The maximum number of P_ℓ copies in P_k-free graphs

Ervin Győri[1*]   Nika Salia[1,2†]   Casey Tompkins[1,3‡]   Oscar Zamora[2,4]

[1] *Alfréd Rényi Institute of Mathematics, Hungarian Academy of Sciences.*
[2] *Central European University, Budapest.*
[3] *Karlsruhe Institute of Technology, Karlsruhe, Germany.*
[4] *Universidad de Costa Rica, San José.*

Generalizing Turán's classical extremal problem, Alon and Shikhelman investigated the problem of maximizing the number of copies of T in an H-free graph, for a pair of graphs T and H. Whereas Alon and Shikhelman were primarily interested in determining the order of magnitude for some classes of graphs H, we focus on the case when T and H are paths, where we find asymptotic and exact results in some cases. We also consider other structures like stars and the set of cycles of length at least k, where we derive asymptotically sharp estimates. Our results generalize well-known extremal theorems of Erdős and Gallai.

Keywords: path, cycle, extremal, generalized Turán

1   Introduction

For a graph G, we let $e(G)$ denote the number of edges in G, and for a given graph H, we let $\mathcal{N}(H, G)$ denote the number of (not necessarily induced) copies of H in G. If there is no copy of H in G, we say that G is H-free. We denote the path with k edges by P_k and the cycle with k edges by C_k. By $C_{\geq k}$ we mean the set of all cycles of length at least k. By S_k we denote the star on $k+1$ vertices. Given a graph G containing a vertex v, we denote the neighborhood of v by $N(v)$. The independence number, minimum degree and number of vertices in G are denoted by $\alpha(G)$, $\delta(G)$ and $v(G)$, respectively. The vertex and edge sets of G are denoted by $V(G)$ and $E(G)$, respectively. Finally, given a set $S \subseteq V(G)$, we denote by $G[S]$ the induced subgraph of G with vertex set S.

Following the notation of Alon and Shikhelman (2016), we let $\mathrm{ex}(n, T, H)$ be the maximum number of (noninduced) copies of T in an H-free graph on n vertices. Observe that we have $\mathrm{ex}(n, P_1, H) =$
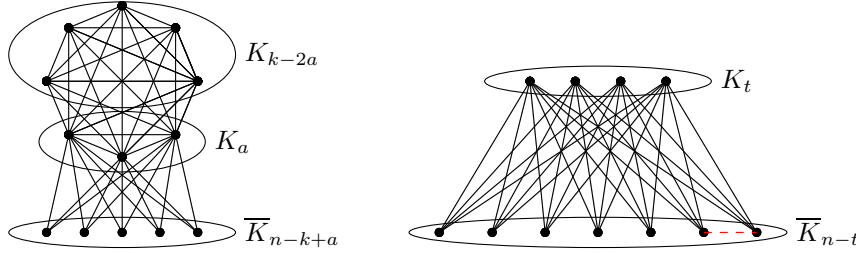
Fig. 1: The graph $G_{n,k,a}$ is pictured on the left, and the special case of $G_{n,k,t}$ is pictured on the right. The dashed edge appears only when k is even.

$\text{ex}(n, H)$, the classical extremal number. If a set of graphs $\mathcal{H}$ is forbidden, then we define $\text{ex}(n, \mathcal{H})$ (and similarly $\text{ex}(n, T, \mathcal{H})$), in the obvious way.

We begin by recalling the famous theorem of Erdős and Gallai on P_k-free graphs as well as some recent generalizations due to Luo, where the number of cliques is considered.

Theorem 1 (Erdős and Gallai (1959))**.** *For all $n \geq k$,*

$$\text{ex}(n, P_k) \leq \frac{(k-1)n}{2},$$

and equality holds if and only if k divides n and G is the disjoint union of cliques of size k.

In their paper, Erdős and Gallai deduced Theorem 1 as a corollary of the following result about graphs with no long cycles.

Theorem 2 (Erdős and Gallai (1959))**.** *For all $n \geq k$,*

$$\text{ex}(n, C_{\geq k}) \leq \frac{(k-1)(n-1)}{2},$$

and equality holds if and only if $k - 2$ divides $n - 1$ and G is a connected graph such that every block of G is a clique of size $k - 1$.

As the extremal examples for Theorem 1 are disconnected, it is natural to consider a version of the problem where the base graph is assumed to be connected. Kopylov (1977) settled this problem, and later Balister et al. (2008) classified the extremal cases.

Definition 1. *We denote by $G_{n,k,a}$ the graph whose vertex set is partitioned into 3 classes, A, B and C with $|A| = a$, $|B| = n - k + a$, $|C| = k - 2a$ such that $A \cup C$ induces a clique, B is an independent set and all possible edges are taken between vertices of A and B. (See Figure 1.)*

Throughout this paper we let $t = \left\lfloor \frac{k-1}{2} \right\rfloor$. In $G_{n,k,t}$, the class C has one vertex when k is odd or two vertices when k is even. By grouping B and C together, we have that $G_{n,k,t}$ is obtained from a complete bipartite graph $K_{t,n-t}$ by adding all edges in the color class of size t, and in the case that k is even, adding one additional edge inside the color class of size $n - t$.

Theorem 3 (Kopylov (1977), Balister et al. (2008))**.** *Let G be a connected n-vertex P_k-free graph, with $n \geq k$, then*

$$e(G) \leq \max(e(G_{n,k,t}), e(G_{n,k,1})).$$

Moreover, the extremal graph is either $G_{n,k,t}$ or $G_{n,k,1}$.

Note that if $n \geq 5k/4$, this maximum is achieved by $G_{n,k,t}$. Also observe that

$$e(G_{n,k,t}) = t(n-t) + \binom{t}{2} + \eta_k,$$

where η_k is 1, if k is even, and 0 otherwise. Thus, Theorems 1 and 3 yield the same bound asymptotically as n tends to infinity.

The following theorem was deduced by Luo (2018) as a corollary of her main result but also follows from Theorem 1 using a simple induction argument. We present this proof here.

Theorem 4 (Luo (2018))**.**

$$\mathrm{ex}(n, K_r, P_k) \leq \frac{n}{k}\binom{k}{r}.$$

Proof: We use induction on r, and the base is Theorem 1. Let G be an n-vertex graph containing no P_k. We have

$$\mathcal{N}(K_r, G) = \frac{1}{r}\sum_{v \in V(G)} \mathcal{N}(K_{r-1}, G[N(v)]) \leq \frac{1}{r}\sum_{v \in V(G)} \frac{v(G[N(v)])}{k-1}\binom{k-1}{r-1} = \frac{1}{k(k-1)}\binom{k}{r}2e(G),$$

since $G[N(v)]$ contains no P_{k-1}. By Theorem 1, we have $e(G) \leq \frac{(k-1)n}{2}$, and the result follows. $\square$

For our results we will need only that $\mathrm{ex}(n, K_r, P_k) \leq c_{k,r}n$ for some constant $c_{k,r}$ depending only on k and r.

If we impose the additional condition that the graph is connected, then the situation is more complicated. Luo proved the following sharp bounds.

Theorem 5 (Luo (2018))**.** *Let $n > k \geq 3$ and G be a connected n-vertex graph with no path of length k, then*

$$\mathcal{N}(K_r, G) \leq \max\left(\mathcal{N}(K_r, G_{n,k,t}), \mathcal{N}(K_r, G_{n,k,1})\right).$$

Theorem 6 (Luo (2018))**.** *Let $n \geq k \geq 4$ and G be a n-vertex graph with no cycle of length k or greater, then*

$$\mathcal{N}(K_r, G) \leq \frac{n-1}{k-2}\binom{k-1}{r}.$$

Some recent generalizations of the Erdős–Gallai theorem and Luo's results can be found in Ning and Peng (2018). In the present paper we focus on results where paths or all sufficiently long cycles are forbidden. The general problem of enumerating cycles of a fixed length when a fixed cycle is forbidden has also been considered recently (see Gishboliner and Shapira (2018) and Gerbner et al. (2017) which generalize earlier results for special cases, e.g., Bollobás and Győri (2008), Győri and Li (2012), Alon and Shikhelman (2016)).

Alon and Shikhelman (2016) considered the problem of maximizing the number of copies of a tree T in a graph which is H-free, for another tree H. Given two trees T and H, they introduced an integer parameter $m(T, H)$ and proved that $\mathrm{ex}(n, T, H) = \Theta(n^{m(T,H)})$, thereby determining the correct order of magnitude for all pairs of trees. A recent result due to Letzter (2018) extends the above result of Alon

and Shikhelman to the case when only H is a tree and T is arbitrary. It is shown that, nonetheless, the order of magnitude of $\mathrm{ex}(n, T, H)$ is a positive integer power of n.

