdot} = \frac{35k^7-455k^6+2505k^5-7644k^4+13980k^3-15240k^2+9120k-2304}{5k^7-35k^5+75k^4-48k^3}.$

- $C_8(k) = c_{\cdot} = \frac{(135/4)k^7-(895/2)k^6+(9967/4)k^5-7631k^4+(27913/2)k^3-15216k^2+9114k-2304}{5k^7-35k^5+75k^4-48k^3}.$

- $C_9(k) = c = \frac{50k^7-610k^6+3129k^5-8902k^4+15326k^3-15956k^2+9264k-2304}{5k^7-35k^5+75k^4-48k^3}$.
- $C_{10}(k) = c = \frac{50k^7-610k^6+3103k^5-8758k^4+15050k^3-15748k^2+9216k-2304}{5k^7-35k^5+75k^4-48k^3}$.

Claim 2.3.1. *For all* $i = 1, \ldots, 10$ *and* $k \geq 4$, $C_1(k) \geq C_i(k)$.

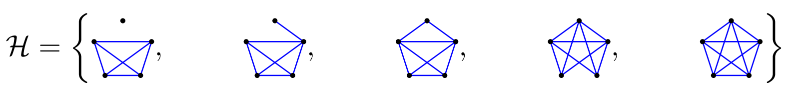
Proof. Observe that for all $i = 1, \ldots, 10$, each polynomial $C_i(k)$ has the same denominator of $5k^7 - 35k^5 + 75k^4 - 48k^3$. It is straightforward to verify that $5k^7 - 35k^5 + 75k^4 - 48k^3$ is positive for all $k \geq 4$. By examining the leading coefficients in the numerator of each polynomial, it is straightforward to check that C_1 is the largest for $k > 1000$. For $4 \leq k \leq 1000$, we have provided SageMath code used to verify the claim in Appendix 2.8.1. ■

By factoring C_1 it follows that

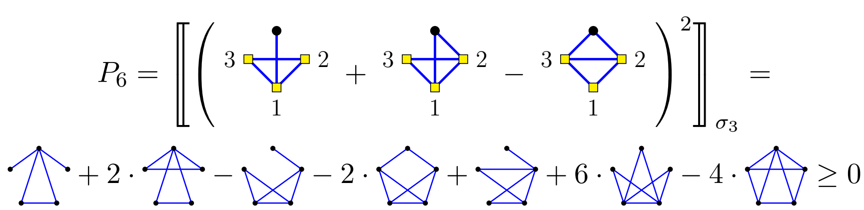
$$\mathrm{OPT}_k(C_5) \leq C_1(k) = \frac{1}{k^4}(12k^4 - 60k^3 + 120k^2 - 120k + 48),$$

completing the proof of Theorem 2.2.4(i) when $k \geq 4$.

Case 2: Suppose that $k = 3$. Assume that $(G_n)_{n \geq 1}$ is a K_4-free convergent sequence of graphs. Each graph in the set $\mathcal{H}$ given below has a limit density equal to zero, and therefore can be removed from our calculations.



In this case, we will use the same polynomials $P_i(k)$ for $i = 1, 2, 3, 4$ that were provided earlier in the proof. We will define one new polynomial P_6, which is nonnegative by (2.7).



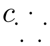
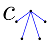
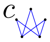
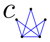
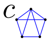
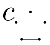
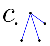
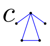
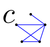
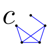
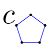
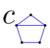
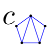
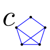
Now suppose that

$p_1 = 1/27$	$p_2 = 13/27$	$p_3 = 8/27$
$p_4 = 2/9$	$p_6 = 17/54.$	

Then

$$d(C_5) = \lim_{n\to\infty} d(C_5, G_n) \leq \sum_{F\in\mathcal{F}_5\setminus\mathcal{H}} \nu(C_5, F)F + \sum_{j=1}^{4} p_j P_j(3) + p_6 P_6.$$

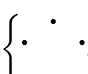
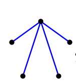
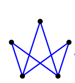
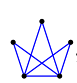
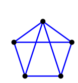
Let c_F denote the coefficient of each graph F after combining each of the two sums. We provide the values of c_F for each graph in $\mathcal{F}_5 \setminus \mathcal{H}$ for which $c_F \neq 0$.

- c $=$ c $=$ c $=$ c $=$ c $= \frac{40}{27}$
- c $= \frac{4}{27}$
- c $= -\frac{2}{27}$
- c $= \frac{11}{18}$
- c $= \frac{1}{54}$
- c $= -\frac{17}{54}$
- c $=$ c $= 1$
- c $= \frac{7}{9}$
- c $= \frac{8}{9}$

SageMath code to verify this calculation can be found in Appendix 2.8.1. It is straightforward to verify that

$$d(C_5) \leq \max\{c_F : F \in \mathcal{F}_5 \setminus \mathcal{H}\} = \frac{40}{27}.$$

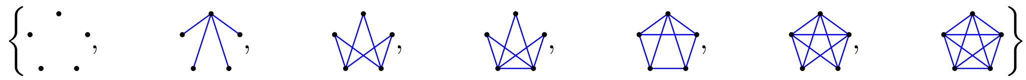
Furthermore, the set T_3 of graphs for which $c_{F_i} = \frac{40}{27}$ is given below.

$T_3 = \Big\{$ , , , ,  $\Big\}$ (2.12)

This completes the proof of Theorem 2.2.4 (i). ■

2.3.1 Finding the Optimal Bound

We will now give a short description on how we found the functions $z(k)Z(k)$ and $p_i(k)P_i(k)$ that were used in the proof of Theorem 2.2.4(i). If $F_j \in \mathcal{F}_5$ is a graph for which $c_{F_j} = \mathrm{OPT}_k(C_5)$, then we call F_j a *tight* subgraph. In our proof of Theorem 2.2.4(i), the set T given below contains the tight subgraphs when $k \geq 4$.

$T =$ 

The set T_3, defined in (2.12), contains the tight subgraphs when $k = 3$. Note that T_3 are the K_4-free graphs from T. The following lemma, which appears as Lemma 2.4.3 in [3], states that any graph appearing with positive probability in the limit of $(G_n)_{n\geq 1}$ must be tight if G_n are extremal graphs.

Lemma 2.3.2 ([3]). *Given $(G_n)_{n\geq 1}$ a convergent sequence of K_{k+1}-free graphs of increasing order, such that $d(C_5, G_n) \to \mathrm{OPT}_k(C_5)$. Let $d(H, G_\infty)$ be the value of $\lim_{n\to\infty} d(H, G_n)$. Then for any exact solution, $d(H, G_\infty) > 0$ implies that H must be a tight subgraph.*

Using semidefinite programming, we verified that the conjectured upper bound of $\mathrm{OPT}_k(C_5)$ was correct for small values of k. In doing so, we were able to guess the correct types and labelled flags to use. It was a greedy process and there may be simpler solutions. This corresponds to the polynomials P_i for $i = 1, \ldots, 6$. Note that each labelled flag is a four-vertex graph appearing in the Turán graph. Next, Lemma 2.3.2 implies that each $F_i \in \mathcal{F}_5$ that is a subgraph of the Turán graph must have the property that $c_{F_i} = \mathrm{OPT}_k(C_5)$. Given this fact, we used SageMath to solve for the correct polynomials $p_i(k)$ and $z(k)$. These agreed with the values calculated by the semidefinite program for small k.

2.4 Stability

In this section we will prove a stability lemma which states that for any K_{k+1}-free graph G on a sufficient number of vertices, if G contains "close" to the extremal number of five-cycles, then G can be made isomorphic to $T_k(n)$ by adding or deleting a small number of edges.

Proposition 2.4.1. *For two positive integers x_1 and x_2, if $x_1 \geq x_2 + 2$, then*

1. $x_1 x_2 < (x_1 - 1)(x_2 + 1)$
2. $x_1 \binom{x_2}{2} < (x_2 + 1)\binom{x_1 - 1}{2}$.

Proof. The first inequality is clear from the equation below:

$$(x_1 - 1)(x_2 + 1) = x_1 x_2 + (x_1 - x_2) - 1 \geq x_1 x_2 + 1.$$

Since $x_2 - 1 < x_1 - 2$, the second inequality follows immediately from the first inequality. ■

The next proposition follows immediately from Lemma 2.4.3, which we will prove next.

Proposition 2.4.2. *For any complete k-partite graph H on n vertices, $\nu(C_5, H)$ is maximized when the sizes of the partite sets are as equal as possible.*

The following lemma will show that if H is a complete k-partite graph with unbalanced partite sets, then we can always increase the number of five-cycles in H by moving the vertices as to make H more balanced. Throughout the proof, we will assume for each $i = 1, \ldots, k$ that $|X_i| = x_i$. For a graph G and a vertex $v \in V(G)$, let $\nu_G(v, C_5)$ denote the number of five-cycles in G that contain v.

Lemma 2.4.3. *Let H be a complete k-partite graph with partite sets $X_1, \ldots, X_k$. Suppose that for two integers i and j,*

$$x_i \geq x_j + 2.$$

Let H' be the graph obtained from H by deleting a vertex in X_i and duplicating a vertex in X_j. Then

$$\nu(C_5, H') > \nu(C_5, H).$$

Proof. By symmetry we may assume that $i = 1$ and $j = 2$. Let H' be obtained from H by removing some vertex $v \in X_1$ and adding a new vertex v' to X_2, where v' is a duplicate of some vertex in X_2. Letting $X_1', \ldots, X_k'$ denote the new partite sets in H', $|X_1'| = x_1 - 1$, $|X_2'| = x_2 + 1$, and $|X_q'| = x_q$ for each $q \in \{3, \ldots, k\}$.

Since the only five-cycles that have been deleted from H are those containing v, we only need to show that $\nu_{H'}(v', C_5) > \nu_H(v, C_5)$. Additionally, there is a one-to-one correspondence between the five cycles in H containing v and no other vertices in $X_1 \cup X_2$ and the five cycles in H' containing v' and no other vertices in $X_1' \cup X_2'$. Because of this, we can focus only on those five-cycles which contained v and at least one other vertex in $X_1 \cup X_2$.

Let $c(v, n_1, n_2)$ denote the number of five cycles in H containing v along with n_1 and n_2 vertices in X_1 and X_2, respectively. We define $c'(v', n_1, n_2)$ in an identical manner, but pertaining to v' and H'. In order to show that $\nu(C_5, H') > \nu(C_5, H)$, it suffices to show the following,

1. $c'(v', 1, 0) + c'(v', 2, 0) + c'(v', 0, 1) > c(v, 0, 1) + c(v, 0, 2) + c(v, 1, 0)$, and
2. $c'(v', 1, 1) + c'(v', 2, 1) > c(v, 1, 1) + c(v, 1, 2)$.

We will prove each of these inequalities as two separate claims. Throughout the proof we will assume that $I = \{3, \ldots, k\}$.

Claim 2.4.4. $c'(v', 1, 0) + c'(v', 2, 0) + c'(v', 0, 1) > c(v, 0, 1) + c(v, 0, 2) + c(v, 1, 0)$.

Proof. Since H is a complete k-partite graph,

$$c(v,0,1) = 6x_2 \cdot \sum_{\{i,j\} \in \binom{I}{2}} \left[\binom{x_i}{2} x_j + \binom{x_j}{2} x_i \right] + 12x_2 \cdot \sum_{\{i,j,h\} \in \binom{I}{3}} x_i x_j x_h,$$

$$c(v,0,2) = 4\binom{x_2}{2} \cdot \sum_{i \in I} \binom{x_i}{2} + 6\binom{x_2}{2} \cdot \sum_{\{i,j\} \in \binom{I}{2}} x_i x_j, \text{ and}$$

$$c(v,1,0) = 4(x_1 - 1) \cdot \sum_{\{i,j\} \in \binom{I}{2}} \left[\binom{x_i}{2} x_j + \binom{x_j}{2} x_i \right] + 6x_1 \cdot \sum_{\{i,j,h\} \in \binom{I}{3}} x_i x_j x_h.$$

By counting in similar way in H',

$$c'(v',1,0) = 6(x_1-1)\cdot \sum_{\{i,j\}\in\binom{I}{2}} \left[\binom{x_i}{2}x_j + \binom{x_j}{2}x_i\right] + 12(x_1-1)\cdot \sum_{\{i,j,h\}\in\binom{I}{3}} x_ix_jx_h,$$

$$c'(v',2,0) = 4\binom{x_1-1}{2}\cdot \sum_{i\in I}\binom{x_i}{2} + 6\binom{x_1-1}{2}\cdot \sum_{\{i,j\}\in\binom{I}{2}} x_ix_j, \text{ and}$$

$$c'(v',0,1) = 4x_2\cdot \sum_{\{i,j\}\in I} \left[\binom{x_i}{2}x_j + \binom{x_j}{2}x_i\right] + 6x_2\cdot \sum_{\{i,j,h\}\in I} x_ix_jx_h.$$

Since $x_1 \geq x_2+2$, it follows that $c'(v',2,0) > c'(v',0,2)$. Thus, it suffices to show that

$$c(v,0,1)+c(v,1,0) \leq c'(v',0,1)+c'(v',1,0).$$

It is straightforward to verify that

$$6\cdot \sum_{\{i,j\}\in\binom{I}{2}} \left[\binom{x_i}{2}x_j + \binom{x_j}{2}x_i\right] + 12\cdot \sum_{\{i,j,h\}\in\binom{I}{3}} x_ix_jx_h \geq 4\cdot \sum_{\{i,j\}\in\binom{I}{2}} \left[\binom{x_i}{2}x_j + \binom{x_j}{2}x_i\right] + 6\cdot \sum_{\{i,j,h\}\in\binom{I}{3}} x_ix_jx_h.$$

This immediately implies that

$$c(v,0,1) - c(v,0,1) \leq c'(v',1,0) - c'(v',1,0),$$

which proves the claim. ■

Claim 2.4.5. $c'(v',1,1)+c'(v',2,1) > c(v,1,1)+c(v,1,2)$.

Proof. For convenience, we will count $c(v,1,1)+c(v,1,2)$ in the following way:

$$c(v,1,1)+c(v,1,2) = x_1x_2f_{11} + \binom{x_2}{2}x_1f_{21}, \tag{2.13}$$

where f_{pq} is a function independent of the values x_1 and x_2 used to count the number of five cycles containing v, p vertices from X_1, and q vertices from X_2. Using the same method to count $c'(v',1,1)+c'(v',2,1)$, we get

$$c'(v',1,1)+c'(v',2,1) = (x_1-1)(x_2+1)f_{11} + \binom{x_1-1}{2}(x_2+1)f_{12}. \tag{2.14}$$

By Proposition 2.4.1,

$$(x_1-1)(x_2+1)f_{11} > x_1x_2f_{11}.$$

Moreover, since the sizes of each set X_j for all $j \in I$ have not changed, $f_{12} = f_{21}$. Therefore,

$$\binom{x_1 - 1}{2}(x_2 + 1)f_{12} > \binom{x_2}{2}x_1 f_{21}$$

by Proposition 2.4.1, completing the proof of the claim. ■

As each of Claims 2.4.4 and 2.4.5 are true, it follows that $\nu(C_5, H') > \nu(C_5, H)$, completing the proof of Lemma 2.4.3. ■

For two graphs G and H of the same order, let $\mathrm{Dist}(G, H)$ equal the minimum number of adjacencies that one needs to change in G in order to obtain a graph isomorphic to H. The parameter $\mathrm{Dist}(G, H)$ is commonly known as the *edit distance* between G and H. Our main goal of this section is to prove the following lemma.

Lemma 2.4.6 (Stability Lemma)**.** *For every $\varepsilon > 0$, there exists an n_0 and $\varepsilon_F > 0$ such that for every K_{k+1}-free graph G of order $n \geq n_0$ with $d(C_5, G) \geq OPT_k(C_5) - \varepsilon_F$, the edit distance between G and $T_k(n)$ is at most εn^2.*

The proof of Lemma 2.4.6 requires the following two lemmas along with Lemma 2.3.2. For a family of graphs $\mathcal{F}$, we say that a graph G is *$\mathcal{F}$-free* if G does not contain any member of $\mathcal{F}$ as an induced subgraph.

Lemma 2.4.7 (Induced Removal Lemma [1])**.** *Let $\mathcal{F}$ be a set of graphs. For each $\varepsilon > 0$, there exist $n_0 \geq 0$ and $\delta > 0$ such that for every graph G of order $n_0 \geq n$, if G contains at most $\delta n^{|V(H)|}$ induced copies of H for every $H \in \mathcal{F}$, then G can be made $\mathcal{F}$-free by removing or adding at most εn^2 edges from G.*

Let $(G_n)_{n\geq 1}$ be a convergent sequence of K_{k+1}-free graphs. In the proof of Theorem 2.2.4(i), we found constants c_{F_i} for each $F_i \in \mathcal{F}_5$ such that

$$d(C_5, G_n) \leq \sum_{i=0}^{33} c_{F_i} F_i \leq \max\{c_{F_i} : F_i \in \mathcal{F}_5\}$$

and

$$\max\{c_{F_i} : F_i \in \mathcal{F}_5\} = \mathrm{OPT}_k(C_5) = \frac{1}{k^4}(12k^4 - 60k^3 + 120k^2 - 120k + 48).$$

Let $\overline{P_3}$ be the three vertex graph with exactly one edge; see Figure 2.1. The goal of Lemma 2.4.8 is to prove that if $\lim_{n\to\infty}(C_5, G_n) = \mathrm{OPT}_k(C_5)$, then $\lim_{n\to\infty}(\overline{P_3}, G_n) = 0$.

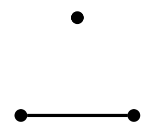
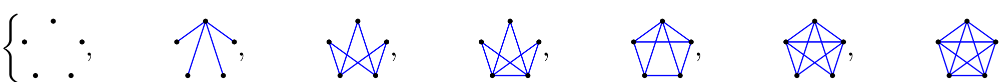


Figure 2.1 $\overline{P_3}$

Lemma 2.4.8. *For each $\delta_F > 0$, there exists $\varepsilon_F > 0$ and $n_0 = n_0(\delta_F)$ such that any K_{k+1}-free graph G on $n \geq n_0$ vertices with $d(C_5, G) > OPT_k(C_5) - \varepsilon_F$ contains at most $\delta_F n^3$ induced copies of $\overline{P_3}$.*

Proof. Let $(G_n)_{n\geq 1}$ be a convergent sequence of K_{k+1}-free graphs maximizing the number of five-cycles. Let T be the set of tight subgraphs in $\mathcal{F}_5$ given by the proof of Theorem 2.2.4(i). This is the same set T provided at the end of Section 2.

$T = \Big\{$  $\Big\}.$

Observe that for each graph $F \in T$,

$$P(\overline{P_3}, F) = 0.$$

Since T contains the set of tight graphs, the following is a consequence of (2.4) for the sequence $(G_n)_{n\geq 1}$,

$$\overline{P_3} = \sum_{i=0}^{33} P(\overline{P_3}, F_i)F_i = 0.$$

It follows that for the sequence $(G_n)_{n\geq 1}$,

$$\lim_{n\to\infty} d(\overline{P_3}, G_n) = 0,$$

which completes the proof of Lemma 2.4.8. ∎

Proof of Lemma 2.4.6. Let $\varepsilon_I > 0$ and $\varepsilon_F > 0$, which we will determine later. By Lemma 2.4.7, there exists a $\delta_F > 0$ and an n_0 such that any K_{k+1}-free graph on $n \geq n_0$ vertices containing at most $\delta_F n^3$ copies of $\overline{P_3}$ can be made $\overline{P_3}$-free after changing at most $\varepsilon_I n^2$ adjacencies. Assume that G is a graph on $n \geq n_0$ vertices such that

$$d(C_5, G) > OPT_k(C_5) - \varepsilon_F,$$

where n_0 is large enough to satisfy the conditions of Lemmas 2.4.7 and 2.4.8 so that G contains at most $\delta_F n^3$ copies of $\overline{P_3}$. Moreover, for sufficiently small ε_I,

$$d(C_5, G) > OPT_{k-1}(C_5) + 2 \cdot 5! \cdot \varepsilon_I.$$

Using Lemma 2.4.7, let G' be a $\overline{P_3}$-free graph obtained from G by changing at most $\varepsilon_I n^2$ edges. Since each edge that was removed in this way was contained in at most n^3 copies of C_5, $\nu(C_5, G') \geq \nu(C_5, G) - \varepsilon_I n^5$. Therefore,

1. $d(C_5, G') > OPT_k(C_5) - 5! \cdot \varepsilon_I - \varepsilon_F$,

2. $d(C_5, G') > OPT_{k-1}(C_5) + 5! \cdot \varepsilon_I$.

Using the previous two inequalities, along with the fact that G' is $\overline{P_3}$-free, we will now show that G' must be a complete k-partite graph.

Claim 2.4.9. *G' is a complete k-partite graph.*

Proof. Since G' does not contain any induced copies of $\overline{P_3}$ as a subgraph, each pair of non-adjacent vertices must have an identical neighborhood. Therefore, we can partition $V(G')$ into independent sets $X_1, \ldots, X_\ell$ such that for all distinct $i, j \in [\ell]$, each vertex in X_i is adjacent to each vertex in X_j. Hence, G' is a complete ℓ-partite graph. Since

$$OPT_{k-1}(C_5) = \lim_{n \to \infty} d(C_5, T_{k-1}(n))$$

and $d(C_5, G') > OPT_{k-1}(C_5) + 5! \cdot \varepsilon_I$, Proposition 2.4.2 implies that G' must be k-partite if n is sufficiently large. ■

At this point, we know that G' only differs from $T_k(n)$ in the sizes of the partite sets $X_1, X_2, \ldots, X_k$. The next claim will show that we can impose that the partite sets in G' must be reasonably close to being balanced.

Claim 2.4.10. *Let G' be a complete k-partite graph with partite sets $X_1, X_2, \ldots, X_k$. Then for any $\varepsilon_T > 0$, there exists $\delta > 0$ such that if*

$$d(C_5, G') > OPT_k(C_5) - \delta,$$

then for each $i = 1, \ldots, k$

$$\frac{n(1-\varepsilon_T)}{k} \leq |X_i| \leq \frac{n(1+\varepsilon_T)}{k}.$$

Proof. For each $i = 1, \ldots, k$ let $x_i = |X_i|$. Let $\varepsilon'(k-1) > \varepsilon_T$ and assume by symmetry that $x_1 = \frac{1+\varepsilon'(k-1)}{k} n$. If we picked $x_1 = \frac{1+\varepsilon'}{k} n$, we would get less pleasant expressions in what follows. We want to calculate an upper bound on $d(C_5, G')$. By Lemma 2.4.3, $d(C_5, G')$ is maximized if all remaining parts are balanced. That is, $x_i = \frac{1-\varepsilon'}{k} n$ for $i = 2, \ldots, k$. With knowing the sizes of all X_i, the following is a straightforward calculation,

$$\begin{aligned} d(C_5, G') \leq OPT_k(C_5) &- 60\varepsilon'^2 \left(1 - \frac{6}{k} + \frac{15}{k^2} - \frac{18}{k^3} + \frac{8}{k^4}\right) + 60\varepsilon'^3 \left(1 - \frac{8}{k} + \frac{25}{k^2} - \frac{34}{k^3} + \frac{16}{k^4}\right) \\ &+ 180\varepsilon'^4 \left(\frac{1}{k} - \frac{5}{k^2} + \frac{8}{k^3} - \frac{4}{k^4}\right) - 12\varepsilon'^5 \left(1 - \frac{15}{k^2} + \frac{30}{k^3} - \frac{16}{k^4}\right) + o(1), \end{aligned}$$

see Appendix 2.8.2 for a code in SageMath.

For all $k \geq 3$, the term $1 - \frac{6}{k} + \frac{15}{k^2} - \frac{18}{k^3} + \frac{8}{k^4}$ is positive with minimum $\frac{8}{81}$ at $k = 3$. For sufficiently small ε' and large n, we get

$$d(C_5, G') \leq OPT_k(C_5) - 5\varepsilon'^2.$$

This implies the statement of the claim. ∎

Returning to the proof of Lemma 2.4.6, suppose that $\varepsilon > 0$, and let $\varepsilon_T = \varepsilon/2$. Next, choose an $\varepsilon_I \leq \varepsilon/2$ small enough so that ε_F and δ_F are sufficiently small. In particular, we must select ε_I, ε_F, and δ_F so that any k-partite graph G' satisfying $d(C_5, G') > OPT_k(C_5) - 5!\varepsilon_I - \varepsilon_F$, must have partite sets $X_1, \ldots, X_k$ that satisfy

$$\frac{n(1-\varepsilon/2)}{k} \leq |X_i| \leq \frac{n(1+\varepsilon/2)}{k}$$

for all $i = 1, \ldots, k$. Then by changing at most $(\varepsilon_I + \varepsilon_T)n^2$ pairs we can obtain $T_k(n)$ from the original graph G, which completes the proof of Lemma 2.4.6. ∎

2.5 Exact Result

In this section we will prove Theorem 2.2.4(ii). First we will give a brief outline. As we have shown, if G is a K_{k+1}-free graph on n vertices for large enough n that contains close to the extremal number of five-cycles, then the edit distance between G and $T_k(n)$ is very small. Given such a graph G, the process of deleting and adding the necessary edges to transform G into the Turán graph actually increases the number of five-cycles. This will prove that $T_k(n)$ is the unique extremal graph for large enough n.

Proof of Theorem 2.2.4(ii). Suppose that $k \geq 3$. By Lemma 2.4.6, there exists an $\varepsilon > 0$ and an integer $n_0 = n(k, \varepsilon)$ so that for any K_{k+1}-free graph G on $n \geq n_0$ vertices satisfying

$$d(C_5, G) > OPT_k(C_5) - \varepsilon,$$

we have that $\mathrm{Dist}(G, T_k(n)) \leq \frac{1}{2k^{10}} n^2$.

Let G be a graph on n vertices, where n is sufficiently large. In particular, $n \geq \max\{n(k, \varepsilon), 2k^5 + 1\}$ and G satisfies

$$d(C_5, G) > OPT_k(C_5) - \varepsilon',$$

where $\varepsilon' \leq \min\{\varepsilon, \frac{1}{k^{10}}\}$. Lemma 2.4.6 gives a partition of $V(G)$ into k sets $X_1, X_2, \ldots, X_k$, where $\lfloor \frac{n}{k} \rfloor \leq |X_i| \leq \lceil \frac{n}{k} \rceil$ for all $i = 1, \ldots, k$, so that by changing at most $\frac{1}{2k^{10}} n^2$ pairs uv for $u, v \in V(G)$, we can construct a new graph G' from G so that G' is isomorphic to $T_k(n)$ and the partite sets of G' are $X_1, X_2, \ldots, X_k$.

Call each edge that is removed in this process a *surplus edge* and call each edge that is added in this process a *missing edge*. For each vertex $v \in V(G)$, let f_v denote the sum of the total number of surplus edges and missing edges incident to v. Define the set X_0 to contain each vertex v with $f_v > \frac{1}{k^6} n$. We will refer to each vertex in X_0 as a *bad vertex*.

Claim 2.5.1. $|X_0| \leq \frac{1}{k^4} n$.

Proof. Since $f_v > \frac{1}{k^6}n$ for each vertex $v \in X_0$ and the combined total of surplus edges and missing edges in G is at most $\frac{1}{2k^{10}}n^2$, it follows that

$$\frac{1}{k^6}n|X_0| \leq \frac{1}{k^{10}}n^2,$$

which proves Claim 2.5.1. ■

For a graph G and a vertex $v \in V(G)$ let $N_G(v)$ denote the neighborhood of v in G. For all $v \in V(G)$, let $d_i(v)$ denote the size of the set $N_G(v) \cap (X_i \setminus X_0)$. Let

$$d^*(v) = \sum_{i=1}^{k} d_i(v).$$

By Claim 2.5.1,

$$\left\lfloor \frac{n}{k} \right\rfloor - \frac{1}{k^4}n \leq |X_i \setminus X_0| \leq \left\lceil \frac{n}{k} \right\rceil,$$

for all $i = 1, \ldots, k$. Thus, for each vertex v not contained in X_0,

$$d^*(v) \geq \left(\frac{k-1}{k} - \frac{1}{k^4} - \frac{1}{k^6} \right) n.$$

For two vertices u and v in a graph G, let $N_G(u, v)$ denote the *common neighborhood* of u and v, which is the set of all vertices in G adjacent to both u and v.

Claim 2.5.2. *There are no surplus edges in $G - X_0$.*

Proof. Assume by way of contradiction that $G - X_0$ contains a surplus edge uv. Our goal is to show that it would be in K_{k+1}. Since uv is removed in the the process of transforming G into the Turán graph, we may assume by symmetry that u and v are contained in the same set X_1. Since neither vertex is contained in X_0,

$$\min\{d_j(v), d_j(u)\} \geq \left(\frac{1}{k} - \frac{1}{k^4} - \frac{1}{k^6} \right) n$$

for each $j = 2, \ldots, k$. Therefore,

$$|N_G(u, v) \cap (X_2 \setminus X_0)| \geq \left(\frac{1}{k} - \frac{1}{k^4} - \frac{2}{k^6} \right) n > 0.$$

Pick one vertex w_2 contained in $N_G(u,v) \cap (X_2 \setminus X_0)$. Since w_2 is not contained in X_0,

$$|N_G(w_2) \cap N_G(u,v) \cap (X_3 \setminus X_0)| \geq \left(\frac{1}{k} - \frac{1}{k^4} - \frac{3}{k^6}\right) n > 0.$$

This implies that we can find some common neighbor, say w_3, of u, v, and w_2, where $w_3 \in X_3 \setminus X_0$. We continue the process of a selecting a vertex $w_j \in X_j \setminus X_0$ in the common neighborhood of the set $\{u, v, w_1, \ldots, w_{j-1}\}$ for all $j = 4, \ldots, k$. This is possible because after selecting w_{j-1}, the common neighborhood of the set $\{u, v, w_2, \ldots, w_{j-1}\}$ contains at least

$$\left(\frac{1}{k} - \frac{1}{k^4} - \frac{j}{k^6}\right) n > 0$$

vertices in $X_j \setminus X_0$ for all $j = 4, \ldots, k$. This implies, however, that the set $\{u, v, w_2, \ldots, w_k\}$ obtained by selecting a vertex in this way from each partite set $X_2, \ldots, X_k$ induces a copy of K_{k+1} in G, which is a contradiction. ■

An immediate consequence of Claim 2.5.2 is that every surplus edge in G is incident to at least one vertex in X_0, implying that $G - X_0$ is a k-partite graph, albeit not necessarily complete k-partite. We will split the vertices of X_0 into two classes. For each vertex $v \in X_0$, one of the following holds.

1. There exists some index $i \in \{1, 2, \ldots, k\}$ such that $d_i(v) = 0$. In this case we will call v a *type 1* vertex, or

2. $d_i(v) > 0$ for all $i = 1, \ldots, n$. In this case we will call v a *type 2* vertex.

