Maximizing five-cycles in K_r-free graphs

Bernard Lidický* Kyle Murphy†

July 9, 2020

Abstract

The Erdős Pentagon problem 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', Norin, and Razborov. Recently, Palmer suggested the 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 begin the extremal graph for large enough n.

1 Introduction

All graphs in this paper are simple. Let G, H, and F be graphs. We define $\nu(H, G)$ as the number of 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 $\mathrm{ex}(n, H, F)$ denote the maximum value of $\nu(H, G)$ among all F-free graphs G on n vertices. The function $\mathrm{ex}(n, H, F)$ is well-studied when H is an edge. As such, it is convention when $H = K_2$ to let $\mathrm{ex}(n, F)$ denote $\mathrm{ex}(n, K_2, F)$. The value of $\mathrm{ex}(n, F)$ for any graph F is called the *Turán number* of F.

One of the first results in extremal graph theory was Mantel's Theorem [26] which states that for all $n \geq 3$, $\mathrm{ex}(n, K_3) \leq \lfloor \frac{n^2}{4} \rfloor$. When $k \geq 4$, the value of $\mathrm{ex}(n, K_r)$ was determined by Turán.

Theorem 1.1 (Turán's Theorem [33]) *For all $k \geq 4$, and all n,*

$$ex(n, K_{k+1}) = \sum_{S \in \binom{k}{2}} \prod_{i \in S} \left\lfloor \frac{n_i + r_i}{k} \right\rfloor = \frac{k-1}{k}\binom{n}{2} + o(n^2),$$

*Department of Mathematics, Iowa State University. Ames, IA, USA. E-mail: `lidicky@iastate.edu`.
Supported in part by NSF grant DMS-1855653.
†Department of Mathematics, Iowa State University. Ames, IA, USA. E-mail: `kylem2@iastate.edu`

where each $r_i \in \{0,1\}$ *so that* $\sum_{i=1}^{k} \left\lfloor \frac{n_i+r_i}{k} \right\rfloor = n$*. Moreover, the Turán graph* $T_k(n)$*, which is the complete* $(k-1)$*-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 [12] 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 $\chi(F) \geq 3$,

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

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 $t < k$, Zykov [34] showed that the Turán graph $T_{k-1}(n)$ is also the unique graph with the maximum number of K_t subgraphs among all K_k-free graphs. The following is a corollary of this result.

Corollary 1.2 (Zykov [34]) *Let* k *and* t *be integers. Then*

$$ex(n, K_t, K_k) = \sum_{0 \leq i_1 \leq \cdots \leq i_t \leq k-2} \prod_{r=1}^{t} \left\lfloor \frac{n+i_r}{k-1} \right\rfloor.$$

In [22], Győri, Pach, and Simonovits studied a handful of cases where $F = K_r$.

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:

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

Gishboliner and Shapira [19] determined the order of magnitude of $\mathrm{ex}(n, C_k, C_\ell)$ for $\ell \geq 1$ and $k \geq 3$, as well as the asymptotic value of $\mathrm{ex}(C_k, C_4)$. 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

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

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, 14]. In [28], 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 [16], Gerbner studied the function $\mathrm{ex}(n, H, F)$ when H and F each have at most 4 vertices. Recently, the authors of [24] 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 $\mathrm{ex}(n, C_5, K_3)$ was known as the Erdős Pentagon problem. In a sense, a graph with $\mathrm{ex}(n, C_5, K_3)$ 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" [13].

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 (max cut).

In 1989, Győri [11] 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', Norin, and Razborov [23] showed that a triangle-free graph on n vertices contains at most $\left(\frac{n}{5}\right)^5 + o(1)$ 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 [23] 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 [25] 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 [27], who showed that the Möbius ladder on 8 vertices contains the same number of five cycles as the balanced C_5 blow-up.

Palmer [29] suggested a generalization to the Erdős Pentagon Problem: 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

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

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

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

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

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

(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 1.3(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 polynomials.

In the next section, we will give a brief overview of the flag algebra method. Section 2 contains the proof of Theorem 1.3(i). Then we prove a stability lemma in Section 3, and use it to prove Theorem 1.3(ii) in Section 4. We will end with some concluding remarks and conjectures concerning the general behavior of the function $\mathrm{ex}(n, H, F)$.

1.1 The Flag Algebra Method

Introduced by Razborov [31], 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, 15, 20, 30], 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 [31]. Another overview of flag algebras can be found in [32].

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 orders 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 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 $\mathrm{im}(\theta)$. A σ-flag is a pair (F, θ) where F is a graph and θ is an injective function of $[k]$ in to $V(F)$ that defines a graph isomorphism of $\mathrm{im}(\theta)$ 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 $|V(F_1)| + |V(F_2)| \leq n$. Let X_1 and X_2 be two disjoint sets of sizes $|V(F_1)| - |\sigma|$ and $|V(F_2)| - |\sigma|$ 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:

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

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

$$F_1 \cdot F_2 = \sum_{F \in \mathcal{F}^\sigma_\ell} P(F_1; F_2; F)F. \tag{2}$$

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

$$F - \sum_{F' \in \mathcal{F}^\sigma_\ell} P(F, F')F' \tag{3}$$

from $\mathbb{R}\mathcal{F}^\sigma$. Note that (3) 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 (3). 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).