In the present paper, we are interested in the case where the forbidden tree is a path, and we find correct asymptotics and sometimes the exact bound for the maximum number of copies of a smaller path (as well as for several other types of graphs). We also obtain asymptotic results for the problem of maximizing copies of T in a graph with no cycles of length at least k, in the case when T is a path.

The paper is organized as follows: In Section 2, we determine asymptotically the maximum number of paths and cycles in a P_k-free graph. For the case when k is even we provide a simple proof using a result of Nikiforov on the spectral radius of P_k free graphs. Then, we give more precise estimates which are also sharp in case when k is odd through double-counting arguments. In Section 3, we determine the order of magnitude of $\mathrm{ex}(n, H, T)$ when T is a tree for the class of graphs H which satisfy the condition that $v(H) - \alpha(H) \leq \left\lfloor \frac{k-1}{2} \right\rfloor$. In Section 4, we determine $\mathrm{ex}(n, H, P_k)$ exactly for several graphs H including 4-cycles, stars and short paths. In Section 5, we consider the problem of enumerating copies of P_{k-1} in a P_k-free graph. We determine the asymptotic result for copies of P_5 in a P_6-free graph and pose a general conjecture.

2   Asymptotic Results

We write $f(n, k) \sim g(n, k)$ when $\displaystyle \lim_{k \to \infty} \left(\lim_{n \to \infty} \frac{f(n, k)}{g(n, k)} \right) = 1$. We estimate on the number of copies of paths and cycles in a P_k-free graph. For a fixed $\ell \in \mathbb{N}$, we prove the following asymptotic results:

Theorem 7.
$$\mathrm{ex}(n, P_{2\ell}, P_k) \sim \frac{k^\ell n^{\ell+1}}{2^{\ell+1}}.$$

Theorem 8.
$$\mathrm{ex}(n, P_{2\ell+1}, P_k) \sim \frac{(\ell+2)k^{\ell+1}n^{\ell+1}}{2^{\ell+2}}.$$

Theorem 9.
$$\mathrm{ex}(n, C_{2\ell}, P_k) \sim \frac{k^\ell n^\ell}{\ell 2^{\ell+1}}.$$

Theorem 10.
$$\mathrm{ex}(n, C_{2\ell+1}, P_k) \sim \frac{k^{\ell+1}n^\ell}{2^{\ell+2}}.$$

The construction showing the lower bounds for Theorems 7 through 10 is the same as the extremal construction for the connected version of the Erdős–Gallai theorem, Theorem 3. Because we are interested in asymptotics, we will omit one edge from this construction which only occurs when k is even. Our n-vertex graph G is defined by taking a clique on a set S of $\left\lfloor \frac{k-1}{2} \right\rfloor$ vertices and connecting every vertex in S to every vertex of an independent set U, defined on $n - \left\lfloor \frac{k-1}{2} \right\rfloor$ vertices. It is easy to see that this graph is P_k-free. In enumerating the copies of $P_{2\ell}$, the only paths which contribute asymptotically alternate between S and U, starting and ending with U (the factor of 2 comes from counting the path in both directions).

When enumerating the copies of $P_{2\ell+1}$, we have two kinds of paths which contribute asymptotically: those that start and end in U, using an edge in S at some step, and those that start in U and end in S, never

using an edge contained in S. For the first type, we condition on which step in the path we use the edge in S (ℓ possibilities). Each such path gets counted twice, hence we divide by two. For the second type, each path is counted once and so we do not have to divide by 2.

We begin by showing how Theorem 7 can be derived from a result about the spectral radius of P_k-free graphs due to Nikiforov (2010). Recall that the spectral radius of a graph G is the maximum of the eigenvalues of the adjacency matrix of G. He determined, for sufficiently large n, the maximal spectral radius of a P_k-free graph on n vertices. We are interested in asymptotics so we will make use of the following corollary which follows directly from the results in Nikiforov (2010).

Corollary 1 (Nikiforov (2010))**.** *If n is sufficiently large and G is a P_k-free graph, then the spectral radius of G is at most $\sqrt{\lfloor (k+1)/2 \rfloor n}$.*

Proof of Theorem 7 using spectral bounds: Let G be a P_k-free graph on n vertices (for n large enough to satisfy Corollary 1). Let A be the adjacency matrix of G, then we have

$$2 \cdot \frac{\mathcal{N}(P_{2\ell}, G)}{n} \leq \frac{\#2\ell\text{-walks in } G}{n} = \frac{\mathbf{1}^t A^{2\ell} \mathbf{1}}{\mathbf{1}^t \mathbf{1}} \leq \left(\sqrt{\lfloor (k+1)/2 \rfloor n} \right)^{2\ell} = (\lfloor (k+1)/2 \rfloor n)^\ell.$$

Where $\mathbf{1}$ is the all 1's vector, and the second inequality comes from the fact that the spectral radius of a Hermitian matrix M is the supremum of the quotient $\frac{x^* M x}{x^* x}$, where x ranges over $\mathbb{C}^n \backslash \{0\}$. Therefore, for every $k \in \mathbb{N}$ and n sufficiently large we have $\text{ex}(n, P_{2\ell}, k) \leq n^{\ell+1} \lfloor (k+1)/2 \rfloor^\ell / 2$. $\qquad\square$

Unfortunately, it does not seem like this approach can be used to prove Theorem 8 as the bound it would yield is off by a factor of order $\sqrt{n}$.

We will now prove the upper bounds from which Theorems 7 and 8 are immediate consequences. We note that the upper bound we obtain for the $P_{2\ell}$-case is sharper than the proof using the spectral radius yields.

Proposition 1. *Let ℓ, k be positive integers with $2\ell < k$, then*

$$\text{ex}(n, P_{2\ell}, P_k) \leq \frac{k^\ell n^{\ell+1}}{2^{\ell+1}} + O(n^\ell).$$

Proposition 2. *Let ℓ, k be positive integers with $2\ell + 1 < k$, then*

$$\text{ex}(n, P_{2\ell+1}, P_k) \leq \frac{(\ell+2)k^{\ell+1} n^{\ell+1}}{2^{\ell+2}} + O(n^\ell).$$

The proofs of the propositions above will use a double-counting argument involving structures defined using matchings. We will begin by estimating the maximum number of certain kinds of matchings occurring in a P_k-free graph.

Let us define M_1^ℓ, M_2^ℓ and M_3^ℓ to be the following graphs: M_1^ℓ is an $(\ell-1)$-matching together with a disjoint triangle, M_2^ℓ is an $(\ell-1)$-matching together with a disjoint K_4 and M_3^ℓ is an $(\ell-2)$-matching with two independent triangles, disjoint from the matching (see Figure 2).

Lemma 1. *The number of copies of M_1^ℓ, M_2^ℓ and M_3^ℓ in an n-vertex P_k-free graph is $O(n^\ell)$.*

Proof: Let G be a P_k-free graph on n vertices. By Theorem 4, the number of triangles in G is at most $O(k^2 n)$. By Theorem 1 the total number of edges in G is at most $(k-1)n/2$. It follows that the number of copies of M_1^ℓ is bounded from above by

$$\binom{\frac{kn}{2}}{\ell - 1} k^2 n = O(k^{\ell+1} n^\ell).$$

The proofs of the bound for M_2^ℓ and M_3^ℓ are similar. $\square$

Proof of Theorem 7: Let G be a P_k-free graph on n vertices. We will consider structures consisting of a matching of ℓ edges and a vertex not contained in these edges. Namely, a *matching structure* is an $(\ell + 1)$-tuple $(e_1, e_2, \ldots, e_\ell, v)$ where $\{e_1, e_2, \ldots, e_\ell\}$ is a matching in G and $v \in V(G) \setminus \cup_{i=1}^\ell e_i$. We say that a path $P_{2\ell}$ *aligns* with a matching structure $(e_1, e_2, \ldots, e_\ell, v)$ if its edges are (consecutively) $e_1, f_1, e_2, f_2, \ldots, e_\ell, f_\ell$ where $v \in f_\ell$. We say that the matching structure *spans* the set of vertices $\cup_{i=1}^\ell e_i \cup \{v\}$.

Let $\mathcal{A} := \{S \subseteq V : |S| = 2\ell + 1, M_1^\ell \subseteq G[S]\}$. By Lemma 1, we have $|\mathcal{A}| = O(n^\ell)$. Let $\mathcal{M}$ be the set of all the matching structures which span a set of vertices not contained in $\mathcal{A}$.

