Counting copies of a fixed subgraph in $F$-free graphs

Dániel Gerbner∗ Cory Palmer†

Abstract

Fix graphs $F$ and $H$ and let $\text{ex}(n, H, F)$ denote the maximum possible number of copies of the graph $H$ in an $n$-vertex $F$-free graph. The systematic study of this function was initiated by Alon and Shikhelman [J. Comb. Theory, B. 121 (2016)]. In this paper, we give new general bounds concerning this generalized Turán function. We also determine $\text{ex}(n, P_k, K_{2,t})$ (where $P_k$ is a path on $k$ vertices) and $\text{ex}(n, C_k, K_{2,t})$ asymptotically for every $k$ and $t$. For example, it is shown that for $t \geq 2$ and $k \geq 5$ we have $\text{ex}(n, C_k, K_{2,t}) = \left(\frac{1}{2k} + o(1)\right)(t-1)^{k/2}n^{k/2}$. We also characterize the graphs $F$ that cause the function $\text{ex}(n, C_k, F)$ to be linear in $n$. In the final section we discuss a connection between the function $\text{ex}(n, H, F)$ and so-called Berge hypergraphs.

Keywords: Turán numbers, generalized Turán numbers
AMS Subj. Class. (2010): 05C35, 05C38

1 Introduction

Let $G$ and $F$ be graphs. We say that a graph $G$ is $F$-free if it contains no copy of $F$ as a subgraph. Following Alon and Shikhelman [1], let us denote the maximum number of copies of the graph $H$ in an $n$-vertex $F$-free graph by

$$\text{ex}(n, H, F).$$

The case when $H$ is a single edge is the classical Turán problem of extremal graph theory. In particular, the Turán number of a graph $F$ is the maximum number of edges possible in an $n$-vertex $F$-free graph $G$. This parameter is denoted $\text{ex}(n, F)$ and thus $\text{ex}(n, K_2, F) = \text{ex}(n, F)$. For more on the ordinary Turán number see, for example, the survey [12]. Recall that the Turán graph $T_{k-1}(n)$ is the complete $(k-1)$-partite graph with $n$ vertices such that the vertex classes are of size as close to each other as possible.

In 1962, Erdős [6] proved that the Turán graph $T_{k-1}(n)$ is the unique graph containing the maximum possible number of copies of $K_t$ in an $n$-vertex $K_k$-free graph (when $t < k$).

*Hungarian Academy of Sciences, Alfrédrényi Institute of Mathematics, P.O.B. 127, Budapest H-1364, Hungary. e-mail: gerbner.daniel@renyi.mta.hu
†Department of Mathematical Sciences, University of Montana, Missoula, Montana 59812, USA. e-mail: cory.palmer@umontana.edu
By counting copies of $K_t$ in $T_{k-1}(n)$ we get the following corollary for \( \text{ex}(n, K_t, K_k) \). Let \( N(H, G) \) denote the number of copies of the subgraph $H$ in the graph $G$.

**Corollary 1** (Erdős, [6]). If $t < k$, then

\[
\text{ex}(n, K_t, K_k) = N(K_t, T_{k-1}(n)) = \binom{k-1}{t} \left( \frac{n}{k-1} \right)^t + o(n^t).
\]

This result also follows from a theorem of Bollobás [2] and the case $t = 3$ and $k = 4$ was known to Moon and Moser [32]. Another proof appears in Alon and Shikhelman [1] modifying a proof of Turán’s theorem.

When $H$ is a pentagon, $C_5$, and $F$ is a triangle, $K_3$, this is a well-known conjecture of Erdős [8]. An upper bound of $1.03\left(\frac{n}{5}\right)^5$ was proved by Győri [20]. The blow-up of a $C_5$ gives a lower-bound of $\left(\frac{n}{5}\right)^5$ when $n$ is divisible by 5. Hatami, Hladký, Král’, Norine and Razborov [24] and independently Grzesik [19] proved

\[
\text{ex}(n, C_5, K_3) \leq \left(\frac{n}{5}\right)^5. \tag{1}
\]

Swapping the role of $C_5$ and $K_3$, we count the number of triangles in a pentagon-free graph. Bollobás and Győri [4] determined

\[
(1 + o(1)) \frac{1}{3\sqrt{3}} n^{3/2} \leq \text{ex}(n, K_3, C_5) \leq (1 + o(1)) \frac{5}{4} n^{3/2}.
\]

The constant in the upper bound was improved to $\frac{\sqrt{3}}{3}$ by Alon and Shikhelman [1] and by Ergemlidze, Győri, Methuku and Salia [10]. Győri and Li [23] give bounds on $\text{ex}(n, K_3, C_{2k+1})$. A particularly interesting case to determine the value of $\text{ex}(n, K_3, K_{r,r,r})$ was posed by Erdős [7] and remains open in general.

The systematic study of the function $\text{ex}(n, H, F)$ was initiated by Alon and Shikhelman [1] who proved a number of different bounds. Two examples from their paper are as follows.

An analogue of the Kővari-Sós-Turán theorem

\[
\text{ex}(n, K_3, K_{s,t}) = O(n^{3-3/s})
\]

which is shown to be sharp in the order of magnitude when $t > (s - 1)!$ (see also [26]).

Another example is an Erdős-Stone-Simonovits-type result that for fixed integers $t < k$ and a $k$-chromatic graph $F$ that

\[
\text{ex}(n, K_t, F) = \binom{k-1}{t} \left( \frac{n}{k-1} \right)^t + o(n^t). \tag{2}
\]

Gishboliner and Shapira [18] determined the order of magnitude of $\text{ex}(n, C_k, C_\ell)$ for every $\ell$ and $k \geq 3$. Moreover, they determined $\text{ex}(n, C_k, C_4)$ asymptotically. We give a theorem (proved independently of the previous authors) that extends this result to $\text{ex}(n, C_k, K_{2,t})$. Other recent results include [16] [13] [31].
The goal of this paper is to determine new bounds $ex(n, H, F)$ and investigate its behavior as a function. In Section 2 we give general bounds using standard extremal graph theory techniques. In particular, we give the following extension of (2) using a slight modification of its proof in [1].

**Theorem 2.** Let $H$ be a graph and $F$ be a graph with chromatic number $k$, then

$$ex(n, H, F) \leq ex(n, H, K_k) + o(n^{\lfloor H \rfloor}).$$

Note that Theorem 2 only gives a useful upper-bound if $ex(n, H, K_k) = \Omega(n^{\lfloor H \rfloor})$. Fortunately, this is often the case. For example, (2) follows by applying Corollary 1 in the case when $H = K_t$. Applying (1) to the case when $H = C_5$ gives

$$ex(n, C_5, F) \leq \left(\frac{n}{5}\right)^5 + o(n^5)$$

for every graph $F$ with chromatic number 3. When $F$ contains a triangle, the construction giving the lower bound in [1] can be used to give an asymptotically equal lower bound on $ex(n, C_5, F)$.

In Section 3 we give bounds on $ex(n, H, F)$ for some specific values of $H$ and $F$. For example,

**Theorem 3.** Fix $t \geq 2$ and $k \geq 5$. Then,

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

In Section 4 we study the behavior of the function $ex(n, C_k, F)$ and determine which graphs $F$ cause this function to be linear in $n$. In Section 5 we investigate for which graphs $H$ the function $ex(n, H, K_k)$ is maximized by the Turán graph. Finally, in Section 6 we establish connections between this counting subgraph problem and so-called Berge hypergraph problems. For notation not defined in this paper, see Bollobás [3].

## 2 General bounds on $ex(n, H, F)$

We begin with a proof of Theorem 2. Our proof mimics the proof of the Erdős-Stone-Simonovits by the regularity lemma. We use the following versions of the regularity lemma and an embedding lemma found in [3].

**Lemma 4** (Embedding lemma). Let $F$ be a $k$-chromatic graph with $f \geq 2$ vertices. Fix $0 < \delta < \frac{1}{k}$, let $G$ be a graph and let $V_1, \ldots, V_k$ be disjoint sets of vertices of $G$. If each $V_i$ has $|V_i| \geq \delta^{-f}$ and each pair of partition classes is $\delta^f$-regular with density $\geq \delta + \delta^f$, then $G$ contains $F$ as a subgraph.
**Lemma 5** (Regularity Lemma). For an integer \(m\) and \(0 < \epsilon < 1/2\) there exists an integer \(M = M(\epsilon, m)\) such that every graph on \(n \geq m\) vertices has a partition \(V_0, V_1, \ldots, V_r\) with \(m \leq r \leq M\) where \(|V_0| < \epsilon n\), \(|V_1| = |V_2| = \cdots = |V_r|\) and all but at most \(\epsilon r^2\) of the pairs \(V_i, V_j, 1 \leq i < j \leq r\) are \(\epsilon\)-regular.