As we are trying to show that every extremal graph is a complete balanced k-partite graph, we will now prove that G cannot contain any type 2 vertices. First in Claim 2.5.3, we will prove that if v is a type 2 vertex, then $d^*(v)$ must be relatively small. In Claim 2.5.4, we will prove a lower bound on the number of five-cycles containing a vertex v. Finally, in Claim 2.5.5, we will show that a type 2 vertex cannot be contained in enough five-cycles to justify the claim that G is an extremal graph.

Claim 2.5.3. *Let $v \in X_0$ be a type 2 vertex. Then there exist distinct integers i and j where $1 \le i, j \le k$ such that*

$$1 \le d_i(v) \le d_j(v) \le \frac{1}{k^5}n.$$

Proof. By symmetry, assume that $1 \le d_1(v)$ and $d_1(v) \le d_q(v)$ for all $q = 2, \dots, k$. For contradiction, assume $d_q(v) > \frac{1}{k^5}n$ for all $q = 2, \dots, k$. Let $w_1 \in X_1 \setminus X_0$ be adjacent to v in G. Since $w_1 \notin X_0$,

$$|N_G(v, w_1) \cap (X_2 \setminus X_0)| \ge \frac{1}{k^5}n - \frac{1}{k^6}n,$$

implying that there exists a vertex $w_2 \in X_2 \setminus X_0$ for which the set $\{v, w_1, w_2\}$ induces a triangle in G. If we continue selecting vertices in this way, then for all $q = 3, \dots, k$, there are at least

$$\frac{1}{k^5}n - \frac{q-1}{k^6}n > 0$$

vertices in $X_q \setminus X_0$ that are adjacent to all of the previously selected vertices $v, w_1, \dots, w_{q-1}$. This implies that we can select k vertices $w_1, \dots, w_k$ so that the set $\{v, w_1, \dots, w_k\}$ induces a copy of K_{k+1} in G, which is a contradiction. Therefore, there exists an index $j \in \{2, \dots, k\}$ for which $1 \le d_1(v) \le d_j(v) \le \frac{1}{k^5}n$, completing the proof of Claim 2.5.3. ∎

Claim 2.5.4. *For all $k \ge 3$, and $v \in V(G)$, $\nu_G(v, C_5) \ge (OPT_k(C_5) - \frac{1}{k^{10}})\binom{n}{4} - \frac{1}{k^5}n^4$.*

Proof. Suppose by way of contradiction that there exists some vertex v for which

$$\nu_G(v, C_5) < \left(OPT_k(C_5) - \frac{1}{k^{10}}\right)\binom{n}{4} - \frac{1}{k^5}n^4.$$

Since $d(C_5, G) > \mathrm{OPT}_k(C_5) - \frac{1}{k^{10}}$, it follows by averaging that there exists some vertex $u \in V(G)$ for which

$$\nu_G(u, C_5) \ge \left(OPT_k(C_5) - \frac{1}{k^{10}}\right)\binom{n}{4}.$$

Let $\nu_G(\{u, v\}, C_5)$ denote the number of five-cycles containing both u and v. Then

$$\nu_G(\{u, v\}, C_5) \le 2n^3.$$

Let G' be the graph obtained from G by deleting v and replacing it with a copy u' of u. Since there is no edge between u' and u, G' is also K_{k+1}-free. As there were previously $\nu(\{u,v\},C_5)$ five-cycles containing u and v,

$$\nu(C_5,G') - \nu(C_5,G) \geq \nu_G(u,C_5) - \nu_G(v,C_5) - \nu_G(\{u,v\},C_5) \geq \frac{1}{k^5}n^4 - 2n^3 > 0$$

since $n > 2k^5$. This, however, contradicts the assumption that G is an extremal graph as $\nu(C_5,G') > \nu(C_5,G)$. Therefore, if n is sufficiently large it follows that for each $v \in V(G)$,

$$\nu_G(v,C_5) \geq \left(OPT_k(C_5) - \frac{1}{k^{10}}\right)\binom{n}{4} - \frac{1}{k^5}n^4,$$

which completes the proof of Claim 2.5.4. ■

In Claims 2.5.5 and 2.5.7 we count the number of 5-cycles containing a particular vertex $v \in X_0$. We use the following argument repeatedly. We want to count the number of 5-cycles $vu_1u_2u_3u_4v$, where v is in X_0, $u_1 \in X_i$, $u_4 \in X_j$ and $u_2, u_3 \in V(G) \setminus X_0$. Assume we already picked u_1 and u_4 and want to count the number of choices for u_2 and u_3. We distinguish two cases.

1. $i = j$: First u_2 can be in any of the remaining $k-1$ parts. Then u_3 has $k-2$ choices for a part to complete the 5-cycle as it needs to avoid the parts containing u_2 and u_4 and these are distinct. After multiplying by n^2/k^2, the number of choices for u_2 and u_3 in each of the selected parts, we get
$$\frac{(k-1)(k-2)}{k^2}n^2$$
choices for u_2 and u_3 together.

2. $i \neq j$: We further distinguish two cases. If $u_2 \notin X_j$, then there are $k-2$ parts which could contain u_2 and $k-2$ parts which could contain u_3. If $u_2 \in X_j$, then there are $k-1$ parts which could contain u_3. After including the number of choices in each part, we get
$$\left(\frac{(k-2)^2}{k^2} + \frac{k-1}{k^2}\right)n^2$$
choices for u_2 and u_3 together.

Claim 2.5.5. *G does not contain any type 2 vertices.*

Proof. Assume for contradiction that $v \in X_0$ is a type 2 vertex. Then by Claim 2.5.3 there are two sets, say X_1 and X_2, such that

$$1 \leq d_1(v) \leq d_2(v) \leq \frac{1}{k^5}n.$$

We will now provide an upper bound on the value of $\nu_G(v, C_5)$. We will count the maximum number of such five-cycles of the form $vu_1u_2u_3u_4v$ based on the locations of u_1 and u_4 as follows:

1. If $u_1, u_4 \in X_1 \setminus X_0$ or $u_1, u_4 \in X_2 \setminus X_0$:
$$2\binom{\frac{n}{k^5}}{2}\frac{(k-1)(k-2)}{k^2}n^2. \tag{2.15}$$

2. $u_1 \in X_1 \setminus X_0$ and $u_4 \in X_2 \setminus X_0$:
$$\left(\frac{n}{k^5}\right)^2\left(\frac{(k-2)^2}{k^2} + \frac{k-1}{k^2}\right)n^2. \tag{2.16}$$

3. $u_1 \in (X_1 \setminus X_0) \cup (X_2 \setminus X_0)$ and $u_4 \notin X_1 \cup X_2$:
$$\frac{2n}{k^5} \cdot \frac{n}{k}\left(\frac{(k-2)^2}{k^2} + \frac{k-1}{k^2}\right)n^2. \tag{2.17}$$

4. $u_1, u_4 \notin X_1 \cup X_2$:
$$(k-2) \cdot \binom{\frac{n}{k}}{2} \cdot \frac{(k-1)(k-2)}{k^2}n^2 + \binom{k-2}{2} \cdot \frac{n^2}{k^2} \cdot \left(\frac{(k-2)^2}{k^2} + \frac{k-1}{k^2}\right)n^2. \tag{2.18}$$

Finally, there are at most $\frac{2n^4}{k^4}$ five-cycles containing v and at least one other vertex in X_0. Combining this, along with the upper bounds obtained in equations (2.15)–(2.18),

$$\nu_G(v, C_5) \leq \frac{n^4}{24}\left(12 - \frac{84}{k} + \frac{228}{k^2} - \frac{300}{k^3} + \frac{216}{k^4} + \frac{48}{k^6} - \frac{144}{k^7} + \frac{144}{k^8} + \frac{48}{k^{10}} - \frac{144}{k^{11}} + \frac{120}{k^{12}}\right).$$

The SageMath code for verifying this fact can be found in Appendix 2.8.3. This implies that for large enough n,

$$\left(OPT_k(C_5) - \frac{1}{k^{10}}\right)\binom{n}{4} - \nu_G(v, C_5) \geq \frac{1}{k^5}n^4.$$

Using SageMath, we verified that this was true for $3 \leq k \leq 1000$. After that, it is straightforward to check the coefficients in order to verify this fact. This contradicts Claim 2.5.4 since G was assumed to be an extremal graph. Therefore, G does not contain any type 2 vertices. ■

Since G does not contain any type 2 vertices, we can place each vertex $v \in X_0$ into the set X_i for which $d_i(v) = 0$. In order to show that G is a complete k-partite graph, we must show that any pair of vertices u and v that were in X_0 and go to the same X_i cannot be adjacent. The next claim will provide an upper bound on the "good degree" of at least one of these adjacent vertices.

Claim 2.5.6. *Suppose that u and v are two adjacent type 1 vertices such that $d_j(u) = d_j(v) = 0$ for some index $j \in \{1, \ldots, k\}$. Then there exists some index $i \in \{1, \ldots, k\}$ such that $i \neq j$ and*

$$d_i(u) \leq \frac{k^2+1}{2k^3}n \text{ or } d_i(v) \leq \frac{k^2+1}{2k^3}n.$$

Proof. By symmetry we may assume that $j = 1$. Assume for contradiction that

$$|N_G(u,v) \cap (X_i \setminus X_0)| > \frac{1}{k^3}n$$

for all $i = 2, \ldots, k$. Using an identical argument to the one made in the proof of Claim 2.5.3, there exists a set $\{w_2, \ldots, w_k\}$ such that $w_i \in (X_i \setminus X_0)$ and the set $\{u, v, w_2, \ldots, w_k\}$ induces a K_{k+1} in G, which is a contradiction. This implies that for at least one index i,

$$|N_G(u,v) \cap (X_i \setminus X_0)| \leq \frac{1}{k^3}n.$$

Without loss of generality assume that $d_i(u) \leq d_i(v)$. Then

$$d_i(u) \leq \frac{n}{2}\left(\frac{1}{k} - \frac{1}{k^3}\right) + \frac{n}{k^3} = \frac{k^2+1}{2k^3}n,$$

which completes the proof of Claim 2.5.6. ∎

We will now show that the vertex u of low degree described in the previous claim cannot be contained in enough five-cycles to justify the assumption that G is an extremal graph. Unlike Claim 2.5.5, we will only show that the two vertices u and v from Claim 2.5.6 cannot be adjacent.

Claim 2.5.7. *Suppose that u and v are type 1 vertices such that $d_j(u) = d_j(v) = 0$ for some $j = 1, \ldots, k$. Then u and v are not adjacent.*

Proof. By symmetry we may assume that $d_1(u) = d_1(v) = 0$. Assume for contradiction that u and v are adjacent. By symmetry and Claim 2.5.6, we may assume that $d_1(u) = 0$ and

$$d_2(u) \leq \frac{k^2+1}{2k^3}n.$$

In a similar manner as in Claim 2.5.5, we will count the number of five-cycles of the form $uv_1v_2v_3v_4u$ incident to u by considering the possibilities for the locations of v_1 and v_4 as follows:

1. $v_1, v_4 \in X_2 \setminus X_0$:
$$\binom{\frac{k^2+1}{2k^3}n}{2}\frac{(k-1)(k-2)}{k^2}n^2. \tag{2.19}$$

2. $v_1 \in X_2 \setminus X_0$ and $v_4 \notin X_2$:
$$\frac{k^2+1}{2k^3}\cdot\frac{k-2}{k}\left(\frac{(k-2)^2}{k^2}+\frac{k-1}{k^2}\right)n^4. \tag{2.20}$$

3. $v_1, v_4 \notin X_2$:
$$(k-2)\cdot\binom{\frac{n}{k}}{2}\cdot\frac{(k-1)(k-2)}{k^2}n^2+\binom{k-2}{2}\cdot\frac{n^2}{k^2}\cdot\left(\frac{(k-2)^2}{k^2}+\frac{k-1}{k^2}\right)n^2. \tag{2.21}$$

There are at most $\frac{2}{k^4}n^4$ five-cycles containing u and at least one other vertex in X_0. Combining this along with equations (2.19)–(2.21),

$$\nu_G(u, C_5) \leq \left(12-\frac{72}{k}+\frac{171}{k^2}-\frac{189}{k^3}+\frac{96}{k^4}+\frac{90}{k^5}+\frac{57}{k^6}-\frac{9}{k^7}+\frac{6}{k^8}\right)\frac{n^4}{24}.$$

For $k > 1000$ it is clear that

$$12-\frac{72}{k}+\frac{171}{k^2}-\frac{189}{k^3}+\frac{96}{k^4}+\frac{90}{k^5}+\frac{57}{k^6}-\frac{9}{k^7}+\frac{6}{k^8} \leq \mathrm{OPT}_k(C_5)-\frac{1}{k^{10}}.$$

The SageMath code for verifying that this is also true for $3 \leq k \leq 1000$ found in Appendix 2.8.4. Given this fact, it is straightforward to verify that

$$\left(\mathrm{OPT}_k(C_5)-\frac{1}{k^{10}}\right)\binom{n}{4}-\nu_G(u, C_5) \geq \frac{1}{k^5}n^4$$

for large enough n. This, however contradicts Claim 2.5.4, which implies that u and v are not adjacent. ■

Claim 2.5.7 implies that if u and v are type 1 vertices for which $d_i(v) = d_i(u) = 0$, then u and v cannot be adjacent. This means that we can place each type 1 vertex v into the set X_i for which $d_i = 0$. Since G does not contain any type 2 vertices, this implies that G is a k-partite graph. Since G maximizes the number of (possibly non-induced) C_5 subgraphs, it follows that G must be a complete k-partite graph. Finally, Proposition 2.4.2 implies that G is isomorphic to $T_k(n)$, implying that for large enough n, the Turán graph $T_k(n)$ is the unique extremal graph maximizing the number of C_5 subgraphs. ■

2.6 Conclusion

In [17], Palmer and Gerbner showed that if H is a graph and F is a graph with chromatic number $k+1$, then

$$\mathrm{ex}(n, H, F) \leq \mathrm{ex}(n, H, K_{k+1}) + o(n^{|H|}).$$

Since the Turán graph $T_k(n)$ does not contain any $(k+1)$-chromatic graph as a subgraph, this immediately implies that for any $(k+1)$-chromatic graph F,

$$\lim_{n\to\infty} d(C_5, F) = \frac{1}{k^4}(12k^4 - 60k^3 + 120k^2 - 120k + 48),$$

which closely resembles the Erdős-Stone-Simonovits theorem.

Let G be a graph with chromatic number k. Then for any $r \geq k$, the Turán graph $T_r(n)$ contains G as a subgraph. When trying to maximize the copies of G among K_{r+1}-free graphs, evidence seems to suggest that $T_r(n)$ is extremal, as we have shown to be the case with five-cycles. While a complete r-partite graph seems to frequently be the best option, it is not always optimal to balance the partite sets. Let S_t be a star with t leaves, also known as $K_{1,t}$. Is it easy to see $\mathrm{ex}(n, S_4, K_3)$ is achieved by an unbalanced bipartite graph. For more detailed treatment of stars, see Cutler, Nir, and Radcliffe [10]. It seems very likely that while the Turán graph is not always extremal, that some complete r-partite graph will be best possible.

Conjecture 2.6.1. *Let G be a graph and let $k > \chi(G)$ be an integer. Then for all $r \geq k$, $ex(n, G, K_r)$ is realized by a complete $(r-1)$-partite graph.*

While an unbalanced r-partite graph might best possible in some cases, we believe that for large enough $r \geq \chi(G)$, the value of $\mathrm{ex}(n, G, K_{r+1})$ is realized by the Turán graph. As r increases, any G-subgraph in $T_r(n)$ can be taken from an increasing number of partite sets. Thus, as r grows larger, the effect of G being unbalanced becomes minimized. The following conjecture also appears in [18].

Conjecture 2.6.2. *Let G be a graph and let $r > |V(G)|$ be an integer. Then $ex(n, G, K_r)$ is realized by the Turán graph $T_{r-1}(n)$.*

Acknowledgement

We would like to thank the anonymous referees for their comments that greatly improved the presentation of the paper.

2.7 References

[1] N. Alon, E. Fischer, M. Krivelevich, and M. Szegedy. Efficient testing of large graphs. *Combinatorica*, 20:451–476, 2000.

[2] N. Alon and C. Shikhelman. Many T copies in H-free graphs. *J. Combin. Theory Ser. B*, 121:146–172, 2016.

[3] R. Baber. Some results in extremal combinatorics. *Dissertation*, 2011.

[4] R. Baber and J. Talbot. Hypergraphs do jump. *Combin. Probab. Comput.*, 20(2):161–171, 2011.

[5] J. Balogh, P. Hu, B. Lidický, O. Pikhurko, B. Udvari, and J. Volec. Minimum number of monotone subsequences of length 4 in permutations. *Combin. Probab. Comput.*, 24(4):658–679, 2014.

[6] P. Bennet, A. Dudek, B. Lidický, and O. Pikhurko. Minimizing the number of 5-cycles in graphs with given edge-density. *Combin. Probab. Comput.*, 29:44–46, 2020.

[7] A. Blumenthal, B. Lidický, Y. Pehova, F. Pfender, O. Pikhurko, and J. Volec. Sharp bounds for decomposing graphs into edges and triangles. *Accepted to Combin. Probab. Comput.*, 2020.

[8] B. Bollobás and E. Győri. Pentagons vs. Triangles. *Discrete Math.*, 308(19):4332–4336, 2008.

[9] I. Choi, B. Lidický, and F. Pfender. Inducibility of directed paths. *Discrete Math.*, 343:112015, 2020.

[10] J. Cutler, J. Nir, and A. Radcliffe. Supersaturation for subgraph counts. Preprint available as arXiv:1903.08059, 2019.

[11] P. Erdős. On some problems in graph theory, combinatorial analysis and combinatorial number theory. *Graph theory and combinatorics (Cambridge, 1983), Academic Press, London*, pages 1–17, 1984.

[12] P. Erdős. On some problems in graph theory, combinatorial analysis and combinatorial number theory. *Graph Theory and Combinatorics, Proc. Conf. Hon. P. Erdős, Cambridge 1983*, pages 1–17, 1984.

[13] B. Ergemlidze, A. Methuku, N. Salia, and E. Győri. A note on the maximum number of triangles in a C_5-free graph. *J. Graph Theory*, 90(3):227–230, 2019.

[14] V. Falgas-Ravry and E. R. Vaughan. Applications of the semi-definite method to the turán density problem for 3-graphs. *Combin. Probab. Comput.*, 22(1):21–54, 2013.

[15] D. Gerbner. Generalized Turán problems for small graphs. Preprint available as arXiv:2006.16150, 2020.

[16] D. Gerbner, E. Győri, A. Methuku, and M. Vizer. Generalized turán problems for even cycles. *Journal of Combinatorial Theory, Series B*, 145:169–213, 2020.

[17] D. Gerbner and C. Palmer. Counting copies of a fixed subgraph in F-free graphs. Preprint available as arXiv:1805.07520, 2018.

[18] D. Gerbner and C. Palmer. Some exact results for generalized Turán problems. Preprint available as arXiv:2006.03756, 2020.

[19] L. Gishboliner and A. Shapira. A generalized Turán problem and its applications. *International Mathematics Research Notices*, 2020(11):3417–3452, 2018.

[20] R. Glebov, D. Král', and J. Volec. A problem of Erdős and Sós on 3-graphs. *The Seventh European Conference on Combinatorics, Graph Theory and Applications*, pages 3–8, 2013.

[21] A. Grzesik. On the maximum number of five-cycles in a triangle-free graph. *J. Combin. Theory Ser. B*, 102(5):1061–1066, 2012.

[22] E. Győri. On the number of C_5's in a triangle-free graph. *Combinatorica*, 9:101–102, 1989.

[23] E. Győri, J. Pach, and M. Simonovits. On the maximial number of certain subgraphs in K_r-free graphs. *Graphs Combin.*, 7(1):31–37, 1991.

[24] H. Hatami, J. Hladký, D. Král', S. Norine, and A. Razborov. On the number of pentagons in triangle-free graphs. *J. Combin. Theory Ser. A*, 120(3):722–732, 2013.

[25] T. Huynh, G. Joret, and D. R. Wood. Subgraph densities in a surface. Preprint available as arXiv:2003.13777, 2020.

[26] B. Lidický and F. Pfender. Pentagons in triangle-free graphs. *European J. Combin.*, 74:85–89, 2018.

[27] W. Mantel. Problem 28. *Winkundige Opgaven*, 10:60–61, 1907.

[28] T. Michael. Cycles of length 5 in triangle-free graphs: a sporadic counterexample to a characterization of equality. *Bulletin of the Institute of Combinatorics and its Applications*, 67, 2013.

[29] D. Mubayi and S. Mukherjee. Triangles in graphs without bipartite suspensions. Preprint available as arXiv:2004.11930, 2020.

[30] C. Palmer. Generalized Turán problems. *Talk given at MMMM in 2018*, 2018.

[31] O. Pikhurko. The minimum size of 3-graphs without a 4-set spanning no or exactly three edges. *European J. Combin.*, 32(7):1142–1155, 2011.

[32] A. A. Razborov. Flag algebras. *J. Symbolic Logic*, 72(4):1239–1282, 2007.

[33] M. K. D. C. Silva, F. M. D. O. Filho, and C. M. Sato. Flag algebras: A first glance. Preprint available as arXiv:1607.04741, 2016.

[34] P. Turán. Eine extremalaufgabe aus der graphentheorie. *Mat. Fiz. Lapok*, 48:436–452, 1941.

[35] A. A. Zykov. On some properties of linear complexes. *Matematicheskii sbornik*, 66:163–188, 1949.

2.8 Appendix

SageMath code on the following pages can be also obtained at

https://arxiv.org/abs/2007.03064.

Table 2.1 Graphs on 5 vertices up to isomorphism.

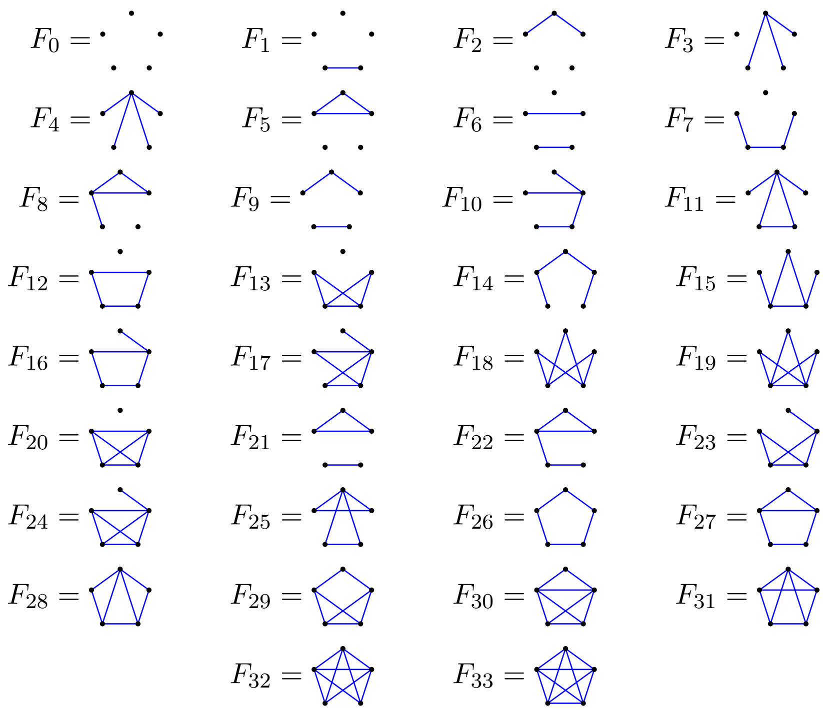
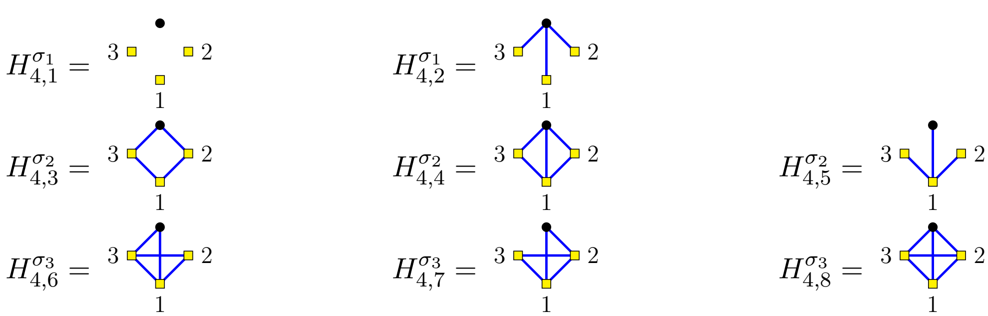


Table 2.2 Labeled graphs on four vertices.



2.8.1 Proof of Claim 2.3.1

```
# SageMath code for Claim 2.1

var('k')

# Vector containing coefficients for each graph in F_5
cFi = [0]*34

# Z(K)
Zk = [(k*k*k*k-10*k*k*k+35*k*k-50*k+24)/k^4]*34
Zk[33] = (-10*k*k*k+35*k*k-50*k+24)/k^4

# P_i(K) first init with 0
P1k = [0]*34
P2k = [0]*34
P3k = [0]*34
P4k = [0]*34
P5k = [0]*34

# P_1(K)
P1k[0] = 10*k^2 - 20*k + 10
P1k[1] = k^2 -2*k +1
P1k[3] =  -k + 1
P1k[4] = -4*k + 4
P1k[18] = 1
P1k[19] = 1
```

```
# P_2(K)
P2k[18] = 3*k^2 - 12*k + 12
P2k[29] = k^2 - 6*k + 8
P2k[31] = -4*k+10
P2k[32] = 3

# P_3(K)
P3k[4] = 6*k^2 - 24*k + 24
P3k[11] = k^2 - 4*k + 4
P3k[17] = -k+2
P3k[19] = -6*k+12
P3k[31] = 2
P3k[32] = 3

# P_4(K)
P4k[19] =  6
P4k[28] = -1
P4k[30] =  2
P4k[31] = -4

# P_5(K)
P5k[19] =  6*k^2-36*k+54
P5k[30] =  2*k^2-20*k+42
P5k[31] =  4*k^2 - 24*k + 36
P5k[32] =  -24*k + 84
P5k[33] =  120
```

```
## Scaling functions z(k) and p_i(k)
zk = 6*(5*k^3 - 20*k^2 + 30*k - 16)/(5*k^3 - 35*k^2 + 75*k - 48)
p1k = 3*(k^5 - 8*k^4 + 22*k^3 - 24*k^2 + 8*k)/ \
      (5*k^7 - 35*k^6 + 75*k^5 - 48*k^4)
p2k = (10*k^5 - 60*k^4 + 109*k^3 - 76*k^2 + 18*k)/ \
      (5*k^7 - 35*k^6 + 75*k^5 - 48*k^4)
p3k = (5*k^5 - 28*k^4 + 45*k^3 - 28*k^2 + 6*k)/ \
      (5*k^7 - 35*k^6 + 75*k^5 - 48*k^4)
p4k = (1/4)*(5*k^7 - 30*k^6 + 53*k^5 - 52*k^4 + 94*k^3 - 96*k^2 + 24*k)/ \
      (5*k^7 - 35*k^6 + 75*k^5 - 48*k^4)
p5k = (1/4)*(15*k^5 - 60*k^4 + 78*k^3 - 40*k^2 + 8*k)/ \
      (5*k^7 - 35*k^6 + 75*k^5 - 48*k^4)

# Number of C_5s  in each of the 34 graphs.
five_cycles = [0]*26+[1,1,1,2,2,4,6,12]

def test_positive_for_small_k(x):
    for r in [4..1000]:
        if x.substitute(k=r) <= 0:
            print ("ERROR:",x,"is not positive for k=",r)
            return

# Denominator for all cFi in the result
den = 5*k^7 - 35*k^6 + 75*k^5 - 48*k^4

# Test if denominator is positive for small k
test_positive_for_small_k(den)
```

```
for i in [0..33]:
        cFi[i] = expand(factor( (zk*(Zk[i]) + p1k*P1k[i] + p2k*P2k[i] +\
      p3k*P3k[i] + p4k*P4k[i] + p5k*P5k[i] + five_cycles[i] )*den ))

# This is the polynomial for each of the tight graphs (when multiplied
# by den)
# It is the coefficient at c_{F_0} = cFi[0]
opt_pol = cFi[0]
#opt_pol = 60*k^7 - 720*k^6 + 3600*k^5 - 9876*k^4 + 16320*k^3 \
#          - 16440*k^2 + 9360*k - 2304

# Printing of all coefficients c_{F_i}
print("Printing the resulting coefficients")
for i in [0..33]:
        print ("cF{}= {} / {}".format(i,cFi[i],den))

# Test that  cFi[0] == optimum value
optimum = -60/k + 120/k^2 - 120/k^3 + 48/k^4 + 12

########## Start of Claim 2.1
# Calculating the differences C_1 - C_i for i=2,..,10.
# We do it by checking c_{F_0} - c_{F_i} for all i
#
cF0_cFi=[0]*34
for i in [0..33]:
        cF0_cFi[i] = expand(cFi[i]- opt_pol)
```

```
# Showing that c_F0 is largest by displaying the difference.
# The leading coefficient at k^7 is negative
# and tests for small values of k by evaluation

def test_not_positive_for_small_k(x):
    for r in [4..1000]:
        if x.substitute(k=r) > 0:
            print ("ERROR:",x,"is positive for k=",r)
            return
print()
print("Showing that c_F0 is largest")
print(" - see difference is 0 or leading coefficient negative")
for i in [0..33]:
    print ("numerator(c_F0 - c_F{})=".format(i),cF0_cFi[i])
    test_not_positive_for_small_k(cF0_cFi[i])



print("all done")
```

```
# SageMath code for Claim 2.1 when k=3

# This code calculates the coefficients cFi
#for the case of k = 3 in Claim 2.1.
# In \mathcal{F}_5: F20, F24, F30, F32, and F33 all contain a K4.
# Here we have removed those graphs and re-indexed the remaining graphs.

cFi = [0]*29

#Counting the number of C5s in each graph.

constants = [0]*29
constants[24] = 1
constants[25] = 1
constants[26] = 1
constants[27] = 2
constants[28] = 4



#  Pi(K) first init with 0

P1k = [0]*34
P2k = [0]*34
P3k = [0]*34
P4k = [0]*34
P6k = [0]*34
```

```
P1k[0] = 40
P1k[1] = 4
P1k[3] =  -2
P1k[4] = -8
P1k[18] = 1
P1k[19] = 1

P2k[18] = 3
P2k[27] = -1
P2k[28] = -2

P3k[4] = 6
P3k[11] = 1
P3k[17] = -1
P3k[19] = -6
P3k[28] = 2

P4k[19] = 6
P4k[26] = -1
P4k[28] = -4

P6k[11] = 1
P6k[23] = 2
P6k[22] = -1
P6k[27] = -2
P6k[17] = 1
```

```
P6k[19] = 6
P6k[28] = -4

#The scaling coefficients for k = 3

p1k = 1/27
p2k = 13/27
p3k = 8/27
p4k = 2/9
p6k = 17/54

# This calculates cFi
for i in [0..28]:
        cFi[i] = p1k*P1k[i] + p2k*P2k[i] + p3k*P3k[i]
        \+ p4k*P4k[i] + p6k*P6k[i] + constants[i]

# This prints the value of cFi along with the (possibly) re-indexed graph.
for i in [0..28]:
        print ('coefficient of F',i,'=',cFi[i])
```