Claim 1. *At most one $P_{2\ell}$ aligns with each matching structure in $\mathcal{M}$.*

Proof: Let $(e_1, e_2, \ldots, e_\ell, v)$ be a matching structure in $\mathcal{M}$ and fix a $P_{2\ell}$ which aligns with it, say $a_1, b_1, a_2, b_2 \ldots, a_\ell, b_\ell, a_{\ell+1}$, where $e_i = \{a_i, b_i\}$ and $v = a_{\ell+1}$. There is no edge from a_i to a_{i+1}, otherwise $e_1, e_2, \ldots, e_{i-1}, \{b_{i+1}, a_{i+2}\}, \{b_{i+2}, a_{i+3}\}, \ldots, \{b_\ell, a_{\ell+1}\}$ together with the triangle $\{a_i, b_i, a_{i+1}\}$ forms an M_1^ℓ. Since there is a unique $P_{2\ell}$ spanning the matching structure and not containing an edge $\{a_i, a_{i+1}\}$, the claim is proved. (See Figure 3.) $\square$

Next, we observe that for every $P_{2\ell}$, there are precisely two matching structures for which that $P_{2\ell}$ is aligned. Indeed, let the vertices of the $P_{2\ell}$ be traversed in the order $v_1, v_2, \ldots, v_{2\ell+1}$, then the two matching structures with which the $P_{2\ell}$ aligns are

$$(\{v_1, v_2\}, \{v_3, v_4\}, \ldots, \{v_{2\ell-1}, v_{2\ell}\}, v_{2\ell+1}) \text{ and } (\{v_{2\ell+1}, v_{2\ell}\}, \{v_{2\ell-1}, v_{2\ell-2}\}, \ldots, \{v_3, v_2\}, v_1).$$

It follows that the if we define $M := |\mathcal{M}|$, then the number of copies of $P_{2\ell}$ is bounded from above by $M/2 + O(n^\ell)$.

By Theorem 1, the number of edges in G is at most $(k-1)n/2$. A matching structure is formed by choosing ℓ edges in order followed by an additional vertex. Thus, we have the following upper bound on
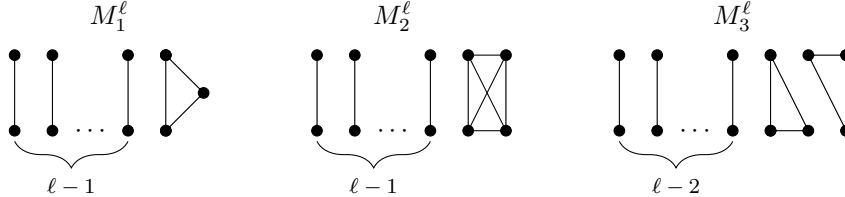

Fig. 2: Matching structures with negligible contribution.

the number of matching structures in $\mathcal{M}$:

$$M \le \binom{\frac{nk}{2}}{\ell} \ell!\, n.$$

Dividing by 2 yields the required bound on the number of copies of $P_{2\ell}$.   □

Proof of Theorem 8: We will now define matching structures in a slightly different way. A *matching structure* is an $(\ell + 1)$-tuple $(e_1, e_2, \ldots, e_{\ell+1})$, where $\{e_1, e_2, \ldots, e_{\ell+1}\}$ is a matching in G. A path $P_{2\ell+1}$ *aligns* with a matching structure $(e_1, e_2, \ldots, e_{\ell+1})$ if its edges are $e_1, f_1, e_2, f_2, \ldots, e_\ell, f_\ell, e_{\ell+1}$, consecutively.

     Let $\mathcal{B} := \{S \subseteq V : |S| = 2\ell + 2, M_2 \subseteq G[S]\}$ and $\mathcal{C} := \{S \subseteq V : |S| = 2\ell + 2, M_3 \subseteq G[S]\}$. By Lemma 1, we have $|\mathcal{B}| = O(n^\ell)$ and $|\mathcal{C}| = O(n^\ell)$. Let $\mathcal{M}$ be the set of matching structures which do not span a vertex set in $\mathcal{B}$ or $\mathcal{C}$.

Claim 2. *There are at most $\ell + 2$ copies of $P_{2\ell+1}$ which align with each matching structure in $\mathcal{M}$.*

Proof: Consider a matching structure $(e_1, e_2, \ldots, e_{\ell+1}) \in \mathcal{M}$. We will consider the edges in the matching structure one by one and show that we can label the vertices of each edge e_j with a_j and b_j in such a way that there is no edge between a_j and a_{j+1}. Thus, every path which aligns with the matching structure will be a subgraph of the graph pictured (on the top) in Figure 4. Given that the matching structure has this form, we may easily upper bound the number of copies of $P_{2\ell+1}$ which can align with it. Indeed, if the $P_{2\ell+1}$ starts with the vertex b_1, there is at most one such path: $b_1, a_1, b_2, a_2, \ldots, b_{\ell+1}, a_{\ell+1}$. If it starts with the vertex a_1, then for at most one i, $1 \le i \le \ell$, the path may use an edge $\{b_i, b_{i+1}\}$; all other choices are forced. Thus, in total there are at most $1 + (\ell + 1) = \ell + 2$ paths which align with such a matching structure. We now prove that the desired labeling of the edges exists.

     We may suppose there is at least one edge from e_i to e_{i+1} for all $i = 1, 2, \ldots, \ell$, otherwise no $P_{2\ell+1}$ aligns with the matching structure. We also know $e_i \cup e_{i+1}$ does not induce a K_4, so there is at least one edge missing among these 4 vertices. Now we may define $e_1 = \{a_1, b_1\}$ in such a way that there is at least one edge missing from a_1 to e_2. Define $e_2 = \{a_2, b_2\}$ such that there is no edge between a_1 and a_2. In general, suppose we have already labeled the edges $e_1, e_2, \ldots, e_j$ in such a way that for $i \in \{1, 2, \ldots, j - 1\}$, a_i is not connected to a_{i+1}. We will show that e_{j+1} can be labeled by a_{j+1} and b_{j+1} such that there is no edge between a_j and a_{j+1}.

     We know there is an edge missing from e_j to e_{j+1}. If there is an edge missing between a_j and e_{j+1}, then define $e_{j+1} = \{a_{j+1}, b_{j+1}\}$ so that there is an edge from a_j to a_{j+1}. Otherwise $\{a_j\} \cup e_{j+1}$ forms a triangle. In this case, there is an edge missing from b_j to e_{j+1}; label $e_{j+1} = \{a_{j+1}, b_{j+1}\}$ so that b_j is not adjacent to b_{j+1}.
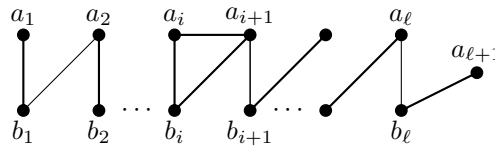

Fig. 3: Matching structure from the proof of Claim 1.

Now if we do not have an edge from a_{j-1} to b_j, then we switch the labels on e_j and e_{j+1}, and we are done. (By switching the labels we mean that the vertex in e_i previously labeled a_i is now labeled b_i, and the vertex previously labeled b_i is now labeled a_i.) Thus, assume we have an edge from a_{j-1} to b_j. Then we have no edge from b_{j-1} to b_j, for this would yield an M_3. Next, consider e_{j-2}. If there is no edge from a_{j-2} to b_{j-1}, then switch the labels on e_{j-1}, e_j and e_{j+1}, and we are done. If there is an edge from a_{j-2} to b_{j-1}, we proceed similarly with e_{j-3}. Continuing this procedure we will reach an edge e_r such that switching the labels of $e_r, e_{r+1}, \ldots, e_{j+1}$ yields no edge between a_i and a_{i+1} for any $1 \le i \le j$. (This procedure is illustrated in Figure 4.) □

We now complete the proof of Theorem 8. Again we set $M := |\mathcal{M}|$. By Theorem 1, there are at most $(k-1)n/2$ total edges in G. Thus,

$$|\mathcal{M}| \le \binom{\frac{nk}{2}}{\ell+1}(\ell+1)!.$$

Since at most $\ell+2$ paths $P_{2\ell+1}$ align with each matching structure from $\mathcal{M}$, and every $P_{2\ell+1}$ aligns with precisely two matching structures. It follows that the total number of copies of $P_{2\ell+1}$ in G is at most

$$\frac{(\ell+2)M}{2} + O(n^\ell) \le \frac{(\ell+2)\binom{\frac{nk}{2}}{\ell+1}(\ell+1)!}{2} + O(n^\ell) = \frac{(\ell+2)k^{\ell+1}n^{\ell+1}}{2^{\ell+2}} + O(n^\ell).$$

□

The lower bound for Theorems 9 and 10 also comes from $G_{n,k,t}$. Similarly as before, the upper bounds are consequence of the following propositions.