**Proof of Proposition 4** Fix \(\delta > 0\) and an integer \(m \geq k\) such that the following inequality holds

\[
\left(\frac{1}{2m} + 2\delta^f + \frac{\delta + \delta^f}{2}\right) N(H, K_{|H|}) < \alpha. \tag{3}
\]

Let us apply the regularity lemma with \(\epsilon = \delta^f\) and \(m\) to get \(M = M(\epsilon, m)\). Let \(G\) be a graph on \(n > M\delta^{-f}\) vertices and more than \(\text{ex}(n, H, K_k) + \alpha n^{|H|}\) copies of \(H\). We will show that \(G\) contains \(F\) as a subgraph.

Let \(V_0, V_1, \ldots, V_r\) be the partition of \(G\) given by the regularity lemma. We will remove the following edges.

1. Remove the edges inside of each \(V_i\). There are at most \(r \binom{n/2}{r} \leq \frac{n^2}{2^r} \leq \frac{1}{2m} n^2\) such edges.
2. Remove the edges between all pairs \(V_i, V_j\) that are not \(\epsilon\)-regular. There are at most \(\epsilon r^2\) such pairs and each has at most \(\binom{n/2}{r}\) edges. So we remove at most \(\epsilon n^2\) such edges.
3. Remove the edges between all pairs \(V_i, V_j\) if the density of the pair \(d(V_i, V_j) < \delta + \delta^f\). There are less than \(\binom{r}{2}(\delta + \delta^f)(\frac{n}{r})^2 < \frac{\delta + \delta^f}{2} n^2\) such edges.
4. Remove all edges incident to \(V_0\). There are at most \(\epsilon n^2\) such edges.

In total we have removed at most

\[
\left(\frac{1}{2m} + 2\delta^f + \frac{\delta + \delta^f}{2}\right) n^2
\]

edges. There are at most \(N(H, K_{|H|}) n^{|H| - 2}\) copies of \(H\) containing a fixed edge. Therefore, by [3] we have removed less than \(\alpha n^{|H|}\) copies of \(H\). Thus, the resulting graph still has more than \(\text{ex}(n, H, K_k)\) copies of \(H\) so it contains \(K_k\) as a subgraph.

The \(k\) classes of the resulting graph that correspond to the vertices of \(K_k\) satisfy the conditions of the embedding lemma so \(G\) contains \(F\). \(\Box\)

Using a standard first-moment argument of Erdős-Rényi [9] we can get a lower-bound on the number of copies of \(H\) in an \(F\)-free graph.

**Proposition 6.** Let \(F\) and \(H\) be graphs such that \(e(F) > e(H)\). Then

\[
\text{ex}(n, H, F) = \Omega\left(n^{|H| - \frac{e(H)(|F|-2)}{e(F)-e(H)}}\right).
\]
Proof. Let $G$ be an $n$-vertex random graph with edge probability

$$p = cn^{-\frac{|F|-2}{e(F) - e(H)}}$$

where $c = |H||e(F) - e(H)| + 1$.

Among $|F|$ vertices in $G$ there are at most $|F|!$ copies of the graph $F$. Therefore, the expected number of copies of $F$ is at most

$$|F|! \left( \frac{n}{|F|} \right)^{p^e(F)} \leq n^{|F|} p^e(F).$$

Fix $|H|$ vertices in $G$. The probability of a particular copy of $H$ appearing among those vertices is $p^e(H)$. Thus, the probability of at least one copy of $H$ appearing among those $|H|$ vertices is at least $p^e(H)$. Therefore, the expected number of copies of $H$ is at least

$$\left( \frac{n}{|H|} \right)^{p^e(H)} \geq \left( \frac{n}{|H|} \right)^{|H|} p^e(H).$$

We remove an edge from each copy of $F$ in $G$ and count the remaining copies of $H$. There are at most $n^{|H|-2}$ copies of $H$ destroyed for each edge removed from $G$.

Let $X$ be the random variable defined by the difference between the number of copies of $H$ and the number of copies of $H$ destroyed by the removal of edges. The expectation of $X$ is

$$E[X] \geq \left( \frac{n}{|H|} \right)^{|H|} p^e(H) - n^{|H|-2} n^{|F|} p^e(F).$$

Which simplifies to

$$E[X] = \Omega \left( n^{|H|-\frac{e(H)|F|-2}{e(F) - e(H)}} \right).$$

This implies that there exists a graph such that after removing an edge from each copy of $F$ we are left with at least $E[X]$ copies of $H$.

We conclude this section with two simple bounds on $ex(n, H, F)$. Neither result is likely to give a sharp bound, but may be useful as simple tools.

**Proposition 7.** $ex(n, H, F) \geq ex(n, F) - ex(n, H)$.

**Proof.** Consider an edge-maximal $n$-vertex $F$-free graph $G$. Remove an edge from each copy of the subgraph in $H$ in $G$. The resulting graph does not contain $H$ and therefore has at most $ex(n, H)$ edges.

The other simple observation is a consequence of the Kruskal-Katona theorem [28, 25]. A hypergraph $\mathcal{H}$ is $k$-uniform if all hyperedges have size $k$. For a $k$-uniform hypergraph $\mathcal{H}$, the $i$-shadow is the $i$-uniform hypergraph $\Delta_i \mathcal{H}$ whose hyperedges are the collection of all subsets of size $i$ of the hyperedges of $\mathcal{H}$. We denote the collection hyperedges of a hypergraph $\mathcal{H}$ by $E(\mathcal{H})$. Here we use a version of the Kruskal-Katona theorem due to Lovász [30].
Theorem 8 (Lovász, [30]). If $H$ is a $k$-uniform hypergraph and

$$|E(H)| = \binom{x}{k} = \frac{x(x-1)\cdots(x-k+1)}{k!}$$

for some real number $x \geq k$, then

$$|E(\Delta H)| \geq \binom{x}{i}.$$ 

This gives the following easy corollary,

Corollary 9.

$$\text{ex}(n, K_t, F) \leq \text{ex}(n, F)^{t/2}.$$ 

Proof. Suppose $G$ is $F$-free and has the maximum number of copies of $K_t$. Let us consider the hypergraph $H$ whose hyperedges are the vertex sets of each copy of $K_k$ in $G$. Pick $x$ such that the number of hyperedges in $H$ is

$$|E(H)| = \binom{x}{t}.$$  \hspace{1cm} (4)

Applying Theorem 8 we get that the 2-uniform hypergraph (i.e., graph) $\Delta_2 H$ has size at least $\binom{x}{2}$.

On the other hand, the family $\Delta_2 H$ is a subgraph of $G$. Therefore,

$$\binom{x}{2} \leq e(G) \leq \text{ex}(n, F).$$ \hspace{1cm} (5)

Combining (4) and (5) gives the corollary. \hfill \Box

3 Counting paths and cycles in $K_{2,t}$-free graphs

The maximum number of edges in a $K_{2,t}$-free graph is

$$\text{ex}(n, K_{2,t}) = \left(\frac{1}{2} + o(1)\right) \sqrt{t-1} n^{3/2}. \hspace{1cm} (6)$$

The upper bound above is given by Kővári, Sós and Turán [27] and the lower bound is given by an algebraic construction of Füredi [11]. We will refer to this construction as the Füredi graph $F_{q,t}$. We recall some well-known properties of $F_{q,t}$ without giving a full description of its construction. For fixed $t$ and $q$ a prime power such that $t-1$ divides $q-1$, the graph $F_{q,t}$ has $n = (q^2 - 1)/(t-1)$ vertices. All but at most $2q$ vertices have degree $q$ and the others have degree $q-1$, thus the number of edges is $(1/2 + o(1))\sqrt{t-1} n^{3/2}$. Furthermore, every pair of vertices has at most $t-1$ common neighbors while every pair of non-adjacent vertices has exactly $t-1$ common neighbors.

Alon and Shikhelman [1] used the Füredi graph to give a lower bound in the following theorem.
Theorem 10 (Alon, Shikhelman, [1]).

\[ \text{ex}(n, K_3, K_{2,t}) = \left( \frac{1}{6} + o(1) \right) (t-1)^{3/2} n^{3/2}. \]

We generalize this theorem to cycles of arbitrary length and paths. We use the notation \( v_1v_2 \cdots v_k \) for the path \( P_k \) with vertices \( v_1, \ldots, v_k \) and edges \( v_iv_{i+1} \) (for \( 1 \leq i \leq k-1 \)). The cycle \( C_k \) that includes this path and the edge \( v_kv_1 \) is denoted \( v_1v_2 \cdots v_kv_1 \).

Proposition 11. For \( t \geq 3 \) we have

\[ \text{ex}(n, C_4, K_{2,t}) = \left( \frac{1}{8} + o(1) \right) (t-1)^2 n^2. \]

Proof. We begin with the upper bound. Consider an \( n \)-vertex graph \( G \) that is \( K_{2,t} \)-free. Fix two vertices \( u \) and \( v \). As \( G \) is \( K_{2,t} \)-free, \( u \) and \( v \) have at most \( t-1 \) common neighbors.