2.8.2 SageMath code for Claim 2.4.10

```
# SageMath code Claim 3.10

var('k,e')

# this is the size of the sets.
# We start by using epsilon*(k-1) for easier counting.

x = (1 + e*(k-1))/k
y = (1 - e)/k

# These count the number of five cycles.
# The first is the one we use.
# The second is a sanity check.
def fivecyclecount(x,y):

    # here we count by picking one vertex in x,
    # then counting the number of possible five cycles.

    neighbors_in_same_sets = \
    x*(k-1)*(y^2/2)*(y^2*(k-2)*(k-3) + x*(k-2)*y*2)

    neighbors_in_diff_sets = \
    x*( (y*(k-1))*((k-2)*y) )/2*( (k-3)*(k-3)*y^2
    + (k-2)*y^2 + x*(k-3)*y)

    # This is counting the number of five-cycles not in X_1.
```

```
    # Note that it is equal to the sanity check but with k-1.

    nobadset_twosame = (y^3*(k-1)*(k-2)/2)*(y^2*(k-2)*(k-3))
    nobadset_nosame = \
    (y^3*(k-1)*(k-2)*(k-3)/2)*( (k-3)*(k-3)*y^2
    + (k-2)*y^2)

    return 120*(neighbors_in_diff_sets + \
    neighbors_in_same_sets) + \
    24*(nobadset_twosame + nobadset_nosame)

def sanity(y):
    nobadset_twosame = (y^3*k*(k-1)/2)*(y^2*(k-1)*(k-2))
    nobadset_nosame = \
    (y^3*k*(k-1)*(k-2)/2)*( (k-2)*(k-2)*y^2 + (k-1)*y^2)
    return 24*(nobadset_twosame + nobadset_nosame)

f = fivecyclecount(x,y)

# to check our count is correct,
# notice that the non-epsilon terms equal OPT.
view(f.collect(e))
view(expand(sanity(1/k)))
```

2.8.3 SageMath code for Claim 4.5

```
# Sage code for claim 4.5 - showing there are no type 2 vertices

var('k')

#This function counts the number of five cycles using equations (15) - (18)

def fivecyclecount(k):
        onebadset_twobadvertices = 2*( 1/(2*k^10) )*( (k-1)*(k-2)/k^2 )
        twobadset_twobadvertices = \
        (1/k^10 + 2/k^6)*( (k-2)^2/k^2 + (k-1)/k^2 )
        nobadset_twosame = ( (k-2)/(2*k^2) )*( (k-1)*(k-2)/k^2 )
        nobadset_nosame = \
        ( (k-2)*(k-3)/(2*k^2) )*( (k-2)^2/k^2 + (k-1)/k^2 )
        with_X0 = 2/k^4
        return 24*(onebadset_twobadvertices \
        + twobadset_twobadvertices + nobadset_twosame \
        + nobadset_nosame + with_X0)


# This gives the sum of equations (15) - (18) factored in a nice way.

expanded_first_check = expand(fivecyclecount(k))
print('five cycles containing a type 2 vertex:\n',expanded_first_check)


# actual upper bound once we account for the vertices in X_0
```

```
def bad_ub(k):
        return  fivecyclecount(k)

# The average "density" of five cycles containing a particular vertex
def good_ub(k):
        return -60/k + 120/k^2 - 120/k^3 + 48/k^4 + 12 - 1/k^10

# The difference between the optimal and the count for type 2.
def epsilon(k):
        return good_ub(k) - bad_ub(k)

print('difference:',factor(epsilon(k)))

# This verifies that for small values of k, count(r) greater than 1/k^5
def count_check(a,b):
        for i in [a..b]:
                if epsilon(i) < 1/(i^5):
                        return "the difference is less than 1/k^5 for k=",i
        return "'difference' is greater than 1/k^5 for all values"+\
        " of "+str(a)+" <= k up to "+str(b)

print(count_check(3,1000))
```

2.8.4 SageMath code for Claim 4.7

```
# SageMath code for Claim 4.7


var('k')


# This function counts the number of five-cycles in equations (19) - (21)
def fivecyclecount(k):
    onebadset_twobadvertices = \
    ( (1/2)*( (k^2+1)/(2*k^3) )^2 )*( (k-1)*(k-2)/k^2 )
    onebadset_onebadvertices = \
    ( (k^2+1)/(2*k^3) )*( (k-2)^3/k^3
    + (k-1)*(k-2)/k^3 )
    nobadset_twosame = \
    ( (k-2)/(2*k^2) )*( (k-1)*(k-2)/k^2 )
    nobadset_nosame = \
    ( (k-2)*(k-3)/(2*k^2) )*( (k-2)^2/k^2 + (k-1)/k^2 )
    with_X0 = 2/k^(4)
    return 24*(onebadset_twobadvertices \
    + onebadset_onebadvertices + nobadset_twosame \
    + nobadset_nosame+with_X0)



# This gives the sum of equations (19) - (21) factored in a nice way


expanded_first_check = expand(fivecyclecount(k))
print('five_cycles_containing_a_type1vertex:\n',expanded_first_check)
```

```
# The upper bound on five cycles containing a suboptimal type 1 vertex

def bad_ub(k):
    return  fivecyclecount(k)

def good_lb(k):
    return -60/k + 120/k^2 - 120/k^3 + 48/k^4 + 12 - 1/k^10

#difference from optimal value
def epsilon(k):
    return good_lb(k) - bad_ub(k)

print('difference:',factor(epsilon(k)))


# This verifies that for small values of k, epsilon(k) is greater than 1/k^5
def count_check(a,b):
    for i in [a..b]:
        if epsilon(i) < 1/(i^(5)):
            return "the difference is less than 1/k^5 for k =",i
    return "'difference' is greater than 1/k^5 for all values"+\
        " of "+str(a)+" <= k up to "+str(b)

print(count_check(3,1000))
```

CHAPTER 3. PATHS OF LENGTH THREE ARE K_{r+1}-TURÁN GOOD

Kyle Murphy, Department of Mathematics, Iowa State University.

J.D. Nir, Department of Mathematics, University of Manitoba.

Modified from a manuscript submitted to *The Electronic Journal of Combinatorics*

3.1 Abstract

The generalized Turán problem $\mathrm{ex}(n, T, F)$ is to determine the maximal number of copies of a graph T that can exist in an F-free graph on n vertices. Recently, Gerbner and Palmer noted that the solution to the generalized Turán problem is often the original Turán graph. They gave the name "F-Turán-good" to graphs T for which, for large enough n, the solution to the generalized Turán problem is realized by a Turán graph. They prove that the path graph on two edges, P_2, is K_{r+1}-Turán-good for all $r \geq 3$, but they conjecture that the same result should hold for all P_ℓ. In this paper, using arguments based in flag algebras, we prove that the path on three edges, P_3, is also K_{r+1}-Turán-good for all $r \geq 3$.

3.2 Introduction

One of extremal graph theory's most celebrated results was introduced in [28] by Turán who asked how many edges a (simple) graph on n vertices can contain if it has no clique containing $r+1$ vertices. Turán's solution, which we denote $\mathrm{ex}(n, K_{r+1})$, is asymptotically $(1 - \frac{1}{r})\binom{n}{2}$. Additionally, Turán showed that the unique extremal graph is the complete r-partite graph on n vertices with parts of size $\lceil \frac{n}{r} \rceil$ or $\lfloor \frac{n}{r} \rfloor$ (so that no pair of parts differs in size by more than one). We call this graph the *Turán graph* and denote it $T_r(n)$.

The first extensions to Turán's theorem considered forbidding graphs other than cliques. For any graph F, we say a graph G is *F-free* if it contains no (not necessarily induced) subgraph isomorphic to F. We use $\text{ex}(n, F)$ to denote the maximal number of edges in an F-free graph on n vertices. The general case is solved asymptotically by the Erdős-Stone-Simonovits Theorem [9] which proves

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

To further generalize the problem, one may consider counting subgraphs other than edges. Let $\nu(T, G)$ denote the number of distinct, not necessarily induced subgraphs of G isomorphic to T. We denote by $\text{ex}(n, T, F)$ the maximum of $\nu(T, G)$ over all F-free graphs G on n vertices. (Here T is the "target" graph while F is "forbidden.") The first question of this form to be resolved was due to Zykov in 1949 [29] who determined the value of the function $\text{ex}(n, K_t, K_r)$ when $t < r$ by proving that the Turán graph is the unique extremal graph.

Theorem 3.2.1 (Zykov [29])**.** *Let r and t be integers such that $t < r$. Then for all n, the Turán graph $T_t(n)$ is the unique K_r-free graph on n vertices containing the maximum number of K_t subgraphs.*

Several sporadic cases were investigated (see, for example, [6, 16]) before 2015 when Alon and Shikhelman introduced a systematic study in [2] in which they determine, among other results, that for forbidden graphs F with $\chi(F) = k + 1 > r$,

$$\text{ex}(n, K_r, F) = (1 + o(1))\binom{k}{r}\left(\frac{n}{k}\right)^r.$$

A more precise result can be found in [23]. Since then, the area has been widely studied; see [8, 12, 17, 21, 22] for an (incomplete) sampling of authors and results.

As in the original Zykov result, for many choices of T and F the Turán graph emerges as the optimal graph, at least for large enough n. In [13], Gerbner and Palmer introduced the term F-Turán-good to describe such target graphs T:

Definition 3.2.2. *Fix an $(r+1)$-chromatic graph F and a graph T that does not contain F as a subgraph. We say that T is F-*Turán-good* if $ex(n, T, F) = \nu(T, T_r(n))$ for every n large enough.*

In the same paper, Gerbner and Palmer prove that the path graph on ℓ edges, P_ℓ, is K_{r+1}-Turán-good for $\ell = 2$ and $r \geq 3$. They conjecture that paths should be Turán-good for all choices of r and ℓ. In this paper we establish that P_3, the path on three edges, is K_{r+1}-Turán-good for all $r \geq 3$.

To be precise, define the *density* of H in G to be

$$d(H,G) = \nu(H,G)\binom{|G|}{|H|}^{-1}$$

and let $\mathcal{F}_{n,r}$ be the family of K_{r+1}-free graphs on n vertices. We define

$$\mathrm{OPT}_r(P_3) = \lim_{n\to\infty} \max_{G_n \in \mathcal{F}_{n,r}} d(P_3, G_n).$$

Then the following theorem is the primary result of this paper:

Theorem 3.2.3. *For any integer $r \geq 3$,*

(i) $\mathrm{OPT}_r(P_3) = 12\left(\frac{r-1}{r}\right)^3$.

(ii) If n is sufficiently large, then P_3 is K_{r+1}-Turán good.

In [13], Gerbner and Palmer provided a proof of part (i) of Theorem 3.2.3. We will re-prove part (i) in the language of flag algebras, since we will require this proof to obtain part (ii). In [11], Gerbner proved that P_3 is K_4-Turán-good. His argument is very specific to K_4, relying on the properties of several other graphs which he proves are K_4-Turán-good. Building on his technique, shortly after the initial announcement of this paper, Qian, Xie, and Ge announced an alternate proof of Theorem 3.2.3 (ii) in [25]. Their proof is distinct from the proof presented here, though both rely on considering what small induced graphs occur in the extremal graph.

In [12], Gerbner and Palmer proved that for two graphs T and F, where $\chi(F) = r$,

$$\mathrm{ex}(n, T, F) \leq \mathrm{ex}(n, T, K_r) + o(n^{|T|}).$$

Combined with Theorem 3.2.3, their theorem implies the following corollary.

Corollary 3.2.4. *For any graph F with chromatic number $r \geq 3$,*

$$ex(n, P_3, F) = \mathrm{OPT}_r(P_3)\binom{n}{4} + o(n^4).$$

In the remainder of this section, we establish the conventions used thorough the paper, reference a few well-known results that will be of use throughout the proof, and then provide a brief introduction to the flag algebra method. Section 3.3 contains the flag algebra calculations we use to establish part (i) of Theorem 3.2.3. In Section 3.4 we establish a stability result, proving that near-extremal graphs have small edit distance from the Turán graph. Then in Section 3.5 we use that stability argument to show that the Turán graph is optimal for large enough n. We conclude in Section 3.6 with some thoughts on what this result means for Gerbner and Palmer's conjecture for general paths P_ℓ.

3.2.1 Background and Conventions

We use P_ℓ to denote the path graph with ℓ edges and $\ell + 1$ vertices. If a copy of P_3 in G is defined by the edges wx, xy and yz, then we will use $wxyz$ to denote it. Note that a set of four vertices in G will frequently give multiple distinct copies of P_3. We use $wxyz$ for that specific ordering.

Figure 3.1 The path $wxyz$

We will need the following corollary of Theorem 3.2.1:

Corollary 3.2.5. *Let G be a K_{r+1}-free graph on n vertices. Then*

$$\nu(K_4, G) \leq \frac{r^3 - 6r^2 + 11r - 6}{r^3} \binom{n}{4} + o(n^4).$$

Proof. In the Turán graph $T_r(n)$, any set of four vertices inducing a copy of K_4 must come from four different partite sets. Thus there are

$$\binom{r}{4} \cdot \frac{n^4}{r^4} + o(n^4)$$

copies of K_4 in $T_r(n)$. The claim immediately follows. ■

We will also need the following lemma from folklore characterizing multipartite graphs:

Lemma 3.2.6. *Define the* co-cherry $\overline{P_2}$ *to be the unique graph on three vertices with one edge. Then G is a complete multipartite graph if and only if it does not contain the co-cherry as an induced subgraph.*

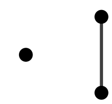


Figure 3.2 The co-cherry

Proof. First, assume G is a complete multipartite graph and let $x, y, z \in V(G)$ such that x is adjacent to y but z is not adjacent to y. As G is complete multipartite, the only way z is not adjacent to y is if they are in the same vertex class. As x is adjacent to y, it must be in a different vertex class. Thus x and z do not share a vertex class and are adjacent, so $G[\{x, y, z\}]$ does not span a co-cherry.

Now let G be a graph that does not contain the co-cherry as an induced subgraph. Define a relation on $V(G)$ by $x \sim y$ if x is not adjacent to y. As G is simple, this relation is reflexive and symmetric, and if x is not adjacent to y and y is not adjacent to z, then x cannot be adjacent to z, as that would form an induced co-cherry, so the relation is transitive as well. Thefore this equivalence relation partitions the vertices of G into classes which contain no internal edges. Furthermore, two vertices from different classes are by definition adjacent and thus every edge between vertex classes is present. We conclude G is complete multipartite. ■

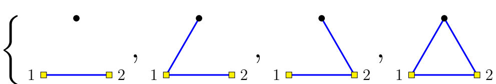
3.2.2 The Flag Algebra Method

Flag algebras were introduced by Razborov [26] as a tool to computationally solve problems in extremal combinatorics. In this section, we will introduce some of the main ideas necessary for our proof. For a complete overview see [26]. Flag algebras have been applied to study a variety of extremal problems on graphs [4, 5, 15, 18, 27] and hypergraphs [10, 14, 24], as well as oriented graphs [7, 19]. These only represent a handful of the many results in combinatorics which were obtained using flag algebras.

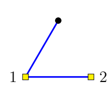
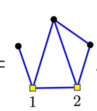
A *type* σ is a graph labelled by $[k]$. An *embedding* of σ into a graph F is an injective map $\theta : [k] \to V(F)$ so that $\text{im}(\theta)$ is isomorphic to σ. A *σ-flag* (F, θ) is a graph F together with an embedding θ of σ into $V(F)$. We will let $\mathcal{F}^{\sigma}$ denote the set of all σ-flags up to isomorphism and $\mathcal{F}^{\sigma}_n$ denote the associated subset containing all σ-flags on n vertices. If σ is the empty graph, then we will drop it from the notation and simply use $\mathcal{F}$ to denote the set of all graphs, or $\mathcal{F}_n$ to denote the set of all graphs on n vertices. As an example, if σ^* is the following labelled graph on two vertices,

$\sigma^* =$

then

$\mathcal{F}^{\sigma^*}_3 = \Big\{$  $\Big\}.$

For a type σ labelled by $[k]$, two σ-flags (H, θ_1) and (G, θ_2), and a set X_1 of size $|V(H)| - k$ selected uniformly at random from $V(G) \setminus \text{im}(\theta_2)$, $P((H, \theta_1), (G, \theta_2))$ is the probability that $X_1 \cup \text{im}(\theta_2)$ is isomorphic to (H, θ_1). For completeness, if $|V(G)| < |V(H)|$, then we let $P(H, G) = 0$. If σ is the empty graph, then we will write $P(H, G)$ to mean $P((H, \theta_1), (G, \theta_2))$. In this case, the definition of $P(H, G)$ coincides with the standard notion of induced density. Using the same type σ^* from the previous example:

If $H =$  and $G =$ , then $P((H, \theta_1), (G, \theta_2)) = \dfrac{1}{3}$.

Now suppose that (J, θ_3) is another σ-flag. Let $X_1, X_2 \subseteq V(G)$ be two disjoint sets of size $|V(H)| - k$ and $|V(J)| - k$, respectively, selected uniformly at random from $V(G) \setminus \text{im}(\theta_2)$. Then $P((H, \theta_1), (J, \theta_3); (G, \theta_2))$ is the probability that $X_1 \cup \text{im}(\theta_2)$ is isomorphic to H and $X_2 \cup \text{im}(\theta_2)$ is isomorphic to J. Once again, if σ is empty, then we write $P(H, J; G)$ in place of $((H, \theta_1), (J, \theta_3); (G, \theta_2))$. Equation 3.1 follows from the definition of $P((H, \theta_1), (J, \theta_3); (G, \theta_2))$.

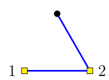
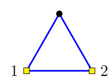
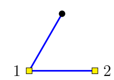
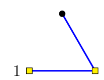
$$|P((H, \theta_1), (J, \theta_3); (G, \theta_2)) - P((H, \theta_1), (G, \theta_2)) \cdot P((J, \theta_3), (G, \theta_2))| \leq O(|V(G)|^{-1}) \tag{3.1}$$

Thus, as the size of G tends toward infinity, we can assume that we select X_1 and X_2 independently.

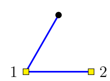
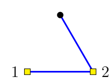
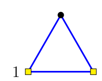
Let $\mathbb{R}\mathcal{F}^\sigma$ be the set of all finite formal linear combinations of elements from $\mathcal{F}^\sigma$. For a given type σ, let $\mathcal{K}^\sigma$ denote the linear subspace of $\mathbb{R}\mathcal{F}^\sigma$ generated by all elements of the form

$$F - \sum_{(H,\theta_2)\in\mathcal{F}_n^\sigma} P((F,\theta_1),(H,\theta_2))\cdot(H,\theta_2)$$

where $|V(F)| < n$. Razborov showed that there exists an algebra $\mathcal{A}^\sigma = \mathbb{R}\mathcal{F}^\sigma/\mathcal{K}^\sigma$ with well defined addition and multiplication. Addition is defined in the natural way by adding coefficients. For example, if $F_1, F_2 \in \mathcal{A}^{\sigma^*}$ such that

$F_1 = 2\cdot$  $+$  and $F_2 =$  $-$ ,

then

$F_1 + F_2 =$  $+$  $+$ .

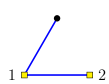
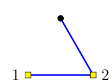
For a fixed type σ of size k, if (F_1,θ_1) and (F_2,θ_2) are two elements in $\mathcal{F}^\sigma$ such that

$$|V(F_1)| + |V(F_2)| - k = n,$$

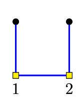
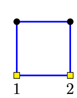
then the product of F_1 and F_2 is defined as

$$(F_1,\theta_1)\cdot(F_2,\theta_2) = \sum_{(H,\theta_3)\in\mathcal{F}_n^\sigma} P((F_1,\theta_1),(F_2,\theta_2);(H,\theta_3))\cdot(H,\theta_3).$$

For example, if

$(F_1,\theta_1) =$  and $(F_2,\theta_2) =$ 

then,

$(F_1,\theta_1)\times(F_1,\theta_2) = \frac{1}{2}\cdot$  $+ \frac{1}{2}\cdot$ .

Observe that the set $\mathcal{F}_4^{\sigma^*}$ contains more than just the two graphs pictured in the previous equation, but in all of these other graphs, $P((F_1,\theta_1),(F_2,\theta_2);(H,\theta_3)) = 0$. Multiplication in $\mathcal{A}^\sigma$ is defined as an extension of multiplication in $\mathcal{F}^\sigma$.

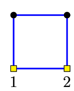
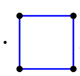
A sequence of graphs $(G_n)_{n\geq 1}$, where $|V(G_n)| = n$, is said to be *convergent* if for every finite graph H, the limit $\lim_{n\to\infty} P(H,G_n)$ exists. Let $\mathrm{Hom}^+(\mathcal{A}^\sigma,\mathbb{R})$ denote the set of all homomorphisms from $\mathcal{A}^\sigma$ to $\mathbb{R}$ such that $\phi(F)\geq 0$ for each element $F\in\mathcal{F}^\sigma$. Razborov showed that each function

$\phi \in \mathrm{Hom}^+(\mathcal{A}^\sigma, \mathbb{R})$ corresponds to some convergent graph sequence $(G_n)_{n\geq 1}$. That is, the values of ϕ correspond to the limits of induced densities in $(G_n)_{n\geq 1}$. It is often more intuitive to think of addition and multiplication operations in $\mathcal{A}^\sigma$ as representing induced densities of subgraphs in some very large graph G_{n_0} with an error term $O(n_0^{-1})$.

For each type σ labelled by $[k]$, Razborov also defined a function $[\![\cdot]\!]_\sigma : \mathbb{R}\mathcal{F}^\sigma \to \mathbb{R}\mathcal{F}$, which we will refer to as the *unlabelling operator*. For a σ-flag (F, θ), let $q_\sigma(F)$ denote the probability that (F, θ') is isomorphic to F, where $\theta' : V(F) \to [k]$ is a randomly chosen injective mapping. Let F' denote the graph isomorphic to F when ignoring labels. Then

$$[\![F]\!]_\sigma = q_\theta(F)F'.$$

As an example,

If, $F =$ , then $[\![F]\!]_\sigma = \frac{4}{6} \cdot$ .

Finally, it can be shown using the Cauchy-Schwarz inequality that if $\alpha \in \mathcal{A}^\sigma$ is some expression and $\phi \in \mathrm{Hom}^+(\mathcal{A}, \mathbb{R})$, then

$$\phi\left([\![\alpha^2]\!]_\sigma\right) \geq 0. \tag{3.2}$$

3.3 Theorem 3.2.3 (i)

First we will prove a lower bound by counting the number of P_3 subgraphs in the Turán graph. After that, the remainder of the section will be devoted to proving the upper bound using flag algebras.

Lemma 3.3.1. *For all $r \geq 3$,*

$$12\left(\frac{r-1}{r}\right)^3 \leq \mathrm{OPT}_r(P_3)$$

Proof. We begin by counting the paths of length three in the Turán graph $T_r(n)$. To do so, we will first choose the central edge of the path and then select two additional vertices and describe how to attach them to the central edge.

As the Turán graph is multipartite, the central edge must fall between two of the r vertex classes. Assume for the moment that n is divisible by r. Then there are $\binom{r}{2}(\frac{n}{r})^2$ choices for the central edge: first choose two vertex classes and select a vertex from each class.

Now we consider two cases. In the first case, the P_3 intersects exactly two of the vertex classes of $T_r(n)$. In this case, as we have already selected the central edge, the two vertex classes are already specified and we need only select an additional vertex from each class. These vertices are each adjacent to a different vertex of our central edge and thus give a unique P_3. There are $(\frac{n}{r}-1)^2$ ways to choose these two vertices.

In the second case, the P_3 intersects at least three vertex classes of $T_r(n)$. (Note that as vertex classes contain no internal edges, the P_3 must contain vertices from more than one vertex class.) We first select this third vertex, for which there are $n-2(\frac{n}{r})$ choices, and then select a fourth unique vertex from the remaining $n-3$ options. If the fourth vertex chosen happens to share a vertex class with either end of the central edge, then there is a unique P_3 containing the four vertices with the given central edge. Otherwise, there are two ways to connect the third and fourth vertices to the central edge. However, we also select pairs of vertices of this form twice as the fourth vertex we selected was an eligible choice when we selected the third in this case. Thus either way, this method produces

$$\left(n-2\left(\frac{n}{r}\right)\right)(n-3)$$

unique copies of P_3.

Putting all of our counts together, for all $r \geq 4$,

$$\nu(P_3, T_r(n)) = \binom{r}{2}\left(\frac{n}{r}\right)^2\left(\left(\frac{n}{r}-1\right)^2+\left(n-2\frac{n}{r}\right)(n-3)\right)+o(n^4),$$

where the error terms accounts for the cases that n is not divisible by r. Factoring out leading leading terms gives

$$\nu(P_3, T_r(n)) = n^4 \cdot \frac{1}{2}\left(1-\frac{1}{r}\right)\left(\left(\frac{1}{r}-\frac{1}{n}\right)^2+\left(1-\frac{2}{r}\right)\left(1-\frac{3}{n}\right)\right)+o(n^4).$$

As $\binom{n}{4} = \frac{1}{24}n^4 + o(n^4)$, it follows that

$$\lim_{n\to\infty} \nu(P_3, T_r(n))\binom{n}{4}^{-1} = \frac{n^4 \cdot \frac{1}{2}\left(1-\frac{1}{r}\right)\left(\left(\frac{1}{r}-\frac{1}{n}\right)^2 + \left(1-\frac{2}{r}\right)\left(1-\frac{3}{n}\right)\right) + o(n^4)}{\frac{1}{24}n^4 + o(n^4)}$$
$$= 12\left(\frac{r-1}{r}\right)^3.$$

Hence, $12\left(\frac{r-1}{r}\right)^3 \leq \mathrm{OPT}_r(P_3)$. ■

We will now prove that $\mathrm{OPT}_r(P_3) \leq 12\left(\frac{r-1}{r}\right)^3$ using the flag algebra method. Unlike many proofs that employ this technique, ours does not require any computer assistance for verification. With that said, this section does require the multiplication and factoring of large polynomials. The authors have included a link to SageMath code used to verify these calculations in the appendix.

Proof of Theorem 3.2.3(i). Let $\mathcal{F}_4 = \{F_i\}_{i=0}^{10}$ denote the set of all unlabeled graphs on 4 vertices up to isomorphism, pictured below.

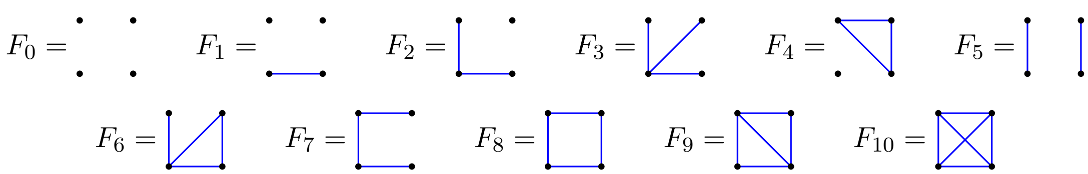
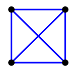


Figure 3.3 Enumeration of all graphs in $\mathcal{F}_4$.

Throughout this section, we will be working with the induced densities of subgraphs in a convergent sequence of K_{r+1}-free graphs $(G_n)_{n\geq 1}$. In order to simplify notation we will let $P(F) = \lim_{n\to\infty} P(F, G_n)$ and similarly $d(F) = \lim_{n\to\infty} d(F, G_n)$. Summing over all of the graphs on $\mathcal{F}_4$, we observe the following:

$$\sum_{i=0}^{10} P(F_i) = 1. \tag{3.3}$$

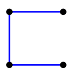
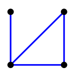
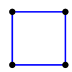
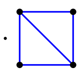
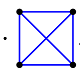
In order to make expressions like this easier to visualize, we will often use a drawing of F in place of $P(F)$ in our computations. For example, if $(G_n)_{n\geq 1}$ was the sequence of complete graphs on n vertices, then $P(K_4) = \lim_{n\to\infty} P(K_4, G_n) = 1$. Using a drawing of K_4 in order to represent this density, we would write:

 $= P(K_4) = 1.$

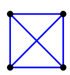
Fix $r \geq 4$ and let $(G_n)_{n\geq 1}$ be an arbitrary convergent sequence of K_{r+1}-free graphs. By the law of total probability, the (non-induced) density of the path P_3 can be expressed as the sum of induced densities of graphs on four vertices in the following way,

$$d(P_3) = \sum_{i=0}^{10} P(F_i) \cdot \nu(P_3, F_i). \tag{3.4}$$

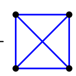
This expression can be simplified, however, as over half of the graphs in $\mathcal{F}_4$ do not contain a P_3 subgraph.

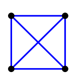
$d(P_3) =$  $+ 2 \cdot$  $+ 4 \cdot$  $+ 6 \cdot$  $+ 12 \cdot$ .

From Corollary 3.2.5 we obtain the following upper bound on $P(K_4)$ in $(G_n)_{n\geq 1}$.

 $\leq \dfrac{r^3 - 6r^2 + 11r - 6}{r^3}.$ (3.5)

Note that

$P_0(r) := \sum_{i=0}^{9} \left(\dfrac{r^3 - 6r^2 + 11r - 6}{r^3}\right) \cdot P(F_i) +$  $\dfrac{-6r^2 + 11r - 6}{r^3}$

$= \left(\dfrac{r^3 - 6r^2 + 11r - 6}{r^3}\right) -$  by (3.3)

≥ 0 by (3.5)

In the following computations, we will use two sets of labeled flags $\mathcal{F}_3^{\sigma_1}$ and $\mathcal{F}_3^{\sigma_2}$, where

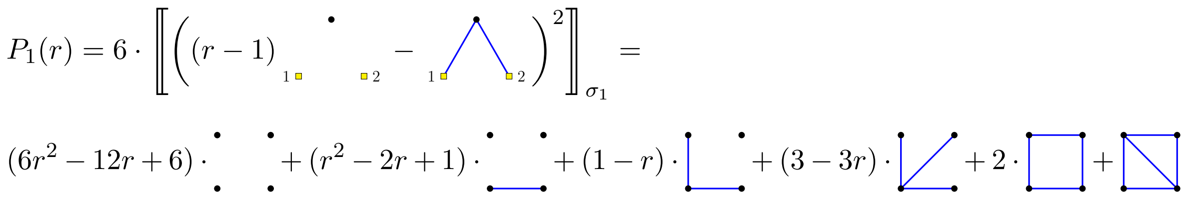
$\sigma_1 =$

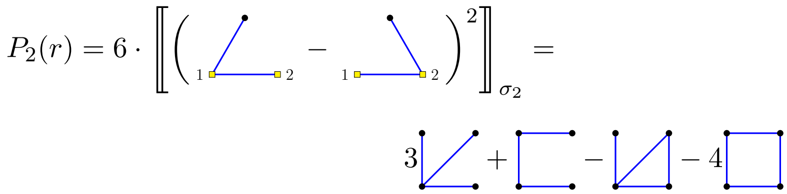
,

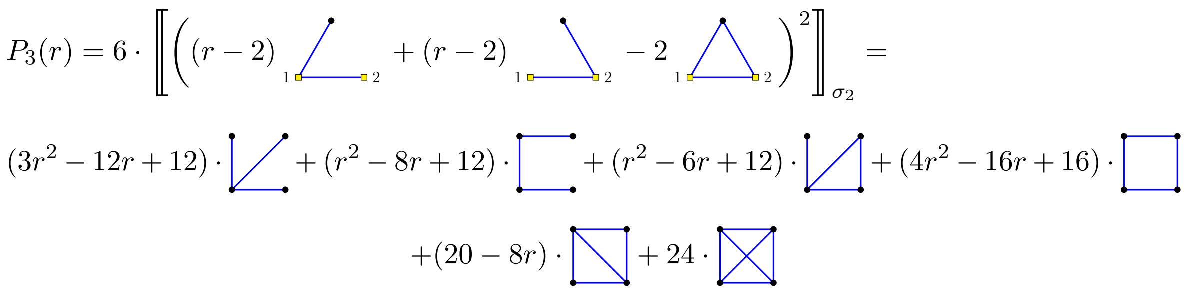
$\sigma_2 =$

.