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 [31]. By construction, the functions in $\mathrm{Hom}^+(\mathcal{A}, \mathbb{R})$ correspond to limits of induced subgraph densities in some convergent sequence. In any fixed graph G, we can express $d(C_5, G)$ as the sum of induced densities in the following way:

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

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 $\phi \in \text{Hom}^+(\mathcal{A}, \mathbb{R})$,

$$\lim_{n \to \infty} d(C_5, G_n) = \sum_{F_i \in \mathcal{F}_5} c_{F_i}^{C_5} \phi(F_i). \tag{4}$$

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 (4) 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 $\text{Hom}^+(\mathcal{A}^\sigma, \mathbb{R})$ to $\text{Hom}^+(\mathcal{A}, \mathbb{R})$ 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, $[\![F]\!]_\sigma = q_\theta(F) d(F, G)$, where $q_\theta(F)$ is equal to the probability that a randomly chosen function from $[k]$ to $V(F)$ induces a σ-flag in F. It can be shown that for any $a \in \mathcal{A}^\sigma$ and any $\phi \in \text{Hom}^+(\mathcal{A}, \mathbb{R})$,

$$\phi\left([\![a \cdot a]\!]_\sigma\right) \geq 0. \tag{5}$$

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

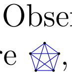
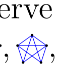
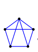
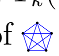
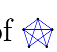
If a flag algebra calculation has a constant number of terms, then it can be interpreted as a calculation in a graph of order n with an error term $O(n^{-1})$ coming from (1).

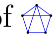
2 Proof of Theorem 1.3(i)

In this section we will prove Theorem 1.3(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 1.3(i). The Turán graph $T_k(n)$ is K_{k+1}-free and

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

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

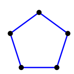
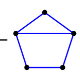
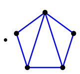
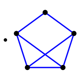
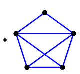
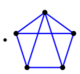
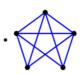
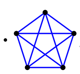
6 distinct C_5 subgraphs. Finally, there are $3\binom{k}{3}\binom{n/k}{2}^2\frac{n}{k}$ copies of  in $T_k(n)$, with every such graph containing 4 distinct C_5 subgraphs. This implies that for all n,

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

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 1 in the Appendix. Observe that for any convergent K_{k+1}-free sequence $(G_n)_{n\geq 1}$,

$$\lim_{n\to\infty} d(C_5, G_n) = \sum_{F_i \in \mathcal{F}_5} c_{F_i}^{C_5} F_i, \tag{6}$$

where $c_{F_i}^{C_5} = \nu(C_5, F_i)$. After removing each $c_{F_i}^{C_5}$ for which $c_{F_i}^{C_5} = 0$ we observe that

$\lim_{n\to\infty} d(C_5, G_n) =$  $+$  $+2\cdot$  $+2\cdot$  $+4\cdot$  $+4\cdot$  $+6\cdot$  $+12\cdot$ . (7)

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

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

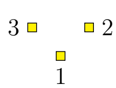
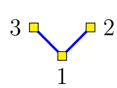
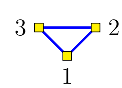
Therefore,

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

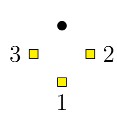
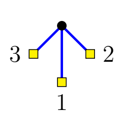
Given this fact, our goal is to find appropriate constants c_{F_i} so that

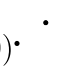
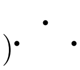
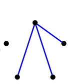
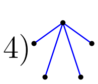
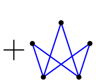
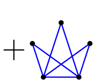
$$\lim_{n\to\infty} d(C_5, G_n) \leq \sum_{F_i \in \mathcal{F}_5} c_{F_i} F_i,$$

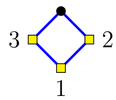
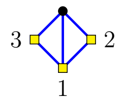
and $\max\{c_{F_i} : F_i \in \mathcal{F}_5\}$ is as small as possible. To do so, we can take advantage of properties that we know must be true of all $\phi \in \mathrm{Hom}^+(\mathcal{A}, \mathbb{R})$ which correspond to K_{k+1}-free convergent sequences of graphs. Additionally, using labeled flags,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

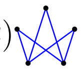
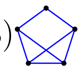
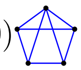
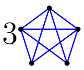
$\sigma_1 =$  $\qquad \sigma_2 =$  $\qquad \sigma_3 =$ 

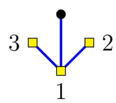
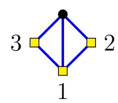
so that $\mathcal{F}_4^{\sigma_1}$, $\mathcal{F}_4^{\sigma_2}$, and $\mathcal{F}_4^{\sigma_3}$ denote three sets of labeled flags on four vertices. By (5), the following expressions are nonnegative for all $k \geq 3$.

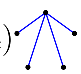
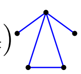
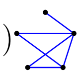
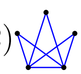
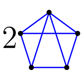
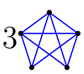
1. $P_1(k) = \left[\!\left[\left((k-1)\right.\right.\right.$  $-$  $\left.\left.\left.\right)^2\right]\!\right]_{\sigma_1} =$

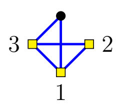
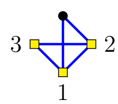
$(10k^2 - 20k + 10)\cdot$  $+(k^2 - 2k + 1)\cdot$  $+(-k+1)\cdot$  $+(-4k+4)\cdot$  $+$  $+$ 

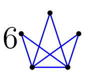
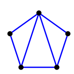
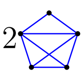
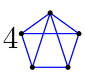
2. $P_2(k) = \left[\!\!\left[\left((k-2)\right.\right.\right.$  $-$  $\left.\left.\left.\right)^2 \right]\!\!\right]_{\sigma_2} =$

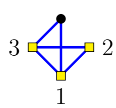
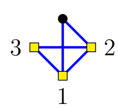
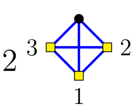
$(3k^2 - 12k + 12)$  $+ (k^2 - 6k + 8)$  $+ (-4k + 10)$  $+ 3$ 