Therefore the number of \( C_4 \)s with \( u \) and \( v \) as non-adjacent vertices is at most \( \binom{t-1}{2} \). Therefore, the number of \( C_4 \)s in \( G \) is at most

\[ \frac{1}{2} \binom{n}{2} \binom{t-1}{2} \leq \frac{1}{8} (t-1)^2 n^2 \]

as each cycle is counted twice.

The lower bound is given by the F"uredi graph \( F_{q,t} \). Every pair of non-adjacent vertices has \( t-1 \) common neighbors, so there are \( \binom{t-1}{2} \) copies of \( C_4 \) containing them. There are \( (1/2 + o(1))n^2 \) pairs of non-adjacent vertices in \( F_{q,t} \). Each \( C_4 \) is counted twice in this way, so the number of \( C_4 \)s in \( F_{q,t} \) is at least

\[ \frac{1}{2} \left( \frac{1}{2} + o(1) \right) n^2 \binom{t-1}{2} \geq \left( \frac{1}{8} + o(1) \right) (t-1)^2 n^2. \]

\[ \square \]

A slightly more sophisticated argument than the proof of Proposition 11 is needed to count longer cycles and paths.

Theorem 12. Fix \( t \geq 2 \). For \( k \geq 5 \),

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

and for \( k \geq 2 \),

\[ \text{ex}(n, P_k, K_{2,t}) = \left( \frac{1}{2} + o(1) \right) (t-1)^{(k-1)/2} n^{(k+1)/2}. \]
Proof. We begin with the upper bound for \( \text{ex}(n, C_k, K_{2,t}) \). Let \( G \) be a \( K_{2,t} \)-free graph. We distinguish two cases based on the parity of \( k \).

**Case 1:** \( k \) is even. Fix a \((k/2)\)-tuple \((x_1, x_2, \ldots, x_{k/2})\) of distinct vertices of \( G \). This can be done in at most \( n^{k/2} \) ways. We count the number of cycles \( v_1v_2 \cdots v_kv_1 \) such that \( x_i = v_{2i} \) for \( 1 \leq i \leq k/2 \). As \( G \) is \( K_{2,t} \)-free, there are at most \( t-1 \) choices for each vertex \( v_{2i+1} \) on the cycle (for \( 0 \leq i \leq (k-2)/2 \)) as \( v_{2i+1} \) must be joined to both \( v_{2i+2} \) and \( v_{2i} \) (where the indicies are modulo \( k \)). Each cycle \( v_1v_2 \cdots v_kv_1 \) is counted by \( 2k \) different \((k/2)\)-tuples, so the number of copies of \( C_k \) is at most

\[
\frac{1}{2k} (t-1)^{k/2} n^{k/2}.
\]

**Case 2:** \( k \) is odd. Fix a \(((k+1)/2)\)-tuple \((x_1, x_2, \ldots, x_{(k-3)/2}, y, z)\) of distinct vertices such that \( yz \) is an edge. This can be done in at most

\[
2e(G)n^{(k-3)/2} \leq (1 + o(1)) (t-1)^{1/2} n^{3/2} n^{(k-3)/2} = (1 + o(1)) (t-1)^{1/2} n^{k/2}
\]

ways by (6). We count the number of cycles \( v_1v_2 \cdots v_kv_1 \) such that \( x_i = v_{2i} \) for \( 1 \leq i \leq (k-3)/2, y = v_{k-1}, \) and \( z = v_k \). Similar to Case 1, as \( G \) is \( K_{2,t} \)-free, there are at most \( t-1 \) choices for each of the \((k-1)/2\) remaining vertices \( v_{2t+1} \) of the cycle. Each cycle \( v_1v_2 \cdots v_kv_1 \) is counted by \( 2k \) different \(((k+1)/2)\)-tuples, so the number of copies of \( C_k \) is at most

\[
\frac{1}{2k} (t-1)^{(k-1)/2} (1 + o(1)) (t-1)^{1/2} n^{k/2} = \left( \frac{1}{2k} + o(1) \right) (t-1)^{k/2} n^{k/2}.
\]

For the upper bound on \( \text{ex}(n, P_k, K_{2,t}) \) we fix a tuple of distinct vertices of \( G \) as above. We sketch the proof and leave the remaining details to the reader. If \( k \) is odd we fix a \(((k+1)/2)\)-tuple \((x_1, x_2, \ldots, x_{(k+1)/2})\) and if \( k \) is even we fix a \(((k+2)/2)\)-tuple \((x_1, x_2, \ldots, x_{(k-2)/2}, y, z)\) such that \( yz \) is an edge. In both cases we count the paths \( v_1v_2 \cdots v_k \) such that \( x_i = v_{2i-1} \) and with the additional conditions that \( y = v_{k-1}, \) and \( z = v_k \) in the case \( k \) even. Similar to the case for cycles there are at most \( t-1 \) choices for each of the remaining vertices of the path. Each path is counted exactly two times in this way.

Both lower bounds are given by the Füredi graph \( F_{q,t} \) for \( q \) large enough compared to \( t \) and \( k \). We begin by counting copies of the path \( P_k = v_1v_2 \cdots v_k \) greedily. The vertex \( v_1 \) can be chosen in \( n \) ways. As the Füredi graph \( F_{q,t} \) has minimum degree \( q-1 \), we can pick vertex \( v_i \) (for \( i > 1 \)) in at least \( q - i + 1 \) ways. Each path is counted twice in this way, therefore, we have at least

\[
\frac{1}{2} n(q-k+1)^{k-1} = \left( \frac{1}{2} + o(1) \right) (t-1)^{(k-1)/2} n^{(k+1)/2}
\]

paths of length \( k \) in the Füredi graph \( F_{q,t} \).

For counting copies of the cycle \( C_k = v_1v_2 \cdots v_kv_1 \) we proceed as above with the addition that \( v_k \) should be adjacent to \( v_1 \). In order to do this, we pick \( v_1 \) arbitrarily and \( v_2, \ldots, v_{k-3} \) greedily as in the case of paths. As \( k \geq 5 \) the vertex \( v_{k-3} \) is distinct from \( v_1 \). From the
neighbors of \(v_{k-3}\) we pick \(v_{k-2}\) that is not adjacent to \(v_1\). The number of choices for \(v_{k-2}\) is at least \(q - k + 3 - (t - 1)\) as \(v_{k-3}\) and \(v_1\) have at most \(t - 1\) common neighbors. From the neighbors of \(v_{k-2}\) we pick \(v_{k-1}\) that is not adjacent to any of the vertices \(v_1, \ldots, v_{k-3}\). Each \(v_i\) has at most \(t - 1\) common neighbors with \(v_{k-2}\) which forbids at most \((k - 3)(t - 1)\) vertices as a choice for \(v_{k-1}\). Therefore, we have at least \(q - k - 2 - (k - 3)(t - 1)\) choices for \(v_{k-1}\).

Since \(v_{k-1}\) is not joined to \(v_1\) by an edge they have \(t - 1\) common neighbors and none of these neighbors are among \(v_1, v_2, \ldots, v_{k-1}\). Hence we can pick any of the common neighbors as \(v_k\). Every copy of \(C_k\) is counted \(2k\) times, thus altogether we have at least

\[
\frac{1}{2k} n(q - t(k - 3))^{k-2}(t - 1) = \left(\frac{1}{2k} + o(1)\right) (t - 1)^{k/2}n^{k/2}
\]

copies of \(C_k\).

\[
\square
\]

4 Linearity of the function \(\text{ex}(n, C_k, F)\)

It is easy to see that \(\text{ex}(n, H, F)\) is never sublinear (except for the obvious case when \(H\) contains \(F\) and it is 0). It is natural to investigate which graphs \(H\) and \(F\) cause \(\text{ex}(n, H, F)\) to be linear. When \(H\) is \(K_3\), Alon and Shikhelman \([1]\) characterized the graphs \(F\) with \(\text{ex}(n, K_3, F) = O(n)\). For trees they also essentially answer the question by determining the order of magnitude of \(\text{ex}(n, T, F)\) where both \(T\) and \(F\) are trees. One can easily see that their proof extends to the case when \(F\) is a forest. On the other hand, if \(F\) contains a cycle and \(T\) is a tree, then \(\text{ex}(n, F)\) is superlinear and \(\text{ex}(n, T)\) is linear. Thus by Proposition 7 we have that \(\text{ex}(n, T, F)\) is superlinear.

![Figure 1: The graphs \(C_{k}^{*r}, C_{4}^{*r}\) and, \(C_{5}^{*r}\)](image)

Now we turn our attention to the case when \(H\) is a cycle. We begin by introducing some notation. Let \(C_{k}^{*r}\) be a cycle \(C_k\) with \(r\) additional vertices adjacent vertex \(x\) of the \(C_k\). For \(k = 4\), let \(C_{4}^{**r}\) be a cycle \(v_1v_2v_3v_4\) with \(2r\) additional vertices; \(r\) are adjacent to \(v_1\) and \(r\) are
Banana graph $B_t^r$ \hspace{1cm} $Q_k^r$-graph \hspace{1cm} $R_k^r(a, b, c, d)$

Figure 2: A banana graph $B_t^r$, a $Q_k^r$-graph, and $R_k^r(a, b, c, d)$

adjacent edge to $v_3$. Similarly, let $C_5^{**r}$ be a cycle $v_1v_2v_3v_4v_5$ with $2r$ additional vertices; $r$ are adjacent to $v_1$ and $r$ are adjacent to $v_3$. See Figure[4] for an example of these graphs.