By the Cauchy-Schwarz inequality, each of the following three expressions is nonnegative for all $r \geq 4$.

1. 

2. 

3. 

Moreover, It can quickly verified that for all $r \geq 4$, the following polynomials are all nonnegative.

1. $p_0(r) = \frac{18(r^2-2r+1)}{3r^2-11r+9}$

2. $p_1(r) = \frac{3r^3-10r^2+7r}{3r^5-11r^4+9r^3}$

3. $p_2(r) = \frac{9r^5-32r^4+25r^3}{4(3r^5-11r^4+9r^3)}$

4. $p_3(r) = \frac{15r^3-24r^2+7r}{4(3r^5-11r^4+9r^3)}$

We can add the sum $\sum_{j=0}^{3} p_j(r)P_j(r)$ to Equation (3.4) to obtain the following upper bound on $d(P_3)$.

$$d(P_3) \leq \sum_{i=0}^{10} P(F_i) \cdot \nu(P_3, F_i) + \sum_{j=0}^{3} p_j(r)P_j(r). \tag{3.6}$$

For each $F_i \in \mathcal{F}_4$, let C_{F_i} denote the coefficient of the graph F_i after combining like-terms in Equation (3.6). This gives the following, simplified upper bound on $d(P_3)$.

$$d(P_3) \leq \sum_{i=0}^{10} C_{F_i} P(F_i).$$

Since $\sum_{i=0}^{10} P(F_i) = 1$, it follows that

$$d(P_3) \le \max\{C_{F_i} : F_i \in \mathcal{F}_4\}. \tag{3.7}$$

The following are the exact values of each C_{F_i}.

- $C_{F_0} = C_{F_3} = C_{F_8} = C_{F_9} = C_{F_{10}} =$
$$12\left(\frac{r-1}{r}\right)^3$$
- $C_{F_1} =$
$$\frac{(21r^2 - 97r + 108)(r-1)^3}{3r^5 - 11r^4 + 9r^3}$$
- $C_{F_2} =$
$$\frac{(18r^3 - 111r^2 + 205r - 108)(r-1)^2}{3r^5 - 11r^4 + 9r^3}$$
- $C_{F_4} = C_{F_5} =$
$$\frac{18(r-1)^3(r-2)(r-3)}{3r^5 - 11r^4 + 9r^3}$$
- $C_{F_6} =$
$$\frac{45r^5 - 351r^4 + 1035r^3 - 1389r^2 + 870r - 216}{2(3r^5 - 11r^4 + 9r^3)}$$
- $C_{F_7} =$
$$\frac{(30r^4 - 180r^3 + 371r^2 - 327r + 108)(r-1)}{3r^5 - 11r^4 + 9r^3}$$

By examining leading coefficients and factoring, it is clear that for all $r > 1000$,

$$\max\{C_{F_i} : F_i \in \mathcal{F}_4\} = 12\left(\frac{r-1}{r}\right)^3. \tag{3.8}$$

We have provided SageMath code which can be used to verify (3.8) for $4 \le r \le 1000$ in the appendix. This fact, together with Equation 3.7 are enough to show that

$$\mathrm{OPT}_r(P_3) \le 12\left(\frac{r-1}{r}\right)^3.$$

Along with Lemma 3.3.1, this completes the proof of Theorem 3.2.3(i). ■

3.4 Stability

For two graphs G and H of the same order, the *edit distance* between G and H, denoted $\mathrm{Dist}(G, H)$, is the minimum number of adjacencies one needs to add or remove in order to change G into a graph isomorphic to H. Our goal in this section is to prove that graphs with P_3 density approaching $\mathrm{OPT}_r(P_3)$ are close in structure to the Turán graph $T_r(n)$. Specifically, we prove the following lemma:

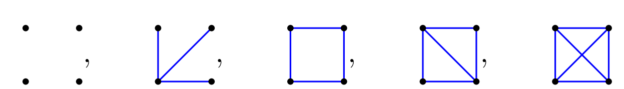
Lemma 3.4.1. *For every $\varepsilon > 0$, there exists an n_0 and $\delta > 0$ such that for every K_{r+1}-free graph G of order $n \geq n_0$, if $d(P_3, G) \geq \mathrm{OPT}_r(n) - \delta$, then $\mathrm{Dist}(G, T_r(n)) \leq \varepsilon n^2$.*

We prepare for the proof of Lemma 3.4.1 with a collection of lemmas. Several of these lemmas use the epsilon-delta paradigm, and so in the interest of legibility we have labelled the lemmas in this section by letter. We adopt the convention that ε_A, for example, will always refer to the ε in Lemma A. The exception to this rule is Lemma 3.4.1 which uses unadorned variables.

The first lemma is the Induced Removal Lemma, proved by Alon, Fischer, Krivelevich and Szegedy [1].

Lemma 3.4.2 (Lemma A, Induced Removal Lemma)**.** *Let $\mathcal{F}$ be a set of graphs. For each $\varepsilon_A > 0$, there exist η_A and $\delta_A > 0$ such that for every graph G of order $n \geq \eta_A$, if G contains at most $\delta_A n^{|V(H)|}$ induced copies of H for every $H \in \mathcal{F}$, then G can be made $\mathcal{F}$-free by removing or adding at most $\varepsilon_A n^2$ edges from G.*

We define the set T to contain all of the graphs $F \in \mathcal{F}_4$ for which $c_F = \mathrm{OPT}_r(P_3)$ in the proof of Theorem 3.2.3.

$T = \{$  $\}$.

The following is a restatement of Lemma 2.4.3 appearing in [3]. For completeness, we will provide a short proof.

Lemma 3.4.3. *[3] Let $(G_n)_{n\geq 1}$ be a sequence of K_{r+1}-free graphs such that*

$$\lim_{n\to\infty} d(P_3, G_n) = \lim_{n\to\infty} \sum_{i=0}^{10} C_{F_i} \cdot P(F_i, G_n) = \mathrm{OPT}_r(P_3),$$

where $F_i \in \mathcal{F}_4$ for all $i = 0, \ldots, 10$. Then for all $F \in \mathcal{F}_4$, $\lim_{n\to\infty} P(F, G_n) > 0$ implies that $F \in T$.

Proof. Let $\mathcal{F}_4^*$ denote the set of graphs F in $\mathcal{F}_4$ for which $\lim_{n\to\infty} P(F, G_n) > 0$. Then $\lim_{n\to\infty} \sum_{F\in\mathcal{F}_4^*} P(F, G_n) = 1$, implying from Theorem 3.2.3(i) that

$$\lim_{n\to\infty} \sum_{F\in\mathcal{F}_4^*} C_F \cdot P(F, G_n) = \mathrm{OPT}_r(P_3).$$

For each graph $H \in \mathcal{F}_4 \setminus T$, we know from the proof of Theorem 3.2.3(i) that $C_H < \mathrm{OPT}_r(P_3)$. Thus, $H \notin \mathcal{F}_4^*$ as otherwise $\lim_{n\to\infty} \sum_{F\in\mathcal{F}_4^*} C_F \cdot P(F, G_n) < \mathrm{OPT}_r(P_3)$. ■

Given the fact that only those graphs in T can appear with positive density in the limit of any extremal sequence, we can now prove the following lemma.

Lemma 3.4.4 (Lemma B)**.** *For each $\varepsilon_B > 0$, there exists a η_B and $\delta_B > 0$ such that any K_{r+1}-free graph G of order $n \geq \eta_B$ satisfying $d(P_3, G) \geq \mathrm{OPT}_r(P_3) - \delta_B$ contains at most $\varepsilon_B n^3$ copies of $\overline{P_2}$.*

Proof. Suppose that $(G_n)_{n\geq 1}$ is some convergent sequence of K_{r+1}-free graphs for which

$$\lim_{n\to\infty} d(P_3, G_n) = \mathrm{OPT}_r(P_3).$$

By inspection, none of the graphs in T contain $\overline{P_2}$ as a subgraph. Thus from Lemma 3.4.3,

$$\lim_{n\to\infty} d(\overline{P_2}, G_n) = 0.$$

This fact immediately implies Lemma 3.4.4. ■

Next we prove that among all complete r-partite graphs on at least four vertices, the Turán graph $T_r(n)$ contains the most P_3 subgraphs.

Lemma 3.4.5. *For $n \geq 4$ and $r \geq 4$, if G is any complete r-partite graph on n vertices then $\nu(P_3, G) \leq \nu(P_3, T_r(n))$.*

Proof. We count the number of P_3 in a complete multipartite graph using a similar approach to that in the proof of Theorem 3.3.1. We sum over each edge and count the number of P_3 with that edge as the center. If $e = xy$ is an edge in the center of P_3 with x in vertex class V_x and y in vertex class V_y, let the other edges of the P_3 be wx and yz. We classify the P_3 into one of four types depending on the location of w and z.

- There are $(|V_x|-1)(|V_y|-1)$ such P_3 with $w \in V_y$ and $z \in V_x$ as we may not reselect x or y.
- When $w \in V_y$ but $z \notin V_x$, there are $(|V_y|-1)(n-|V_x|-|V_y|)$ choices for the P_3 as z falls in some vertex class other than V_x or V_y.
- Similarly, when $w \notin V_y$ and $z \in V_x$, there are $(n-|V_x|-|V_y|)(|V_x|-1)$ many such P_3.
- Finally, if $w \notin V_y$ and $z \notin V_x$, then we must take care to select them uniquely. Choosing w first and then z gives $(n-|V_x|-|V_y|)(n-|V_x|-|V_y|-1)$ many such P_3.

Thus in total, for complete multipartite graphs G,

$$\begin{aligned}\nu(P_3,G) = \sum_{e=xy} (|V_x|-1)(|V_y|-1) + (|V_y|-1)(n-|V_x|-|V_y|) \quad + \\ + \quad (n-|V_x|-|V_y|)(|V_x|-1) + (n-|V_x|-|V_y|)(n-|V_x|-|V_y|-1) \\ = \sum_{e=xy} (|V_x|-1)(|V_y|-1) + (n-|V_x|-|V_y|)(n-3)\end{aligned}$$

Now suppose that G has r parts $V_1,\ldots,V_r$. There are $|V_i||V_j|$ edges between parts V_i and V_j, each of which contributes the same term in the sum above. Thus we may also write

$$\nu(P_3,G) = \sum_{1\le i<j\le r} |V_i||V_j|\big((|V_i|-1)(|V_j|-1) + (n-|V_i|-|V_j|)(n-3)\big). \tag{3.9}$$

Let G be a complete r-partite graph on n vertices with parts $V_1,\ldots,V_r$ such that $|V_1| \ge |V_2|+2$. If G has no edges but at least four vertices, it cannot be extremal, so assume G contains at least one edge. Define G' to be the complete multipartite graph on n vertices with parts $V_1', V_2', \ldots, V_r'$ where $|V_1'| = |V_1|-1$, $|V_2'| = |V_2|+1$, and $|V_i'| = |V_i|$ for $i \ge 3$.

After straightforward, if tedious, calculation, we use (3.9) to see $\nu(P_3,G') - \nu(P_3,G) = \Delta_{P_3}$ where

$$\Delta_{P_3} = (|V_1|-|V_2|-1)\Big((n-|V_1|-|V_2|)(n-3) + 2(|V_1|-1)|V_2| + \sum_{j=3}^{r} |V_j|(n-2-|V_j|)\Big)$$

Note that by assumption $|V_1| \geq |V_2| + 2$ and $n \geq 4$, that $n \geq |V_1| + |V_2|$ as $V_1, V_2 \subseteq V(G)$ and that $n - 2 \geq |V_j|$ for $j \geq 3$ because $|V_j| \leq n - |V_1|$ and V_1 must have at least two vertices to satisfy $|V_1| \geq |V_2| + 2$. Thus $|V_1| - |V_2| - 1$ is strictly positive and $(n - |V_1| - |V_2|)(n - 3)$, $2(|V_1| - 1)|V_2|$, and $\sum_{j=3}^{r} |V_j|(n - 2 - |V_j|)$ are each nonnegative. If $\Delta_{P_3} = 0$, then each term must be exactly zero. This means $n = |V_1| + |V_2|$ and $|V_2| = 0$. But then $n = |V_1|$, so all of the vertices of G are in one part which contradicts that G has at least one edge. We conclude $\Delta_{P_3} > 0$ and thus G' contains more P_3 than G.

Thus we see G was not extremal and therefore the Turán graph, the unique complete r-partite graph in which no pair of vertex classes differs in size by more than one, is the complete r-partite graph with the greatest number of P_3. ■

In the next lemma, we prove that if G has large P_3-density, it is close in edit distance to a nearly balanced complete r-partite graph.

Lemma 3.4.6 (Lemma C). *For any two independent parameters $\varepsilon_C > 0$ and $\gamma_C > 0$ there are η_C and $\delta_C > 0$ such that if G is a K_{r+1}-free graph with order $n \geq \eta_C$ satisfying $d(P_3, G) > \mathrm{OPT}_r(P_3) - \delta_C$, then there is a complete r-partite graph G' with parts $X_1, \ldots, X_r$ satisfying $\mathrm{Dist}(G, G') \leq \gamma_C n^2$ and, for each $1 \leq i \leq r$,*

$$\frac{1 - \varepsilon_C}{r} n \leq |X_i| \leq \frac{1 + \varepsilon_C}{r} n.$$

Proof. Let $\varepsilon_C, \gamma_C > 0$ be given. We require a $\gamma'_C > 0$ but defer its exact definition until later. Take η_A and δ_A to be as in Lemma 3.4.2 so that any graph G of order $n \geq \eta_A$ containing at most $\delta_A n^3$ copies of $\overline{P_2}$ can be made $\overline{P_2}$-free by editing at most $\gamma'_C n^2$ edges. Then take η_B and δ_B to be as in Lemma 3.4.4 so that for any graph G of order $n \geq \eta_B$ which satisfies $d(P_3, G) \geq \mathrm{OPT}_r(P_3) - \delta_B$ contains at most $\delta_A n^3$ copies of $\overline{P_2}$ (that is, apply Lemma 3.4.4 with $\varepsilon_B = \delta_A$).

Though we are not ready to define them yet, we will ensure $\eta_C \geq \max(\eta_A, \eta_B)$ and $\delta_C \leq \min(\delta_A, \delta_B)$. Let G be a graph of order $n \geq \eta_C$ satisfying $d(P_3, G) \geq \mathrm{OPT}_r(n) - \delta_C$. By Lemma 3.4.4, G has at most $\delta_A n^3$ copies of $\overline{P_2}$ and thus by Lemma 3.4.2 we may edit at most

$\gamma'_C n^2$ edges of G to get a $\overline{P_2}$-free graph, G'. It follows from Lemma 3.2.6 that G' is a complete r-partite graph as it is both K_{r+1}-free and $\overline{P_2}$-free. Let $X_1, \ldots, X_r$ denote the partite sets of G'. We complete the proof by demonstrating these partite sets all have size nearly $\frac{n}{r}$.

There is a constant $c > 0$ such that each edge removed from G is contained in at most cn^2 copies of P_3. (The constant c counts the number of ways to extend an edge and two other vertices into a copy of P_3.) Thus

$$d(P_3, G') \geq \mathrm{OPT}_r(P_3) - \delta_C - c\gamma'_C$$

as the P_3-density of the removed edges is at most

$$\frac{\gamma'_C n^2 \cdot cn^2}{n^{|V(P_3)|}} = c\gamma'_C.$$

To prove that the partite sets have bounded size, we will show that if they do not, we may alter G' to increase its P_3 density beyond $\mathrm{OPT}_r(P_3)$. As $\mathrm{OPT}_r(P_3)$ is, by definition, a limit, we can, for large enough η_C, get upper bounds on the P_3-density of such graphs that are as close to $\mathrm{OPT}_r(P_3)$ as necessary to arrive at a contradiction.

We require a partial result from the proof of Lemma 3.4.5. Recall that when moving one vertex from vertex class V_1 to vertex class V_2 the change in the number of P_3 subgraphs was

$$\Delta_{P_3} = (|V_1| - |V_2| - 1)\big((n - |V_1| - |V_2|)(n-3) + 2(|V_1| - 1)|V_2| + \sum_{j=3}^{r} |V_j|(n - 2 - |V_j|)\big)$$

Assume first that there is a partite set that is too large. Specifically, assume, without loss of generality, that $|X_1| > \frac{1+\varepsilon_C}{r} n$. We consider two cases.

First, assume

$$\frac{1+\varepsilon_C}{r} n < |X_1| \leq \frac{n}{2}.$$

There must be a partite set of G', say X_2, that satisfies $|X_2| < \frac{n}{r}$; if not,

$$n = \sum_{i=1}^{r} |X_i| \geq \frac{1+\varepsilon_C}{r} n + (r-1)\frac{n}{r} = n + \frac{\varepsilon_C}{r} n$$

is a contradiction. Consider the process of moving one vertex from X_1 to X_2 repeated $\frac{\varepsilon_C}{3r}n$ times. At each step of this process,

$$|X_1| - |X_2| > \left(\frac{1+\varepsilon_C}{r}n - \frac{\varepsilon_C}{3r}n\right) - \left(\frac{1}{r}n + \frac{\varepsilon_C}{3r}n\right) = \frac{\varepsilon_C}{3r}n.$$

We take η_C large enough that this value is always at least 2 so that number of P_3 subgraphs increases at every step. In particular, as $|X_1| + |X_2|$ stays constant and

$$|X_1| + |X_2| < \frac{n}{2} + \frac{n}{r} \leq \frac{3}{4}n$$

we have

$$\Delta_{P_3} \geq (|X_1| - |X_2| - 1)((n - |X_1| - |X_2|)(n-3)) \geq \left(\frac{\varepsilon}{3r}n - 1\right)\left((n - \tfrac{3}{4}n)(n-3)\right).$$

Take η_C large enough so that $n \geq \eta_C$ implies $n - 3 \geq \frac{n}{2}$ and $\frac{\varepsilon_C}{3r}n - 1 \geq \frac{\varepsilon_C}{4r}n$, giving

$$\Delta_{P_3} \geq \frac{\varepsilon_C}{4r}n \cdot \frac{n}{4} \cdot \frac{n}{2} = \frac{\varepsilon_C}{32r}n^3.$$

Now, as we repeat this process $\frac{\varepsilon_C}{3r}n$ times, the total increase in the number of copies of P_3 is at least

$$\frac{\varepsilon_C}{3r}n \cdot \frac{\varepsilon_C}{32r}n^3 = \frac{\varepsilon_C^2}{96r^2}n^4.$$

As $|V(P_3)| = 4$, this increases the P_3 density of G' by at least $\frac{\varepsilon_C^2}{96r^2}$. By choosing δ_C and γ'_C such that

$$\delta_C + c\gamma'_C < \frac{\varepsilon_C^2}{96r^2}$$

we arrive at a graph G'' with

$$d(P_3, G'') \geq d(P_3, G') + \frac{\varepsilon_C^2}{96r^2} \geq \mathrm{OPT}_r(P_3) - \delta_C - c\gamma'_C + \frac{\varepsilon_C^2}{96r^2} > \mathrm{OPT}_r(P_3),$$

a contradiction for large enough η_C.

Otherwise we have $|X_1| \geq \frac{n}{2}$. We wish to use a similar approach to the first case, but we must assure that the lower bound on Δ_{P_3} is cubic in n at each step of the process. Note that for $2 \leq i \leq r$, we must have $|X_i| \leq \frac{1}{2(r-1)}n \leq \frac{1}{6}n$ (recall $r \geq 4$). We start by moving $\frac{n}{12}$ vertices from

X_1 to X_2. These moves increase the number of copies of P_3, but we disregard those increases. Then we have $|X_1| > \frac{n}{2} - \frac{n}{12} = \frac{5}{12}n$ and

$$\frac{n}{12} \leq |X_2| \leq \frac{n}{6} + \frac{n}{12} = \frac{3}{12}n.$$

Starting from this modified graph we can move $\frac{n}{24}$ additional vertices from X_1 to X_2. For each such move, we have

$$(|X_1| - |X_2| - 1) \geq \left(\frac{5}{12}n - \frac{1}{24}n\right) - \left(\frac{3}{12}n + \frac{1}{24}n\right) - 1 = \frac{n}{12} - 1 \geq \frac{n}{13}$$

by choosing η_C large enough, and

$$2(|X_1| - 1)|X_2| \geq 2\left(\frac{5}{12}n - 1\right)\left(\frac{n}{12}\right) > \frac{n^2}{18},$$

again with η_C large enough. Thus

$$\Delta_{P_3} \geq \frac{n}{13} \cdot \frac{n^2}{18} = \frac{n^3}{234}$$

and repeating this process $\frac{n}{24}$ times increases the total number of P_3 subgraphs by at least $\frac{n^4}{5616}$, increasing the P_3 density of G' by $\frac{1}{5616}$. By taking $\delta_C + c\gamma'_C < \frac{1}{5616}$ we again get a graph with P_3 density larger than the optimal density, a contradiction when η_C is sufficiently large.

Finally, we now assume for contradiction that $|X_1| < \frac{1-\varepsilon_C}{r}n$. If $|X_1| < \frac{1-(r-1)\varepsilon_C}{r}n$, then there must be another partite set X_i with $|X_i| > \frac{1+\varepsilon_C}{r}n$ as otherwise

$$n = \sum_{i=1}^{r} |X_i| < \frac{1-(r-1)\varepsilon_C}{r}n + (r-1)\frac{1+\varepsilon_C}{r}n = n$$

is a contradiction. As we have already handled cases with a too large part, we may assume

$$\frac{1-(r-1)\varepsilon_C}{r}n \leq |X_1| < \frac{1-\varepsilon_C}{r}n.$$

There must be a partite set X_i with $|X_i| > \frac{n}{r}$, again because otherwise the parts combined cannot contain n vertices. Then we move a vertex from X_i to X_1 and repeat the move $\frac{\varepsilon}{3r}n$ times. Then as before at every step of the process

$$|X_i| - |X_1| \geq \frac{1-\varepsilon}{3r}n > 0$$

and, using very rough bounds,

$$|X_1| + |X_i| \le \frac{1-\varepsilon_C}{r}n + \frac{1+\varepsilon_C}{r}n < \frac{n}{r} + \frac{n}{2} \le \frac{3}{4}n.$$

Therefore this process also increases the P_3 density of G' by at least $\frac{\varepsilon_C^2}{96r^2}$, a contradiction for δ_C small enough. We conclude each partite set $X_1, \ldots, X_r$ must be within the specified bounds.

For completeness, we explicitly specify our choices of η_C, δ_C, and γ'_C. We set

$$\delta_C = \min\left(\delta_A, \delta_B, \frac{1}{20000}, \frac{\varepsilon_C^2}{200r^2}\right)$$
$$\gamma'_C = \min\left(\gamma_C, \frac{1}{20000c}, \frac{\varepsilon_C^2}{200cr^2}\right)$$
$$\eta_C \ge \max\left(\eta_A, \eta_B, \frac{12r}{\varepsilon_C}, 144\right)$$

where η_C is also large enough to guarantee all graphs of this form are sufficiently close to $\text{OPT}_r(P_3)$.

These choices assure that we can combine Lemmas 3.4.2 and 3.4.4 to produce a G' with $\text{Dist}(G, G') \le \gamma'_C n^2 \le \gamma_C n^2$ also that

$$\delta_C + c\gamma'_C \le \frac{1}{20000} + \frac{c}{20000c} = \frac{1}{10000} < \frac{1}{5616}$$

and

$$\delta_C + c\gamma'_C \le \frac{\varepsilon_C^2}{200r^2} + \frac{c\varepsilon_C^2}{200cr^2} = \frac{\varepsilon_C^2}{100r^2} < \frac{\varepsilon_C^2}{96r^2},$$

as well as the bounds we use on n, all hold. ■

We are now ready to prove Lemma 3.4.1.

Proof of Lemma 3.4.1. Let $\varepsilon > 0$ be given. Set $n_0 = \eta_C$ and $\delta = \delta_C$ from Lemma 3.4.6 with $\gamma_C = \varepsilon/2$ and $\varepsilon_C = \varepsilon/2r$. Then given a graph G of order $n \ge n_0$ that satisfies $d(P_3, G) \ge \text{OPT}_r(P_3) - \delta$, we get a complete r-partite graph G' satisfying $\text{Dist}(G, G') \le \frac{\varepsilon}{2}n^2$ and with parts $X_1, \ldots, X_r$ satisfying

$$\frac{1-\frac{\varepsilon}{2r}}{r}n \le |X_i| \le \frac{1+\frac{\varepsilon}{2r}}{r}n.$$

We claim $\mathrm{Dist}(G', T_r(n)) \le \frac{\varepsilon}{2}n^2$. From each of the r parts, at most $\frac{\varepsilon}{2r}n$ vertices must be added to or removed from that part. Thus in total, $\frac{\varepsilon}{2}n$ vertices are altered. Each vertex requires changing at most n adjacencies, so the total edit distance is bounded above by $\frac{\varepsilon}{2}n^2$.

Finally, by first making the at most $\frac{\varepsilon}{2}n^2$ edits to change G into G' and then making the at most $\frac{\varepsilon}{2}n^2$ edits to change G' into $T_r(n)$, we have demonstrated $\mathrm{Dist}(G, T_r(n)) \le \varepsilon n^2$, completing the proof. ■

3.5 Exact Result

In this section we will prove Theorem 3.2.3(ii). We now know that for large enough n, if G is an n-vertex K_{r+1}-free graph that is close to being extremal, then G is close in edit-distance to $T_r(n)$. As we will show in this section, the process of adding or removing the necessary edges in order to transform G into $T_r(n)$ must increase the number of P_3-subgraphs in G. First we need the following proposition, which shows that in any extremal graph each pair of vertices must be contained in approximately the same number of P_3-subgraphs. We define $\nu_G(v, T)$ as the number of (not necessarily induced) subgraphs of a graph G isomorphic to T containing v.

Proposition 3.5.1. *Fix $r \ge 4$. Then there exists an $n_0 = n_0(r)$ such that if G a K_{r+1}-free graph on $n \ge n_0$ vertices for which $\nu(P_3, G) = ex(n, P_3, K_{r+1})$, then for every vertex $v \in V(G)$*

$$\nu_G(v, P_3) \ge \left(\mathrm{OPT}_r(P_3) - \frac{1}{r^{10}}\right)\binom{n-1}{3} - \frac{1}{r^4}n^3.$$

Proof. From the proof of Theorem 3.2.3(i), there must exist some n_0 such that

$$\nu(P_3, G) \ge \left(\mathrm{OPT}_r(P_3) - \frac{1}{r^{10}}\right)\binom{n}{4}$$

for every extremal graph G on $n \ge n_0$ vertices. Suppose that G is such a graph on $n \ge \max\{n_0, 2r^4\}$ vertices. We count the copies of P_3 in G in two ways to see

$$\sum_{v \in V(G)} \nu_G(v, P_3) = 4\nu(P_3, G) \ge 4\left(\mathrm{OPT}_r(P_3) - \frac{1}{r^{10}}\right)\binom{n}{4}.$$

Thus, by averaging there must exist some vertex $u \in V(G)$ for which

$$\nu_G(u, P_3) \ge \left(\mathrm{OPT}_r(P_3) - \frac{1}{r^{10}}\right)\binom{n-1}{3}.$$

Suppose for contradiction that for some vertex $v \in V(G)$,

$$\nu_G(v, P_3) < \left(\mathrm{OPT}_r(P_3) - \frac{1}{r^{10}}\right)\binom{n-1}{3} - \frac{1}{r^4}n^3.$$

Let G' be the graph obtained from G by deleting v and replacing it with a vertex u' so that $N(u') = N(u)$. We claim that G' is K_{r+1}-free. Suppose for contradiction that it is not. Then u' must be contained in every copy of K_{r+1} in G'. As u is not adjacent to u', none of these K_{r+1} contain u. However, since $N(u) = N(u')$, this implies that we can replace u' with u in each $(r+1)$-clique. Since $V(G') - \{u'\} = V(G) - \{v\}$, this implies the existence of an $(r+1)$-clique in G, which is a contradiciton.

Let $\nu_G(u, v, P_3)$ denote the number of P_3 subgraphs containing both u and v in G. Then since $\nu_G(u', P_3) = \nu_G(u, P_3)$, we have added at least $\nu_G(u, P_3) - \nu_G(u, v, P_3)$ subgraphs and removed at most $\nu_G(v, P_3)$ subgraphs. Hence,

$$\nu(P_3, G') = \nu(P_3, G) + \nu_G(u, P_3) - \nu_G(u, v, P_3) - \nu_G(v, P_3)$$

Since $\nu_G(u, v, P_3) \leq 2n^2$,

$$\nu(P_3, G') > \nu(P_3, G) + \frac{1}{r^4}n^3 - 2n^2.$$

By assumption, $\frac{1}{r^4}n^3 - 2n^2 > 0$. This would imply that $\nu(P_3, G') > \nu(P_3, G)$ which contradicts the assumption that G was extremal, completing the proof. ■

We will also require the following proposition much later in the proof of Theorem 3.2.3(ii), where we will provide more explanation of why it is required. For completeness, we will state it here.