Proposition 3. *Let* $2\ell < k$, *then*

$$\mathrm{ex}(n, C_{2\ell}, P_k) \le \frac{k^\ell n^\ell}{\ell 2^{\ell+1}} + O(n^{\ell-1}).$$

Proposition 4. *Let* $2\ell + 1 < k$, *then*

$$\mathrm{ex}(n, C_{2\ell+1}, P_k) \le \frac{k^{\ell+1}n^\ell}{2^{\ell+2}} + O(n^{\ell-1}).$$
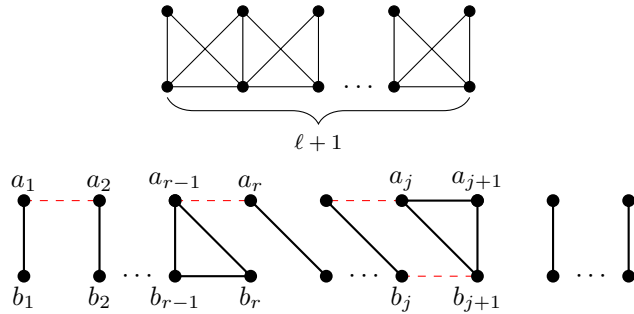

Fig. 4: The structure of paths aligning with matching structures from $\mathcal{M}$.

It is enough to prove the following claims from which the propositions above follow (a proof of this implication is included after the proof of the claim). Their proofs are similar, so we just give the proof of the first claim.

Claim 3. *For every $k, \ell \in \mathbb{N}$, there exists $n_0 \in \mathbb{N}$ such that if $n \geq n_0$,*

$$\mathrm{ex}(n+1, C_{2\ell}, P_k) - \mathrm{ex}(n, C_{2\ell}, P_k) \leq \frac{k^\ell n^{\ell-1}}{2^{\ell+1}} + O(n^{\ell-2}).$$

Claim 4. *For every $k, \ell \in \mathbb{N}$, there exists $n_0 \in \mathbb{N}$ such that if $n \geq n_0$,*

$$\mathrm{ex}(n+1, C_{2\ell+1}, P_k) - \mathrm{ex}(n, C_{2\ell+1}, P_k) \leq \frac{\ell k^{\ell+1} n^{\ell-1}}{2^{\ell+2}} + O(n^{\ell-2}).$$

Proof Proof of Claim 3: Let G be a P_k-free graph on $n+1$ vertices with maximum number of copies of $C_{2\ell}$. If $\delta(G) > t$, then by the classical argument of Dirac (1952), every connected component must have size at most k, and then $\mathcal{N}(C_{2\ell}, G) \leq k^{2\ell-1} n$. So assume $\delta(G) \leq t$, and let v be a vertex of minimum degree. Then every $C_{2\ell}$ using v can be divided into two paths: v together with the vertex preceding it and following it in the cycle (forming a P_2), and the remaining $2\ell - 3$ vertices (forming a $P_{2(\ell-2)}$). Note that every P_2 and $P_{2(\ell-2)}$ can be joined in at most two ways to make a $C_{2\ell}$; therefore, the number of copies of $C_{2\ell}$ containing v is at most

$$2\binom{d(v)}{2} \mathrm{ex}(n, P_{2(\ell-2)}, P_k) \leq 2\binom{t}{2}\frac{k^{\ell-2}n^{\ell-1}}{2^{\ell-1}} + O(n^{\ell-2}) \leq \frac{k^\ell n^{\ell-1}}{2^{\ell+1}} + O(n^{\ell-2}).$$

$\square$

We include a proof that Proposition 3 follows from Claim 3. Similar ideas are used throughout the paper.

Proof Proof that Claim 3 implies Proposition 3: We have

$$\mathrm{ex}(n, C_{2\ell}, P_k) = \mathrm{ex}(n_0, C_{2\ell+1}, P_k) + \sum_{s=n_0+1}^{n} (\mathrm{ex}(s, C_{2\ell+1}, P_k) - \mathrm{ex}(s-1, C_{2\ell+1}))$$

$$\leq \mathrm{ex}(n_0, C_{2\ell+1}, P_k) + \frac{k^\ell}{2^{\ell+1}} \sum_{s=1}^{n} \left(s^{\ell-1} + O(n^{\ell-2})\right)$$

$$\leq \mathrm{ex}(n_0, C_{2\ell+1}, P_k) + \frac{k^\ell}{2^{\ell+1}} \sum_{s=1}^{n} \left(\frac{(s+1)^\ell}{\ell} - \frac{s^\ell}{\ell}\right) + O(n^{\ell-1})$$

$$\leq \frac{k^\ell n^\ell}{\ell 2^{\ell+1}} + O(n^{\ell-1}),$$

where in the first inequality we used Claim 3 and pulled the constant out of the sum, and the second inequality follows from $(s+1)^\ell = s^\ell + \ell s^{\ell-1} + O(s^{\ell-2})$. The final inequality follows from the telescoping sum. $\square$

3   The number of copies of H in graphs without a certain tree

Alon and Shikhelman, while considering the case when H is a bipartite graph and T is a tree, mention that $\operatorname{ex}(n, H, T) = O(n^{\alpha(H)})$ is a consequence of a theorem from Alon (1981). We prove that, in fact, this holds for general graphs H.

Theorem 11. *Let H be any graph and let T be any tree, then $\operatorname{ex}(n, H, T) = O(n^{\alpha(H)})$.*

Corollary 2. *For any graph H such that $v(H) - \alpha(H) \leq \lfloor \frac{k-1}{2} \rfloor$, we have $\operatorname{ex}(n, H, P_k) = \Theta(n^{\alpha(H)})$.*

A construction yielding the lower bound in Corollary 2 is $G_{n,k,t}$. Indeed, for every subset of size $\alpha(H)$ of the independent set in $G_{n,k,t}$ we can find a copy of H by joining the $t = \lfloor \frac{k-1}{2} \rfloor$ vertices involved in the clique in $G_{n,k,t}$.

Theorem 11 follows as a simple consequence of the following lemma which will be proven by induction on $\alpha(H)$.

Lemma 2. *For any graph H and any tree T,*

$$\operatorname{ex}(n+1, H, T) - \operatorname{ex}(n, H, T) = O(n^{\alpha(H)-1}).$$

Here, the constant given by the O notation depends only on H and T.

We start by proving the following well-known fact.

Proposition 5. *Let H be a graph, and let u be a vertex of H. If H' is the graph obtained by removing u together with its neighborhood, then $\alpha(H') \leq \alpha(H) - 1$.*

Proof: If X is a maximal independent set in H', then since no neighbor of u is in X, the set $X \cup \{u\}$ is independent in H and so $\alpha(H') + 1 \leq \alpha(H)$. □

We are now ready to prove Lemma 2.

Proof Proof of Lemma 2: For the base case of the induction, note that if $\alpha(H) = 1$, then H is a clique and it is easy to see that $\operatorname{ex}(n, K_s, T) = O(n)$ for any s and T. (We may, for example, use the simple bound of $\operatorname{ex}(n, T) \leq v(T)n$, for any tree T, and apply an induction argument similar to the proof of Theorem 4.)

To estimate $\operatorname{ex}(n+1, H, T) - \operatorname{ex}(n, H, T)$, we will start with a graph G on $n+1$ vertices which is T-free with maximum number of copies of H. We know that $\delta(G) < v(T)$, otherwise $T \subseteq G$. Let v be a vertex of minimum degree in G, and we will count the number of copies of H in G containing v as a vertex.

Let $V(H) = \{u_1, u_2, \ldots, u_{v(H)}\}$, and let H_i be the graph obtained by removing u_i together with its neighbors. By Proposition 5, we know that $\alpha(H_i) \leq \alpha(H) - 1$. Now for each copy of H using v as a vertex, v must play the role of some u_i, and the neighbors of u_i must be embedded in the neighborhood of v. Then the other vertices of H, that is the vertices of H_i, must be embedded in some way in the remaining vertices of G. We have to choose $d_H(u_i)$ vertices in $N(v)$, so the number of copies of H using v is at most

$$\sum_{i=1}^{v(H)} d(v)^{d_H(u_i)} \mathcal{N}(H_i, G) \leq \sum_{i=1}^{v(H)} v(T)^{d_H(u_i)} \mathcal{N}(H_i, G) = \sum_{i=1}^{v(H)} O_{H_i}(n^{\alpha(H_i)}) = O(n^{\alpha(H)-1}).$$

Thus, if G' is the graph obtained from G by removing v, we have that

$$\text{ex}(n + 1, H, T) = \mathcal{N}(H, G) = \mathcal{N}(H, G') + O(n^{\alpha(H)-1}) \leq \text{ex}(n, H, T) + O(n^{\alpha(H)-1}).$$

$\square$

For some particular graphs H, by studying more carefully the number of copies of H that use some fixed vertex, we can find a better recursion than the one from Lemma 2. In the following section, we improve the recursion for several specific classes of graphs. For these graphs we will find an integer valued function $f(n)$ which is a lower bound of the extremal number $\text{ex}(n, H, T)$, such that $f(n)$ grows faster than $\text{ex}(n, H, T)$ (when they do not agree). Since both functions are integer valued they must coincide eventually.