A banana graph $B_t^r$ is the union of $r$ internally-disjoint $u-v$ paths of length $t$. We call the vertices $u, v$ the main vertices of $B_t^r$ and the $u-v$ paths of $B_t^r$ are its internal paths.

A $Q_k^r$-graph is a graph consisting of a banana graph $B_t^r$ (for some $t < k$) with main vertices $u, v$ and a $u-v$ path of length $k-t$ that is otherwise disjoint from $B_t^r$. Alternatively, a $Q_k^r$-graph is a $C_k$ with $r-1$ additional paths of length $t$ (for some $t < k$) between two vertices that are joined by a path of length $t$ in the $C_k$. The internal paths and main vertices of a $Q_k^r$-graph are simply the internal paths and main vertices of the associated banana graph $B_t^r$. The main path of a $Q_k^r$-graph is the associated $u-v$ path of length $k-t$.

For $a, c \geq 2$ and $b, d \geq 0$ such that $a + b + c + d = k$, let $R_k^r(a, b, c, d)$ be the graph formed by a copy of $B^r_a$ with main vertices $u, v$ and a copy of $B^r_c$ with main vertices $u', v'$ together with a $v-v'$ path of length $b$ and a $u-u'$ path of length $d = k - (a + b + c)$. When $b = 0$ we identify the vertices $v$ and $u'$ and when $d = 0$ we identify the vertices $u$ and $v'$. Note that the last parameter $d$ is redundant, but we include it for ease of visualizing individual instances of this graph. For simplicity, we call any graph $R_k^r(a, b, c, d)$ an $R_k^r$-graph. Finally, let the family forests which are subgraphs of every $R_k^r$-graph be the $F_k^r$-graphs (i.e., the forests that are subgraphs of every $R_k^r(a, b, c, d)$ for all permissible values of $a, b, c, d$).

We now characterize those graphs $F$ for which the function $\text{ex}(n, C_k, F)$ is linear.

**Theorem 13.** For $k = 4$ and $k = 5$, if $F$ is a subgraph of $C_k^{***r}$ (for some $r$ large enough), then $\text{ex}(n, C_k, F) = O(n)$. For $k > 5$, if $F$ is a subgraph of $C_k^{**r}$ or an $F_k^r$-graph (for some $r$ large enough), then $\text{ex}(n, C_k, F) = O(n)$. On the other hand, for every $k > 3$ and every other $F$ we have $\text{ex}(n, C_k, F) = \Omega(n^3)$.

It is difficult to give a simple characterization of $F_k^r$-graphs. However, the following lemma
gives some basic properties of these forests. For simplicity, the term high degree refers to a vertex of degree greater than 2. A star is a single high degree vertex joined to vertices of degree 1. A broom is a path (possibly of a single vertex) with additional leaves attached to one of its end-vertices. Finally, let $c(F)$ be the sum of the number of vertices in the longest path in each component of $F$ (excluding the isolated vertex components).

![Figure 3: An $F_k^r$-graph with a non-broom component and an $F_k^r$-graph $F$ with $c(F) = k + 4$.]

**Proposition 14.** Let $F$ be an $F_k^r$-graph, i.e., $F$ is a subforest of every $R_k^r$-graph. Then the following properties hold when $k > 5$:

1. $F$ has at most two vertices of high degree. This implies that all but at most two components of $F$ are paths.

2. Each component of $F$ has at most one vertex of high degree.

3. Each vertex of high degree in $F$ is adjacent to at most two vertices of degree 2.

4. If $F$ has two high degree vertices, then at least one of them is contained in a component that is a broom.

5. The number of vertices in the longest path in $F$ is at most $k$.

6. $c(F) \leq k + 4$.

7. If $c(F) = k + 4$, then $F$ contains three components that are stars on at least 3 vertices. Furthermore, each component of $F$ with a high degree vertex is a star.

**Proof.** The first property follows as $F$ is a subgraph of the graph $R_k^r(2, 0, k - 2, 0)$ which has exactly two high degree vertices.
For property two, consider the graphs $R^r_k(2,0,k-2,0)$ and $R^r_k(3,0,k-3,0)$. Each graph has two high degree vertices and they are at distance 2 and 3, respectively. If $F$ had a component with two high degree vertices, then these vertices would be at distance 2 and 3 simultaneously; a contradiction. Note that we use $k > 5$ here.

For property three, consider the graph $R^r_k(2,0,2,k-4)$. This graph contains three high degree vertices $x, y, z$ such that every vertex adjacent to $y$ is adjacent to either $x$ or $z$. If $F$ has a component with a high degree vertex adjacent to more than two vertices of degree 2, then that component contains a cycle; a contradiction.

For property four, again consider the graph $R^r_k(2,0,2,k-4)$ and define the three high degree vertices $x, y, z$ as before. If $F$ has two components each with a high degree vertex, then without loss of generality one of these high degree vertices is $x$. If $x$ is adjacent to two vertices of degree 2 in $F$, then one of these vertices is $y$. Therefore, the other high degree vertex in $F$ is $z$. That component cannot contain $y$, so $z$ is adjacent to at most one vertex of degree 2, i.e., that component is a broom.

For property five, observe that the number of vertices in a longest path in $R^r_k(2,0,2,k-4)$ is $k$.

For property six and seven we can assume that all the components of $F$ are paths (by deleting unnecessary leaves) and that each component contains at least two vertices.

Consider again the graph $R^r_k(2,0,2,k-4)$ with high degree vertices $x, y, z$ as above. Note that this graph contains an $x$-$z$ path on $k - 3$ vertices. The components of $F$ containing $x$ or $z$ have at most 2 additional vertices not on this path. Moreover, the component of $F$ containing $y$ has at most 3 vertices not on this path (this includes $y$ itself). Therefore, $c(F) \leq k - 3 + 2 + 2 + 3 = k + 4$. This proves property six. In order to achieve equality $c(F) = k + 4$ there must be three distinct components containing $x, y$ and $z$ and each of these components has 3 vertices in their longest path, i.e., each such component is a star. This proves property seven.

The next lemma establishes another class of graphs that contains each $F^r_k$-graph as a subgraph.

**Lemma 15.** Let $k > 5$ and $H$ be a graph formed by two $Q^r_k$-graphs $Q_1, Q_2$ such that $Q_1$ and $Q_2$ share at most one vertex and such a vertex is on the main path of both $Q_1$ and $Q_2$. Then each $F^r_k$-graph is a subgraph of $H$.

**Proof.** Let $F$ be an $F^r_k$-graph. We will show that $F$ can be embedded in $H$. Suppose $Q_1, Q_2$ share a vertex $x$ on their main paths as this is the more difficult case.

Let $F'$ be a graph formed by components of $F$ such that $c(F') \leq k$ and there is at most one vertex of high degree in $F'$. We claim that $F'$ can be embedded into $Q_2$. Indeed, first we embed the component of $F'$ containing the high degree vertex using a main vertex of $Q_2$. The remaining (path) components of $F'$ can be embedded into the remaining vertices of the $C_k$ in $Q_2$ greedily. Now, if we can embed components of $F$ into $Q_1$ without using the vertex $x$ such that the remaining components satisfy the conditions of $F'$ above, then we are done.

First suppose that $c(F) = k + 4$. By property seven of Proposition 14 let $T$ and $T'$ be distinct star components of $F$ such that $T$ has exactly 3 vertices. It is easy to see that $T$
and $T'$ can both be embedded to $Q_1$ without using the vertex $x$. Therefore, the remaining components of $F$ can be embedded into $Q_2$.

We may now assume $c(F) \leq k + 3$. If $F$ contains a single component, then it can be embedded into $Q_1$ by property five of Proposition 14. If $F$ has no high degree vertex, then every component is a path. In this case it is easy to embed $F$ into $Q_1$ and $Q_2$. So let us assume that $F$ contains at least two components and at least one high degree vertex.

The graph $Q_1$ has two high degree vertices. Therefore, one of them is connected to $x$ by a path $P_\ell$ with $\ell > (k + 2)/2$.

Suppose $F$ contains two high degree vertices, then let $T$ be a component containing a high degree vertex. We may assume that the number of vertices on the longest path in $T$ is at most $(k + 3)/2$ (as there are two components with a high degree vertex). Therefore, we may embed $T$ into $Q_1$ without using the vertex $x$. The remaining components of $F$ can be embedded into $Q_2$.