Proposition 3.5.2. *For all integers $r \geq 4$, there exists an $n_0 = n_0(r)$ such that for all $n \geq n_0$,*

(i) $\mathrm{OPT}_r(P_3)\binom{n-1}{3} - \delta_1(r)\frac{n^3}{6} \geq \left(\frac{9}{r} - \frac{39}{2r^2}\right)n^3$, *where*

$$\delta_1(r) = 12 - \frac{45}{r} + \frac{111}{2r^2} - \frac{27}{2r^3} - \frac{21}{r^4} + \frac{24}{r^5} + \frac{3}{2r^6} - \frac{3}{2r^7}.$$

(ii) $\mathrm{OPT}_r(P_3)\binom{n-1}{3} - \delta_2(r)\frac{n^3}{6} \geq \left(\frac{18}{r} - \frac{42}{r^2} - \frac{12}{r^3}\right)n^3$, *where*

$$\delta_2(r) = 12 - \frac{54}{r} + \frac{78}{r^2} - \frac{96}{r^4} + \frac{72}{r^5} + \frac{24}{r^6} - \frac{24}{r^7}.$$

Proof. Part (i) immediately follows from the inequality below, which is true for all $r \geq 4$.

$$\mathrm{OPT}_r(P_3) - \delta_1(r) = \frac{9}{r} - \frac{39}{2r^2} + \frac{3}{2r^3} + \frac{21}{r^4} - \frac{24}{r^5} - \frac{3}{2r^6} + \frac{3}{2r^7} > \frac{9}{r} - \frac{39}{2r^2},$$

In an identical manner, part (ii) is implied from the following, which is true for all $r \geq 4$.

$$\mathrm{OPT}_r(P_3) - \delta_2(r) = \frac{18}{r} - \frac{42}{r^2} - \frac{12}{r^3} + \frac{96}{r^4} - \frac{72}{r^5} - \frac{24}{r^6} + \frac{24}{r^7} > \frac{18}{r} - \frac{42}{r^2} - \frac{12}{r^3},$$

completing the proof. ∎

After assuming that G is an extremal graph, and therefore close in edit-distance to $T_r(n)$, we will show that most vertices in G must closely resemble a vertex appearing in the Turán graph. Given this fact, we will use Proposition 3.5.1 to show that any vertex that does not look like this cannot be contained in enough copies of P_3 to justify G being extremal. This will ultimately show that G must be isomorphic to $T_r(n)$, since the removal/duplication process described in the proof of Proposition 3.5.1 would otherwise increase the number of P_3 copies in G.

Proof of Theorem 3.2.3(ii). Let $r \geq 4$. Fix $\varepsilon > 0$ and assume that $n_0 = n_0(r)$ is large enough to satisfy the following conditions.

(i) Any K_{r+1}-free graph G on $n \geq n_0$ vertices with

$$d(P_3, G) > OPT_r(P_3) - \varepsilon$$

must satisfy $\mathrm{Dist}(G, T_r(n)) \leq \frac{2}{r^{10}} n^2$.

(ii) $n_0 \geq 2r^4$ and is large enough to satisfy the conditions of Proposition 3.5.1.

(iii) n_0 is large enough to satisfy the conditions of Proposition 3.5.2.

Let G be an extremal graph on $n \geq n_0$ vertices. Recall that (i) means that we can transform G into $T_r(n)$ by changing at most $\frac{2}{r^{10}} n^2$ adjacencies. We will call each edge removed in the process of transforming G into $T_r(n)$ a *surplus edge*, and each added edge a *missing edge*. Let $b(v)$ denote the total number of surplus edges and missing edges incident with a vertex v. If v is a vertex for which $b(v) > \frac{1}{r^5} n$, then we say that v is a *bad vertex*.

Partition the vertex set of G into sets $X_1, X_2, \ldots, X_r$ so that after changing all required adjancencies in G the sets $X_1, X_2, \ldots, X_r$ are the partite sets of $T_r(n)$. For the moment, move each bad vertex from its original set and place it into a new set X_0.

Claim 3.5.3. $|X_0| \leq \frac{1}{r^5}n$.

Proof. Since $\mathrm{Dist}(G, T_r(n)) \leq \frac{2}{r^{10}}n^2$ and each vertex $v \in X_0$ satisfies $b(v) > \frac{1}{r^5}n$,

$$|X_0| \cdot \frac{1}{r^5}n \leq \frac{1}{r^{10}}n^2.$$

Claim 3.5.3 follows immediately. ■

Now we will show that all surplus edges must be incident with at least one vertex in X_0. This will allow us to focus only on the bad vertices. For a finite collection of vertices $x_1, x_2, \ldots, x_\ell \in V(G)$ let $N(x_1, x_2, \ldots, x_\ell)$ denote the *common neighborhood* of $x_1, x_2, \ldots, x_\ell$, which is the set of vertices in $V(G)$ adjacent to each of $x_1, x_2, \ldots, x_\ell$.

Claim 3.5.4. *There are no surplus edges in* $V(G) \setminus X_0$.

Proof. Suppose for contradiction that for two vertices u and v in $X_j \setminus X_0$ are adjacent for some integer $j \in [r]$. By symmetry we may assume that $j = 1$. Since neither vertex is contained in X_0, both u and v are incident with at most $\frac{1}{r^5}n$ missing edges in $X_2 \setminus X_0$. This implies that there are at most $\frac{2}{r^5}n$ vertices in $X_2 \setminus X_0$ not contained in $N(u, v)$. Since Claim 3.5.3 implies that we have moved at most $\frac{1}{r^5}n$ vertices from X_2 to X_0,

$$|(N(u, v) \cap X_2) \setminus X_0| \geq \left\lfloor \frac{n}{r} \right\rfloor - \left\lfloor \frac{3n}{r^5} \right\rfloor > 0.$$

Let w_2 be one of the vertices contained in the set $(N(u, v) \cap X_2) \setminus X_0$. Then uvw_2 induces a triangle in G. Since w_2 is also only incident with $\frac{1}{r^5}n$ missing edges, we can apply an identical argument using u, v, w_2 and the set X_3 to show:

$$|(N(u, v, w_2) \cap X_3) \setminus X_0| \geq \left\lfloor \frac{n}{r} \right\rfloor - \left\lfloor \frac{4n}{r^5} \right\rfloor > 0,$$

implying that we can find some $w_3 \in X_3$ such that uvw_2w_3 induces a K_4 in G. Continuing this process for each $j \in \{4, \ldots, r\}$, we can always select one vertex $w_j \in X_j$ in an identical manner so that $uvw_2 \dots w_j$ induces a copy of K_{j+1} in G. This is possible since

$$|(N(u,v,w_2,\dots,w_{j-1}) \cap X_j) \setminus X_0| \geq \left\lfloor \frac{n}{r} \right\rfloor - \left\lfloor \frac{(j+1)n}{r^5} \right\rfloor > 0$$

for each j. This, however, would imply that after selecting vertices $u, v, w_2, \dots, w_{r-1}$ that induce a copy of K_r,

$$|(N(u,v,w_2,\dots,w_{r-1}) \cap X_r) \setminus X_0| \geq \left\lfloor \frac{n}{r} \right\rfloor - \left\lfloor \frac{(r+1)n}{r^5} \right\rfloor > 0.$$

Thus, we can select a vertex in X_r that is adjacent to each of $u, v, w_2, \dots, w_{r-1}$. This, however, induces a copy of K_{r+1} in G which is a contradiction. ■

For each $i \in [r]$, let $d_i(v) = |(N(v) \cap X_i) \setminus X_0|$. We say that $v \in X_0$ is a *type* 2 *vertex* if $d_i(v) > 0$ for all $i = 1, \dots, r$. Otherwise, if there exists some $i \in [r]$ for which $d_i(v) = 0$, then v is a *type* 1 *vertex.*

Claim 3.5.5. *If v is a type 2 vertex, then there exist $i, j \in [r]$ for which*

$$1 \leq d_i(v) \leq d_j(v) \leq \frac{1}{r^3} n.$$

Proof. Suppose for contradiction that for all $i \in [r]$, $d_i(v) > \frac{1}{r^3}n$. By symmetry, we may assume that

$$\frac{1}{r^3} n < d_1(v) \leq d_2(v) \leq \dots \leq d_r(v).$$

Let $w_1 \in X_1$ be a neighbor of v. Then $d_i(w_1) \geq \frac{n}{r} - \frac{n}{r^5}$ for all integers $i \geq 2$ since $w \notin X_0$. This, along with Claim 3.5.3, implies that

$$|(N(v,w_1) \cap X_2) \setminus X_0| \geq \left\lfloor \frac{n}{r^3} \right\rfloor - \left\lfloor \frac{2n}{r^5} \right\rfloor > 0.$$

Using an argument identical to that in Claim 3.5.4, we can continue selecting vertices $w_j \in X_j$ for each $j \in \{3, \dots, r\}$ so that $vw_1w_2 \dots w_j$ induces a copy of K_{j+1}. This is possible since for each $j \in \{3, \dots, r-1\}$,

$$|(N(v,w_1,w_2 \dots, w_{j-1}) \cap X_j) \setminus X_0| \geq \left\lfloor \frac{n}{r^3} \right\rfloor - \left\lfloor \frac{jn}{r^5} \right\rfloor > 0.$$

This would imply, however, that $vw_1 \ldots w_r$ induces a copy of K_{r+1}. Since the above argument only relied on $d_1(v)$ being nonzero, and v is a type 2 vertex, this implies that $d_2(v) < \frac{1}{r^3}n$ completing the proof of Claim 3.5.5. ■

Given $v \in G$ and a path $P = vxyz$ or $P = xvyz$ of P_3 containing v, we say that P is v-*good* if none x, y, or z is contained in X_0. The next claim will show that a type 2 vertex in G would not be contained in enough copies of P_3 to justify G being extremal.

Claim 3.5.6. *G does not contain any type 2 vertices.*

Proof. Suppose that $v \in X_0$ is a type 2 vertex. Then by symmetry, $d_1(v) \leq d_2(v) < \frac{1}{r^3}n$. Let $vu_1u_2u_3$ be a v-good path. We can count the number of these paths by considering the possible locations of u_1. The following list will provide the location of u_1, followed by the maximum number of paths of the form $vu_1u_2u_3$.

1. If $u_1 \in X_1$ or $u_1 \in X_2$, then there are at most $\frac{2n}{r^3}$ ways to select u_1. Otherwise, there are at most $\frac{r-2}{r}n$ ways to select u_1. There are $\frac{(r-1)^2}{r^2}n^2$ ways to select u_2 and u_3 since the only requirement is that each vertex cannot be in the same set as its predecessor. This gives
$$\left(\frac{2}{r^3} + \frac{r-2}{r}\right) \cdot \frac{(r-1)^2}{r^2}n^3$$
v-good copies of P_3 where v is an end point.

Next suppose that $u_1vu_2u_3$ is a v-good path. The maximum number of such paths can be counted by considering the locations of u_1 and u_2. In each case below, we give the location of u_1 and u_2, followed by the corresponding maximum number of P_3 subgraphs.

1. If $u_1, u_2 \in X_1$ or $u_1, u_2 \in X_2$, then there are at most $\frac{n^2}{r^6}$ ways to select each of u_1 and u_2 from either of the two sets. There are at most $\frac{r-1}{r}n$ ways to select u_3. If $u_1 \in X_1$ and $u_2 \in X_2$ or $u_1 \in X_2$ and $u_2 \in X_1$, then there are at most then there are at most $\frac{n^2}{r^6}$ ways to select u_1 and u_2 from each of their given sets. Again, there are at most $\frac{r-1}{r}n$ ways to select u_3. This accounts for at most
$$\frac{4}{r^6} \cdot \frac{(r-1)}{r} \cdot n^3.$$

copies of P_3.

2. If exactly one of u_1 or u_2 is contained in $X_1 \cup X_2$, then there are at most $\frac{n}{r^3}$ ways to select that particular vertex. The vertex not in $X_1 \cup X_2$ can be selected from $(r-2)$ possible sets. Thus, there are $\frac{4(r-2)n^2}{r^4}$ ways to select u_1 and u_2. Finally, there are at most $\frac{r-1}{r}n$ ways to select u_3. This accounts for at most

$$\frac{4}{r^3} \cdot \frac{(r-1)(r-2)}{r^2} \cdot n^3$$

copies of P_3.

3. If $u_1, u_2 \notin X_1 \cup X_2$, then there are at most $\frac{(r-2)(r-3)}{r^2}n^2$ ways to choose u_1 and u_2 if they are in different sets, and $\frac{(r-2)}{r^2}n^2$ ways to choose u_1 and u_2 if they are in the same set. As there are at most $\frac{r-1}{r}n$ ways to select u_3, this accounts for at most

$$\left(\frac{(r-1)(r-2)(r-3)}{r^3} + \frac{(r-1)(r-2)}{r^3}\right) n^3$$

copies of P_3.

There are at most $\frac{2}{r^5}n^3$ subgraphs containing v and at least one other vertex in X_0. Thus, combining each of the terms we have calculated, we get the following upper bound:

$$\nu(v, P_3) \leq \delta_2(r)\frac{n^3}{6}.$$

where $\delta_2(r)$ is taken from Proposition 3.5.2, which then implies the following:

$$\mathrm{OPT}_r(P_3)\binom{n-1}{3} - \nu(v, P_3) \geq \left(\frac{18}{r} - \frac{42}{r^2} - \frac{12}{r^3}\right) n^3. \tag{3.10}$$

It is straightforward to verify that for all $r \geq 4$,

$$\frac{18}{r} - \frac{48}{r^2} - \frac{12}{r^3} > \frac{1}{r^4}.$$

Since (3.10) must be true of each type 2 vertex and G is assumed to be an extremal graph, Proposition 3.5.1 implies that G cannot contain any type 2 vertices. ■

By Claim 3.5.6, each $v \in X_0$ is a type 1 vertex. We will now show that if u and v are two type 1 vertices for which $d_i(v) = d_i(u) = 0$, then u and v cannot be adjacent. Specifically, we will prove that if u and v are adjacent, then one or the other is not contained in sufficiently many P_3 subgraphs to justify G being extremal. Note this is slightly different from our approach to type 2 vertices, as we will not disprove the existence of type 1 vertices.

Claim 3.5.7. *Suppose that u and v are two adjacent type 1 vertices for which $d_i(v) = d_i(u) = 0$. Then there exists some index $j \neq i$ for which*

$$|N(u,v) \cap (X_j \setminus X_0)| \leq \frac{1}{r^3}n.$$

Proof. By symmetry we may assume that $i = 1$. Suppose for contradiction that $|N(u,v) \cap (X_j \setminus X_0)| > \frac{1}{r^3}n$ for all $j \in \{2,\dots,r\}$. Using an argument identical to those in Claims 3.5.4 and 3.5.5, select one vertex w_j in X_j for all $j = 2,\dots,r$, starting with X_2, so that $w_j \in N(u,v,w_2,\dots,w_{j-1}) \cap X_j$. This is possible since

$$|(N(u,v,w_2,\dots,w_{j-1}) \cap X_j) \setminus X_0| \geq \left\lfloor \frac{n}{r^3} \right\rfloor - \left\lfloor \frac{(j-1)n}{r^5} \right\rfloor > 0$$

for all $j \in \{2,\dots,r-1\}$. After selecting vertices $w_2,\dots,w_r$ in this way, we once again obtain a copy of K_{r+1} in G which is a contradiction. ■

Claim 3.5.8. *If u and v are two type one vertices for which $d_i(v) = d_i(u) = 0$, then u and v are not adjacent.*

Proof. By symmetry, Claim 3.5.7 implies that

$$N(u,v) \cap (X_2 \setminus X_0) \leq \frac{1}{r^3}n.$$

Therefore without loss of generality,

$$|(N(v) \cap (X_2 \setminus X_0)) \setminus N(u)| \leq \frac{r^2-1}{2r^3}n.$$

Hence,

$$d_2(v) \leq \frac{r^2+1}{2r^3}n.$$

Suppose that $vu_1u_2u_3$ is a v-good path. Similar to Claim 3.5.6 we can count the number of such paths by considering the location of u_1.

1. If $u_1 \in X_2$ then there are $\frac{r^2+1}{2r^3}n$ ways to choose u_1. Otherwise, there are $\frac{r-2}{r}n$ ways to choose u_1. Similar to before, there are $\frac{(r-1)^2}{r^2}n^2$ ways to choose u_2 and u_3. This accounts for at most
$$\left(\frac{r^2+1}{2r^3} + \frac{r-2}{r}\right)\frac{(r-1)^2}{r^2}n^3$$
copies of P_3 where v is an end-vertex.

Next we can count the number of v-good paths of the form $u_1vu_2u_3$ by considering the locations of u_1 and u_2.

1. If $u_1, u_2 \in X_2$, then there are at most $\left(\frac{r^2+1}{2r^3}n\right)^2$ ways to select u_1 and u_2. There are $\frac{(r-1)}{r}n$ ways to select u_3. This gives an upper bound of
$$\left(\frac{r^2+1}{2r^3}\right)^2 \cdot \frac{(r-1)}{r} \cdot n^3$$
copies of P_3.

2. If exactly one of u_1 or u_2 is contained in X_2, then there are $\frac{r^2+1}{2r^3}n$ to choose that specific vertex. Since neither of the remaining vertices can be contained in the same set as its neighbors, there are at most $\frac{(r-1)(r-2)}{r^2}n^2$ ways to choose the remaining vertices on the path. This gives at most
$$2 \cdot \frac{r^2+1}{2r^3} \cdot \frac{(r-1)(r-2)}{r^2} \cdot n^3$$
copies of P_3.

3. If $u_1, u_2 \notin X_2$, then there are at most $\frac{(r-2)(r-3)}{r^2}n^2$ ways to choose u_1 and u_2 if they are in a different set and $\frac{(r-2)}{r^2}n^2$ ways if they are in the same set. There are $\frac{(r-1)}{r}n$ ways to select u_3, giving an upper bound of
$$\left(\frac{(r-1)(r-2)(r-3)}{r^3} + \frac{(r-1)(r-2)}{r^3}\right)n^3$$
copies of P_3.

Since there are at most $\frac{2}{r^5}n^3$ copies of P_3 containing v and at least one other vertex in X_0,

$$\nu(v, P_3) \leq \delta_1(r)\frac{n^3}{6}.$$

Where $\delta_1(r)$ is taken from Proposition 3.5.2, which implies the following:

$$\text{OPT}_r(P_3)\binom{n-1}{3} - \nu(v, P_3) \geq \left(\frac{9}{r} - \frac{39}{2r^2}\right)n^3.$$

It is straightforward to verify that for all $r \geq 4$,

$$\frac{9}{r} - \frac{39}{2r^2} > \frac{1}{r^4}.$$

Thus, by Proposition 3.5.1, vertex v cannot exist in G under the assumption that G is extremal. Since u and v were arbitrarily chosen, this completes the proof of Claim 3.5.8. ■

Proof of Theorem 1.3(ii), continued. From Claim 3.5.8, if two vertices u and v in X_0 have the property that $d_i(u) = d_i(v) = 0$ for some $i \in [j]$, then u and v cannot be adjacent. Thus, we can take each vertex in X_0 (since each vertex is a type 1 vertex) and place it in some partite set so that G is a an r-partite graph. Adding the necessary edges to make G a complete r-partite graph, however, would increase the number of P_3 subgraphs in G. As we have already shown by Proposition 3.4.5 that the Turán graph is best possible among all complete r-partite graphs, this completes the proof of Theorem 3.2.3(ii). ■

3.6 Concluding Remarks

The main result in this paper follows a similar approach to that used in [20], which determined that the five cycle C_5 is also K_{r+1}-Turán-good for $r \geq 3$. It is likely that this method could be applied to other graphs, perhaps including P_4 or C_6. However, as the number of vertices in the target graph increases, the number of graphs considered in the flag algebra step grow exponentially and the number of cases in the stability result increase as well. This difficulty is noted in [25] as well; there, the authors were able to extend their technique to show that P_4 is K_{r+1}-Turán-good but did not progress further. Therefore, the authors believe a different method will need to be used to investigate the conjecture of Gerbner and Palmer that P_ℓ is K_{r+1}-Turán-good for all values of ℓ.

3.7 Acknowledgements

The authors would like to thank Bernard Lidický for the use of his flag algebra program.

3.8 References

[1] N. Alon, E. Fischer, M. Krivelevich, and M. Szegedy. Efficient testing of large graphs. *Combinatorica*, 20(4):451–476, 2000.

[2] N. Alon and C. Shikhelman. Many T copies in H-free graphs. *J. Combin. Theory Ser. B*, 121:146–172, 2016.

[3] R. Baber. Some results in extremal combinatorics. *Dissertation*, 2011.

[4] J. Balogh, P. Hu, B. Lidický, and F. Pfender. Maximum density of induced 5-cycle is achieved by an iterated blow-up of 5-cycle. *European J. Combin.*, 52(part A):47–58, 2016.

[5] P. Bennett, A. Dudek, B. Lidický, and O. Pikhurko. Minimizing the number of 5-cycles in graphs with given edge-density. *Combin. Probab. Comput.*, 29(1):44–67, 2020.

[6] B. Bollobás and E. Györi. Pentagons vs. triangles. *Discrete Mathematics*, 308:4332–4336, 2008.

[7] I. Choi, B. Lidický, and F. Pfender. Inducibility of directed paths. *Discrete Math.*, 343(10):112015, 11, 2020.

[8] J. Cutler, J. Nir, and A. J. Radcliffe. Supersaturation for subgraph counts. *arXiv e-prints*, page arXiv:1903.08059, Mar 2019.

[9] P. Erdős and M. Simonovits. A limit theorem in graph theory. *STUDIA SCI. MATH. HUNG*, 1:51–57, 1965.

[10] V. Falgas-Ravry and E. R. Vaughan. Turán H-densities for 3-graphs. *Electron. J. Combin.*, 19(3):Paper 40, 26, 2012.

[11] D. Gerbner. On turán-good graphs, 2020.

[12] D. Gerbner and C. Palmer. Counting copies of a fixed subgraph in f-free graphs. *European Journal of Combinatorics*, 82:103001, 2019.

[13] D. Gerbner and C. Palmer. Some exact results for generalized Turán problems. *arXiv e-prints*, page arXiv:2006.03756, June 2020.

[14] R. Glebov, D. Král', and J. Volec. A problem of Erdős and Sós on 3-graphs. *Israel J. Math.*, 211(1):349–366, 2016.

[15] A. Grzesik. On the maximum number of five-cycles in a triangle-free graph. *J. Combin. Theory Ser. B*, 102(5):1061–1066, 2012.

[16] E. Györi. On the number of c5's in a triangle-free graph. *Combinatorica*, 9:101–102, 1989.

[17] A. Halfpap and C. Palmer. On supersaturation and stability for generalized Turán problems. *arXiv e-prints*, page arXiv:1909.13043, Sep 2019.

[18] H. Hatami, J. Hladký, D. Král', S. Norine, and A. Razborov. On the number of pentagons in triangle-free graphs. *J. Combin. Theory Ser. A*, 120(3):722–732, 2013.

[19] J. Hladký, D. Král', and S. Norin. Counting flags in triangle-free digraphs. *Combinatorica*, 37(1):49–76, 2017.

[20] B. Lidický and K. Murphy. Maximizing five-cycles in K_r-free graphs. *arXiv e-prints*, page arXiv:2007.03064, July 2020.

[21] P.-S. Loh, M. Tait, C. Timmons, and R. M. Zhou. Induced turán numbers. *Combinatorics, Probability and Computing*, 27(2):274–288, 2018.

[22] R. Luo. The maximum number of cliques in graphs without long cycles. *Journal of Combinatorial Theory, Series B*, 128:219–226, Jan 2018.

[23] J. Ma and Y. Qiu. Some sharp results on the generalized turán numbers. *European Journal of Combinatorics*, 84:103026, 2020.

[24] O. Pikhurko. The minimum size of 3-graphs without a 4-set spanning no or exactly three edges. *European J. Combin.*, 32(7):1142–1155, 2011.

[25] B. Qian, C. Xie, and G. Ge. Some results on k-turán-good graphs, 2021.

[26] A. A. Razborov. Flag algebras. *J. Symbolic Logic*, 72(4):1239–1282, 2007.

[27] A. A. Razborov. On the minimal density of triangles in graphs. *Combin. Probab. Comput.*, 17(4):603–618, 2008.

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

[29] A. A. Zykov. On some properties of linear complexes. *Matematicheskii sbornik*, 66:163–188, 1949.

3.9 Appendix

```
# This code can be used to calculate the values of C_Fi for each graph in F_4,
#and then verify that equation (8) is true for all 4 <= r <= 1000

var('k')

# Each entry in this list corresponds to a different graph in F_4.
#The enumeration of these graphs matches Figure 3 at the beginning
#of the proof of Theorem 1.3.
C_F = [0]*11

# This list contains the number of P3 subgraphs in each graph in F_4.
#This corresponds to equation (4).
P3count = [0]*11
P3count[6] = 1
P3count[7] = 2
P3count[8] = 4
P3count[9] = 6
P3count[10] = 12


# These are the functions for P_0,
#obtained using Corollary 1.5 of Zykov's Theorem.
P_0 = [(k^3 - 6*k^2 + 11*k - 6)/k^3]*11
P_0[10] = (-6*k^2 + 11*k - 6)/k^3
```

```
# These are the coefficient values of P_1, P_2, and P_3.
P_1 = [0]*34
P_2 = [0]*34
P_3 = [0]*34

P_1[0] = 6*k^2 - 12*k + 6
P_1[1] = k^2 - 2*k + 1
P_1[2] = 1 - k
P_1[3] = 3-3*k
P_1[8] = 2
P_1[9] = 1

P_2[3] = 3
P_2[7] = 1
P_2[6] = -1
P_2[8] = -4

P_3[3] = 3*k^2 - 12*k+12
P_3[7] = k^2 - 8*k + 12
P_3[6] = k^2 - 6*k + 12
P_3[8] = 4*k^2 - 16*k + 16
P_3[9] = 20-8*k
P_3[10] = 24


# Values of the polynomials p_0,p_1,p_2,p_3.
```

```
p_0 = 18*(k^2 - 2*k + 1)/(3*k^2 - 11*k + 9)
p_1 = (3*k^3 - 10*k^2 + 7*k)/( (3*k^2 - 11*k + 9)*k^3 )
p_2 = 1/4*(9*k^5 - 32*k^4 + 25*k^3)/( (3*k^2 - 11*k + 9)*k^3)
p_3 = 1/4*(15*k^3 - 24*k^2 + 7*k)/( (3*k^2 - 11*k + 9)*k^3 )


#This combines the terms from equation (6) and calculates the values of C_Fi.
for i in [0..10]:
        C_F[i] = factor(expand(p_0*P_0[i] + p_1*P_1[i]\
        + p_2*P_2[i] + p_3*P_3[i] + P3count[i]))
        print('the coefficient of','F_',i,'equals',C_F[i])

# Now we will verify that coefficients 0,3,8,9,10 are maximum.
#In order to do so, we will check each C_Fi against C_F0.

#This function will verify that an input function x,
#which will be C_F[0] - C_F[i],
#is nonnegative for all r = 4 up to 1000.

def test_positive(x):
        for r in [4..1000]:
                if x.substitute(k=r) < 0:
                        return('C_F_0 is not maximum,\
                        coeffcient','fails when','r = ',r)
        return('for all r from 4 to',1000,'opt is max')

this verifies equation (8) for all r from 4 to 1000.
```

```
for i in [1..10]:
    print('Testing C_F',i,':',test_positive(C_F[0] - C_F[i]))
```

CHAPTER 4. ON WEAK FLEXIBILITY OF PLANAR GRAPHS

Bernard Lidický, Department of Mathematics, Iowa State University.

Tomáš Masařík, Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Poland & Department of Mathematics, Simon Fraser University, Burnaby, BC, Canada.

Kyle Murphy, Department of Mathematics, Iowa State University.

Shira Zerbib, Department of Mathematics, Iowa State University.

Modified from a manuscript currently under review in *Graphs and Combinatorics*

4.1 Abstract

Recently, Dvořák, Norin, and Postle introduced flexibility as an extension of list coloring on graphs [JGT 19']. In this new setting, each vertex v in some subset of $V(G)$ has a request for a certain color $r(v)$ in its list of colors $L(v)$. The goal is to find an L coloring satisfying many, but not necessarily all, of the requests.

The main studied question is whether there exists a universal constant $\varepsilon > 0$ such that any graph G in some graph class $\mathcal{C}$ satisfies at least ε proportion of the requests. More formally, for $k > 0$ the goal is to prove that for any graph $G \in \mathcal{C}$ on vertex set V, with any list assignment L of size k for each vertex, and for every $R \subseteq V$ and a request vector $(r(v) : v \in R,\ r(v) \in L(v))$, there exists an L-coloring of G satisfying at least $\varepsilon|R|$ requests. If this is true, then $\mathcal{C}$ is called *ε-flexible for lists of size k*.

Choi et al. [2] introduced the notion of *weak flexibility*, where $R = V$. We further develop this direction by introducing a tool to handle weak flexibility. We demonstrate this new tool by showing that for every positive integer b there exists $\varepsilon(b) > 0$ so that the class of planar graphs without K_4, C_5, C_6, C_7, B_b is weakly $\varepsilon(b)$-flexible for lists of size 4 (here K_n, C_n and B_n are the complete graph, a cycle, and a book on n vertices, respectively). We also show that the class of

planar graphs without K_4, C_5, C_6, C_7, B_5 is ε-flexible for lists of size 4. The results are tight as these graph classes are not even 3-colorable.

4.2 Introduction

A *k-coloring* of a graph G is a function $f : V(G) \to S$, where $|S| = k$. The elements of S are often called *colors*. A k-coloring of G is called *proper* if adjacent vertices are assigned different colors. Suppose that for each vertex v in G, we gave v a list $L(v)$ of available colors. A *list coloring* of a graph G is a proper coloring of G where each vertex v is assigned a color from $L(v)$. In particular, for two distinct vertices u and v, $L(u)$ and $L(v)$ might be different. A graph is *k-choosable* if every assignment L of at least k colors to each vertex guarantees an L-coloring. The *choosability* of a graph G is the minimum k such that G is k-choosable.

In many applications of list coloring, such as scheduling, some vertices may have preferences which are not directly captured by the lists themselves. For example, a professor may be willing to teach classes X,Y, or Z but prefers to teach X. Ideally, the scheduler can satisfy the specific requests of each professor, but it is often the case that they cannot. The goal is then to satisfy as many requests as possible. This idea motivates the following definitions.

A *weighted request* is a function w that assigns a nonnegative real number to each pair (v, c) where $v \in V(G)$ and $c \in L(v)$. For $\varepsilon > 0$, we say that w is *ε-satisfiable* if there exists an L-coloring φ of G such that

$$\sum_{v \in V(G)} w(v, \varphi(v)) \geq \varepsilon \cdot \sum_{v \in V(G), c \in L(v)} w(v, c).$$

The unweighted variant is defined as follows. A *request* for a graph G with a list assignment L is a function r with domain $\operatorname{dom}(r) \subseteq V(G)$ such that $r(v) \in L(v)$ for all $v \in \operatorname{dom}(r)$. In the special case that each vertex requests a color, i.e., $\operatorname{dom}(r) = V(G)$, we call such a request *widespread*. Analogously, for $\varepsilon > 0$, a request r is *ε-satisfiable* if there exists an L-coloring φ of G such that at least $\varepsilon|\operatorname{dom}(r)|$ vertices v in $\operatorname{dom}(r)$ receive color $r(v)$. We say that a graph G with list assignment L is *ε-flexible*, *weakly ε-flexible*, or *weighted ε-flexible* if every request, widespread request, or weighted request, respectively, is ε-satisfiable. Note that weak flexibility does not make sense in the weighted setting since one can set some weights to 0 to turn off the requests for these vertices. If G is (weighted/weakly) ε-flexible for every list assignment with lists of length k, we say that G is *(weighted/weakly) ε-flexible for lists of size k.* Note that for k-colorable graphs, if the lists are exactly the same the problem becomes trivial as by permuting the colors we can achieve $\frac{1}{k}$-flexibility [6].