4 Exact Results

We now turn our attention to proving some exact results. Recall that we are using the notation $t = \lfloor \frac{k-1}{2} \rfloor$.

4.1 Number of copies of C_4

We begin by determining the maximal number of copies of C_4 in a P_k-free graph.

Theorem 12. *For every integer $k \geq 5$, there exists $n_1 \in \mathbb{N}$ such that if $n \geq n_1$,*

$$\text{ex}(n, C_4, P_k) = \mathcal{N}(C_4, G_{n,k,t}) = \binom{n-t}{2}\binom{t}{2} + 3(n-t)\binom{t}{3} + 3\binom{t}{4} + 2\eta_k\binom{t}{2},$$

where $\eta_k = 1$, if k is even, and 0 otherwise. Moreover, the only extremal graph is $G_{n,k,t}$.

To prove Theorem 12, we will prove the following claim from which the theorem follows simply by induction on n.

Claim 5. *There exists $n_0 \in \mathbb{N}$ such that if $n \geq n_0$, $\text{ex}(n + 1, C_4, P_k) - \text{ex}(n, C_4, P_k) \leq \binom{t}{2}(n - 2)$. Equality can hold for this difference only if the unique extremal graph with $n + 1$ vertices is $G_{n+1,k,t}$.*

It is easy to see that $\mathcal{N}(C_4, G_{n+1,k,t}) = \mathcal{N}(C_4, G_{n,k,t}) + \binom{t}{2}(n-2)$. By Claim 5, $\text{ex}(n+1, C_4, P_k) \leq \text{ex}(n, C_4, P_k) + \binom{t}{2}(n-2)$ with equality only if the unique extremal graph with $n+1$ vertices is $G_{n+1,k,t}$. It follows that $\text{ex}(n+1, C_4, P_k) - \mathcal{N}(C_4, G_{n+1,k,t}) \leq \text{ex}(n, C_4, P_k) - \mathcal{N}(C_4, G_{n,k,t})$ and so the sequence $\text{ex}(n, C_4, P_k) - \mathcal{N}(C_4, G_{n,k,t})$ is a non-increasing sequence of non-negative integers that is strictly decreasing after every non-zero term. Thus, this sequence is eventually the constant 0 sequence, and hence, moreover, $G_{n,k,t}$ is eventually the unique extremal graph.

We now prove Claim 5.

Proof: Let G be a P_k-free graph on $n + 1$ vertices with the maximum number of copies of C_4; that is, $\mathcal{N}(C_4, G) = \text{ex}(n + 1, C_4, P_k)$.

If $\delta(G) > t$, then by the classical argument of Dirac (1952), every connected component must have size at most k, and therefore $N(C_4, G) \leq \frac{n+1}{k} 3\binom{k}{4} = \frac{(n+1)(k-1)(k-2)(k-3)}{8}$. Then we can choose n_0 so that this number is less than $\mathcal{N}(C_4, G_{n,k,t})$ for $n \geq n_0$. Thus, we can assume $\delta(G) \leq t$.

Let v be a vertex of minimum degree. By removing v, we are removing at most $\binom{d(v)}{2}(n - 2) \leq \binom{t}{2}(n - 2)$ copies of C_4. Equality can hold only if $d(v) = t$ and the neighbors of v have full degree. It

follows that if equality holds, then G contains a complete bipartite graph with color classes of size t and $n + 1 - t$ respectively such that the size t class is a clique. If k is odd, we have that $G = G_{n+1,k,t}$. If k is even, since G contains the maximum number of C_4's, it follows that G has an additional edge (it cannot have 2 more for otherwise we would have a P_k). Thus, if k is even we also have $G = G_{n+1,k,t}$.

Therefore, either $G = G_{n+1,k,t}$ or by removing a minimum degree vertex v we obtain a graph G' with $\mathcal{N}(C_4, G') > \mathcal{N}(C_4, G) - \binom{t}{2}(n-2) = \mathrm{ex}(n + 1, C_4, P_k) - \binom{t}{2}(n-2)$. Since $\mathrm{ex}(n, C_4, P_k) \geq \mathcal{N}(C_4, G')$, we have that $\mathrm{ex}(n + 1, C_4, P_k) - \mathrm{ex}(n, C_4, P_k) < \binom{t}{2}(n-2)$. $\qquad\square$

The same argument proves the following.

Theorem 13. *For every positive integer $k \geq 5$, there exists $n_1 \in \mathbb{N}$ such that if $n \geq n_1$*

$$\mathrm{ex}(n, C_4, C_{\geq k}) = \mathcal{N}(C_4, G_{n,k,t}) = \binom{n-t}{2}\binom{t}{2} + 3(n-t)\binom{t}{3} + 3\binom{t}{4} + 2\eta_k\binom{t}{2},$$

where $\eta_k = 1$, if k is even, and 0 otherwise. Moreover, the only extremal graph is $G_{n,k,t}$.

4.2 Number of copies of S_r

We will prove the following theorem about the number of copies of P_2. However, it will follow as a consequence of a more general result about stars.

Theorem 14. *For every positive integer $k \geq 3$, there exists $n_1 \in \mathbb{N}$ such that if $n \geq n_1$*

$$\mathrm{ex}(n, P_2, P_k) = \mathcal{N}(P_2, G_{n,k,t}) = t\binom{n-1}{2} + (n-t)\binom{t}{2} + 2t\eta_k,$$

where $\eta_k = 1$, if k is even, and 0 otherwise. Moreover, the only extremal graph is $G_{n,k,t}$.

More generally we have,

Theorem 15. *For every positive integer $k \geq 3$ and $r \geq 2$, there exists $n_1 \in \mathbb{N}$ such that if $n \geq n_1$,*

$$\mathrm{ex}(n, S_r, P_k) = \mathcal{N}(S_r, G_{n,k,t}) = t\binom{n-1}{r} + (n-t)\binom{t}{r} + 2\eta_k\binom{t}{r-1},$$

where $\eta_k = 1$, if k is even, and 0 otherwise. Moreover, the only extremal graph is $G_{n,k,t}$, unless k is even and $t \leq r - 2$ in which case the only extremal graphs are $G_{n,k,t}$ and $G_{n,k-1,t}$.

Again, the result follows from a claim about the difference of the values of two consecutive extremal numbers. Let $a_n = \mathcal{N}(S_r, G_{n+1,k,t}) - \mathcal{N}(S_r, G_{n,k,t}) = \binom{t}{r} + t\binom{n-1}{r-1}$.

Claim 6. *There exists $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$, $\mathrm{ex}(n + 1, S_r, P_k) - \mathrm{ex}(n, S_r, P_k) \leq a_n$ and equality can hold only if either $G_{n+1,k,t}$ is the only extremal graph on $n + 1$ vertices or k is even, $t \leq r - 2$ and the only extremal graphs are $G_{n+1,k,t}$ and $G_{n+1,k-1,t}$.*

Proof: For any graph G, we have that $\mathcal{N}(S_r, G) = \sum_{v \in V(G)} \binom{d(v)}{r}$. Let G be a P_k-free graph with $n + 1$ vertices and maximum number of copies of S_r; that is, $\mathcal{N}(S_r, G) = \mathrm{ex}(n + 1, S_r, P_k)$. We will consider cases depending on the minimum degree of G.

If $\delta(G) > t$, then every connected component of G must have at most k vertices. So the number of copies of S_r is bounded by $n\binom{k-1}{r}$, then we choose n_0 such that this number is less than $\mathcal{N}(S_r, G_{n,k,t})$ for $n \geq n_0$.

If $\delta(G) \leq t$, then by removing v a vertex of minimum degree, we remove at most

$$\binom{d(v)}{r} + \sum_{u \in N(v)} \binom{d(u)-1}{r-1} \leq \binom{t}{r} + t\binom{n-1}{r-1}$$

copies of S_r. Equality can hold only if $\delta(G) = t$ and t vertices have degree n, so G contains a complete bipartite graph with color classes of size t and $n+1-t$ such that class of size t is a clique. Then the characterization of the extremal cases again follows from the maximality of G. $\square$

Remark 1. *By checking more carefully the difference between the number of r-stars using v and the number a_n, we can find a bound for n_1 of order $k^{3/2}$.*

Similarly to before, the same method proves the following two results.

Theorem 16. *For every positive integer $k \geq 5$, there exists $n_1 \in \mathbb{N}$ such that if $n \geq n_1$,*

$$\mathrm{ex}(n, P_2, C_{\geq k}) = \mathcal{N}(P_2, G_{n,k,t}) = t\binom{n-1}{2} + (n-t)\binom{t}{2} + 2t\eta_k,$$

where $\eta_k = 1$, if k is even, and 0 otherwise. Moreover the only extremal graph is $G_{n,k,t}$.