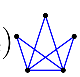
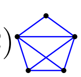
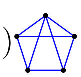
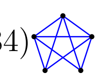
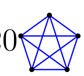
3. $P_3(k) = \left[\!\!\left[\left((k-2)\right.\right.\right.$  $-$  $\left.\left.\left.\right)^2 \right]\!\!\right]_{\sigma_2} =$

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

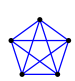
4. $P_4(k) = \left[\!\!\left[\left(\right.\right.\right.$  $-$  $\left.\left.\left.\right)^2 \right]\!\!\right]_{\sigma_3} =$

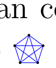
6  $-$  $+ 2$  $- 4$ 

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

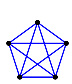
$(6k^2 - 36k + 54)$  $+ (2k^2 - 20k + 42)$  $+ (4k^2 - 24k + 36)$  $+ (-24k + 84)$  $+$ 120 

Additionally, we can apply Theorem 1.2, which states that for any K_{k+1}-free convergent sequence of graphs,

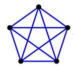
 $$\leq \frac{k^4 - 10k^3 + 35k^2 - 50k + 24}{k^4}.$$

At this point we will split the proof in to 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 $k \geq 4$.

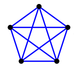
Case 1: Suppose that $k \geq 4$. Since $\sum_{i=0}^{33} F_i = 1$,

 $$\leq \sum_{i=0}^{33} F_i \left(\frac{k^4 - 10k^3 + 35k^2 - 50k + 24}{k^4} \right). \tag{8}$$

After rearranging the terms from (8) we obtain the following constraint on the elements of $\mathcal{F}_5$:

$0 \leq \sum_{i=0}^{32} F_i \left(\frac{k^4 - 10k^3 + 35k^2 - 50k + 24}{k^4} \right) +$  $(-10k^3 + 35k^2 - 50k + 24). \qquad (9)$

In particular observe that equation (9) is non-negative for all $k \geq 4$. Let

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

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

$$p_1(k) = \frac{3(k^5-8k^4+22k^3-24k^2+8k)}{5k^3-35k^2+75k-48} \qquad p_2(k) = \frac{10k^5-60k^4+109k^3-76k^2+18k}{5k^3-35k^2+75k-48}$$
$$p_3(k) = \frac{5k^5-28k^4+45k^3-28k^2+6k}{5k^3-35k^2+75k-48} \qquad p_4(k) = \frac{5k^7-30k^6+53k^5-52k^4+94k^3-96k^2+24k}{4(5k^3-35k^2+75k-48)}$$
$$p_5(k) = \frac{15k^5-60k^4+78k^3-40k^2+8k}{4(5k^3-35k^2+75k-48)} \qquad z(k) = \frac{120(5k^3-20k^2+30k-16)}{5k^3-35k^2+75k-48}$$

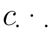
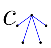
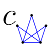
Thus, for any convergent sequence $(G_n)_{n\geq 1}$ of K_{k+1}-free graphs with $k \geq 4$,

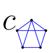
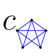
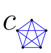
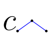
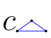
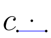
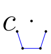
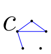
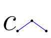
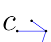
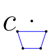
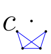
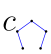
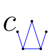
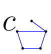
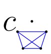
$$\lim_{n\to\infty} d(C_5, G_n) \leq \sum_{i=0}^{33} c_{F_i}^{C_5} F_i + z(k)Z(k) + \sum_{j=1}^{5} p_j(k)P_j(k) = \sum_{i=0}^{33} c_{F_i} F_i, \qquad (10)$$

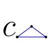
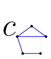
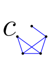
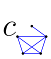
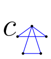
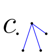
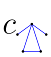
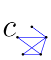
where c_{F_i} is the coefficient of F_i after all expansions. Then

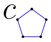
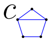
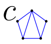
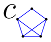
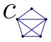
$$\mathrm{OPT}_k(C_5) \leq \max\{c_{F_i} : F_i \in \mathcal{F}_5\}.$$

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

- $C_1 = c$ $= c$ $= c$ $= c$

  $= c$ $= c$ $= c$ $= \frac{60k^7-720k^6+3600k^5-9876k^4+16320k^3-16440k^2+9360k-2304}{5k^3-35k^2+75k-48}$.
- $C_2 = c$ $= \frac{33k^7-450k^6+2547k^5-7824k^4+14214k^3-15360k^2+9144k-2304}{5k^3-35k^2+75k-48}$.
- $C_3 = c$ $= c$ $= c$ $= c$ $= c$ $= c$ $= c$ $= c$ $= c$ $= c$ $= c$ $= c$ $=$ c

  $= c$ $= c$ $= c$ $= c$ $= c$ $= \frac{30k^7-420k^6+2430k^5-7596k^4+13980k^3-15240k^2+9120k-2304}{5k^3-35k^2+75k-48}$.
- $C_4 = c$ $= \frac{30k^7-423k^6+2457k^5-7686k^4+14118k^3-15336k^2+9144k-2304}{5k^3-35k^2+75k-48}$.
- $C_5 = c$ $= \frac{35k^7-468k^6+2607k^5-7916k^4+14278k^3-15367k^2+9144k-2304}{5k^3-35k^2+75k-48}$.
- $C_6 = c$ $= \frac{30k^7-425k^6+2468k^5-7697k^4+14098k^3-15302k^2+9132k-2304}{5k^3-35k^2+75k-48}$.

- $C_7 = c$ $=c$ $= \frac{35k^7-455k^6+2505k^5-7644k^4+13980k^3-15240k^2+9120k-2304}{5k^3-35k^2+75k-48}$.
- $C_8 = c$ $= \frac{(135/4)k^7-(895/2)k^6+(9967/4)k^5-7631k^4+(27913/2)k^3-15216k^2+9114k-2304}{5k^3-35k^2+75k-48}$.
- $C_9 = c$ $= \frac{50k^7-610k^6+3129k^5-8902k^4+15326k^3-15956k^2+9264k-2304}{5k^3-35k^2+75k-48}$.
- $C_{10} = c$ $= \frac{50k^7-610k^6+3103k^5-8758k^4+15050k^3-15748k^2+9216k-2304}{5k^3-35k^2+75k-48}$.