The concept of ε-flexibility was introduced by Dvořák, Norin, and Postle [6]. Subsequently, it was studied for various sub-classes of planar graphs, e.g., triangle-free [4], girth six [5], or C_4-free [9]. Graphs of bounded maximum degree were subsequently characterized in terms of flexibility [1].

A central notion in graph coloring is that of *reducible configurations*, which are local subgraphs that cannot appear in a smallest counterexample because their presence implies that the graph can be colored from a smaller subgraph by induction. Reducible configurations for flexibility are slightly more delicate as we explain in Section 4.3. Recently, Choi et al. [2] proposed a strengthened tool (see Lemma 4.4.4 below) for designing reducible configurations for flexibility. The authors of [2] also introduced the notion of weak flexibility defined above. They demonstrated that the weak setting allows one to create stronger reducible configurations.

We further develop this direction by strengthening the tools for handling weak flexibility; see Lemma 4.4.7 in Section 4.4. We exhibit our new tool by showing the following results for subclasses of planar graphs.

For an integer $n \geq 3$ let B_n denote the book on n vertices, i.e., the graph consisting of $n-2$ triangles sharing an edge. Let C_n and K_n denote a cycle and a clique on n vertices, respectively. Given a set of graphs $\mathcal{F}$ and a graph H, we say that H is *$\mathcal{F}$-free* if there is no subgraph of H isomorphic to any of the graphs in $\mathcal{F}$.

Theorem 4.2.1. *There exists $\varepsilon > 0$ such that every planar $\{K_4, C_5, C_6, C_7, B_5\}$-free graph is weighted ε-flexible for lists of size 4.*

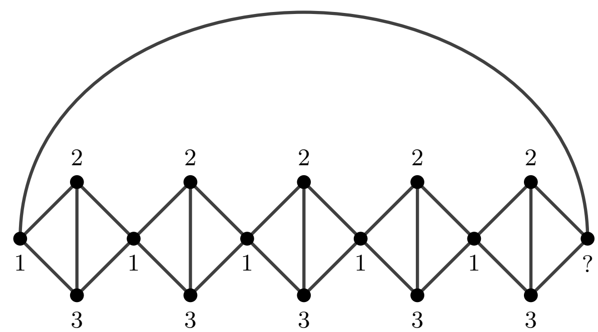


Figure 4.1 A construction proving Observation 4.2.3 with an attempt for a 3-coloring that fails.

Theorem 4.2.2. *There exists $\varepsilon = \varepsilon(b) > 0$ such that every planar $\{K_4, C_5, C_6, C_7, B_b\}$-free graph is weakly ε-flexible for lists of size 4.*

The results in Theorems 4.2.1 and 4.2.2 are tight as in general such graphs are not even 3-colorable. This is exemplified by the construction in Figure 4.2. This construction implies:

Observation 4.2.3. *For every $\ell, b \geq 5$ exists a $\{K_4, B_b\}$-free planar graph G that does not contain any cycle C_k of length $5 \leq k \leq \ell$, such that G is not 3-colorable.*

Furthermore, our results follow a recent line of research trying to narrow the gap between known degeneracy upper-bounds and choosability lower-bounds, in particular on subclasses of planar graphs, as is described below. We say that a graph G is *d-degenerate* if each induced subgraph of G contains a vertex of degree at most d. It is easy to observe that d-degenerate graphs are $(d+1)$-choosable. A similar statement holds for flexibility as well: in [6] it was proved that d-degenerate graphs with lists of size $d+2$ are weighted ε-flexible. Therefore, as C_5-free planar graphs are 3-degenerate [10], they are ε-flexible for lists of size 5. The same is true for C_6-free planar graphs [7]. For C_3-free graphs, Dvořák, Masařík, Musílek, and Pangrác [4] showed that they are weighted ε-flexible for lists of size 4 and that the result is tight. Surprisingly, the discharging proof in [4] is quite involved compared to the easy observation that C_3-free planar graphs are 3-degenerate, which implies 4-choosability. An analogous result holds for

$\{C_3, C_4, C_5\}$-free graphs, where list of size 3 are sufficient for weighted ε-flexibility and the result is tight [5].

When only C_4 is forbidden, Masařík [9] proved that lists of size 5 are sufficient for weighted ε-flexibility. However, it is unknown whether the result is tight as those graphs are 4-choosable [8] (but not necessarily 3-degenerate). There were attempts to bring down the list size to 4 but so far only partial results are known in this direction: planar graphs that do not contain C_4 and C_3 at distance at most 1 [2] or $\{C_4, C_5\}$-free planar graphs [11]. See [2, Table 1] for a comprehensive overview of known results for various subclasses of planar graphs. Our results aim to improve this narrow gap as they show that lists of size 4 are sufficient even for planar graphs in which some copies of C_3 and C_4 are allowed.

4.3 Methods - informal discussion

The purpose of this section is to informally describe some of the difficulties one faces when trying to extend a list-coloring proof to a flexibility proof. This discussion serves as the intuition behind the formal definitions in the next section.

As in previous related papers mentioned above, we use the discharging method to obtain our results. For an introduction to the discharging method see [3]. A typical discharging proof that a graph G is L-list-colorable gives a list of unavoidable *reducible configurations*, which are subgraphs of G that cannot appear in a minimal counterxample. The goal is to decompose G into subgraphs $R_1, \ldots, R_N$ such that R_i is a reducible configuration in $G[R_i \cup \cdots \cup R_N]$ (this will be defined as a *resolution* later), so that any L-coloring of $G[R_{i+1} \cup \cdots \cup R_N]$ can be extended to an L-coloring of $G[R_i \cup \cdots \cup R_N]$. Extending the coloring in a descending order from R_N to R_1 gives an L-coloring of G.

When requests are introduced, this method becomes more difficult. To explain this, assume for simplicity that every vertex has a request. If we manage to accommodate one request from each R_i and each R_i has at most b vertices, then we would satisfy n/b requests, showing that G is

ε-flexible for $\varepsilon = 1/b$, and our job would be done. However, this is not necessarily possible. Indeed, suppose $v \in R_i$ has some request $r(v)$ and let φ be an L-coloring of $G[R_{i+1} \cup \cdots \cup R_N]$. Suppose further that v has one neighbor u in $R_{i+1} \cup \cdots \cup R_N$. If $\varphi(u) = r(v)$, there is no way to simply extend φ and accommodate the request of v. Thus more changes to φ, such as recoloring u, would have to occur to accommodate $r(v)$. In addition to this issue, it may also be the case that $r(v)$ cannot be satisfied because R_i itself prevents it.

This means that reducible configurations for flexibility need to provide slightly more freedom in the colorings they allow. The easier problem to deal with is that $r(v)$ cannot be satisfied because R_i itself prevents it. This can be patched by adding a requirement that for any one vertex x in R_i, the coloring φ extends to R_i even if x has a list of size 1 after removing the colors of already colored neighbors of x in $R_{i+1} \cup \cdots \cup R_N$. For v, this would be used in case $L(v) = \{r(v)\}$ and $\varphi(u) \neq r(v)$. This requirement will be called (FIX) in the formal definitions.

The problem occurring when $r(v)$ cannot be satisfied because its neighbor u in $R_{i+1} \cup \cdots \cup R_N$ is already colored $r(v)$ is more complicated to solve. The idea is the following. Instead of constructing just one L-coloring φ, one needs to construct L-colorings $\varphi_1, \ldots, \varphi_\ell$ and in some of them, u gets colored by a color different than $r(v)$. Then $\varphi_1, \ldots, \varphi_\ell$ can be extended to $\varphi'_1, \ldots, \varphi'_{\ell'}$, where $r(v)$ is satisfied in some of them. At the end, this process gives a set of L-colorings of G and at least one of them satisfies a positive fraction of the requests. Formally, this is done by creating a probability distribution on L-colorings of G.

In order to make this idea work, there must be a sufficient variety of proper colorings for each reducible configuration. In our example, if we want to color v by $r(v)$, we cannot use $r(v)$ on u. We need to address this when we are coloring u and remove $r(v)$ from its list. Further, we would need to do this for each neighbor of v in $R_{i+1} \cup \cdots \cup R_N$. This is achieved in the following way. When we are L-coloring R_i, we look at all subsets $I \subseteq V(R_i)$ of vertices that could form a neighborhood of a vertex in $R_1 \cup \ldots \cup R_{i-1}$, i.e. in the set of not yet colored vertices. Individually for each I, we show that any proper L-coloring φ of $R_{i+1} \cup \cdots \cup R_N$ can still be extended to R_i

even if we decrease the sizes of the lists of vertices in I by 1. This will be called (FORB) in the formal definitions.

To summarize, the reducible configurations for flexibility must have size bounded by a constant, any one vertex can be precolored (FIX), and for different subsets of vertices, reducing their lists sizes by 1 does not break the extendability of the coloring (FORB).

The main feature of weak flexibility is that instead of demanding in (FIX) that "any one vertex in R_i can be precolored", it is enough to ask for "at least one vertex in R_i can be precolored". It is then easier to satisfy this version of (FIX).

We introduce an additional trick, where we ask "at least one vertex in R_i with few outside neighbors can be precolored". This makes satisfying (FIX) more difficult since the the reducible configurations needs a vertex of small degree but it helps a lot with checking (FORB).

4.4 Methods - definitions and lemmas

We use some of the notation and tools introduced in [2, 6, 4, 5]. In particular, our definitions are quite similar to those used in [2].

Let 1_I denote the characteristic function of I, i.e., $1_I(v) = 1$ if $v \in I$ and $1_I(v) = 0$ otherwise. Let G be a graph. Given a function $f : V(G) \to \mathbb{Z}$ and a vertex $v \in V(G)$, let $f \downarrow v$ denote the function satisfying $(f \downarrow v)(w) = f(w)$ for $w \neq v$ and $(f \downarrow v)(v) = 1$. We will use $f \downarrow v$ to indicate that the list size at vertex v has been reduced to 1. In other words, $f \downarrow v$ means that v has been "precolored". A list assignment L is an *f-assignment* if $|L(v)| \geq f(v)$ for all $v \in V(G)$. We will let $\deg_G$ be the function from $V(G)$ to $\mathbb{Z}$ which maps each vertex to its degree. If $X \subset V(G)$, then we let $\deg_X$ equal $\deg_{G[X]}$, where $G[X]$ is the induced subgraph of G consisting of the vertices in X.

Given a set of graphs $\mathcal{F}$ and a graph H, a set $I \subseteq V(H)$ is *$\mathcal{F}$-free* if the graph H together with one additional vertex u adjacent to all of the vertices in I does not contain any subgraph isomorphic to a graph in $\mathcal{F}$. Throughout the following definitions, let H be an induced subgraph of a graph G, let $\mathcal{F}$ be a set of graphs, and let k be a positive integer.

Definition 4.4.1 ($(\mathcal{F}, k)$-boundary-reducibility)**.** *We say that H is an $(\mathcal{F}, k)$*-boundary-reducible *subgraph if there exists a set $R \subseteq V(H)$ such that $R \neq \emptyset$ and*

(FIX) *for every $v \in R$, $H[R]$ is L-colorable for every $((k - \deg_G + \deg_R) \downarrow v)$-assignment L, and*

(FORB) *for every $\mathcal{F}$-free set $I \subseteq R$ of size at most $k-2$, $H[R]$ is L-colorable for every $(k - \deg_G + \deg_R - 1_I)$-assignment L.*

Definition 4.4.2 (weak $(\mathcal{F}, k)$-boundary-reducibility)**.** *We say that H is* weakly $(\mathcal{F}, k)$-boundary-reducible *if it satisfies* (FORB) *and there exists at least one vertex v satisfying* (FIX) *from Definition 4.4.1. In this case, we denote v by* $\mathrm{Fix}(H)$.

In both of the preceding definitions, we will occasionally refer to the set $V(H) \setminus R$ as the *boundary* of the configuration and the set R as the *reduced part* of the configuration. Note that (FORB) in particular implies that $\deg_G - \deg_R \leq k-2$ for all $v \in R$.

Definition 4.4.3 ($(\mathcal{F}, k, b)$-resolution)**.** *Let G be an $\mathcal{F}$-free graph with lists of size k. An* $(\mathcal{F}, k, b)$-resolution *of G is a set $\{G_0, G_1, \ldots, G_M\}$ of subgraphs of G such that for $i \geq 1$, H_i is an induced $(\mathcal{F}, k)$-boundary-reducible subgraph of G_{i-1} with reduced part R_i and*

$$G_i := G - \bigcup_{j=1}^{i} R_j.$$

Additionally, for each $i \geq 1$ $|R_i| \leq b$ and G_M is itself a $(\mathcal{F}, k)$-boundary-reducible graph with empty boundary and order at most b. For technical reasons, let $G_{M+1} := \emptyset$.

A *weak $(\mathcal{F}, k, b)$-resolution* is defined analogously, save that it uses weak $(\mathcal{F}, k)$-boundary-reducibility in the place of $(\mathcal{F}, k)$-boundary-reducibility.

The following lemma is the main tool we use for proving weighted ε-flexibility.

Lemma 4.4.4 (Lemma 13 in [2])**.** *For integers $k \geq 3$ and $b \geq 1$ and for a set $\mathcal{F}$ of forbidden subgraphs, let G be an $\mathcal{F}$-free graph with an $(\mathcal{F}, k, b)$-resolution. Then there exists an $\varepsilon > 0$ such that G is weighted ε-flexible for lists of size k. Furthermore, if the request is widespread and G has a weak $(\mathcal{F}, k, b)$-resolution, then G is weakly $\left(\varepsilon \cdot \frac{1}{b}\right)$-flexible for lists of size k.*

For the proof of Theorem 4.2.2 we prove a stronger version of Lemma 4.4.4 tailored to the setting of weak flexibility. For this, we define new "enhanced" versions of weak $(\mathcal{F}, k)$-boundary-reducibility and of a weak $(\mathcal{F}, k, b)$-resolution. We will now require $\mathrm{Fix}(H)$ to contain only vertices v satisfying $\deg_G(v) - \deg_R(v) \leq k - 3$. This change will allows us to consider smaller sets for the (FORB) condition.

Definition 4.4.5 (enhanced weak $(\mathcal{F}, k)$-boundary-reducibility)**.** *A graph H is* enhanced weakly $(\mathcal{F}, k)$-boundary-reducible *if there exist non-empty sets $\mathrm{Fix}(H) \subseteq R \subseteq V(H)$ such that*

(FIX) *for every $v \in \mathrm{Fix}(H)$, $\deg_G(v) - \deg_R(v) \leq k - 3$ and $H[R]$ is L-colorable for every $((k - \deg_G + \deg_R) \downarrow v)$-assignment L, and*

(FORB) *for every $\mathcal{F}$-free set $I \subseteq R$ of size at most $k - 3$, $H[R]$ is L-colorable for every $(k - \deg_G + \deg_R - 1_I)$-assignment L.*

Before proceeding further, observe that (FORB) in the enhanced version is easier to check because I is of size at most $k - 3$, instead of $k - 2$ in the non-enhanced version. However, (FIX) in the enhanced version has an additional restriction on the degree of vertices in $\mathrm{Fix}(H)$, which makes it more difficult to satisfy. Note that in general, the (FORB) condition on a single vertex v implies $\deg_G - \deg_R \leq k - 2$. However for vertices in $\mathrm{Fix}(H)$, the (FIX) condition implies $\deg_G - \deg_R \leq k - 3$. In particular, a vertex of degree $k - 2$ is no longer reducible under the enhanced definition. We overcome this obstacle by allowing $(k - 2)$-vertices in a resolution under special conditions. Forbidding books B_ℓ helps with satisfying these special conditions. By doing this we can have both: a vertex of degree $k - 2$ is reducible in our setting, and in addition we obtain subgraphs H that are reducible under the enhanced weak $(\mathcal{F}, k)$-boundary-reducibility definition, given that certain special circumstances occur. This rather technical improvement helps substantially in reducing the complexity of the analysis of the discharging process for the graph classes studied in this paper. Note that further generalization of this idea may be possible, but for lack of use in this paper we will not aim to formulate this in the full generality.

For a subgraph H of a graph G, let $N_G(H)$ be the induced subgraph of G on all neighbors of the vertices in H.

If G is a graph satisfying the conditions of Definition 4.4.5 and $I \subseteq R$ is an $\mathcal{F}$-free set of size $k-2$ so that $G[R]$ is L-colorable for every $(k - \deg_G + \deg_R - 1_I)$-assignment L, then we call I *loose*.

Let G be a graph, H its subgraph and $v \in V(G - H)$. We say that v is *H-tight* if $\deg_G(v) = k-2$, $N_G(v) \subseteq V(H)$, and $N_G(v)$ is not loose in H.

Definition 4.4.6 (enhanced weak $(\mathcal{F}, k, b, \beta)$-resolution)**.** *Let G be an $\mathcal{F}$-free graph with lists of size k. An* enhanced weak $(\mathcal{F}, k, b, \beta)$-resolution *of G is a set $\{G_0, G_1, \ldots, G_M\}$ of subgraphs of G, such that all the following three conditions hold:*

1. *For $1 \leq i \leq M$, there exists a subgraph H_i of G_{i-1} satisfying that*
   - *H_i is an induced enhanced weak $(\mathcal{F}, k)$-boundary-reducible subgraph of G_{i-1} with reducible part R_i such that $|R_i| \leq b$, or*
   - *H_i is an induced weak $(\mathcal{F}, k)$-boundary-reducible subgraph of G_{i-1} with reducible part R_i, such that $|R_i| \leq b$ and for all $v \in Fix(H_i)$ either $|N_{G_{i-1}}(v) \cap H_j| \leq k-3$ or $N_{G_{i-1}}(v) \cap H_j$ is a loose set in H_j for all $j > i$, or*
   - *H_i is a single vertex with $\deg_{G_{i-1}}(v) = k-2$.*
2. *For every $1 \leq i \leq M-1$,*
$$G_i := G - \bigcup_{j=1}^{i} R_j,$$
   G_M is a weak $(\mathcal{F}, k)$-boundary-reducible graph with empty boundary and order at most b, $G_{M+1} := \emptyset$, and $H_{M+1} := G_M$.
3. *The following is satisfied:*

(TIGHT) *For every $1 \leq j \leq M$, there are at most β different H_j-tight vertices v_i, where $V(H_i) = \{v_i\}$ with $i < j$.*

Note that in Definition 4.4.6 H_i can be H_j-tight only if H_i is a single vertex with $\deg_{G_{i-1}}(v) = k-2$. A natural way to satisfy (TIGHT) condition is to show that whenever there is an H_j such that more than β subgraphs $H_{a_1}, \ldots, H_{a_\beta}, H_{a_{\beta+1}}$ are H_j-tight, then

$$H_j \cup \bigcup_{i \in \{1,\ldots,\beta,\beta+1\}} H_{a_i} \in \mathcal{F}.$$

If two adjacent vertices have many common neighbors, we get a book, which will be in $\mathcal{F}$.

We are now ready to state and proof our main lemma.

Lemma 4.4.7 (Reducible configurations for weak flexibility)**.** *For integers $k \geq 4$, $b \geq 1$, $\beta \geq 0$, and for a set $\mathcal{F}$ of forbidden subgraphs, let G be a $\mathcal{F}$-free graph with an enhanced weak $(\mathcal{F}, k, b, \beta)$-resolution. Then, there exists an $\varepsilon > 0$ such that G is weakly ε-flexible for lists of size k.*

The proof of the lemma is similar the proof of Lemma 4.4.4 in [2]. In particular, we explicitly formulate a few arguments in their proof as a separate claim (Claim 4.4.9 below) that we use in our proof. We will also need the following Lemma 4.4.8, which is Lemma 12 in [2].

Let G be a graph with a weak $(\mathcal{F}, k, b)$-resolution $\mathcal{R}$. Let AllFix(G) denote the union of all Fix(H) over all reducible subgraphs H in the resolution $\mathcal{R}$.

Lemma 4.4.8 (Lemma 12 in [2])**.** *Let b be an integer. Let G be a graph with list assignment L of size k on $V(G)$. Suppose G has a weak $(\mathcal{F}, k, b)$-resolution, G is L-colorable, and there exists a probability distribution on the L-colorings φ of G such that for every $v \in$ AllFix(G) and $c \in L(v)$,* $\mathbf{Prob}[\varphi(v) = c] \geq \varepsilon$. *Then G with L is weakly $\left(\varepsilon \cdot \frac{1}{b}\right)$-flexible.*

Proof of Lemma 4.4.7. For $1 \leq j \leq M+1$, let $\mathcal{H}_j$ be the set of all H_j-tight subgraphs where the (TIGHT) property applied. Let $H_i \in \mathcal{H}_j$ for some i and j. This means that H_i is a single vertex with $k - 2 \geq 2$ neighbors in H_j. Hence $\mathcal{H}_i = \emptyset$.

Now, we refactor the enhanced weak $(\mathcal{F}, k, b, \beta)$-resolution $\mathcal{R}$ into an enhanced weak $(\mathcal{F}, k, b+\beta, 0)$-resolution $\mathcal{R}'$. To do so, we attach all H_j-tight subgraphs to H_j and thus we create a larger configuration H'_j. The vertices in tight subgraphs are not part of any Fix set. Formally

$$H'_j := \begin{cases} \emptyset & \text{if exists } i \text{ such that } H_j \in \mathcal{H}_i \\ H_j \cup \bigcup_{H \in \mathcal{H}_j} H & \text{otherwise} \end{cases}$$

and $\text{Fix}(H'_i) = \text{Fix}(H_i)$ if $H'_i \neq \emptyset$ and $\text{Fix}(H'_i) = \emptyset$ otherwise. Observe that by the (TIGHT) property, the size of the resulting H'_j will be upper-bounded by $b + \beta$ and that H'_j is enhanced weakly $(\mathcal{F}, k)$-boundary-reducible or only weakly $(\mathcal{F}, k)$-boundary-reducible (provided its neighbourhood is always a loose or small set) if it is not empty. We simultaneously remember both $\mathcal{R}$ and $\mathcal{R}'$, since each time we are using H'_j (or G'_j) we are referring to $\mathcal{R}'$ and each time we are using H_j (or G_j) we are referring to $\mathcal{R}$.

The next step is to create a probability distribution on L-colorings φ of G_i for all i starting with G'_M. Let $p = k^{-(b+\beta)}$ and $\varepsilon' = p^{k-1}$. We are going to show that each i satisfies the following properties:

(i) for every $v \in \text{AllFix}(G'_i)$ and a color $c \in L(v)$, the probability that $\varphi(v) = c$ is at least ε', and

(ii) for every color c and every $\mathcal{F}$-free set I in G'_i of size at most $k - 3$, the probability that $\varphi(v) \neq c$ for all $v \in I$ is at least $p^{|I|}$.

(iii) for every color c and every loose $\mathcal{F}$-free set I in G'_i of size exactly $k - 2$, the probability that $\varphi(v) \neq c$ for all $v \in I$ is at least $p^{|I|}$.

Note that for G'_{M+1} all of the properties trivially hold. Note that Property (i) on $G'_0 = G_0 = G$ immediately implies that G with L is weakly $\left(\varepsilon' \cdot \frac{1}{b}\right)$-flexible by Lemma 4.4.8 and therefore weakly ε-flexible for $\varepsilon = \frac{\varepsilon'}{b}$.

We will make use of the following claim proven implicitly in [2].

Claim 4.4.9 (Implicit in the proof Lemma 13 in [2])**.** *Suppose that we have an enhanced weak $(\mathcal{F}, k, b + \beta, 0)$-resolution and a probability distribution on L-colorings of G'_{i+1} satisfying Properties (i), (ii), and (iii) on G'_{i+1}. If for each vertex $v \in Fix(H'_i)$ and for each $I = N(v) \cap H'_j$ where $j > i$ one of the following holds:*

(a) $|I| = k-2$ *and* I *is loose in* H'_j*, or*

(b) $|I| < k-2$

then there exists a probability distribution on L-colorings of G'_i *such that Properties (i), (ii), and (iii) are satisfied on* G'_i*.*

In order to use Claim 4.4.9, we need verify (a) and (b). If H'_i is not a single vertex v with $\deg_{G_i}(b) = k-2$, then (a) or (b) hold by the definition of H'_i. Hence we need to check the case of H'_i being a single vertex v with $\deg_{G_i}(v) = k-2$. We do it by showing v is not H'_j tight for any $j \geq i$ in the following claim. It implies that for H'_i, either (a) or (b) is satisfied. In particular, we will show that we got rid of all tight subgraphs when we refactored $\mathcal{R}$ into $\mathcal{R}'$.

Claim 4.4.10. *There are no* $i < j$ *such that* H'_i *is* H'_j*-tight.*

Proof. Suppose for contradiction that H'_i is H'_j-tight for some $i < j$. By the definition, H'_i is one vertex v with degree $k-2$ in G_{i-1}. By the definition of $\mathcal{R}'$, v is not H_ℓ-tight for any $\ell > i$. In particular, v is not H_j-tight. Since v is H'_j-tight, $\mathcal{H}_j$ is not empty. Hence H'_j is a union of H_j and vertices W, where every $w \in W$ has $k-2$ neighbors in H_j. Since v is not H_j-tight, it has at most $k-3$ neighbors in H_j and at least one in W. Notice that every vertex w in W has $k-2$ neighbors in $\mathcal{H}_j$ hence a list of $k-1$ colors suffices for extending any coloring of $\mathcal{H}_j$ to w greedily. This and the (FORB) property for H_j imply that v is not H'_j-tight because $N(v)$ in H'_j is loose, which is a contradiction. ■

We conclude that Claim 4.4.10 enables us to use Claim 4.4.9 directly on $\mathcal{R}'$. This finishes the proof of Lemma 4.4.7. ■

For a positive integer d, a *d-vertex*, a *d^+-vertex*, and a *d^--vertex* are a vertex of degree d, at least d, and at most d, respectively. A *d-face*, a *d^+-face*, and a *d^--face* are defined analogously. A *(d_1, d_2, d_3)-face* is a 3-face where the degrees of the vertices on the face are d_1, d_2, d_3. We will sometimes call 3-faces *triangles*. A *diamond* D is a graph isomorphic to K_4 minus an edge. The 2-vertices of D are be called the *side vertices*, and the 3-vertices will be called the *middle vertices* of D. For a vertex v, denote by $d(v)$ the degree of v. Let G be a graph. By $T(a, b, c)$ we denote a triangle in G vertices of degree a, b, and c in G, and by $Dia(a-b,c,d)$ a diamond in G with middle vertices of degrees a and b and side vertices of degrees c and d.

Lemma 4.4.11. *Let G be a plane $\{C_5, C_6, C_7\}$-free graph. Suppose that v is the middle vertex of k distinct diamonds, and v is adjacent to m faces of size 3 or 4 that are not part of a diamond in which v is a middle vertex. Then $d(v) \geq 3k+2m$ and $k \leq \lfloor \frac{d(v)}{3} \rfloor$.*

Proof. If not, then v is adjacent to three faces f, g, h, each of them of size at most 4, such that f, g share an edge and g, h share an edge. But this induces a cycle C_i with $5 \leq i \leq 7$, a contradiction. ■

In all the figures in the paper, black vertices have all their incident edges drawn, whereas a white vertex may have more edges incident than drawn (since white vertices are in the boundary).

4.5 Proof of Theorem 4.2.1

4.5.1 Reducible Configurations

Let $\mathcal{F} = \{K_4, C_5, C_6, C_7, B_5\}$. In this section we will provide a handful of $(\mathcal{F}, 4)$-boundary-reducible configurations.

Lemma 4.5.1. *The following configurations are $(\mathcal{F}, 4)$-boundary-reducible. See Figure 4.3 for reference. If boundary is not mentioned, it is empty.*

(C1) *A vertex of degree at most 2.*

(C2) *Three 3-vertices appearing on a path of length 2.*

(C3) *The triangle $T(3,3,3)$.*

(C4) *Let u be a 3-vertex adjacent to the middle 4-vertex of the diamond $D = Dia(4-3, 4, 5^+)$. Let v denote the 5^+-vertex that is a side vertex of D. Then $D \cup \{u\}$ is reducible with boundary v.*

(C5) *$Dia(3-3, 5^+, 5^+)$ with 5^+-vertices in the boundary.*

(C6) *$Dia(3-5^+,3,5^+)$ with 5^+-vertices in the boundary.*

(C7) *The diamond $Dia(5-4,3,3)$.*

(C8) *Let $D_1 = Dia(4-4,5,3)$ and $D_2 = (5-3,4,4^+)$ be two diamonds sharing the same 5-vertex. Let v denote the 4^+-vertex that is a side vertex of D_2. Then the subgraph $D_1 \cup D_2$ is reducible with v in the boundary.*

(C9) *Let $D_1 = D_2 = Dia(3-4,4,5^+)$ be two diamonds whose middle 4-vertices are connected by an edge. Let v_1 and v_2 denote the two 5^+-vertices that are the side vertices of D_1 and D_2. Then the subgraph $D_1 \cup D_2$ is reducible with v_1 and v_2 in the boundary.*

(C10) *Let $D_1 = Dia(4-3,5^+,5)$ and $D_2 = Dia(5-3,4,4^+)$ be two diamonds sharing a middle 5-vertex. Let v_1 denote the 5^+-vertex that is a side vertex of D_1 and let v_2 denote the 4^+-vertex that is a side vertex of D_2. Then the subgraph $D_1 \cup D_2$ is reducible with v_1 and v_2 in the boundary.*

(C11) *A diamond $Dia(4-4,3,4)$ along with a 3-vertex adjacent to one of the middle 4-vertices.*

(C12) *Let $D_1 = Dia(3-4,4,5^+)$ and $D_2 = Dia(4-4,4,3)$ be two diamonds whose middle 4-vertices are connected by an edge. Let v denote the 5^+-vertex that is a side vertex of D_1. Then the subgraph $D_1 \cup D_2$ is reducible is reducible with v in the boundary.*

(C13) *Let $D_1 = Dia(3-5,5,5)$ and $D_2, D_3 = Dia(5-3,4,5+)$ be two diamonds where the two side 5-vertices of D_1 are middle vertices of D_2 and D_3. Let v_1 and v_2 denote the two side 5^+-vertices of D_2 and D_3, respectively. Then $D_1 \cup D_2 \cup D_3$ is reducible with v_1 and v_2 in the boundary.*

We wish to point out that the reducible configurations are meant to be induced subgraphs by definition, and we will use them as such in the discharging part of the proof. The only configuration, where two external edges can be identified is (C2) and it gives (C3), which we explicitly list. It can be straightforwardly checked that no identification of vertices in (C1)–(C15) is possible since otherwise, it creates a forbidden subgraph.