Or more generally,

Theorem 17. *For every positive integer $k \geq 5$, there exists $n_1 \in \mathbb{N}$ such that if $n \geq n_1$,*

$$\mathrm{ex}(n, S_r, C_{\geq k}) = \mathcal{N}(S_r, G_{n,k,t}) = \mathcal{N}(P_2, G_{n,k,t}) = t\binom{n-1}{r} + (n-t)\binom{t}{r} + 2\eta_k\binom{t}{r-1},$$

where $\eta_k = 1$, if k is even, and 0 otherwise. Moreover the only extremal graph is $G_{n,k,t}$, unless k is even and $t \leq r-2$ in which case the only extremal graphs are $G_{n,k,t}$ and $G_{n,k-1,t}$.

Remark 2. *For $k = 3$, Theorem 16 also holds. Since G must be a tree and by convexity the number of stars is maximized in a star of n vertices, we have $G_{n,3,1} = K_{1,n-1}$, and this graph has $\binom{n-1}{r}$ stars. For $k = 4$, a star with a perfect matching or almost perfect matching in the neighborhood of the center vertex maximizes the number of copies of P_2, with $\binom{n-1}{2} + (n-1)$, when n is odd or $\binom{n-1}{2} + (n-2)$, when n is even. Any graph containing the n vertex star maximizes the number of copies of S_r for $r \geq 3$.*

4.3   Number of copies of P_3

Theorem 18. *For every positive integer $k \geq 5$, there exists $n_1 \in \mathbb{N}$ such that if $n \geq n_1$*

$$\mathrm{ex}(n, P_3, P_k) = \mathcal{N}(P_3, G_{n,k,t}) = \frac{3t(t-1)}{2}n^2 + O(n).$$

Moreover the only extremal graph is $G_{n,k,t}$.

Let $a_n = \mathcal{N}(P_3, G_{n+1,k,t}) - \mathcal{N}(P_3, G_{n,k,t}) = 2t\left(\binom{t-1}{2} + (n-t)(t-1) + \eta_k\right) + t(t-1)(n-2).$

As in the previous results it is enough to prove the following claim.

Claim 7. *There exists $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$, $\mathrm{ex}(n+1, P_3, P_k) - \mathrm{ex}(n, P_3, P_k) \leq a_n$ and equality can hold only if $G_{n+1,k,t}$ is the only extremal graph on n vertices.*

Proof: Let G be an $(n+1)$-vertex graph with maximum number of copies of P_3. We may assume that $\delta(G) \geq 2$. (If a vertex has degree 1, then it is in at most $2(t(n-t) + \binom{t}{2} + \eta_k)$ copies of P_3.)

If $\delta(G) > t$, then each connected component of G must have size at most k (by Dirac's argument) and so $\mathcal{N}(P_3, G) \leq 3\binom{k}{3}\frac{n+1}{k}$. In this case, we can choose n_0 such that for $n \geq n_0$, this number is less than $\mathcal{N}(P_3, G_{n,k,t})$. Thus, we assume that $\delta(G) \leq t$.

Let v be a vertex in G with minimum degree, and consider the copies of P_3 beginning containing v as their second vertex. We may suppose G is connected and has enough vertices to apply Theorem 3. Then the number of copies of P_3 whose second vertex is v is bounded from above by

$$d(v)(d(v)-1)(n-2) - 2(d(v)-2)\left(\binom{d(v)}{2} - e(N(v))\right).$$

Indeed, the first term is the trivial upper bound $2\binom{d(v)}{2}(n-2)$ obtained if every pair of neighbors of v could be extended to path of length 3 in any possible way. The subtraction comes from the fact that each non-edge $\{a,b\}$ in the neighborhood of v along with a third neighbor $c \in N(v)$ uniquely forbids 2 copies of P_3 namely $cvab$ and $cvba$. We have bounded from above the number of copies of P_3 containing v as a second vertex. Now we will obtain an estimate on the number of copies of P_3 starting at v. We consider the number of ways to take distinct $u \in N(v)$, $w \in N(u)$ and $x \in N(w)$:

$$\sum_{u \in N(v)} \sum_{\substack{w \in N(u) \\ w \neq v}} \left(d(w) - 1 - \mathbb{1}_{w \in N(v)}\right) = \sum_{u \in N(v)} \left(\sum_{\substack{w \in N(u) \\ w \neq v}} \left(d(w)\right) - d(u) + 1\right) - 2e(N(v))$$

$$= \sum_{u \in N(v)} \left(\sum_{w \in V(G)} \left(d(w)\right) - \sum_{\substack{w \notin N(u) \\ w \neq u}} \left(d(w)\right) - d(v) - 2d(u)\right) - 2e(N(v)) + d(v)$$

$$= \sum_{u \in N(v)} \left(2e(G) - \sum_{\substack{w \notin N(u) \\ w \neq u}} \left(d(w)\right) - 2d(u)\right) - 2e(N(v)) - d(v)(d(v)-1)$$

$$\leq \sum_{u \in N(v)} \left(2e(G) - 2(n - d(u)) - 2d(u)\right) - 2e(N(v)) - d(v)(d(v)-1)$$

$$= 2d(v)(e(G) - n) - 2e(N(v)) - d(v)(d(v)-1),$$

where the inequality uses that $\delta(G) \geq 2$.

The above sum is maximized when $d(v) = t$, and to achieve this maximum it is necessary that for every neighbor u of v, we have that the non-neighbors of u have degree 2. Moreover, if $t = 2$ it is simple to check that to maximize the expression the two neighbors of v must be adjacent. From here it follows that the number of copies of P_3 containing v is at most a_n.
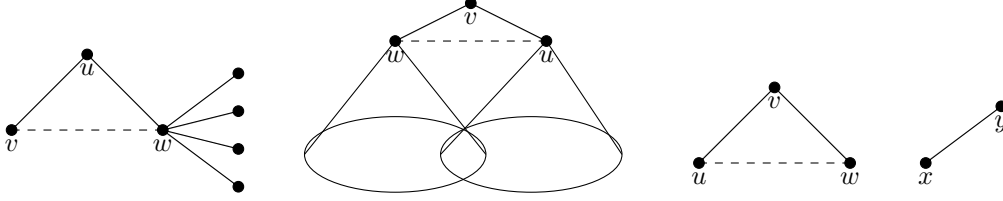
Fig. 5: Constructing paths using v.

When $t \geq 3$, we have the bound (conditioning on whether v is at the beginning or middle of the path)

$$2t(e(G) - n) - 2e(N(v)) - t(t-1) + t(t-1)(n-2) - 2(t-2)\left(\binom{t}{2} - e(N(v))\right),$$

$$\leq 2t(e(G) - n) - 2t(t-1) + t(t-1)(n-2).$$

From Theorem 3 it follows that this number is at most a_n. To obtain equality, in both cases it is necessary that every neighbor of v has full degree and so by maximality we have that $G = G_{n+1,k,t}$. □

Remark 3. *For $k = 4$, it is simple to check that the only extremal graph is a balanced double star on n vertices, which has $\lfloor \frac{n-1}{2} \rfloor \lceil \frac{n-1}{2} \rceil$ copies of P_3.*

Next we consider paths of length 4.

4.4 Number of copies of P_4

Theorem 19. *For every positive integer $k \geq 5$, there exists $n_1 \in \mathbb{N}$ such that if $n \geq n_1$, then*

$$\mathrm{ex}(n, P_4, P_k) = \mathcal{N}(P_4, G_{n,k,t}) = \frac{t(t-1)}{2}n^3 + O(n^2).$$

Moreover the only extremal graph is $G_{n,k,t}$.

Let $a_n := \mathcal{N}(P_4, G_{n+1,k,t}) - \mathcal{N}(P_4, G_{n,k,t})$, defined similarly as before. We can check that $a_n = 2t\mathcal{N}(P_2, G_{n-1,k-2,t-1}) + 2t(t-1)e(G_{n-2,k-4,t-2}) + \binom{t}{2}(n-2)(n-3)$.

As in the previous results it is enough to prove the following Claim.

Claim 8. *There exists $n_0 \in \mathbb{N}$ such that for every $n \geq n_0$, $\mathrm{ex}(n+1, P_4, P_k) - \mathrm{ex}(n, P_4, P_k) \leq a_n$ and equality can hold only if $G_{n+1,k,t}$ is the only extremal graph on n vertices.*

Proof: Let G be an $(n+1)$-vertex graph with maximum number of P_4 copies. Assume that $\delta(G) \geq 2$. (If a vertex v has degree 1, then it is in at most $2\mathcal{N}(P_2, G_{n,k,t}) < a_n$ copies of P_4.)

If $\delta(G) > t$ then each component of G must have size at most k and so $\mathcal{N}(P_4, G) \leq 12\binom{k}{4}\frac{n+1}{k}$ so we can choose n_0 such that if $n \geq n_0$ this number is less than $\mathcal{N}(P_4, G_{n,k,t})$.