Claim 2.1 *For all $i = 1, \dots, 10$ and $k \geq 4$, $C_1(k) \geq C_i(k)$.*

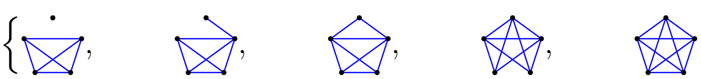
Proof. Observe that for all $i = 1, \dots, 10$, each polynomial $C_i(k)$ has the same denominator of $5k^3 - 35k^2 + 75k - 48$. It is straightforward to verify that $5k^3 - 35k^2 + 75k - 48$ is positive for all $k \geq 4$. By examining the leading coefficients in the numerator of each polynomial, it is straightforward to check that C_1 is the largest for $k > 1000$. For $4 \leq k \leq 1000$, we have provided Sage code used to verify the claim in Appendix 6.1. ∎

By factoring C_1 it follows that

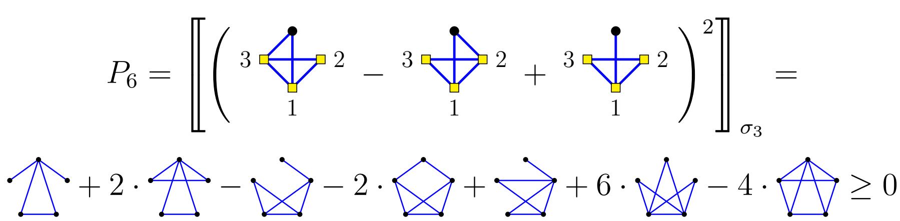
$$\mathrm{OPT}_k(C_5) \leq C_1(k) = \frac{1}{k^4}(12k^4 - 60k^3 + 120k^2 - 120k + 48),$$

completing the proof of Theorem 1.3(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 of zero, and therefore can be removed from our calculations.

$\mathcal{H} = \{$  $\}$.

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 (5).



Now suppose that

$p_1 = 1/27$ $\quad p_2 = 13/27$ $\quad p_3 = 8/27$

$p_4 = 2/9$ $\quad p_6 = 17/54.$

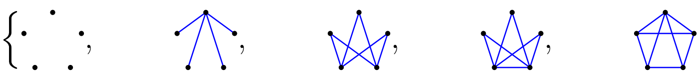
Then

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

Let c_{F_i} denote the coefficient of each graph F_i after combining each of the two sums. It is straightforward to verify that

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

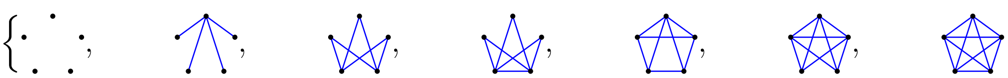
Furthermore, the set T_3 of graphs for which $c_{F_i} = \frac{40}{27}$ is given below.

$T_3 = \{$  $\}$

This completes the proof of Theorem 1.3 (i). ■

2.1 Finding the Optimal Bound

We will now give a short description on how we found the polynomials $z(k)Z(k)$ and $p_i(k)P_i(k)$ that were used in the proof of Theorem 1.3(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 1.3(i), the set T given below contains the tight subgraphs.

$T = \{$  $\}.$

The set T_3 contains the tight subgraphs when $k = 3$. 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.

Lemma 2.2 ([3]) *Given $(G_n)_{n\geq 1}$ a convergent sequence of K_{k+1}-free graphs of increasing order, such that $d(C_5, G_n) \to OPT_k(C_5)$. 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 2.2 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.

3 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 3.1 *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:

$$(x_i - 1)(x_j + 1) = x_i x_j + (x_i - x_j) - 1 \geq x_i x_j + 1.$$

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

The next Proposition follows immediately from Lemma 3.3, which we will prove next.

Proposition 3.2 *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 $\nu(v, C_5)$ denote the number of five-cycles in G that contain v. As the graph G containing v will be clear from context, we do not need to specify G in this parameter.

Lemma 3.3 *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 deleting a vertex in X_i and replacing it with a copy of a vertex in X_j. Then

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

Proof. Let H be a complete k-partite graph on n vertices. Let $X_1, X_2, \ldots, X_k$ denote the partite sets of H. Suppose that there exist $i, j \in \{1, 2, \ldots, k\}$ for which $x_i \geq x_j + 2$. By symmetry we may assume that $i = 1$ and $j = 2$. We will construct a new graph H' from H by removing some vertex $v \in X_1$ and replacing it with a new vertex $v' \in X_2$ so that H' is also a complete k-partite graph. Let $X_1', \ldots, X_k'$ denote the new partite sets in H, then $|X_1'| = x_1 - 1$, $|X_2'| = x_2 + 1$, 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 $\nu(v', C_5) > \nu(v, C_5)$. 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 $I=\{3,\dots,k\}$.

Claim 3.4 $c'(v',1,0)+c'(v',2,0)+c'(v',0,1) > c(v,0,1)+c(v,0,2)+c(v,1,0)$.

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

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

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

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

By counting in similar way in H',

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

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

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

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

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

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 3.5 $c'(v',1,1)+c'(v',2,1) > c(v,1,1)+c(v,1,2)$.

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

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

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

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

By Proposition 3.1,

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

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

$$\binom{x_1-1}{2}(x_2+1)f_{12} > \binom{x_2}{2}x_1f_{21}$$

by Proposition 3.1, completing the proof of the claim. ■

As each of Claims 3.4 and 3.5 are true, it follows that $\nu(C_5,H') > \nu(C_5,H)$, completing the proof of Lemma 3.3. ■

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 3.6 (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 3.6 requires the following two lemmas along with Lemma 2.2.