Now suppose $F$ contains exactly one high degree vertex. If $F$ contains a component with longest path on $k$ vertices, then it can be embedded into $Q_2$ and the remaining component of $F$ can be embedded into $Q_1$ without using vertex $x$. So we may assume all components in $F$ have longest paths with less than $k$ vertices. If there is a (path) component on at least three vertices, then it can be embedded into $Q_1$ without using the vertex $x$ and the remaining components of $F$ can be embedded into $Q_2$. If there is no such path, then all path components are single edges. Two such edges can be embedded into $Q_1$ without using the vertex $x$ and the remaining components can be embedded into $Q_2$ as before.  

A version of the next lemma has already appeared in a slightly different form in [14].

\begin{figure}[h]
\centering
\includegraphics[width=0.5\textwidth]{figure4.png}
\caption{The graph $B$ from Lemma 16}
\end{figure}

\textbf{Lemma 16.} Fix integers $s \geq 2$ and $i \geq 2$. Let $G$ be a graph containing a family $P$ of $(s_i)^{2i-2}$ $u$–$v$ paths of length $i$. Then $G$ contains a subgraph $B$ consisting of a banana graph $B_t^s$ (for some $t \leq i$) with main vertices $u', v'$ together with a $u$–$u'$ path and and $v'$–$v$ path that are disjoint from each other and otherwise disjoint from $B$ (we allow that the additional paths be of length 0, i.e., $u = u'$ and $v = v'$), such that each $u$–$v$ path in $B$ is a sub-path of some member of $P$. Moreover, if each member of $P$ is a sub-path of some copy of $C_k$ in $G$, then $G$ contains a $Q_k^l$-graph where $s' = s - k$.

\textbf{Proof.} We prove the first part of the lemma by induction on $i$. The statement clearly holds for $i = 2$, as such a collection of paths is a banana graph. Let $i > 2$ and suppose the lemma holds for smaller values of $i$. If there are $s$ disjoint paths of length $i$ between $u$ and $v$ then
we are done. So we may assume that there are at most $s - 1$ disjoint paths of length $i$ from $u$ to $v$. The union of a set of disjoint paths of length $i$ from $u$ to $v$ has at most $si$ vertices. Furthermore, every other $u$-$v$ path of length $i$ must intersect this set of vertices. Therefore, there is a vertex $w$ that is contained in at least $(si)^{2i-3}$ of these paths. This $w$ can be in different positions in those paths, but there are at least $(si)^{2i-4}$ paths where $w$ is the $(p + 1)$st vertex (counting from $u$) with $1 \leq p < i$. Then there are either at least $(si)^{2p-2} > (sp)^{2p-2}$ sub-paths of length $p$ from $u$ to $w$ or at least $(si)^{2(i-p)-2} > (s(i-p))^{2(i-p)-2}$ sub-paths of length $i - p$ from $w$ to $v$. Without loss of generality, suppose there are at least $(si)^{2p-2} > (sp)^{2p-2}$ sub-paths of length $p$ from $u$ to $w$. Then, by induction on this collection of paths of length $p < i$, we find a banana graph $B^*_i$ with main vertices $u'$, $v'$ together with a $u'$-$u$ path and a $v'$-$w$ path (that are disjoint from each other). As there is a path from $w$ to $v$ we have the desired subgraph $B$.

Now it remains to show that if each member of $P$ is a sub-path of some copy of $C_k$ in $G$, then $G$ contains a $Q'_k$-graph. Suppose we have a graph $B$ from the first part of the lemma. Let $C$ be a cycle of length $k$ that contains any $u$-$v$ path of length $i$ in $B$. Note that $C$ also contains a $u$-$v$ path $P$ of length $k - i$. The internal vertices of $P$ intersect at most $k$ of the internal paths of the banana graph $B^*_i$ in $B$. Remove these internal paths from $B^*_i$ and let $B'$ be the resulting subgraph of $B$. Now $B'$ together with $P$ forms a $Q'_k$-graph.

**Proof of Theorem 13** First let us suppose that $F$ is a graph such that $ex(n, C_k, F) = o(n^2)$. Therefore, $F$ must be a subgraph of every graph with $\Omega(n^2)$ copies of $C_k$. It is easy to see that each $R^r_k$-graph contains $\Omega(n^2)$ copies of $C_k$. Thus, $F$ is a subgraph of every $R^r_k$-graph.

The Füredi graph $F_{q,2}$ does not contain a copy of $C_4$ and contains $\Omega(n^2)$ copies of $C_k$ for $k \geq 5$. Furthermore, $F_{q,3}$ contains $\Omega(n^2)$ copies of $C_4$. This follows from the proof of the lower bound in Theorem 12. Therefore, when $k \geq 5$ and $r$ and $q$ are large enough, the graph $F$ is a subgraph of $F_{q,2}$. When $k = 4$, and $r$ and $q$ are large enough the graph $F$ is a subgraph of $F_{q,3}$.

**Claim 17.** The graph $F$ contains at most one cycle and it is of length $k$.

**Proof.** As $F$ is a subgraph of $R^r_k(2,0,2,k - 4)$, every cycle in $F$ is of length $k$ or 4. If $k > 4$, as $F$ is a subgraph of $F_{q,2}$, it does not contain cycles of length 4. Therefore, all cycles in $F$ are of length $k$.

Suppose there is more than one copy of $C_k$ in $F$. For $k = 4$, as $F$ is a subgraph of $R^r_4(2,0,2,0)$ it is easy to see that any two copies of $C_4$ in $F$ form a $K_{2,3}$ or $K_{2,4}$. This contradicts the fact that $F$ is also a subgraph of $F_{q,3}$. For $k = 5$, as $F$ is a subgraph of $R^r_5(2,0,3,0)$ it is easy to see that any two copies of $C_5$ in $F$ form a $C_4$ or $C_6$. This contradicts the fact that all cycles are of length $k$. For $k > 5$, as $F$ is a subgraph of $R^r_k(2,0,2,k - 4)$, every pair of $C_k$s in $F$ share $k - 3$ or $k - 1$ vertices. On the other hand, as $F$ is a subgraph of $R^r_k(3,0,3,k - 6)$, every pair of $C_k$s in $F$ share $k - 4$ or $k - 2$ vertices; a contradiction. □

We now distinguish three cases based on the value of $k$.

**Case 1:** $k = 4$. The graph $F$ is a subgraph of $R^r_4(2,0,2,0)$. By Claim 17, $F$ has at most one cycle. The subgraphs of $R^r_4(2,0,2,0)$ with at most one cycle are clearly subgraphs of $C^{*rr}_4$. 

14
Case 2: $k = 5$. The graph $F$ is a subgraph of $R_5^r(2, 0, 3, 0)$ and therefore has at most 2 vertices of degree greater than 2 and they are non-adjacent. Furthermore, $F$ is a subgraph of $R_5^r(2, 0, 2, 1)$. By the above claim, $F$ has at most one cycle. The subgraphs of $R_5^r(2, 0, 2, 1)$ with at most one cycle that are simultaneously subgraphs of $R_5^r(2, 0, 3, 0)$ are subgraphs of $C_5^{w,r}$.

Case 3: $k > 5$. First assume that $F$ is a forest. As every $R_k^r$-graph contains $\Omega(n^2)$ copies of $C_k$, each must contain $F$ as a subgraph. Therefore, $F$ is an $F_k^r$-graph by definition.

Now consider the remaining case when $F$ contains a cycle $C$. As $F$ is a subgraph of $R_k^r(2, 0, k - 2)$, every edge of $F$ is incident to $C$. If $F$ has at least two vertices of degree greater than 2 on $C$, then as $F$ is a subgraph of both $R_k^r(2, 0, k - 2)$ and $R_k^r(3, 0, k - 3)$, we have that these two vertices should be at distance 2 and 3 from each other in $F$; a contradiction. Thus, there is only one vertex of degree greater than 2 on $C$. Therefore, $F$ is a subgraph of $C_k^{w,r}$. This completes the first part of the proof that if $F$ is a graph such that $\text{ex}(n, C_k, F) = O(n)$, then $F$ is as characterized in the theorem.

Now it remains to show that if $F$ is as characterized in the theorem, then $\text{ex}(n, C_k, F) < cn$ for some constant $c$. The constants $k$ and $r$ are given by the statement of the theorem. Fix constants $r''$, $r'$, $\gamma$, $c', c$ in the given order such that each is large enough to $k$ and $r$ and the previously fixed constants.

Let $G$ be a vertex-minimal counterexample, i.e, $G$ is an $n$-vertex graph with at least $cn$ copies of $C_k$ and no copy of $F$ such that $n$ is minimal. We may assume every vertex in $G$ is contained in at least $c$ copies of $C_k$, otherwise we can delete such a vertex (destroying fewer than $c$ copies of $C_k$) to obtain a smaller counterexample.

Case 1: $F$ contains a cycle. Thus, for $k = 4, 5$ we have that $F$ is a subgraph of $C_k^{w,r}$ and for $k > 5$ we have that $F$ is a subgraph of $C_k^{w,r}$. If every vertex of $G$ has degree at least $2r + k$, then on any $C_k$ in $G$ we can build a copy of $C_k^{w,r}$ or $C_k^{w,r}$ greedily. These graphs contain $F$; a contradiction.