Proof of Lemma 4.5.1. It is straightforward to check that each configuration (C1)–(C15) satisfies the (FIX) and (FORB) conditions in Definition 4.4.1. However, checking all the cases is rather tedious. Hence we developed a simple computer program that does it, see http://lidicky.name/pub/flexibility[1]. In particular, a greedy coloring works in all cases. For an interested reader who wishes to check some cases by hand, we added list sizes to Figure 4.3. We also provide here proofs showing that (C2) and (C5) are $(\mathcal{F}, 4)$-boundary-reducible. Together, these two configurations demonstrate how to prove that the remaining configurations are reducible.

The two reducible configurations H_1 and H_2 corresponding to (C2) and (C5), respectively are depicted in Figure 4.2. The reduced parts $R_1 = H_1$ and $R_2 \subset H_2$ are provided as well. Finally, we have labeled each vertex in the figure with the value of the function $4 - \deg_{H_i} + \deg_{R_i}$ for $i \in \{1, 2\}$.

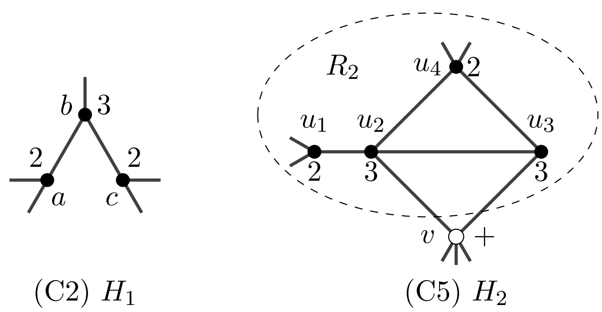


Figure 4.2 Reducible configurations (C2) and (C5). The reduced parts consist of the black vertices.

By definition, checking the (FIX) condition for any subgraph H with reducible part R is equivalent to showing that for each $v \in V(R)$, R can be properly colored after assigning each vertex a list of size $((4 - \deg_H + \deg_R) \downarrow v)$. It is clear by inspection that this is the case for $(C2) = H_1 = R_1$, and hence we only need to check the (FORB) condition for (C2). Since (FIX) is already verified, it implies (FORB) for subsets of size one in R. It remains to verify (FORB) for subsets of size two in R.

[1]This program is also available as a part of the sources in our arXiv submission.

If we apply (FORB) to a and c, then both a and b will be left with one available color in their lists. Vertex b still has three colors in it's list. Therefore, we can greedily color a, c, and b in this order to obtain a proper coloring for (C2). If we apply (FORB) to a and b, then the color for a will be fixed, and each of b and c will be left with two possible colors. Therefore, we can greedily color a, b, and c in this order to obtain a proper coloring for (C2). By symmetry, the case of applying (FORB) to b and c is also verified, implying that (C2) is reducible.

Let H_2 be a subgraph of G isomorphic to (C5). Let $R_2 \subset H_2$ denote the reducible part of (C5), i.e. the subgraph of H_2 induced by vertices $u_1, \ldots, u_4$. For each $i = 1, \ldots, 4$, we will check the (FIX) condition for u_i. Let L_i be an arbitrary list assignment where each vertex in R is assigned a list of size $((4 - \deg_H + \deg_R) \downarrow u_i)$. We will now show that R can be properly colored. In each case we list the order of vertices in greedy coloring.

- $L_1 : u_1, u_2, u_4, u_3$.
- $L_2 : u_2, u_1, u_4, u_3$.
- $L_3 : u_3, u_4, u_2, u_1$.
- $L_4 : u_4, u_3, u_2, u_1$.

Next we need to verify that H satisfies the (FORB) condition. However, only one subset of R of size two is $\mathcal{F}$-free: $\{u_1, u_2\}$. In that case R can be colored greedily in the following order u_1, u_2, u_4, u_3. Thus, (C5) is a reducible configuration. ■

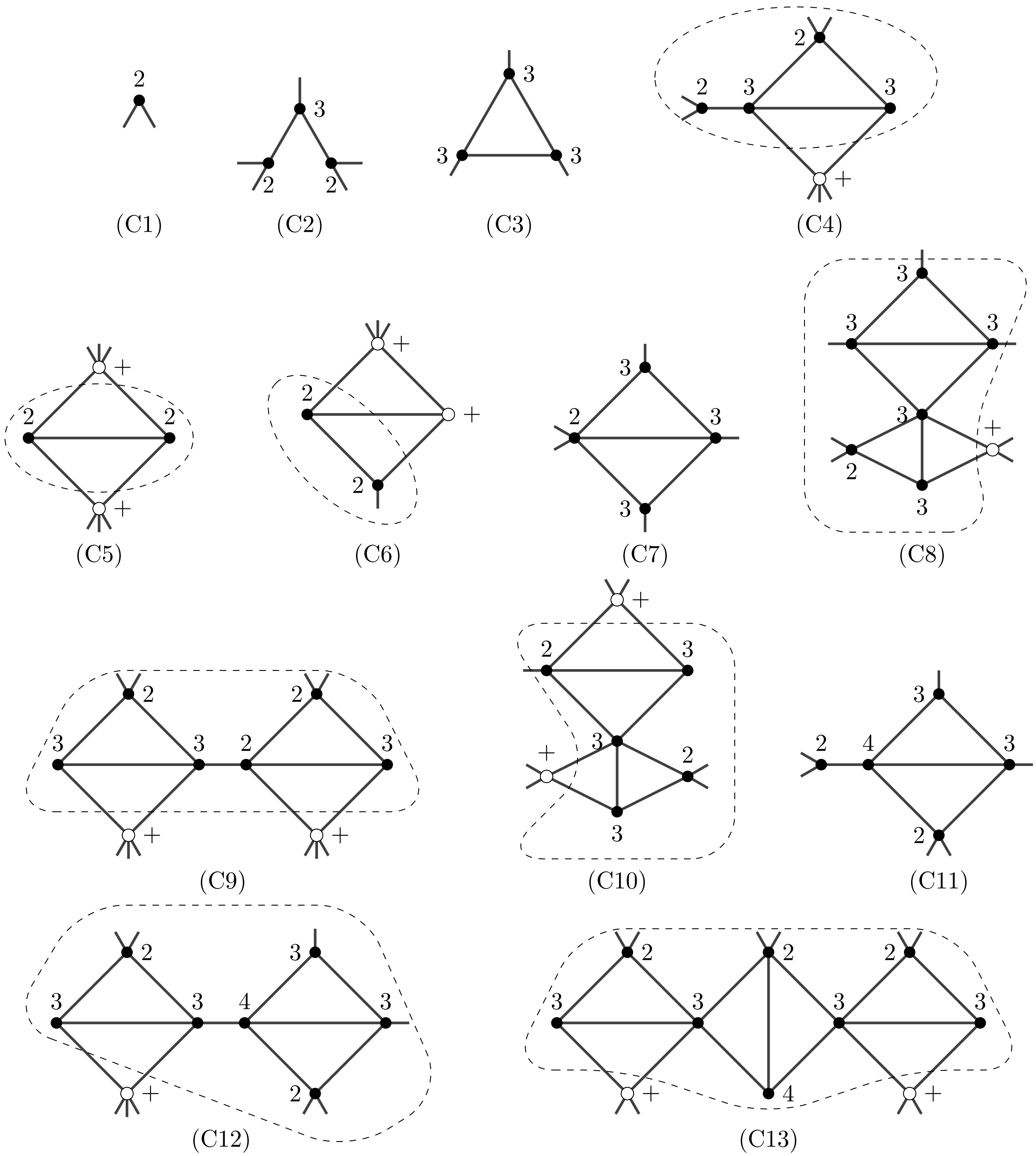


Figure 4.3 Reducible Configurations for Theorem 4.2.1. The labels give the list sizes remaining after accounting for the external neighbors and boundary vertices.

4.5.2 Discharging

In this section we prove the following lemma, which by Lemma 4.4.4 implies Theorem 4.2.1.

Lemma 4.5.2. *Let G be a connected $\{K_4, C_5, C_6, C_7, B_5\}$-free plane graph. Then G contains at least one of the reducible configurations* (C1)–(C13)*.*

Proof. Suppose for contradiction that G is a connected $\{K_4, C_5, C_6, C_7, B_5\}$-free plane graph that contains none of the configurations (C1)–(C13). We use discharging to obtain a contradiction with Euler's formula.

We denote the initial charge by ch. For every vertex v, we let $ch(v) = \deg(v) - 4$, and every face f we let $ch(f) = \ell(f) - 4$, where $\ell(F)$ is the length of the facial walk. For convenience we will also assign charge to the edges of G. The initial charge is 0 for each edge. By Euler's formula, the total sum of initial charges is -8.

We sequentially apply the following rules that move the charge around, while keeping the sum of charges unchanged. The charge at the end is called the *final charge*. The final charges will be all nonnegative, contradicting that their sum is -8.

(R1) Every 8^+-face sends charge $1/2$ to every incident 3-face and 4-face for every edge they have in common.

(R2) For every edge e that is not incident with any 3-face or 4-face the following applies. If e is a bridge, e receives charge 1 from the unique face incident with e. If e is not a bridge, e receives charge $1/2$ for each of the two faces incident to e.

(R3) For every vertex u and an incident edge $e = uz$ with charge 1:

- (R3a) If u and z are both 3-vertices, then e sends charge $1/2$ to u.
- (R3b) If u is a 3-vertex and z is 4^+-vertex, then e sends charge 1 to u.
- (R3c) If z is a 4^+-vertex, u is the middle 4-vertex of the diamond $Dia(4-3, 4, 4^+)$, and v is the 3-vertex on this diamond, then e sends 1 to v.

(R3d) If z is a 4^+-vertex, u is one of the middle 4-vertices of the diamond $Dia(4-4,3,4)$, and v is the 3-vertex on this diamond, then e sends 1/2 to v.

(R4) Every 4-face sends charge 1 to each incident 3-vertex.

(R5) Let f be a 3-face that is not part of a diamond. If exactly one vertex u of f has degree 3, then f sends 1/2 to u.

(R6) The following rules apply for a 5-vertex u. If u is a middle vertex in

(R6a) $Dia(5-3,4,4)$ or $Dia(5-3,5,4)$, then u sends 1 to the middle 3-vertex;

(R6b) $Dia(5-3,6^+,4^+)$ or $Dia(5-3,5,5)$, then u sends 1/2 to the middle 3-vertex;

(R6c) $Dia(5-5^+,3,3)$, then u sends 1/4 to each of the two side 3-vertices;

(R6d) $Dia(5-4^+,4^+,3)$, then u sends 1/2 to the side 3-vertex.

(R7) The following rules apply for every 5-vertex u and a diamond D, where v is a side vertex of D. If D is

(R7a) $Dia(4-4,5,3)$, then u sends 1/2 to the side 3-vertex;

(R7b) $Dia(4-3,5,5^+)$, then u sends 1/2 to the middle 3-vertex;

(R7c) $Dia(5-3,5,4^+)$ and u has not already sent 1 to another diamond under rule (R6a), then u sends 1/2 to the middle 3-vertex.

(R8) The following rules apply for every 6^+-vertex u and a diamond D, where u is a side vertex of D. If D is

(R8a) $Dia(5-3,6^+,4^+)$, then u sends 1/2 to the middle 3-vertex;

(R8b) $Dia(4-4,6^+,3)$, then u sends 1/2 to the side 3-vertex;

(R8c) $Dia(4-3,6^+,4^+)$, then u sends 1/2 to the middle 3-vertex.

(R9) The following rules apply for every 6^+-vertex u and a diamond D, where u is a middle vertex of D. If D is

(R9a) $Dia(6^+ - 4^+, 3, 3)$ then u sends 1/2 to each of the side vertices;

(R9b) $Dia(6^+ - 4^+, 4^+, 3)$, then u sends 1 to the side 3-vertex;

(R9c) $Dia(6^+ - 3, 4^+, 4^+)$ then u sends 1 to the other middle 3-vertex.

Claim 4.5.3. *The final charge of every face of G is nonnegative.*

Proof. Given that G does not contain any faces of length $5, 6$ or 7, we consider 8^+-faces, 4-faces, and 3-faces as three separate cases covering everything.

Suppose that f is an 8^+-face. Then the initial charge of f is equal to $\ell(f) - 4$. By (R1) and (R2), f sends at most $\frac{1}{2}$ for each of these edge that is not a bridge and charge 1 to each bridge by (R2). This means that f sends at most $\left\lceil \frac{\ell(f)}{2} \right\rceil \leq \ell(f) - 4$ total charge. Since (R1) and (R2) are the only rules requiring an 8^+-face to send out charge, every 8^+-face has nonnegative final charge.

Suppose that f is a 4-face. Then f has its initial charge 0. Since C_5, C_6, and C_7 are forbidden subgraphs, f must be incident with four 8^+-faces. By (R1), each face sharing an edge with f sends charge $\frac{1}{2}$ to f for every edge they have in common, leaving f with a total charge of 2 before applying (R2)–(R9). Given that (C2) is a reducible configuration, f cannot contain more than two 3-vertices. Thus, (R4) applies to f at most twice, which decreases the charge at f by at most 2. Since no other rules apply to 4-faces, f has nonnegative final charge.

Next suppose that f is a 3-face that is not contained in a diamond. Every face incident to f must be an 8^+-face since C_5, C_6, and C_7 are forbidden subgraphs. This means that after applying (R1), f has charge $\frac{1}{2}$. Among rules (R2)–(R9), only (R5) requires a 3-face to send out charge. If (R5) applies to f, then it only requires f to send a charge of $\frac{1}{2}$. This means that f has nonnegative final charge.

Lastly, assume that f is a 3-face contained in a diamond D. Then f shares one edge with another 3-face. Since C_5, C_6, and C_7 are forbidden subgraphs, f shares its other two edges with 8^+-faces. By (R1), f receives charge at least $\frac{1}{2}$ for each edge it shares with an 8^+-face, leaving f with charge at least 0 before applying (R2)–(R9). None of these rules, however, demand charge

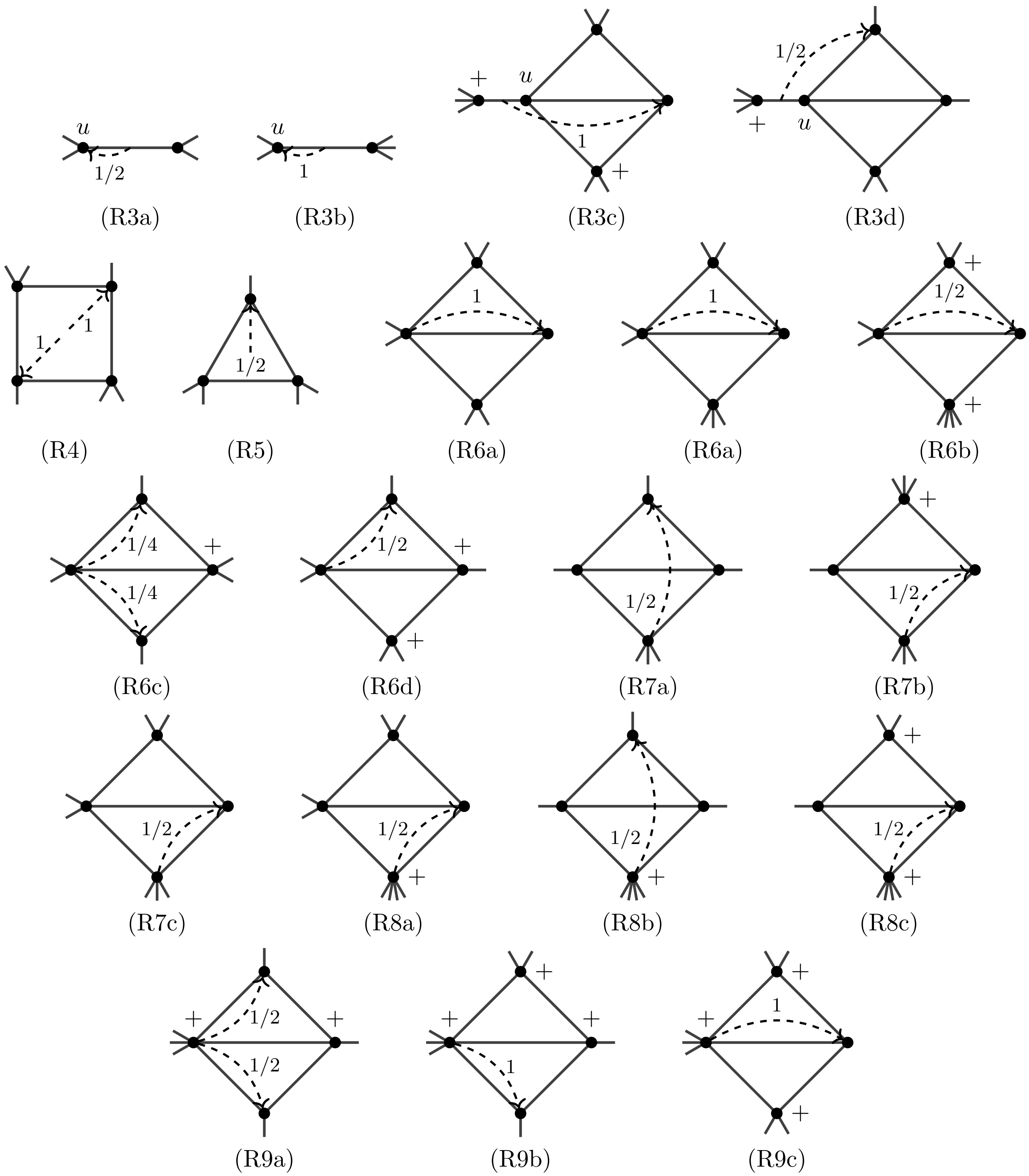


Figure 4.4 Discharging rules for Theorem 4.2.1

from a 3-face that is contained in a diamond, implying that f will end with nonnegative charge. As we have considered all possible faces in G, this completes the proof of Claim 4.5.3. ■

Claim 4.5.4. *The final charge of every edge of G is nonnegative.*

Proof. Let $e = uz$ be an edge of G. If e is incident with a 3-face or a 4-face, then none of the rules apply to e and there is nothing to prove. Otherwise, e has charge 1 after applying (R2). As (R3) is the only rule that requires any edge to send out charge, it suffices to verify that e will never be asked to give more than 1 charge under (R3).

If u and z are both 3-vertices, then only (R3a) applies to e and the edge sends exactly $\frac{1}{2}$ to each of u and z. If u is a 3-vertex and z is a 4^+-vertex, then only (R3b) applies, and e send exactly 1 to u.

If u is a 4^+-vertex and z is a 5^+-vertex, then (R3) does not apply with z and e sends charge at most 1 using either (R3c) or (R3d).

The remaining case is that both u and z are 4-vertices. The rules demand e to send charge more than 1 if by symmetry (R3c) applies with u and one of (R3c) and (R3d) applies with z. However, this would give reducible configurations (C9) and (C12), respectively. Therefore, no edge in G that begins with charge 1 will ever be asked to send out more than 1 total charge, completing the proof of Claim 4.5.4. ■

Claim 4.5.5. *The final charge of every 4^+-vertex is nonnegative.*

Proof. Suppose that v is a 4-vertex. The initial charge of v is 0, and there are no rules requiring v to send out charge, so v will end with nonnegative charge.

Next suppose that v is a 5-vertex. Then the initial charge of v is 1. Only (R6) and (R7) require a 5-vertex to distribute charge. Therefore, we may assume that v is incident with at least one diamond. Given that G does not contain any C_5, C_6, C_7, or B_5 subgraphs, v is incident with at most two diamonds.

First suppose that v is incident with exactly one diamond D. If v is a middle vertex of D then only (R6) applies to v, and if v is a side vertex of D then only (R7) applies to v. As neither of these two rules will require v to send out charge more than 1, v will end with nonnegative charge.

Next suppose that v is incident with two diamonds D_1 and D_2. Since that G does not contain any C_5, C_6, C_7, or B_5 subgraphs, D_1 and D_2 must be edge disjoint. Since $d(v) = 5$, v cannot be a middle vertex of both diamonds. If v is a side vertex of both diamonds, then only (R7) applies to v. As (R7) will not require v to send charge more than $\frac{1}{2}$ to either diamond, v will end with nonnegative charge.

Therefore, we may assume that v is a middle vertex of D_1 and a side vertex of D_2. In this case, it is possible that both (R6) and (R7) apply to v. Among the subcases of (R6), only (R6a) requires v to send out charge for more than $\frac{1}{2}$, and (R7) will never ask v send out charge more than $\frac{1}{2}$. Given that configuration (C8) is reducible, (R6a) cannot apply with (R7a). Next, given that configuration (C10) is reducible, (R6a) cannot apply with (R7b). By assumption of (R7c), (R6a) cannot apply with (R7c). Therefore v is never asked to send more than 1, implying that v will end with nonnegative charge.

Now suppose that v is a 6^+-vertex. The only rules that apply to v are (R8) and (R9). Under these rules, v sends at most 1 to all diamonds that contain v as a middle vertex, and v sends at most 1/2 to all diamonds that contain v as a side vertex. Assume that v is the middle vertex of k distinct diamonds, and incident to m other faces of size 3. By Lemma 4.4.11, the final charge of v is at least

$$d(v) - 4 - k - \frac{m}{2} = d(v) - 4 - \frac{3k + 2m}{4} - \frac{k}{4} \geq \frac{3d(v)}{4} - 4 - \frac{1}{4}\left\lfloor \frac{d(v)}{3} \right\rfloor \geq \frac{2d(v)}{3} - 4,$$

and $\frac{2d(v)}{3} - 4$ is nonnegative whenever $d(v) \geq 6$. This completes the proof of Claim 4.5.5. ■

Claim 4.5.6. *The final charge of every* 3*-vertex that is not contained in a diamond is nonnegative.*

Proof. Let v be a 3-vertex that is not contained in a diamond. Then the initial charge of v is -1. As there are no rules requiring v to send out charge, we only need to verify that v will receive charge at least 1. First suppose that v is not incident to any 3-faces or 4-faces. Then each of the three edges incident to v receive charge 1 under (R2). Next, each of these edges sends $\frac{1}{2}$ to v by (R3), leaving v with a charge of $\frac{1}{2}$.

Now suppose that v is incident to at least one 4-face f. By (R4), v receives 1 from f and we are done. Therefore, we may assume that v is not incident to any 4-face, and that v is incident to at least one 3-face T. By assumption, T is not contained in a diamond.

- Case 1: T contains another 3-vertex. In this case, v must be adjacent to a 4^+-vertex u that is not contained T, since (C2) is reducible. Then by (R3b), v receives a charge of 1 from the edge uv.

- Case 2: T contains two 4^+-vertices. Again, let u be the neighbor of v that in not contained in T. Here, v will receive at least $\frac{1}{2}$ from (R3). With that being said, T will send the remaining $\frac{1}{2}$ to v under (R5).

This completes the proof of Claim 4.5.6. ■

Claim 4.5.7. *The final charge of every* 3*-vertex that is incident to a diamond is nonnegative.*

Proof. Assume that v is a 3-vertex incident to a diamond D. Since (C2) and (C3) are reducible, there is at most one other 3-vertex incident to D. Since the initial charge of v is -1, and there is no rule requiring a 3-vertex to send charge, it suffices to show that v will always receive charge at least 1 after applying rules (R1)–(R9). We consider the following cases.

Case 1: v is the only 3-vertex incident to D and v is a side vertex of D. Given the list of forbidden subgraphs in G, the other two faces incident to v must be 8^+-faces. Hence by (R3), v receives charge at least $\frac{1}{2}$ from the only edge incident to v that is not a part of D. There are three subcases to Case 1 showing how v gets another $\frac{1}{2}$ of charge.

1. $D = Dia(4-4, 4, 3)$

   Let x and y denote the two middle vertices of D. Since (C11) is reducible, each of the neighbors of x and y that are not contained in D must be 4^+-vertices. Therefore by (R3d), v receives $\frac{1}{2}$ from the each of two edges incident to x and y that are not contained in D.

2. $D = Dia(4-4, 5^+, 3)$

Let u denote the other side vertex in D. If u is a 5-vertex, then v receives $\frac{1}{2}$ from u by (R7a). If u is a 6^+-vertex, then v receives $\frac{1}{2}$ from u by (R8b).

3. $D = Dia(5^+ - 4^+, 4^+, 3)$

   Let u denote the 5^+-vertex that is the middle vertex of D. If $d(u) = 5$, then u sends charge $\frac{1}{2}$ to v by (R6d). If $d(u) \geq 6$, then u sends charge $\frac{1}{2}$ by (R9b).

In all three cases, the final charge of v is nonnegative.

Case 2: v is the only 3-vertex incident to D and v is a middle vertex of D. There are four subcases to Case 2. In each v receives charge 1 which leads to nonnegative final charge.

1. $D = Dia(4 - 3, 4, 4^+)$

   Let u denote the middle 4-vertex of D. Since (C4) is reducible, the unique neighbor z of u not contained in D must be a 4^+-vertex. Therefore, the edge uz will send charge 1 to v by (R3c), leaving v with nonnegative charge.

2. $D = Dia(4 - 3, 5^+, 5^+)$

   Let x and y denote the two side vertices of D. If $d(x) = 5$, then x will send charge $\frac{1}{2}$ to v by (R7b). If $d(x) \geq 6$, then x will send charge $\frac{1}{2}$ to v by (R8c). As the rules apply to y identically as they do to x, it follows that v will end with nonnegative charge.

3. $D = Dia(5 - 3, 4^+, 4^+)$.

   Let u denote the middle 5-vertex of D and let x and y denote each of the side 4^+-vertices of D with $d(x) \leq d(y)$. If $d(x) = d(y) = 4$, then u sends charge 1 to v by (R6a). If $d(x) = 4$ and $d(y) = 5$, then u sends charge 1 to v by (R6a). If $d(x) = 4$ and $d(y) \geq 6$, then u sends charge $\frac{1}{2}$ to v by (R6b) and y sends $\frac{1}{2}$ to v by (R8a).

   If $d(x) = d(y) = 5$, then u sends charge $\frac{1}{2}$ to v by (R6b). Since (C13) is reducible, x and y will not both send charge 1 to a vertex of a different diamond under rule (R6a). Therefore, v receives charge $\frac{1}{2}$ from either x or y by (R7c). If $d(y) \geq 6$, y sends charge $\frac{1}{2}$ to v by (R8a) and u sends charge $\frac{1}{2}$ to v by (R6b). This leaves v with nonnegative charge.

4. $D = Dia(6^+ - 3, 4^+, 4^+)$

   Let u denote the middle vertex of D. Then u sends charge 1 to v by (R9c). This leaves v with nonnegative charge.

Case 3: There are two 3-vertices incident to D, one of which is v. Let x denote the other 3-vertex incident to D. Since (C5) and (C6) are reducible configurations, both x and v are side verties of D. Since (C7) is reducible, we may assume that if one of the middle vertices of D is a 4-vertex, then the other middle vertex is a 6^+-vertex. There are two subcases to Case 3.

1. $D = Dia(6^+ - 4, 3, 3)$.

   Let u denote the 6^+-vertex incident to D. By (R9a), u sends charge $\frac{1}{2}$ to v. Since v is a side vertex of D, and $d(v) = 3$, it follows that v is incident to exactly one edge that is not contained in D. By (R3), v will receive charge at least $\frac{1}{2}$ from this edge, leaving v with nonnegative charge. The case of x is symmetric.

2. $D = Dia(5^+ - 5^+, 3, 3)$

   Since v is a side vertex of D, and $d(v) = 3$, it follows that v is incident to exactly one edge that is not contained in D. By (R3), v will receive charge at least $\frac{1}{2}$ from this edge.

   Let a and b denote the middle vertices of D. If $d(a) = 5$, then v receives charge $\frac{1}{4}$ from a by (R6c). If $d(a) \geq 6$, then v receives charge $\frac{1}{2}$ from a by (R9a). As the rules apply to b identically as they do to a, it follows that v will end with nonnegative charge. Again, the case of x is symmetric.

Since we have covered all cases where v is contained in a diamond, this completes the proof of Claim 4.5.7. ■

Claims 4.5.3–4.5.7 show that the final charge of every vertex, face, and edge is nonnegative. Hence the sum of the charges is also nonnegative, which is a contradiction with the sum of the initial charges being -8. This finishes the proof of Lemma 4.5.2. ■

4.6 Proof of Theorem 4.2.2

4.6.1 Reducible configurations

We will use the following list of enhanced weakly reducible configurations. See Figure 4.6 for illustration of these configurations.

(D1) A vertex of degree at most 2.

(D2) $T(3,3,3)$ and $T(3,3,4)$.

(D3) Two diamonds $D_1 = Dia(6-3,4,3)$ and $D_2 = Dia(6-3,4,4)$ sharing a middle 6-vertex.

(D4) $Dia(3-3,4^+,4^+)$ where the side vertices are in the boundary.

(D5) $Dia(4-5,3,3)$.

(D6) $Dia(4-3,4,4)$.

(D7) $Dia(5-3,4,4)$ with another 3-vertex adjacent to the 5-vertex.

(D8) $Dia(5-5,3,3)$ with another 3-vertex adjacent to one of the 5-vertices.

(D9) Three 3-vertices u, v, w such that uv and vw are edges, and u and w are independent.

(D10) $Dia(5-3,4,3)$.

(D11) $T(5,3,3)$ with another 3-vertex adjacent to the 5-vertex.

(D12) Two triangles $T_1 = T_2 = T(6,3,3)$ sharing the 6-vertex.

In the next section we will prove the following theorem, showing that (D1)–(D12) are unavoidable. We remark that no identification of vertices in (D1)–(D12) is possible since otherwise, it creates a forbidden subgraph. It is possible that some external edges can be identified in (D8), (D9), and (D11). We explicitly list those cases in Figure 4.6 as (D8'), (D9'), (D11'), and (D11").

Theorem 4.6.1. *Every $\{K_4, C_5, C_6, C_7\}$-free planar graph contains one of* (D1)–(D12)*.*

Let $\mathcal{F} = \{K_4, C_5, C_6, C_7, B_\ell\}$ for any fixed ℓ. We will show that (D2), (D3), and (D5)–(D12) are enhanced weakly $(\mathcal{F}, 4)$-boundary-reducible configurations. We will also show that (D4) is only a weakly $(\mathcal{F}, 4)$-boundary-reducible configuration. However, the only neighbors of the vertices in the reducible part of (D4) are the non-adjacent vertices in the boundary, and all non-adjacent pairs that do not forbid $\mathcal{F}$ in each of (D1)–(D12) form a loose sets, implying that the condition required by the enhanced weak resolution is satisfied.

In order to use Lemma 4.4.7, we want to have an enhanced weak $(\mathcal{F}, 4, b, \beta)$-resolution for some β and b. First, we check the condition (TIGHT) in the following lemma.

Lemma 4.6.2. *Let G be an $\mathcal{F}$-free graph containing H, where H is one of (D2)–(D12). The number of H-tight vertices is at most $\beta \leq 10\ell$.*

Proof. Let H be one of (D2)–(D12) and v be an H-tight vertex adjacent to u and w in H. First, suppose that u and w are not adjacent. There are no non-edges in (D2) and (D4). The non-edge in (D9) forms a loose set. By inspection of each pair of non-adjacent verticese in (D5)–(D8) and (D10)–(D12), we observed that if v was adjacent to any of these pairs, we would obtain a C_5 or a C_6, contradicting that G is $\mathcal{F}$-free.