Lemma 3.7 (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 1.3(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 1. The goal of Lemma 3.8 is to prove that if $\lim_{n\to\infty}(C_5,G_n) = \mathrm{OPT}_k(C_5)$ is an extremal sequence, then $\lim_{n\to\infty}(\overline{P_3},G_n) = 0$.

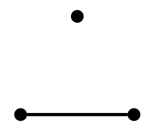
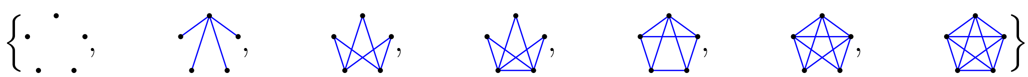


Figure 1: $\overline{P_3}$

Lemma 3.8 *For each $\delta_F > 0$, there exists $\varepsilon_F > 0$ and n_0 such that any K_{k+1}-free graph G on at least $n_0 = n_0(\delta_F)$ vertices with $d(C_5, G) > OPT_k(C_5) - \varepsilon_F$, G 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 1.3(i). This is the same set T provided at the end of Section 2.

$T =$ 

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

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

Since T contains the set of tight graphs, it follows that for the sequence $(G_n)_{n\geq 1}$,

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

Thus, (3) implies that for $(G_n)_{n\geq 1}$,

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

which completes the proof of Lemma 3.8. ∎

Proof of Lemma 3.6. Let $\varepsilon_I > 0$ and $\varepsilon_F > 0$, which we will determine later. By Lemma 3.7, there exists a $\delta_F > 0$ and an n_0 such that any K_{k+1}-free graph G on at least n_0 vertices containing at most $\delta_F n^3$ copies of $\overline{P_3}$ can be made $\overline{P_3}$-free after the removal of at most $\varepsilon_I n^2$ total edges. Assume that G is a graph on at least 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 3.7 and 3.8 so that G contains at most $\delta_F n^3$ copies of $\overline{P_3}$. Moreover, for sufficiently small ε_I,

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

By Lemma 3.7, 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 3.9 *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 pairs $i, j \in [1, \ell]$, each vertex in X_i is adjacent to each vertex in X_j. By definition, G' is a complete ℓ-partite graph. Since $d(C_5, G') > OPT_{k-1}(C_5) + 5! \cdot \varepsilon_I$, Lemma 3.8 implies that G' must be k-partite. ■

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 3.10 *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 exist $\varepsilon_I > 0$ and $\varepsilon_F > 0$ such that if*

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

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

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

Proof. For each $i = 1, \ldots, k$ let $x_i = |X_i|$. Let $\varepsilon' > 0$ and assume by symmetry that $x_1 = \frac{(1+\frac{\varepsilon'}{k-1})n}{k}$. We want to calculate an upper bound on $d(C_5, G')$. By Lemma 3.3, $d(C_5, G')$ is maximized it all remaining parts are balanced. That is, $x_i = \frac{(1-\varepsilon')n}{k}$ for $i = 2, \ldots, k$. If we picked $x_1 = \frac{(1+\varepsilon')n}{k}$, we would get less pleasant expressions in what follows. With knowing the sizes of all X_i's, the following is a straightforward calculation,

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

see Appendix 6.2.

For all $k \ge 3$, 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

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

This implies the statement of the claim. ■

Given an $\varepsilon > 0$, let $\varepsilon_T = \varepsilon/2$. Next, choose an $\varepsilon_I \le \varepsilon/2$ 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 3.6. ■

4 Exact Result

In this section we will prove Theorem 1.3(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 1.3(ii). Suppose that $k \geq 3$. By Lemma 3.6, there exists an $\varepsilon > 0$ and an integer $n = n(k, \varepsilon)$ so that for any K_{k+1}-free graph G on at least n vertices satisfying

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

we have that $\text{Dist}(G, T_k(n)) \leq \frac{1}{k^{10}}n^2$. This defines 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 $\frac{1}{k^{10}}n^2$ 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 4.1 *$|X_0| \leq \frac{1}{k^4}n$.*

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 $\frac{1}{k^{10}}n^2$, it follows that

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

which proves Claim 4.1. ■

For all $v \in V(G)$, let $d_i(v)$ denote the size of the set $N(v) \cap (X_i \setminus X_0)$. Let

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

By Claim 4.1,

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

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

Claim 4.2 *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,

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

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

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

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

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

This implies that we can find some common neighbor, say w_3, of u, v, and w_2, where $w_3 \in X_3 \setminus X_0$. 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 4.2 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 4.3, we will prove that if v is a type 2 vertex, then $d^*(v)$ must be relatively small. In Claim 4.4, we will prove a lower bound on the number of five-cycles containing a vertex v. Finally, in Claim 4.5, 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. Recall that

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

Claim 4.3 *Let $v \in X_0$ be a type 2 vertex. Then there exists integers i and j where $1 \leq i, j \leq k$ such that*

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

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

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

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 $q = 3, \ldots, k$, 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 $v, w_1, \ldots, w_{q-1}$. This implies that we can select k vertices $w_1, \ldots, w_k$ so that the set $\{v, w_1, \ldots, w_k\}$ induces a copy of K_{k+1} in G, which is a contradiction. Therefore, there exists an index $j \in \{2, \ldots, k\}$ for which $1 \leq d_1(v) \leq d_j(v) \leq \frac{1}{k^5}n$, completing the proof of Claim 4.3. ■

Claim 4.4 *For all $k \geq 3$, and $v \in V(G)$, $\nu(v, C_5) \geq (OPT_k(C_5) - \frac{1}{k^{10}})\binom{n}{4} - \frac{1}{k^5}n^4$.*