Now let $x$ be a vertex with degree less than $2r + k$. This implies that there is an edge $xy$ contained in at least $c/(2r + k)$ copies of $C_k$. Therefore, that there are at least $c/(2r + k)$ $x$--$y$ paths of length $k - 1$. As $c$ is large enough compared to $k$ and $r$, we may apply Lemma 16 to this collection of paths of length $k - 1$ (each a subgraph of a $C_k$) to get a $Q_k^r$-graph $Q$. The graph $Q$ contains $C_k^{w,r}$ when $k > 5$ and $C_k^{w,r}$ when $k = 4, 5$. This implies that $G$ contains $F$; a contradiction.

Case 2: $F$ is a forest. Note that if $k \leq 4$, then $F$ is a subgraph of $C_k^{w,r}$, and we are done. Thus, we may assume $k > 5$.

Claim 18. Suppose $G$ contains a collection $\mathcal{C}$ of at least $c'n$ copies of $C_k$. Then there is an integer $\ell < k$ such that $G$ contains a $Q_k^r$-graph $Q$ with main vertices $x, y$ and internal paths of length $\ell$ such that less than $c'n$ members of $\mathcal{C}$ contain $x, y$ at distance $\ell$.

Proof. We distinguish two cases.

Case 1: There exists two vertices $u, v$ of $G$ in at least $c'n$ members of $\mathcal{C}$. Then there are at least $(c'/k)n$ members of $\mathcal{C}$ that contain a $u$--$v$ sub-path of length $i < k$. Let us suppose that $u$ and $v$ are chosen such that $i$ is minimal. Among these $u$--$v$ paths of length $i$ we can find a collection $\mathcal{P}$ of $(c'/k)n/(ni) \geq c'/k^2$ of them that contain some fixed vertex.
Applying Lemma 16, this collection \( P \) of \( u-w \) paths of length \( j \) gives a \( Q_k' \)-graph \( Q \). Let \( x, y \) be the main vertices of \( Q \) and let \( \ell \leq j < i \) be the length of the main paths in \( Q \). By the minimality of \( i \), there are less than \( c'n \) members of \( C \) that contain \( x, y \) at distance \( \ell \).

**Case 2:** The graph \( G \) does not contain two vertices in \( c'n \) members of \( C \). As \( G \) is \( F \)-free and \( F \) is a forest there are at most \( 2|V(F)| \) edges in \( G \). Thus, there is an edge \( uv \) contained in at least \( c'/(2|V(F)|) \) members of \( C \). Let \( P \) be a collection of \( c'/(2|V(F)|) \) \( u-v \) paths of length \( k-1 \) defined by these members of \( C \). Applying Lemma 16 to \( P \) gives a \( Q_k' \)-graph \( Q \). Let \( x, y \) be the main vertices of \( Q \). By the assumption in Case 2, the vertices \( x, y \) are contained in less than \( c'n \) total copies of \( C_k \) in \( G \).

Now let us apply Claim 18 repeatedly in the following way. Let \( C_0 \) be the collection of all \( cn \) copies of \( C_k \) in \( G \). We may apply Claim 18 to \( C_0 \) to find a \( Q_k' \)-graph \( Q_1 \) with main vertices \( x_1, y_1 \) at distance \( \ell_1 \) in \( Q_1 \). Now remove from \( C_0 \) the copies of \( C_k \) that contain \( x, y \) at distance \( \ell_1 \) and let \( C_1 \) be the remaining copies of \( C_k \) in \( C_0 \). Note that \( |C_1| \geq (c-c')n \) and that none of the copies of \( C_k \) in \( Q_1 \) are present in \( C_1 \). Repeating the argument above on \( C_1 \) in place of \( C_0 \) gives another \( Q_k' \)-graph \( Q_2 \) with main vertices \( x_2, y_2 \). We can continue this argument until we have \( k\gamma \) different \( Q_k' \)-graphs (as \( c \) is large enough compared to \( c' \)).

A pair of vertices \( x, y \) can appear as main vertices in at most \( k \) of the graphs \( Q_1, Q_2, \ldots, Q_{k\gamma} \). Indeed, as once they appear as main vertices at distance \( \ell \leq k \) in some \( Q_k' \)-graph we remove all copies of \( C_k \) that have \( x, y \) at distance \( \ell \). Therefore, there is a collection of \( \gamma = k\gamma/k \) different \( Q_k' \)-graphs such that no two of the \( Q_k' \)-graphs have the same two main vertices.

Let \( Q_1', Q_2', \ldots, Q_{\gamma}' \) be this collection of \( Q_k' \)-graphs.

The internal paths of any \( Q_i' \) may share vertices with \( Q_j' \) (for \( j \neq i \)). However, for \( r' \) large enough compared to \( r'' \), we may remove internal paths from each of the \( Q_i' \)'s to construct a collection of \( Q_k'' \)-graphs \( Q_1'', Q_2'', \ldots, Q_{\gamma}'' \) such that any two graphs \( Q_i'' \) and \( Q_j'' \) only share vertices on their respective main paths (for \( i \neq j \)).

Now let \( M_1, M_2, \ldots, M_{\gamma} \) be the collection of main paths of the \( Q_k'' \)-graphs \( Q_1'', Q_2'', \ldots, Q_{\gamma}'' \). If there are two paths \( M_i \) and \( M_j \) that share at most one vertex, then we may apply Lemma 15 to \( Q_i'' \) and \( Q_j'' \) to find a copy of \( F \) in \( G \); a contradiction.

So we may assume that each \( M_i \) shares at least two vertices with each other \( M_j \). Recall that they can share at most one of their main vertices. Therefore, there is a vertex \( u \in M_1 \) that is contained in at least \( \gamma/k \) of the paths \( M_2, M_3, \ldots, M_{\gamma} \). Moreover, \( u \) is the \( i \)th vertex in at least \( \gamma/k^2 \) of those paths. Each of these paths contain another vertex from \( M_1 \). At least \( \gamma/k^3 \) of them contains the same vertex \( v \), and it is the \( j \)th vertex in at least \( \gamma/k^4 \) of them. Thus, there are at least \( \gamma/k^4 \) vertices \( u \) and \( v \) such that the \( j \)th vertex of \( v \) is contained in at least \( \gamma/k^4 \) of them. As \( \gamma/k^4 \) is large enough we may apply Lemma 16 to this collection of \( u-v \) paths of length \( j-i \) to get a subgraph \( B \) consisting of a banana graph \( B'_t \) (for some \( t < k \)) with main vertices \( u, v \) together with a \( u-u' \) path and a \( v-v' \) path. Each \( u-v \) path of \( B \) is a sub-path of some \( Q_i'' \). Pick any such \( Q_i'' \) and take its union with \( B \). The vertices of \( B \) intersect at most \( kr \) internal paths of \( Q_i'' \). As \( r'' \) is large enough compared to \( r \), we may remove internal paths of \( Q_i'' \) that intersect the vertices of \( B \) to get a graph containing an \( R_k^r \)-graph. As \( F \) is a subgraph of every \( R_k^r \)-graph, we have that \( G \) contains \( F \); a contradiction.
5 Maximizing copies of $H$ in $K_k$-free graphs

In [1] it is shown that if we forbid $K_k$ and want to maximize the number of copies of some $(k-1)$-partite graph $H$, then the graph with the maximum number of copies of $H$ is itself a complete $(k-1)$-partite graph, but it is not necessarily the Turán graph $T_{k-1}(n)$. A theorem of Ma and Qiu [31] shows that the Turán graph gives the maximum if the $(k-1)$-partite graph has $k-2$ parts of size $s$ and one part of size $t$ with $s \leq t < s + 1/2 + \sqrt{2s + 1}/4$ and $n$ is large enough.

Now we investigate for which graphs $H$ the function $\text{ex}(n, H, K_k)$ is maximized by the Turán graph $T_{k-1}(n)$. A graph $H$ is $k$-Turán-good if $\text{ex}(n, H, K_k) = \mathcal{N}(H, T_{k-1}(n))$ for every $n$. Theorem [1] shows that complete graphs are $k$-Turán-good for any $k$.

**Lemma 19.** Let $H$ be a $k$-Turán-good graph. Let $H'$ be any graph constructed from $H$ in the following way. Choose a complete subgraph of $H$ with vertex set $X$, add a vertex-disjoint copy of $K_{k-1}$ to $H$ and join the vertices in $X$ to the vertices of $K_{k-1}$ by edges arbitrarily. Then $H'$ is $k$-Turán-good.

**Proof.** By Theorem [1] the maximum number of copies of $K_{k-1}$ in a $K_k$-free graph is achieved by the Turán graph $T_{k-1}(n)$. Since $H$ is $k$-Turán-good, the Turán graph $T_{k-1}(n-k+1)$ has the maximum number of copies of $H$ among $K_k$-free graphs on $n-k+1$ vertices. We will show that $T_{k-1}(n)$ has the maximum number of copies of $H'$.