Second, suppose that uw is an edge. As B_ℓ is in $\mathcal{F}$, the number of H-tight vertices for uv is at most $\ell - 3$. Since H is one of (D2)–(D12), it has at most 10 edges. This bounds that the total number of H-tight vertices as $\beta \leq 10(\ell - 3) \leq 10\ell$. ■

Lemma 4.6.3. *The configurations (D2), (D3), (D5)–(D12) are enhanced weakly $(\mathcal{F}, 4)$-boundary-reducible. See Figure 4.6 for illustration.*

Proof. Given the rules for enhanced weak $(\mathcal{F}, k)$-boundary reducibility, it is straightforward to verify that configurations (D2)–(D12) are reducible. We also provide a computer program at http://lidicky.name/pub/flexibility[2] to do so. One notable difference is that the greedy algorithm is not always sufficient. We also added test for Gallai tree, which helped. Just greedy

[2]This program is also available as a part of the sources in our arXiv submission.

algorithm could end with a diamond, where middle vertices have lists L of size 3 and side vertices would have lists of size 2, which is not a Gallai tree and hence it is L-colorable, but not in a greedy way.

In order to highlight the difference between regular and weak reducibility, we will give a short proof that (D10) is weakly $(\mathcal{F}, k)$-boundary reducible, but not $(\mathcal{F}, k)$-boundary reducible. Let R be a subgraph of a graph G defined by configuration (D10). Let a, b, c and d be vertices of R. The initial list sizes of a list assignment L as defined by the function $(4 - (\deg_G + \deg_R))$ are given in Figure 4.5.

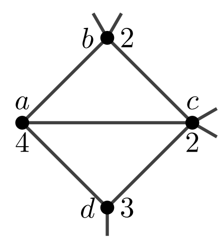


Figure 4.5 Configuration (D10)

First we will show that we cannot fix the color of a and still properly color R. Indeed, if the lists of b and c are identical and both contained the color assigned to a, there would be no proper L-coloring of R.

That being said, if we fix the color of any other vertex in R, then we will still be able to properly L-color R. Therefore, we can only apply (FIX) to a subset of the vertices of R. Given the graphs in $\mathcal{F}$, it immediately follows that the graph $H = R$ is weakly $(\mathcal{F}, k)$-boundary reducible, but as we have show, it is not $(\mathcal{F}, k)$-boundary reducible.

In the enhanced version, the main trick is that we never need to check (FORB) on two adjacent vertices. We do that by allowing (FIX) only on vertices, where their external neighbors are non-adjacent. The easiest way to do so is to use (FIX) only vertices that have at most 1 outside neighbor. In case of (D10), the only option for (FIX) is the vertex d. ■

The above works in all cases except (D1) and (D4). As the case (D1) was already discussed, we now justify the usage of the configuration (D4).

Lemma 4.6.4. *The configuration (D4) is weakly* $(\mathcal{F}, 4)$*-boundary-reducible and for all vertices* x *in its reducible part holds that* $|N(x) \cap R_j| \leq 1$ *or* $N(x) \cap R_j$ *is a loose set, where* R_j *is a reducible part of some other configuration.*

Proof. The check of the reducibility is straightforward.

The only vertices adjacent to the vertices in the reducible part are the two vertices in the boundary which are non-adjacent as $K_4 \in \mathcal{F}$. Therefore, it is sufficient to check non-adjacent non-$\mathcal{F}$-forbidding vertices in configurations (D1)–(D12). As was already discussed the only such non-edge is the non-edge in (D9) which forms a loose set. ■

4.6.2 Discharging rules

In this section, we prove the following Lemma 4.6.5, that makes Theorem 4.2.2 a corollary of Lemma 4.4.7.

Lemma 4.6.5. *Let* G *be a connected* $\{K_4, C_5, C_6, C_7\}$*-free plane graph. Then* G *contains at least one of the reducible configurations* (D1)–(D13)*.*

Proof. Assume for contradiction that G is a $\{K_4, C_5, C_6, C_7\}$-free plane graph with no (D1)–(D12). We will use discharging to arrive to a contradiction.

For every vertex v assign the initial charge $ch(v) := 2\deg(v) - 6$, and every face f assign $ch(f) := \ell(f) - 6$, where $\ell(F)$ is the length of the facial walk around f. By Euler's formula, the total initial charges of all vertices and faces is -12. We sequentially apply the following rules that transfer charge. The charge after applying all the rules is called the *final charge*. We will show that the final charge is nonnegative for every vertex and every face, which is a contradiction with the total sum of all charges being -12.

(R1) Every 8^{+}-face sends charge $\frac{1}{4}$ to every incident 3-face and 4-face for every edge they have in common.

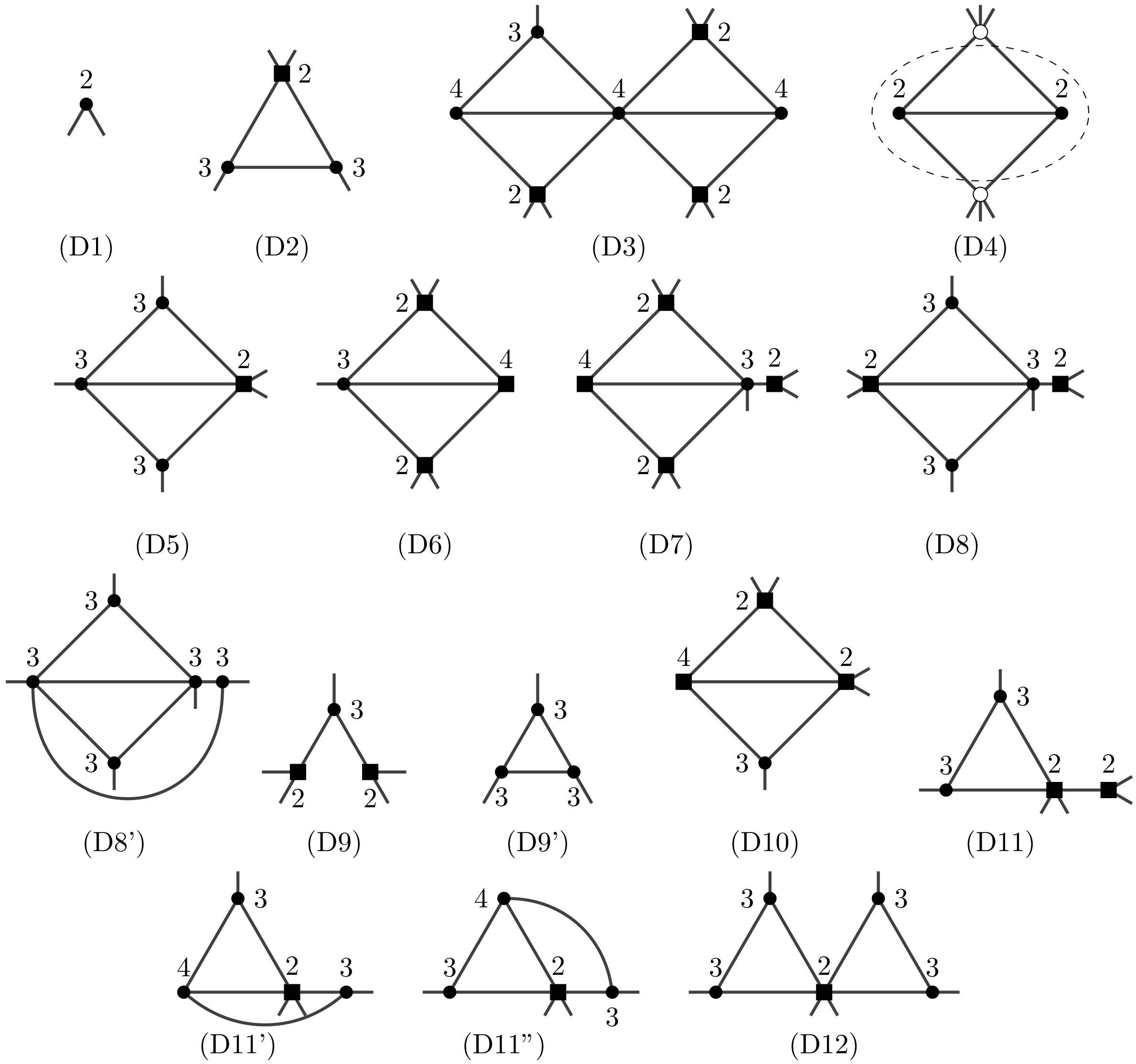


Figure 4.6 Reducible configurations for Theorem 4.2.2. The labels give the list sizes remaining after accounting for the external neighbors and boundary vertices. The vertices whose colors cannot be fixed are drawn as squares. These vertices cannot be fixed because either their coloring does not extend or they have two external neighbors, with the one exception being (D4).

(R2) For every 3-vertex v that is incident to a triangle t and an edge uv that is not part of any triangle, the following applies. The two faces[3] that are incident to uv, each send the following charge to t:

(R2a) $\frac{1}{8}$ if $\deg(u) = 3$,

(R2b) $\frac{1}{4}$ if $\deg(u) \geq 4$.

(R3) Every 4-vertex sends charge 1 to every 3-face and 4-face adjacent to it.

(R4) Every 5-vertex sends charge 1 to every 4-face adjacent to it.

(R5) Every 5-vertex that is a middle vertex in $Dia(5-3,4,4)$, $Dia(5-3,3,5^+)$, or $Dia(5-5,3,3)$ sends charge 1.5 to every 3-face of such diamond.

(R6) Every 5-vertex v, where rule (R5) does not apply, sends charge 1 to every 3-face of a diamond having v as a middle vertex.

(R7) For every 3-face $f = \{v, u, w\}$ and 5-vertex v such that f is not part of a diamond having v as a middle vertex, the following applies.

(R7a) If both $\deg(u) \geq 4$ and $\deg(w) \geq 4$, then v sends charge 1 to f.

(R7b) Otherwise v sends charge 2 to f.

(R8) Every 6^+-vertex v sends charge 1 to every 4-face adjacent to it, and 2 to every 3-face f adjacent to it, unless f is part of a diamond having v as a middle vertex.

(R9) Every 6-vertex v sends charge 1.75 to every 3-faces of every $Dia(6-3,3,4)$ that contains v.

(R10) Every 6-vertex v sends charge 1.5 to each 3-face of $Dia(6-3,4,4)$ that contains v.

(R11) Every 6-vertex v sends charge 1.25 to every 3-face of any diamond d having v as a middle vertex, where (R9) and (R10) did not apply.

[3] may be the same face twice if uv is a bridge

(R12) Every 7^+-vertex v sends charge 1.75 to every 3-face of any diamond having v as a middle vertex.

(R13) For every two 3-faces f, g that form a diamond, if g has positive charge while f has negative charge, then g gives f all its positive charge.

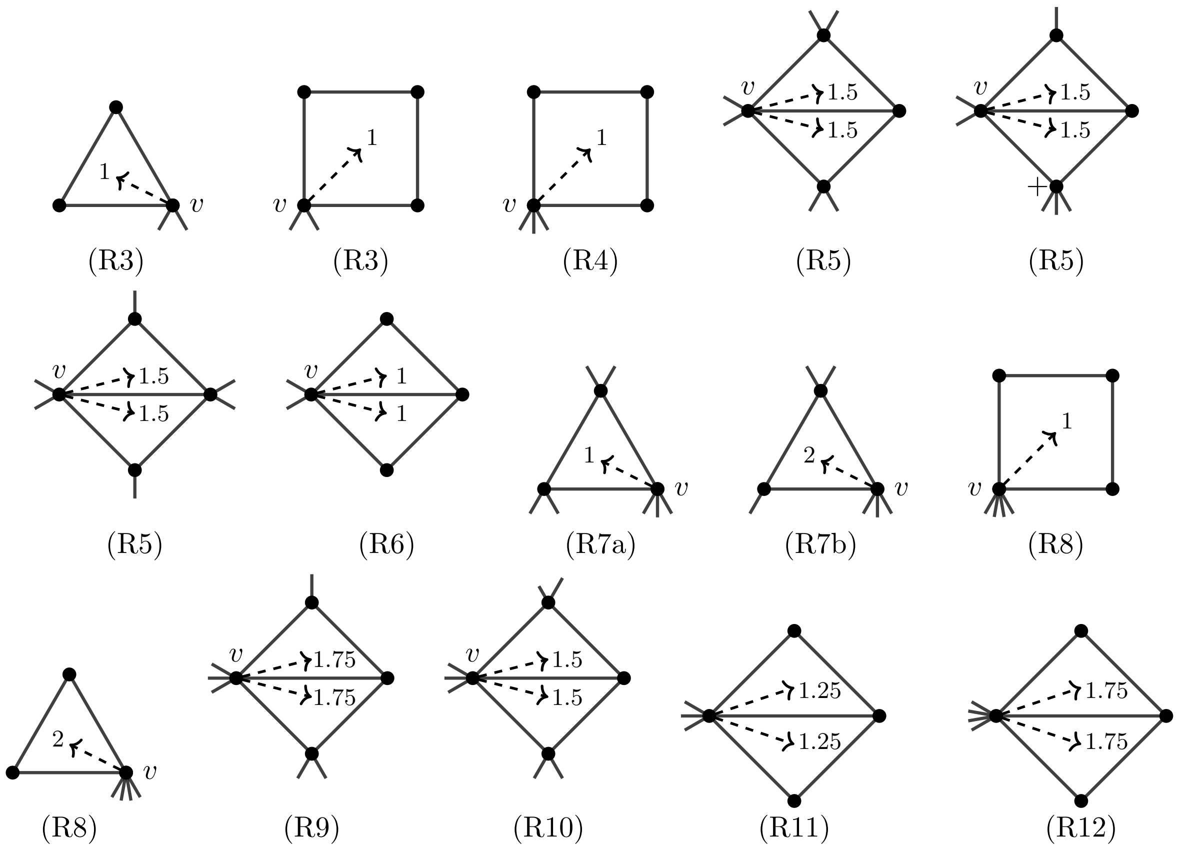


Figure 4.7 Discharging Rules for Theorem 4.2.2

Claim 4.6.6. *The final charge of every vertex is nonnegative.*

Proof. There are no vertices of degree less than 3, by (D1). The initial charge of a 3-vertex is 0, and this does not change in the discharging process. A 4-vertex v has initial charge 2. It can be adjacent to at most two 4^--faces, or otherwise a cycle C_k with $5 \leq k \leq 7$ is created. Therefore (R3) applies on v at most twice and no other rules apply. Hence v has a nonnegative final charge.

Let v be a 5-vertex that is not a middle vertex of a diamond. Note that v can be adjacent to at most two faces of size at most 4, or otherwise a cycle C_k with $5 \leq k \leq 7$ is created. Thus, the initial charge of v is 4, and (R4) and (R7) are applied together at most twice, implying that v has nonnegative final charge.

Let v be a 5-vertex that is a middle vertex of a diamond d. Then v is adjacent to at most one more face f of size at most 4, and f does not share any edge with d, or otherwise a cycle C_k with $5 \leq k \leq 7$ is created. If (R5) does not apply to v, then by (R4), (R6) and (R7), v sends 1 to each of the two 3-faces in d and at most 2 to f, leaving v with final nonnegative charge. Suppose (R5), where v sends charge 3 to the faces in d, applies to v. If v sends charge of at most 1 to f, then it has final nonnegative charge. So by (R4) and (R7a) we may assume that f is a triangle $\{v, u, w\}$ with $d(u) = 3$ (and $d(w) \leq 4$). See Figure 4.8 for an illustration. But then G contains (D7), (D11), or (D8) as d is $Dia(5-3, 4, 4)$, $Dia(5-3, 3, 5^+)$, or $Dia(5-5, 3, 3)$, respectively. Hence (R7b) does not apply to v and the final charge is nonnegative.

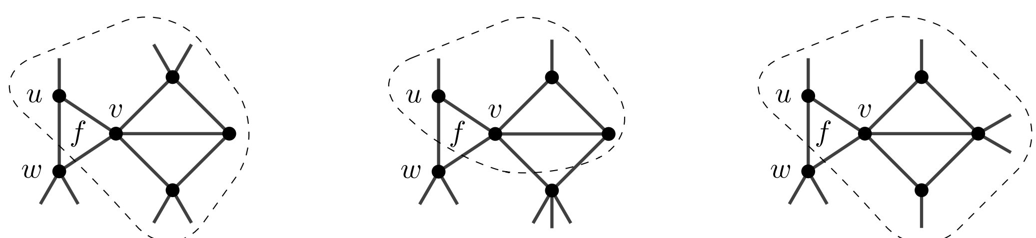


Figure 4.8 Three cases in Claim 4.6.6.

Let v be a 6-vertex that is the middle vertex of k diamonds and it is adjacent to m faces of size 3 or 4 that are not part of a diamond in which v is a middle vertex. By Lemma 4.4.11, $6 \geq 3k + 2m$. Recall that $ch(v) = 6$. Suppose $k = 2$, then $m = 0$. By by (D12) and (D3), (R9) cannot apply twice and (R9) cannot apply at the same time as (R10). Then by (R9)–(R11), the final charge of v is at least $6 - 3.5 - 2.5 = 0$ or $6 - 3 - 3 = 0$. If $k = 1$ and $m \leq 1$, then by (R8)–(R11), the final charge of v is at least $6 - 3.5 - 2 > 0$. Finally, if $k = 0$ and $m \leq 3$ then by (R8), the final charge of v is at least $6 - 3 \cdot 2 = 0$.

Let v be a 7^+-vertex that is the middle vertex of k distinct diamonds, and v is adjacent to m faces of sizes 3 and 4 that are not part of a diamond in which v is a middle vertex. Then by Lemma 4.4.11, $\deg(v) \geq 3k + 2m$. By (R12) v sends total weight of $3.5k$ to the k diamonds in which v is a middle vertex, and by (R8) it sends at most $2m$ to the other faces of size at most 4 it is adjacent to. Altogether, the final charge of v is at least

$$2\deg(v) - 6 - 3.5k - 2m = 2\deg(v) - 6 - (3k + 2m) - k/2 \geq \deg(v) - 6 - \frac{1}{2} \cdot \left\lfloor \frac{\deg(v)}{3} \right\rfloor,$$

where the last inequality follows from Lemma 4.4.11, and $\deg(v) - 6 - \frac{1}{2} \cdot \left\lfloor \frac{\deg(v)}{3} \right\rfloor \geq 0$ whenever $\deg(v) \geq 7$. ■

Claim 4.6.7. *The final charge of every face that is not contained in a diamond is nonnegative.*

Proof. By (R1) and (R2), an 8^+-face f sends out a total charge of at most $\frac{\ell(f)}{4}$. Thus the final charge of f is at least $\ell(f) - 6 - \frac{\ell(f)}{4} = \frac{3\ell(f)}{4} - 6$ which is nonnegative if $\ell(f) \geq 8$.

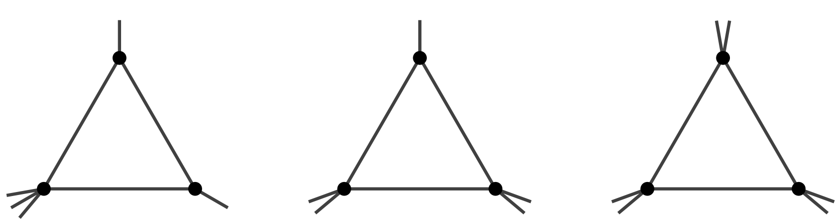
Let f be a 3-face that is not part of any diamond. Then the faces sharing an edge with f must be of size at least 8, since otherwise one of them is of size at most 4, which forces a diamond or a cycle C_i with $5 \leq i \leq 7$ together with f. Hence (R1) applies three times with f and f has charge $-3 + \frac{3}{4} = -2.25$ after (R1).

By (D1) and (D2), one of the following holds (see Figure 4.9):

(1) f is $T(3, 3, 5^+)$

(2) f is $T(3, 4^+, 4^+)$, or

(3) f if $T(4^+, 4^+, 4^+)$.



$T(3, 3, 5^+)$ $T(3, 4^+, 4^+)$ $T(4^+, 4^+, 4^+)$

Figure 4.9 Three possible triangles in Claim 4.6.7

In case (1), (R2a) applies twice giving charge $\frac{4}{8}$ to f. In addition, (R7b) or (R8) applies and the final charge of f is at least $-3 + \frac{3}{4} + \frac{1}{2} + 2 \geq 0$.

In case (2), (R2) applies once, giving charge $\frac{2}{8}$ to f. Rules (R3), (R7) and (R8) apply together twice with f, each time f receives charge at least 1, and thus the final charge of f is at least $-3 + \frac{3}{4} + \frac{1}{4} + 2 \geq 0$.

In case (3), rules (R3), (R7), and (R8) together apply three times to f and thus the final charge of f is at least $-3 + \frac{3}{4} + 3 > 0$.

If f is a 4-face, the faces sharing an edge with f must be of size at least 8, since otherwise one of them is of size at most 4, which forces a cycle C_i of size $5 \leq i \leq 7$ with f. Hence (R1) applies four times with f contributing charge $\frac{4}{4}$. By (D9) f has at least two 4^+-vertex. Thus at least one of (R3), (R4), and (R8) applies to f, giving charge 1 to f. Hence the final charge of f is at least $-2 + 1 + 1 \geq 0$. ■

Claim 4.6.8. *The final charge of every 3-face that is contained in a diamond is nonnegative.*

Proof. In the light of (R13), we will consider the faces that form a diamond together in pairs and show that as a pair, they receive sufficient charge. Let f and g be 3-faces sharing an edge, i.e. they form a diamond. Observe that in this case the other faces sharing edges with f and g must be of size at least 8, for otherwise one of them is of size at most 4, which forces a cycle C_i of size $5 \leq i \leq 7$ with f and g. Therefore, (R1) applies twice to each f and g and $ch_1(f) = ch_1(g) = -3 + \frac{2}{4} = -2.5$. Hence we aim to show that f and g together receive at least 5 more charge. We denote the the vertices of f by u, v, x where u, v are shared with g, and by y the third vertex of g.

By symmetry, we assume that $\deg(u) \geq \deg(v)$ and $\deg(y) \geq \deg(x)$. Note that by (D4), $\deg(u) \geq 4$ and by (D1) the degree of each of the other vertex is at least 3.

We split into cases based on the type of diamond f and g form.

- $Dia(4-3, \star, \star)$

  By (D2) and (D6), $\deg(x) \geq 4$ and $\deg(y) \geq 5$. Hence we are in case $Dia(4-3, 4^+, 5^+)$. For u, (R3) applies twice, for x one of (R3), (R7b), or (R8) applies, and for y one of (R7b), or

(R8) applies. Thus the charge f and g receive using these rules is at least $3 \cdot 1 + 2 = 5$. Hence the final charges of f and g are nonnegative.

- $Dia(5-3,3,3)$, $Dia(5-3,3,4)$

  Reducible by (D9) and (D10).

- $Dia(5-3,3,5^+)$

  In this case, (R5) applies to u and (R7b) or (R8) applies to y. This gives charge $2 \cdot 1.5 + 2 = 5$. Hence the final charges of f and g are nonnegative.

- $Dia(5-3,4,4)$

  In this case, (R5) applies to u. In addition (R3) applies to both x and y. This gives charge $2 \cdot 1.5 + 1 + 1 = 5$. Hence the final charges of f and g are nonnegative.

- $Dia(5-3,4^+,5^+)$

  In this case, (R6) applies to u. In addition (R3), (R7b), or (R8) applies to x and (R7b), or (R8) to y. This gives charge at least $2 \cdot 1 + 1 + 2 = 5$. Hence the final charges of f and g are nonnegative.

- $Dia(6-3,3,3)$

  Reducible by (D9).

- $Dia(6-3,3,4)$

  By (R9), u contributes charge 3.5 and by (R4), y contributes charge 1. Let z be a neighbor of x that is not u or v. By (D9), $\deg(z) \geq 4$ Hence the application of (R2b) around x contributes charge 1/2. This gives total charge $3.5 + 1 + 0.5 = 5$. Hence the final charges of f and g are nonnegative.

- $Dia(6-3,3,5^+)$

  By (R11), u contributes charge 2.5 and by (R7b) or (R8), y contributes charge 2. Let z be a neighbor of x that is not u or v. By (D9), $\deg(z) \geq 4$ Hence the application of (R2) around

x contributes charge 1/2. This gives total charge $2.5 + 2 + 0.5 = 5$. Hence the final charges of f and g are nonnegative.

- $Dia(6 - 3, 4, 4)$

  By (R10), u contributes charge 3 and by (R3), x and y each contribute charge 1. This gives total charge $3 + 1 + 1 = 5$. Hence the final charges of f and g are nonnegative.

- $Dia(6 - 3, 4^+, 5^+)$

  By (R11), u contributes charge 2.5, by (R3), (R7b) or (R8), x contributes charge at least 1, and by (R7b) or (R8), y contributes charge 2. This gives total charge at least $2.5 + 1 + 2 = 5$. Hence the final charges of f and g are nonnegative.

- $Dia(7^+ - 3, 3, 3)$

  Reducible by (D9).

- $Dia(7^+ - 3, 3, 4^+)$

  By (R12), u contributes charge 3.5, by (R3), (R7b) or (R8), y contributes charge at least 1. Let z be the neighbor of x that is not u or v. By (D9), $\deg(z) \geq 4$ Hence the application of (R2b) around x contributes charge 1/2. This gives total charge at least $3.5 + 1 + 0.5 = 5$. Hence the final charges of f and g are nonnegative.

- $Dia(7^+ - 3, 4^+, 4^+)$

  By (R12), u contributes charge 3.5, by (R3), (R7b) or (R8), x and y each contribute charge at least 1. This gives total charge at least $3.5 + 1 + 1 = 5.5$. Hence the final charges of f and g are nonnegative.

- $Dia(4 - 4, \star, \star)$ and $Dia(5 - 4, \star, \star)$

  By (D5), $\deg(y) \geq 4$. By (R3) or (R6), u and v together contribute charge 4, by (R3), y each contributes charge 1. This gives total charge at least $4 + 1 = 5$. Hence the final charges of f and g are nonnegative.

- $Dia(6^+ - 4^+, \star, \star)$

By (R3), (R6), (R11), and (R12), u and v together contribute charge at least $1.25 + 1.25 + 1 + 1$. If (R3), (R7), or (R8) applies to y, then the total charge is at least 5.5. Hence we can assume the case $Dia(6^+ - 4^+, 3, 3)$. Then (R2a) or (R2b) applies at each x and y and the total contribution is at least 0.5. This gives total charge at least $4.5 + 0.5 = 5$. Hence the final charges of f and g are nonnegative.

- $Dia(5 - 5, 3, 3)$

  By (R5), u and v together contribute charge at least $4 \times 1.5 = 6$. Hence the final charges of f and g are nonnegative.

- $Dia(5 - 5, 3^+, 4^+)$

  By (R6), u and v together contribute charge at least $4 \times 1 = 4$. By (R3), (R7a), or (R8), y contributes charge at least 1. This gives total charge at least $4 + 1 = 5$. Hence the final charges of f and g are nonnegative.

This concludes the proof of Claim 4.6.8. ■

Since all final charges are nonnegative, this concludes the proof of Lemma 4.6.5. ■

Acknowledgements. The work was initiated at Dagstuhl Seminar on "Graph Colouring: from Structure to Algorithms"(19271). Part of the work was carried out when T. Masařík was visiting Iowa State University. He thanks Steve Butler for kind hospitality. We would like to thank Ilkyoo Choi for fruitful discussions at the beginning of this project.

4.7 References

[1] Peter Bradshaw, Tomáš Masařík, and Ladislav Stacho. Flexible list colorings in graphs with special degeneracy conditions. In *31st International Symposium on Algorithms and Computation (ISAAC 2020)*, volume 181 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 24:1–24:14, 2020. doi:10.4230/LIPIcs.ISAAC.2020.24.

[2] Ilkyoo Choi, Felix Christian Clemen, Michael Ferrara, Paul Horn, Fuhong Ma, and Tomáš Masařík. Flexibility of planar graphs – sharpening the tools to get lists of size four, 2020. arXiv:2004.10917.

[3] Daniel W. Cranston and Douglas B. West. An introduction to the discharging method via graph coloring. *Discrete Mathematics*, 340(4):766–793, 2017. doi:10.1016/j.disc.2016.11.022.

[4] Zdeněk Dvořák, Tomáš Masařík, Jan Musílek, and Ondřej Pangrác. Flexibility of triangle-free planar graphs, February 2019. arXiv:1902.02971.

[5] Zdeněk Dvořák, Tomáš Masařík, Jan Musílek, and Ondřej Pangrác. Flexibility of planar graphs of girth at least six. *Journal of Graph Theory*, 95(3):457–466, November 2020. doi:10.1002/jgt.22567.

[6] Zdeněk Dvořák, Sergey Norin, and Luke Postle. List coloring with requests. *Journal of Graph Theory*, 92(3):191–206, January 2019. doi:10.1002/jgt.22447.

[7] Gašper Fijavž, Martin Juvan, Bojan Mohar, and Riste Škrekovski. Planar graphs without cycles of specific lengths. *European Journal of Combinatorics*, 23(4):377–388, May 2002. doi:10.1006/eujc.2002.0570.

[8] Peter Che Bor Lam, Baogang Xu, and Jiazhuang Liu. The 4-choosability of plane graphs without 4-cycles. *Journal of Combinatorial Theory, Series B*, 76(1):117–126, May 1999. doi:10.1006/jctb.1998.1893.

[9] Tomáš Masařík. Flexibility of planar graphs without 4-cycles. *Acta Mathematica Universitatis Comenianae*, 88(3):935–940, August 2019. URL: http://www.iam.fmph.uniba.sk/amuc/ojs/index.php/amuc/article/view/1182.

[10] Weifan Wang and Ko-Wei Lih. Choosability and edge choosability of planar graphs without five cycles. *Applied Mathematics Letters*, 15(5):561–565, July 2002. doi:10.1016/s0893-9659(02)80007-6.

[11] Donglei Yang and Fan Yang. Flexibility of planar graphs without C_4 and C_5, 2020. arXiv:2006.05243.

CHAPTER 5. GENERAL CONCLUSIONS

Perhaps the most interesting question related to Generalized Turán Problems is whether or not $\text{ex}(n, H, K_{r+1})$ is always uniquely achieved by the Turán graph for large enough n and r. The results in this dissertation provide some evidence in favor of this being true. The methods used in Chapters 2 and 3 could reasonably be used to test graphs on fewer than seven vertices. However, to find a more definitive answer to this question, one would need to apply different methods.

Similarly, while the discharging method used in Chapter 4 has been used successfully to discover many results, it is likely that more powerful tools will need to be developed in order to make significant progress in graph flexibility. It would be fascinating to either find a planar graph which is not ε-flexible with lists of size 5, or to expand the tools to show that all planar graphs are ε-flexible with lists of size 5. Additionally, it would be fascinating to see the problem studied on different types of graphs.