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

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

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

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

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

$$\nu(\{u, v\}, C_5) \leq n^3.$$

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,

$$\nu(C_5,G') - \nu(C_5,G) \geq \nu(u,C_5) - \nu(v,C_5) - \nu(\{u,v\},C_5)$$

$$\geq \frac{1}{k^5}n^4 - n^3 > 0$$

for all $n > k^5$. 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 $v \in V(G)$,

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

which completes the proof of Claim 4.4. ■

Claim 4.5 *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 4.3 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 $\nu(v,C_5)$. Suppose that C is a five-cycle defined by the edges $vu_1, u_1u_2, u_2u_3, u_3, u_4$, and u_4v. First we will count the maximum number of such five-cycles based on the locations of u_1 and u_4 as follows:

1. $u_1, u_4 \in X_1 \setminus X_0$ or $u_1, u_4 \in X_2 \setminus X_0$:

$$2\binom{\frac{n}{k^5}}{2}\frac{(k-1)(k-2)}{k^2}n^2. \tag{13}$$

2. $u_1 \in X_1 \setminus X_0$ and $u_4 \in X_2 \setminus X_0$:

$$\frac{1}{k^{10}}\left(\frac{(k-2)^2}{k^2} + \frac{k-1}{k^2}\right)n^4. \tag{14}$$

3. $u_1 \in (X_1 \setminus X_0) \cup (X_2 \setminus X_0)$ and $u_4 \notin X_1 \cup X_2$:

$$\frac{2}{k^6}\left(\frac{(k-2)^2}{k^2} + \frac{k-1}{k^2}\right)n^4. \tag{15}$$

4. $u_1, u_4 \notin X_1 \cup X_2$:

$$\frac{k-2}{2k^2} \cdot \frac{(k-1)(k-2)}{k^2}n^4 + \left(\frac{(k-2)(k-3)}{2k^2}\right)\left(\frac{(k-2)^2}{k^2} + \frac{k-1}{k^2}\right)n^4. \tag{16}$$

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 (13)–(16),

$$\nu(v, C_5) \leq \frac{n^4}{24}\left(12 - \frac{84}{k} + \frac{228}{k^2} - \frac{300}{k^3} + \frac{168}{k^4} + \frac{48}{k^6} - \frac{144}{k^7} + \frac{144}{k^8} + \frac{48}{k^{10}} - \frac{144}{k^{11}} + \frac{120}{k^{12}}\right).$$

The SageMath code for verifying this fact can be found in Appendix 6.3. This implies that for large enough n,

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

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 4.4 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 4.6 *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 without loss of generality there exists some index $i \in \{1, \ldots, k\}$ such that $i \neq j$ and*

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

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

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

for all $i = 2, \ldots, k$. Using an identical argument to the one made in the proof of Claim 4.3, 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,

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

Then without loss of generality,

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

which completes the proof of Claim 4.6. ■

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 4.5, we will only show that the two vertices u and v from Claim 4.6 cannot be adjacent.

Claim 4.7 *Suppose that u and v are adjacent 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. We will consider the cases where $k = 3$ and $k \geq 4$ separately.

Case 1: $k = 3$. Then by Claim 4.6 we may assume that $d_1(u) = 0$ and $d_2(u) = \frac{10}{54}n$. First we will count the five cycles containing u and no vertices in X_0. Note that since $G - X_0$ is a 3-partite graph, each of these five-cycles must contain at least one vertex in X_2. There are at most

$$\binom{\frac{10}{54}n}{2} \cdot \frac{2n}{3} \cdot \frac{n}{3}$$

five cycles in $G - X_0$ where both neighbors of u are contained in X_2. Next, there are at most

$$\frac{10n}{54} \cdot \frac{n}{3} \cdot \frac{2n}{3} \cdot \frac{n}{3}$$

five cycles in $G - X_0$ where exactly one neighbor of u is contained in X_2. Finally, there are at most

$$\binom{\frac{n}{3}}{2} \cdot \left(\frac{10n}{54} \cdot \left(\frac{10n}{54} + \frac{n}{3}\right) + \frac{10n}{54} \cdot \frac{n}{3}\right)$$

five cycles in $G - X_0$ where both neighbors of u are contained in X_3. Since there are at most $\frac{2n^4}{81}$ five cycles containing u and at least one other vertex in X_0,

$$\nu(u, C_5) \leq \frac{772}{729} \cdot \frac{n^4}{24}$$

Since

$$\mathrm{OPT}_3(C_5) - \frac{1}{3^{10}} = \frac{40}{27} - \frac{1}{3^{10}},$$

it is straightforward to verify that for large enough n

$$\left(\mathrm{OPT}_3(C_5) - \frac{1}{3^{10}}\right)\binom{n}{4} - \nu(u, C_5) \geq \frac{1}{3^5}n^4.$$

This contradicts Claim 4.4, completing the proof of Case 1.