Suppose now that $\delta(G) \leq t$. Let v be a vertex of minimum degree. As before suppose G is connected. To count the number of paths of length 4 starting at v, fix $u \in N(v)$ and let G' be the subgraph of G obtained by removing v and u. Any path $vuu_1u_2u_3$ can be decomposed as the edge vu together with the ordered path $u_1u_2u_3$ in G' so the number of paths of the form $vuu_1u_2u_3$ is at most $2\mathcal{N}(P_2, G')$, since there are two orderings of any P_2. It is easy to check that G' cannot contain a cycle of length at least
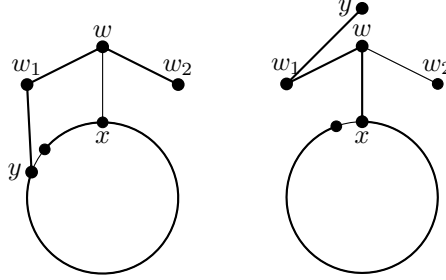
Fig. 6: Lemma of the cycle in Theorem 19.

$k - 1$, otherwise together with the edge uv we would have a copy of P_k. Thus, there are two cases:

a) Suppose first that G' is does not contain a C_{k-2}. Then G' is $C_{\geq k-2}$-free and so $\mathcal{N}(P_2, G') \leq ex(n-1, P_2, C_{\geq k}) = \mathcal{N}(G_{n-1,k-2,t-1}, P_2)$. We will take n_0 bigger than the constant from Theorem 16, when $k = 5$, or $k \geq 7$. When $k = 6$ we use the following lemma.

Lemma 3. *If H is a graph on n vertices containing no cycle of length at least 4, then either H contains a vertex of degree $n-1$ or $\mathcal{N}(P_2, H) < \binom{n-1}{2} + 2$.*

Proof: Suppose H has no vertex of degree $n-1$. If H has degree 1 vertices, then the number of copies of P_2 is maximized when all these vertices are adjacent to the vertex of maximum degree, so suppose H has no vertex of degree 1. Then $\sum_{v \in V(H)} d(v) \leq 3(n-1)$ and $2 \leq d(v) \leq n-2$. Therefore by convexity the number of copies of P_2 is maximized when there is one vertex of degree $n-2$, one of degree 3 and $n-2$ of degree 2. This yields $\binom{n-2}{2} + n + 1 = \binom{n-1}{2} + 3$, however a graph with such a degree sequence must have a cycle of length bigger than four. Thus we consider the second best is a graph with one vertex of degree $n-2$ and $n-1$ vextex of degree 2 (if possible) which has $\binom{n-1}{2} + 1$ copies of P_2.   □

Now according to this lemma for $k = 6$, either $\mathcal{N}(P_2, G') < \binom{n-2}{2} + 2 = \mathcal{N}(P_2, G_{n-1,4,1})$ or G' has a vertex of degree $n-2$, call it w, and some edges t in $N_{G'}(w)$. The vertex u cannot be connected to two different edges in $N_{G'}(w)$, otherwise G would contain a P_6, and if u is connected to both vertices of one of these edges and to all other vertices of G', then the number of copies of P_4 starting with vu would be $(n - 2t)(n - 3) + 2 + 2t \leq (n - 2)(n - 1) + 4$.

b) Now suppose that G' contains a cycle of length $k - 2$, C. In this case we have the following.

Claim 9. *If w is a vertex which is not in the cycle and $w \in N(x)$ where x is a vertex of the cycle, then w has at most one neighbor outside of C.*

Proof: Suppose w_1 and w_2 are two neighbors of w. Since $\delta(G) > 1$, w_1 has a neighbor y. If y is in C, then C together with $w_1 w w_2$ is a length k path. If y is outside of C, then C together with $w w_1 y$ is a P_k.   □

As a corollary we have,

Claim 10. *If w is a vertex not in the cycle and $w \in N(x)$ where x is a vertex of the cycle, then $d(w) < k$.*

The edge uv is connected to C by some path. If there is an intermediate vertex from C to uv, then clearly this path will have length at least k and so this is not possible. Hence C is connected to either u or v. If C is connected to v, then every neighbor of u must be in C for otherwise we have a k path. If C is connected to u, then by Claim 9, all neighbors of u, except for v, are in C. So for every path $vuu_1u_2u_3$ we have that $u_1 \in C$, if $u_2 \in C$, then we have less than k choices for both u_1 and u_2 and at most n choices for u_3. If u_2 is not in C, then by Claim 10, since u_2 is a neighbor of $u_1 \in V(C)$, we have $d(u_2) < k$ and so there are at most k choices for u_3 and less than n choices for u_2. Hence we have less than $2k^2n$ such paths in total and we can take n_0 such that if $n \geq n_0$, then this number is less than $2\mathcal{N}(G_{n-1,k-2,t-1}, P_2)$.

It follows that the number of paths starting with v is at most $2d(v)\mathcal{N}(P_2, G_{n-1,k-2,t-1})$.

Now if $d(v) \leq t-1$, then the trivial bound on the number of copies of P_4 with middle vertex v is $d(v)(d(v)-1)\binom{n-2}{2}$ and the bound on the number of P_4 cpoies with v as a second vertex is $2d(v)(d(v)-1)e(G)$. Thus, we would have that the number of copies of P_4 using v is less than a_n. So we will now suppose $d(v) = t$. To simplify the notation in the following calculations let $S := \sum_{u \in N(v)} d(u)$.

To count paths with v in the middle, we will count in order paths of the form $xuvwy$, where u, w can be any neighbors of v and then we have to choose a neighbor x of u and a neighbor y of w with $y \neq x$. Hence the number of ordered paths with v as the middle vertex is

$$\sum_{u \in N(v)} \left(\sum_{\substack{w \in N(v) \\ w \neq u}} (d(u) - 1 - \mathbb{1}_{u \in N(w)})(d(w) - 1 - \mathbb{1}_{u \in N(w)}) - |N(u) \cap N(w)| + 1 \right)$$

$$\leq \sum_{u \in N(v)} \left(\sum_{\substack{w \in N(v) \\ w \neq u}} d(u)d(w) - 2d(w) - 2d(u) - \mathbb{1}_{u \in N(w)}(d(w) + d(u)) + 5 \cdot \mathbb{1}_{u \in N(w)} + n + 1 \right)$$

$$= S^2 - \left(\sum_{u \in N(v)} d(u)^2 \right) - 4(t-1)S - 2\left(\sum_{u \in N(v)} d(u)|N(u) \cap N(v)| \right) + 10e(N(v)) + t(t-1)(n+1)$$

$$\leq S^2 - \left(\sum_{u \in N(v)} d(u)^2 \right) - 4(t-1)S - 2\left(\sum_{u \in N(v)} d(u)(d(u) + t - n - 1) \right) + t(t-1)(n+6)$$

$$= S^2 - 3\left(\sum_{u \in N(v)} d(u)^2 \right) + (2n - 6(t-1))S + t(t-1)(n+6)$$

$$\leq S^2 - \frac{3S^2}{t} + (2n - 6(t-1))S + t(t-1)(n+6)$$

$$= \frac{t-3}{t}S^2 + (2n - 6(t-1))S + t(t-1)(n+6),$$

where in the first and second inequality we use the fact that for every pair of vertices x, y of the graph $|N(x) \cap N(y)| \geq d(x) + d(y) - n + 1 - 2 \cdot \mathbb{1}_{x \in N(y)}$ together with $e(N(v)) \leq \binom{t}{2}$. The last inequality was obtain by applying the Cauchy-Schwarz inequality to $\sum_{u \in N(v)} d(u)^2$. Since any path can have two distinct orders we divide this expression by 2.

To count the number of paths with v as the second vertex, we will decompose the path $uvwxy$ into uvw together with $e = xy$. First we choose in order two neighbors of v, then an edge not using u, v or w.