Let $G$ be a $K_k$-free graph on $n$ vertices with the maximum number of copies of $H'$. Since $H'$ contains a copy of $K_{k-1}$, the graph $G$ must contain a copy of $K_{k-1}$. Let $K$ be this copy of $K_{k-1}$ in $G$. Every other vertex of $G$ is adjacent to at most $k-2$ vertices of $K$. Let $Y$ be a complete graph that is disjoint from $K$.

Consider a bipartite graph with classes formed by the vertices of $Y$ and $K$, respectively and join two vertices by an edge if they are non-adjacent in $G$.

Suppose this bipartite graph does not have a matching saturating the class $Y$, i.e., a matching that uses every vertex of $Y$. Then, by Hall’s theorem, there exists a non-empty subset $Y'$ of $Y$ whose neighborhood in $K$ has size less than $|Y'|$. In the original graph $G$ this means that all of the vertices in $Y'$ are connected to a fixed set of more than $|K| - |Y'|$ vertices in $K$. As $Y'$ and $K$ are complete graphs, this gives a copy of $K_k$ in $G$; a contradiction. Therefore, this bipartite graph has a matching saturating $Y$ which implies that in $G$ the edges between $Y$ and $K$ are a subgraph of a complete bipartite graph minus a matching saturating $Y$.

On the other hand, in a $(k-1)$-partite Turán graph the edges between $K_{k-1}$ and a clique of size $|Y|$ form a complete bipartite graph minus a matching saturating the clique of size $|Y|$. This implies that there are at least as many ways to join the vertices of a copy of $H$ with a copy of $K_{k-1}$ in a Turán graph as in $G$.

The number of copies of $H'$ is the number of copies of $K_{k-1}$, multiplied by the number of copies of $H$ on the remaining $n-k+1$ vertices, multiplied by the number of ways to join the vertices of $K_{k-1}$ and $H$, divided by how many times a copy of $H'$ was counted. The first three quantities are maximized by the Turán graph, while the last quantity depends only on $H'$. This implies that the number of copies of $H'$ is maximized by $T_{k-1}(n)$. \qed
We do not characterize the graphs that can be built this way from complete graphs. Instead we give three simple consequences.

**Corollary 20.**
1. Every path is 3-Turán good.
2. Every Turán graph $T_{k-1}(\ell)$ is $k$-Turán-good.
3. The cycle $C_4$ is 3-Turán good.

We conclude this section with a simple proposition.

**Proposition 21.** The path $P_3$ is $k$-Turán-good.

Proof. Fix a graph $G$ and let $a$ be the number of induced copies of $P_3$. Let us count the number of pairs $(e,v)$ where $e$ is an edge in $G$ and $v$ is a vertex in $G$ that is disjoint from $e$. Clearly, there are $e(G)(n-2)$ such pairs. On the other hand, on any set of three vertices there is at most one triangle or one induced $P_3$ and each triangle consists of three such pairs $(e,v)$ and every induced $P_3$ consists of two such pairs $(e,v)$. Thus

$$2a + 3N(K_3,G) \leq e(G)(n-2).$$

(7)

If $G$ is a complete multi-partite graph, then we have equality in (7) as for any edge $e$ and disjoint vertex $v$, there is at least one edge incident to $e$ and $v$.

Now suppose that $G$ is an $n$-vertex $K_k$-free graph with the maximum number of copies of $P_3$. By Turán’s theorem, we have that $e(G) \leq e(T_{k-1}(n))$ and by Theorem 1 we have that $N(K_3,G) \leq N(K_3,T_{k-1}(n))$. Counting copies of $P_3$ in $G$ we have

$$N(P_3,G) = a + 3N(K_3,G) = (a + \frac{3}{2}N(K_3,G)) + \frac{3}{2}N(K_3,G)$$

$$\leq (a + \frac{3}{2}N(K_3,G)) + \frac{3}{2}N(K_3,T_{k-1}(n))$$

$$\leq \frac{1}{2}e(G)(n-2) + \frac{3}{2}N(K_3,T_{k-1}(n))$$

$$\leq \frac{1}{2}e(T_{k-1}(n))(n-2) + \frac{3}{2}N(K_3,T_{k-1}(n)) = N(P_3,T_{k-1}(n)).$$

\[\square\]

6 Connection to Berge-hypergraphs

The problem of counting copies of a graph $H$ in an $n$-vertex $F$-free graph is closely related to the study of so-called Berge hypergraphs. Generalizing the notion of hypergraph cycles due to Berge, the authors introduced the notion of Berge copies of any graph. Let $F$ be a graph. We say that a hypergraph $\mathcal{H}$ is a Berge-$F$ if there is a bijection $f : E(F) \rightarrow E(\mathcal{H})$ such that $e \subseteq f(e)$ for every $e \in E(F)$. Note that Berge-$F$ actually denotes a class of
hypergraphs. The maximum number of hyperedges in an \( n \)-vertex hypergraph with no sub-hypergraph isomorphic to any Berge-\( F \) is denoted \( \text{ex}(n, \text{Berge-} F) \). When we restrict ourselves to \( r \)-uniform hypergraphs, this maximum is denoted \( \text{ex}_r(n, \text{Berge-} F) \).

Results of Győri, Katona and Lemons [21] together with Davoodi, Győri, Methuku and Tompkins [5] give tight bounds on \( \text{ex}_r(n, \text{Berge-} P_\ell) \). Upper-bounds on \( \text{ex}_r(n, \text{Berge-} C_\ell) \) are given by Győri and Lemons [22] when \( r \geq 3 \). A brief survey of Turán-type results for Berge-hypergraphs can be found in the introduction of [17].

An early link between counting subgraphs and Berge-hypergraph problems was established by Bollobás and Győri [4] who investigated both \( \text{ex}_3(n, \text{Berge-} C_5) \) and \( \text{ex}(n, K_3, C_5) \). The connection between these two parameters is also examined in two recent manuscripts [15, 33]. In this section we prove two new relationships between these problems.

**Proposition 22.** Let \( F \) be a graph. Then

\[
\text{ex}(n, K_r, F) \leq \text{ex}_r(n, \text{Berge-} F) \leq \text{ex}(n, K_r, F) + \text{ex}(n, F).
\]

and

\[
\text{ex}(n, \text{Berge-} F) = \max_G \left\{ \sum_{i=0}^{n} N(K_i, G) \right\} \leq \sum_{i=0}^{n} \text{ex}(n, K_i, F)
\]

where the maximum is over all \( n \)-vertex \( F \)-free graphs \( G \).

**Proof.** Given an \( F \)-free graph \( G \), let us construct a hypergraph \( \mathcal{H} \) on the vertex set of \( G \) by replacing each clique of \( G \) by a hyperedge containing exactly the vertices of that clique. The hypergraph \( \mathcal{H} \) contains no copy of a Berge-\( F \). This gives \( \text{ex}(n, K_r, F) \leq \text{ex}_r(n, \text{Berge-} F) \) and

\[
\max_G \left\{ \sum_{i=0}^{n} N(K_i, G) \right\} \leq \text{ex}(n, \text{Berge-} F)
\]

where the maximum is over all \( n \)-vertex \( F \)-free graphs \( G \).

Given an \( n \)-vertex hypergraph \( \mathcal{H} \) with no Berge-\( F \) subhypergraph, we construct a graph \( G \) on the vertex set of \( \mathcal{H} \) as follows. Consider an order \( h_1, \ldots, h_k \) of the hyperedges of \( \mathcal{H} \) such that the hyperedges of size two appear first. We proceed through the hyperedges in order and at each step try to choose a pair of vertices in \( h_i \) to be an edge in \( G \). If no such pair is available, then each pair of vertices in \( h_i \) is already adjacent in \( G \). In this case, we add no edge to \( G \). A copy of \( F \) in \( G \) would correspond exactly to a Berge-\( F \) in \( \mathcal{H} \), so \( G \) is \( F \)-free.

For each hyperedge \( h_i \) where we did not add an edge to \( G \), there is a clique on the vertices of \( h_i \) in \( G \). Thus, the number of hyperedges of \( \mathcal{H} \) is at most the number of cliques in \( G \). If \( \mathcal{H} \) is \( r \)-uniform, then each hyperedge \( h_i \) of \( \mathcal{H} \) corresponds to either an edge in \( G \) or a clique \( K_r \) on the vertices of \( h_i \) (when we could not add an edge to \( G \)). Therefore, the number of hyperedges in \( \mathcal{H} \) is at most \( \text{ex}(n, K_r, F) \) + \( \text{ex}(n, F) \). \( \square \)

As in the case of traditional Turán numbers we may forbid multiple hypergraphs. In particular, let \( \text{ex}_r(n, \{\text{Berge-} F_1, \text{Berge-} F_2, \ldots, \text{Berge-} F_k\}) \) denote the maximum number of
hyperedges in an $r$-uniform $n$-vertex hypergraph with no subhypergraph isomorphic to any Berge-$F_i$ for all $1 \leq i \leq k$. Similarly, $\text{ex}(n, H, \{F_1, F_2, \ldots, F_k\})$ denotes the maximum number of copies of the graph $H$ in an $n$-vertex graph that contains no subgraph $F_i$ for all $1 \leq i \leq k$.