Case 2: $k \geq 4$. By symmetry, we may assume that $d_1(u) = 0$ and

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

Suppose that C is a five-cycle defined by the edges $uv_1, v_1v_2, v_2v_3, v_3, v_4$, and v_4u. In a similar manner as in Claim 4.5, we will count the number of five-cycles incident to u by considering the possibilities for the locations of v_1 and v_4 as follows:

1. $v_1, v_4 \in X_2 \setminus X_0$:

$$\binom{\frac{k^2+1}{k^3}n}{2} \frac{(k-1)(k-2)}{k^2} n^2. \tag{17}$$

2. $v_1 \in X_2 \setminus X_0$ and $v_4 \notin X_2$:

$$\frac{k^2+1}{k^3} n \left(\frac{(k-1)^2(k-2)}{k^3} + \frac{(k-1)(k-2)}{k^2} \right) n^3. \tag{18}$$

3. u if $v_1, v_4 \notin X_2$:

$$\left(\frac{k-2}{2k^2} n^2 \right) \frac{(k-1)(k-2)}{k^2} n^2 + \left(\frac{(k-2)(k-3)}{2k^2} n^2 \right) \left(\frac{(k-2)^2}{k^2} + \frac{k-1}{k^2} \right) n^2. \tag{19}$$

As before, 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 (17)-(19),

$$\nu(u, C_5) \leq \left(12 - \frac{72}{k} + \frac{195}{k^2} - \frac{273}{k^3} + \frac{144}{k^4} + \frac{6}{k^5} + \frac{15}{k^6} - \frac{9}{k^7} + \frac{6}{k^8} \right) \frac{n^4}{24}.$$

For $k > 1000$ it is clear that

$$12 - \frac{72}{k} + \frac{195}{k^2} - \frac{273}{k^3} + \frac{144}{k^4} + \frac{6}{k^5} + \frac{15}{k^6} - \frac{9}{k^7} + \frac{6}{k^8} \leq \mathrm{OPT}_k(C_5) - \frac{1}{k^{10}}.$$

The SageMath code for verifying that this is also true for $4 \leq k \leq 100$ found in Appendix 6.4. Given this fact, it is straightforward to verify that

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

for large enough n. This, however contradicts Claim 4.4, which implies that u and v are not adjacent. ■

Claim 4.7 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 compelte k-partite graph. Finally, Proposition 3.2 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. ■

5 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. In [10], Cutler, Nir, and Radcliffe showed that the value of $\mathrm{ex}(n, S_t, K_{r+1})$ is achieved by an unbalanced r-partite graph when $r = 7$ and $t = 13$. It seems very likely that while the Turán graph is not always extemal, that some complete r-partite graph will be best possible.

Conjecture 5.1 *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 5.2 *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 $T_{r-1}(n)$.*

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

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] D. Gerbner. Generalized turán problems for small graphs, 2020.

[17] D. Gerbner and C. Palmer. Counting copies of a fixed subgraph in f-free graphs. 2018.

[18] D. Gerbner and C. Palmer. Some exact results for generalized Turán problems, 2020.

[19] L. Gishboliner and A. Shapira. A generalized Turán problem and its applications. *Proceeding sof the 50th Annual ACM SIGACT Symposium on Theorey of Computing*, pages 760–772, 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, J. Pach, and M. Simonovits. On the maximial number of certain subgraphs in K_r-free graphs. *Graphs Combin.*, 7(1):31–37, 1991.

[23] 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.

[24] T. Huynh, G. Joret, and D. R. Wood. Subgraph densities in a surface, 2020.

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

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

[27] T. Michael. Cycles of length 5 in triangle-free graphs: A sporadic counterexample to a characterization of equality. 2014.

[28] D. Mubayi and S. Mukherjee. Triangles in graphs without bipartite suspensions, 2020.

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

[30] 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.

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

[32] 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.

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

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

6 Appendix

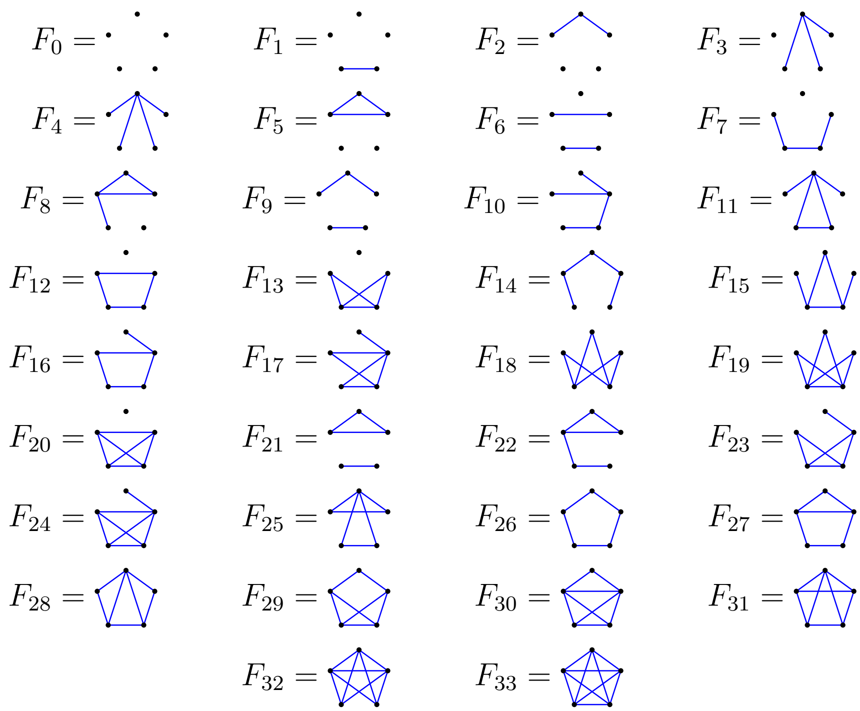


Table 1: Graphs on 5 vertices up to isomorphism.

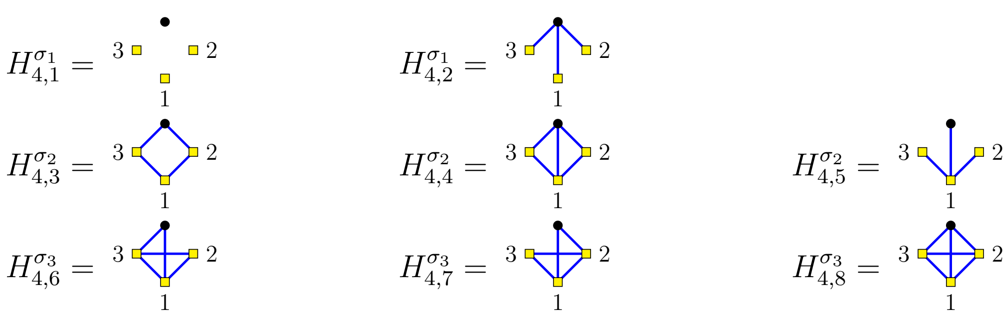


Table 2: Labeled graphs on four vertices.

6.1 Proof of Claim 2.1

```
# SageMath code for Claim 2.1

# Calculating the differences C_1 - C_i for i=2,..,10.
def differences(k):
        #Since the denominators in C_1,...,C_10 are
        #the same we clear those.

        C1 = 60*k^7 - 720*k^6 + 3600*k^5 -9876*k^4 + \
        16320*k^3 - 16440*k^2 + 9360*k - 2304
        C2 = 33*k^7 - 450*k^6 + 2547*k^5 - 7824*k^4 + \
        14214*k^3 - 15360*k^2 + 9144*k - 2304
        C3 = 30*k^7 - 420*k^6 + 2430*k^5 - 7596*k^4 + \
        13980*k^3 - 15240*k^2 + 9120*k - 2304
        C4 = 30*k^7 - 423*k^6 + 2457*k^5 - 7686*k^4 + \
        14118*k^3 - 15336*k^2 + 9144*k - 2304
        C5 = 35*k^7 - 468*k^6 + 2607*k^5 - 7916*k^4 + \
        14278*k^3 - 15376*k^2 + 9144*k - 2304
        C6 = 30*k^7 - 425*k^6 + 2468*k^5 - 7697*k^4 + \
        14098*k^3 - 15302*k^2 + 9132*k - 2304
        C7 = 35*k^7 - 455*k^6 + 2505*k^5 - 7644*k^4 + \
        13980*k^3 - 15240*k^2 + 9120*k - 2304
        C8 = 135/4*k^7 - 895/2*k^6 + 9967/4*k^5 - 7631*k^4 + \
        27913/2*k^3 - 15216*k^2 + 9114*k - 2304
        C9 = 50*k^7 - 610*k^6 + 3129*k^5 - 8902*k^4 + \
        15326*k^3 - 15956*k^2 + 9264*k - 2304
        C10 = 50*k^7 - 610*k^6 + 3103*k^5 - 8758*k^4 + \
        15050*k^3 - 15748*k^2 + 9216*k - 2304

        return [C2 - C1, C3 - C1, C4 - C1, C5 - C1,
           C6 - C1, C7 - C1, C8 - C1, C9 - C1, C10 - C1]


for k in [4..1000]:
        if max(differences(k)) > 0:
                print("k=",k," failed")

print("all values of k up to 1000 checked and failed reported if any")

var('k')
print("Differences as functions of k, these should be negative")
print("See the leading coefficients of the polynomials.")
for x in differences(k):
    print(x)
```

6.2 SageMath code for Claim 3.10

```
# 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)))
```

6.3 SageMath code for Claim 4.5

```
# SageMath code for Claim 4.5 - showing there are no type 2 vertices

var('k')

#This function counts the number of five cycles using equations (9) - (12)
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 )

        return 24*(onebadset_twobadvertices + twobadset_twobadvertices
        + nobadset_twosame + nobadset_nosame)


# Use the following function to verify that the method of
# counting five-cycles is correct
def counting_check(k):
        nobadset_twosame = ( (k-1)/(2*k^2) )*( (k-1)*(k-2)/k^2 )

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

        return 24*(nobadset_twosame + nobadset_nosame)

# This gives the sum of equations (9) - (12) factored in a nice way.

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

# actual upper bound once we account for the vertices in X_0
def bad_ub(k):
        return  expanded_first_check + 2/k^4

# 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 average and the count for type 2.
# This should be positive.
def epsilon(k):
        return factor(good_ub(k) - bad_ub(k))

print('difference between an average five cycle',epsilon(k))

# This is the polynomial representing the asymptotic difference in C_5
#density between an average vertex and a type 2 vertex.
#Again, we want to show this is positive.

#This function matches the difference output. We want this to be positive.
def count(r):
        return (24*r^11 - 108*r^10 + 180*r^9 - 168*r^8 - 49*r^2
                + 144*r - 120)/r^12

#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 count(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 k up to 1000"

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

6.4 SageMath code for Claim 4.7

```
# SageMath code for Claim 4.7

var('k')

# This function counts the number of five-cycles in equations (13) - (15)
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-1)^2*(k-2)/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 )
        return 24*(onebadset_twobadvertices
        + onebadset_onebadvertices + nobadset_twosame
        + nobadset_nosame)


# This gives the sum of equations (13) - (15) factored in a nice way

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

# The upper bound on five cycles containing a suboptimal type 1 vertex
def bad_ub(k):
        return  expanded_first_check + 48/k^4

def good_ub(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 factor(good_ub(k) - bad_ub(k))

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

# this is the polynomial representing the asymptotic difference
#in density between an ideal vertex,
#and a type one vertex of low degree.
def count(r):
        return (12*r^9 - 75*r^8 + 153*r^7 - 144*r^6
```

```
        - 6*r^5 - 15*r^4 + 9*r^3
        - 6*r^2 - 1)/(r^10)

# This verifies that for small values of k, epsilon(k) is greater than 1/k^20
def count_check(a,b):
        for i in [a..b]:
                if count(i) < 1/(i^(10)):
                        return "the difference is less than 1/k^5 for k =",i
        return "'difference' is greater than 1/k^20 for all values\
        of k up to 1000"

print(count_check(4,1000))
```