**Proposition 23.** Let $k \geq 4$. Then

$$\text{ex}_3(n, \{\text{Berge-}C_2, \ldots, \text{Berge-}C_k\}) = \text{ex}(n, K_3, \{C_4, \ldots, C_k\}).$$

**Proof.** Let $\mathcal{H}$ be an $n$-vertex 3-uniform hypergraph with no Berge-$C_i$ for $i = 2, 3, \ldots, k$ and the maximum number of hyperedges. Consider the graph $G$ on the vertex set of $\mathcal{H}$ where a pair of vertices are adjacent if and only if they are contained in a hyperedge of $\mathcal{H}$. As $\mathcal{H}$ is $C_2$-free (i.e., each pair of hyperedges share at most one vertex) each edge of $G$ is contained in exactly one hyperedge of $\mathcal{H}$.

Each hyperedge of $\mathcal{H}$ contributes a triangle to $G$. We claim that $G$ contains no other cycles of length $i$ for $i = 3, 4, 5, \ldots, k$. That is, $G$ contains no cycle with two edges coming from different hyperedges of $\mathcal{H}$. Suppose (to the contrary) that $G$ does contain such a cycle $C$. If two edges of $C$ come from the same hyperedge, then they are incident in $C$. Therefore, these two edges can be replaced by the edge between their disjoint endpoints (which is contained in the same hyperedge) to get a shorter cycle. We may repeat this process until we are left with a cycle such that each edge comes from a different hyperedge of $\mathcal{H}$. Then this cycle corresponds exactly to a Berge-cycle of at most $k$ hyperedges in $\mathcal{H}$; a contradiction. Thus, $\text{ex}_3(n, \{\text{Berge-}C_2, \ldots, \text{Berge-}C_k\}) \leq \text{ex}(n, K_3, \{C_4, \ldots, C_k\})$.

On the other hand, let $G$ be an $n$-vertex graph with no cycle $C_4, C_5, \ldots, C_k$ and the maximum number of triangles. Construct a hypergraph $\mathcal{H}$ on the vertex set of $G$ where the hyperedges of $\mathcal{H}$ are the triangles of $G$. The graph $G$ is $C_4$-free, so each pair of triangles share at most one vertex, i.e., $\mathcal{H}$ contains no Berge-$C_2$. If $\mathcal{H}$ contains a Berge-$C_3$, then it is easy to see that $G$ contains a $C_4$; a contradiction.

Therefore, if $\mathcal{H}$ contains any Berge-$C_i$ for $i = 4, \ldots, k$, then $G$ contains a cycle $C_i$; a contradiction. Thus, $\text{ex}_3(n, \{\text{Berge-}C_2, \ldots, \text{Berge-}C_k\}) \geq \text{ex}(n, K_3, \{C_4, \ldots, C_k\})$. \hfill \QED

Alon and Shikhelman [11] showed that for every $k > 3$, $\text{ex}(n, K_3, \{C_4, \ldots, C_k\}) \geq \Omega(n^{1 + \frac{1}{k-1}})$. For $k = 4$ they showed that $\text{ex}_3(n, K_3, C_4) = (1 + o(1)) \frac{1}{6} n^{3/2}$.
Lazebnik and Verstraëte [29] proved $\text{ex}_3(n, \{\text{Berge-}C_2, \text{Berge-}C_3, \text{Berge-}C_4\}) = (1 + o(1)) \frac{1}{6} n^{3/2}$. By Proposition 23 these two statements are equivalent.

**References**

[1] N. Alon, C. Shikhelman, Many $T$ copies in $H$-free graphs. *Journal of Combinatorial Theory, Series B* 121 (2016) 146–172.

[2] B. Bollobás, On complete subgraphs of different orders. *Math. Proc. Cambridge Philos. Soc.* 79 (1976) 19–24.
[3] B. Bollobás, Modern Graph Theory, Graduate Texts in Mathematics, 184. Springer-Verlag, New York, 1998. xiv+394 pp. ISBN: 0-387-98488-7

[4] B. Bollobás, E. Győri, Pentagons vs. triangles. Discrete Mathematics 308 (2008) 4332-4336.

[5] A. Davoodi, E. Győri, A. Methuku, C. Tompkins, An Erdős-Gallai type theorem for hypergraphs. arXiv:1608.03241 (2016).

[6] P. Erdős, On the number of complete subgraphs contained in certain graphs. Magyar Tud. Akad. Mat. KutatóInt. Közl. 7 (1962) 459–464.

[7] P. Erdős, Problems and results in graph theory and combinatorial analysis, Proceedings of the Fifth British Combinatorial Conference (Univ. Aberdeen, Aberdeen, 1975), Congress. Numer. XV , pp. 169–192, Utilitas Math., Winnipeg, Man., 1976 MR53 #13006; Zentralblatt 335.05002.

[8] 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 (1984) 1–17.

[9] P. Erdős, A. Rényi, On the evolution of random graphs. Magyar Tud. Akad. Mat. KutatóInt. Közl. 5 (1960) 17–61.

[10] B. Ergemlidze, E. Győri, A. Methuku, N. Salia, A note on the maximum number of triangles in a $C_5$-free graph. arXiv:1706.02830, 2017.

[11] Z. Füredi. New asymptotics for bipartite Turán numbers. Journal of Combinatorial Theory, Series A, 75 (1996) 141–144.

[12] Z. Füredi, M. Simonovits, The history of degenerate (bipartite) extremal graph problems. Erdős Centennial. Springer, Berlin, Heidelberg, (2013) 169–264.

[13] D. Gerbner, E. Győri, A. Methuku, M. Vizer, Generalized Turán problems for even cycles. arXiv:1712.07079 (2017).

[14] D. Gerbner, B. Keszegh, C. Palmer, B. Patkós, On the number of cycles in a graph with restricted cycle lengths. SIAM Journal on Discrete Mathematics 32 (2018) 266–279.

[15] D. Gerbner, A. Methuku, M. Vizer, Asymptotics for the Turán number of Berge-$K_{2,t}$. arXiv:1705.04134 (2017).

[16] D. Gerbner, A. Methuku, M. Vizer, Generalized Turán problems for disjoint copies of graphs. arXiv:1712.07072 (2017).

[17] D. Gerbner, C. Palmer, Extremal Results for Berge Hypergraphs. SIAM Journal on Discrete Mathematics 31 (2017) 2314–2327.
[18] L. Gishboliner, A. Shapira, A Generalized Turán Problem and its Applications. arXiv:1712.00831 (2017).

[19] A. Grzesik, On the maximum number of five-cycles in a triangle-free graph. Journal of Combinatorial Theory, Series B, 102 (2012) 1061–1066.

[20] E. Győri, On the number of $C_5$’s in a triangle-free graph. Combinatorica 9 (1989) 101–102.

[21] E. Győri, G. Y. Katona, N. Lemons, Hypergraph extensions of the Erdős-Gallai Theorem, European Journal of Combinatorics 58 (2016) 238–246.

[22] E. Győri, N. Lemons, Hypergraphs with no cycle of a given length. Combinatorics, Probability and Computing 21 (2012) 193–201.

[23] E. Győri, H. Li, The maximum number of triangles in $C_{2k+1}$-free graphs. Combinatorics, Probability and Computing 21 (2011) 187–191.

[24] H. Hatami, J. Hladký, D. Král’, D. Norine, A. Razborov. On the number of pentagons in triangle-free graphs. Journal of Combinatorial Theory, Series A 120 (2012) 722–732.

[25] G.O.H. Katona, A theorem for finite sets. Theory of Graphs, Proc. Colloq., Tihany, 1966, Academic Press, New York, (1968) 187–207.

[26] A. Kostochka, D. Mubayi and J. Verstraëte, Turán problems and shadows III: expansions of graphs, SIAM Journal on Discrete Math. 29 (2015) 868-876.

[27] T. Kővári, V. Sós, P. Turán, On a problem of K. Zarankiewicz. In Colloquium Mathematicae Vol. 3, No. 1, (1954) 50–57.

[28] J. Kruskal. The optimal number of simplices in a complex. Mathematical Optimization Techniques (1963) 251–268.

[29] F. Lazebnik, J. Verstraëte, On hypergraphs of girth five. Electronic Journal of Combinatorics 10 (2003) #R25.

[30] L. Lovász, Combinatorial Problems and Exercises. Akadémiai Kiadó, North Holland, 1979.

[31] Jie Ma and Yu Qiu, Some sharp results on the generalized Turán numbers. arXiv:1802.01091 (2018).

[32] I. W. Moon, L. Moser, On a problem of Turán. Magyar Tud. Akad. Mat. KutatóInt. Közl. 7 (1962) 311–314.

[33] C. Palmer, M. Tait, C. Timmons, A. Z. Wagner, Turán numbers for Berge-hypergraphs and related extremal problems. arXiv:1706.04249 (2017)