There are at most 2 ways to connected the edge to w, so the number of these paths is at most

$$2 \sum_{u \in N(v)} \left(\sum_{\substack{w \in N(u) \\ w \neq u}} e(G) - d(w) - d(u) - t + 2 + \mathbb{1}_{u \in N(w)} \right)$$

$$= 2t(t-1)(e(G) - t + 2) + 4e(N(v)) - 4(t-1)S$$

$$\leq 2t(t-1)(e(G_{n,k,t}) + 2) + 4\binom{t}{2} - 4(t-1)S.$$

By summing the previous bounds, we have that the number of paths using v is at most

$$2t\mathcal{N}(P_2, G_{n,k-2,t-1}) + \frac{t-3}{2t}S^2 + (n - 7(t-1))S + t(t-1)(n+6) + 2t(t-1)(e(G_{n,k,t}) + 2) + 4\binom{t}{2}.$$

The value of this expression when $S = tn$ is precisely a_n. By considering this expression as a quadratic in S, we can check that if $t \geq 3$ the maximum is attained only when $S = nt$. This means that every neighbor of v must have degree n, so this is only possible if $G = G_{n+1,k,t}$. If $t = 2$, the expression attains its maximum when $S = 2n - 14$, hence if $S < 2n - 28$. This value would be less than a_n, but now with the condition $S \geq 2n - 28$. It is simple to check that for $k = 5$, G must be either $K_{2,n-2}$ or $G_{n,5,2}$, and if $k = 6$, then G must be $G_{n,6,2}$.   □

5   The number of copies of P_{k-1} in P_k-free graphs

If k is odd, it seems likely that the graph $G_{n,k,t}$ attains the value $\mathrm{ex}(n, P_{k-1}, P_k)$. However, for k even the situation changes. We have that $\mathcal{N}(P_{k-1}, G_{n,k,t}) = \Theta(n^t)$, but there is another graph $H_{n,k}$ such that $\mathcal{N}(P_{k-1}, H_{n,k}) = \Theta(n^{t+1})$. In order to define this graph, first for $r \geq 2$ and a, b positive integers, let $S_{a,b}^{(r)}$ be the $(a + b + r)$-vertex graph consisting of a clique on r vertices and two independent sets A and B on a and b vertices, respectively. Let v be a vertex of the clique and join v to every vertex in B, then join every vertex of the clique except v to every vertex in A. Let $\mathcal{S}_n^{(r)}$ be the family of all such graphs on n vertices. For even k, let $H_{n,k}$ be the graph in $\mathcal{S}_n^{(t+1)}$ which maximizes the number of P_{k-1} copies. In this case we conjecture that the graph $H_{n,k}$ is extremal.

Conjecture 1. *If k is even and $k \geq 4$, the extremal number $\mathrm{ex}(n, P_{k-1}, P_k)$ is attained by the the graph $H_{n,k}$.*

Remark 4. *For $r \geq 2$ the graphs $S_{a,b}^{(r)}$ are P_{2r}-free and have $(r - 1)! \, ba(a - 1) \cdots (a - r + 2)$ copies of P_{2r-1}. In $\mathcal{S}_n^{(r)}$ this number is maximized when a is roughly $\frac{r-1}{r}n$, and this maximum approaches $(r - 1)! \, (\frac{(r-1)^{r-1}}{r^r})n^r$ as n tends to infinity. In particular by taking $r = 3$ we have a P_6-free graph with $8n^3/27 + O(n^2)$ copies of P_5.*

Remark 5. *Note that the only edges of the clique in $S_{a,b}^{(r)}$ that a P_{2r-1} uses are the ones that are incident with the vertex v. So we have several graphs for which we conjecture the number of copies of P_{2r-1} is maximal, namely those subgraphs of $S_{a,b}^{(r)}$ formed by removing edges from the clique not incident to v.*

Conjecture 1 can be easily checked for $k = 4$, and the following theorem says that this conjecture is also true for $k = 6$.

Theorem 20. *There exists $n_1 \in \mathbb{N}$ such that if $n \geq n_1$, then*

$$\mathrm{ex}(n, P_5, P_6) = \mathcal{N}(P_5, H_{n,6}) = \frac{8n^3}{27} + O(n^2).$$

Proof:

Let G be a P_6 free graph, and suppose $n \geq 7$. It is enough to bound the copies of P_5 in each connected component, so assume G is connected.

Let C be the largest cycle in G and let G' be the graph obtained by removing C from G; clearly C cannot be a 6-cycle, otherwise G would contain a P_6. We will consider cases based on the length of C.

a) Suppose C is a 5-cycle with vertices v_1, v_2, v_3, v_4, v_5 appearing consecutively. Then every vertex in G' is connected to a vertex of C. Suppose that v_1 has a neighbor in G', if v_1 is the only vertex of C with an edge to G', then $\mathcal{N}(P_5, G) < 24n$. So suppose this is not the case, v_2 and v_5 cannot have neighbors in G'. Thus, without loss of generality, we may suppose v_3 has a neighbor in G', then v_4 cannot have neighbors in G', also note that v_2 cannot be connected with v_4 or with v_5 and so $G \subseteq G_{n,6,2}$ (where the v_1 and v_3 take the role of the high degree vertices, and the edge $v_4 v_5$ is the only edge that is not incident with one of v_1 or v_3), hence $\mathcal{N}(P_5, G) = O(n^2)$.

b) Now suppose C is a 4-cycle defined by v_1, v_2, v_3, v_4, consecutively. Then G' cannot contain a P_3, otherwise by connectivity we would have a path of length at least 6. Consider the set X of vertices of G' that have at least one neighbor in both C and G'. Note that if $y \in G'$ is a neighbor of $x \in X$, then y cannot have any other neighbor in G'. Also note that the only possible neighbor of y in C is the neighbor of x in C (y cannot have a neighbor in C if x has more than one neighbor in C).

If $|X| > 1$, then every vertex in X must be adjacent to the same vertex in C, say v_1. Then v_2 and v_4 cannot have a neighbor outside of C. If v_2 and v_4 are adjacent, then it also holds that v_3 cannot have neighbors in G'. It is then easy to check that $\mathcal{N}(P_5, G) < 6n$. So suppose v_2 and v_4 are not connected (see Figure 5), then every P_5 in G is of the form xyv_1vv_3u, where $x \in X$, y is a neighbor of x, and both of v, u are common neighbors of v_1 and v_3. If $a = |N(v_1) \cap N(u) \cap N(v)|$ and b is the number vertices in G' with a neighbor in X, then we have that $\mathcal{N}(P_5, G) \leq ba(a-1)$ which is half $\mathcal{N}(P_5, S_{a,b}^{(3)})$ but $S_{a,b}^{(3)}$ can have at most one more vertex than G.

If $X = \{x\}$, then something similar holds, except that both v_1 and v_3 are allow to be connected to x and there is the extra possibility of G being a subgraph of $S_{a,b}^{(3)}$.

Now suppose $X = \emptyset$. If no two vertices of C share a common vertex in G', then $\mathcal{N}(P_5, G)$ is quadratic. So suppose two non-consecutive vertices, say v_1, and v_3, share a common neighbor, then it is not possible for the other two vertices in C to have a neighbor in G'. Thus, our graph is again a subgraph of $S_{a,b}^{(3)}$.

c) Suppose C is a triangle, then every pair of vertices are the end vertices of at most one P_5. If two different paths of length 5 have the same end vertices, then either G would contain a cycle of length at least four or a P_6. Thus, $\mathcal{N}(P_5, G) < \binom{n}{2}$. □
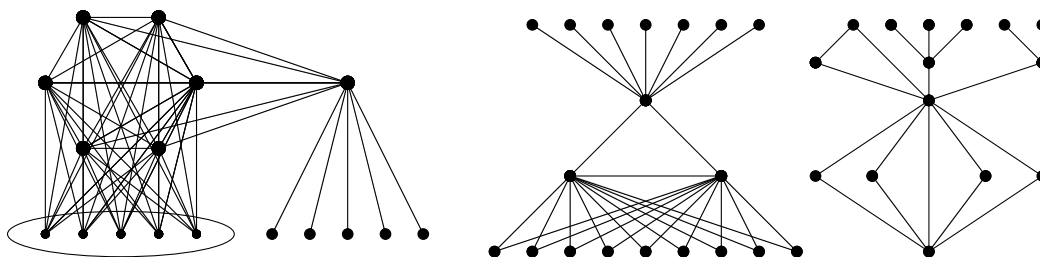
Fig. 7: The family $\mathcal{S}_{a,b}^{(r)}$ from which the conjectured extremal graph $H_{n,k}$ is obtained. And a graph that appears in Case b) of the proof of the Theorem 20.

References

N. Alon. On the number of subgraphs of prescribed type of graphs with a given number of edges. *Israel Journal of Mathematics*, 38(1-2):116–130, 1981.

N. Alon and C. Shikhelman. Many T copies in H-free graphs. *Journal of Combinatorial Theory, Series B*, 121:146–172, 2016.

P. N. Balister, E. Győri, J. Lehel, and R. H. Schelp. Connected graphs without long paths. *Discrete Mathematics*, 308(19):4487–4494, 2008.

B. Bollobás and E. Győri. Pentagons vs. triangles. *Discrete Mathematics*, 308(19):4332–4336, 2008.

G. A. Dirac. Some theorems on abstract graphs. *Proceedings of the London Mathematical Society*, 3(1): 69–81, 1952.

P. Erdős and T. Gallai. On maximal paths and circuits of graphs. *Acta Mathematica Hungarica*, 10(3-4): 337–356, 1959.

D. Gerbner, E. Győri, A. Methuku, and M. Vizer. Generalized turán problems for even cycles. *arXiv preprint arXiv:1712.07079*, 2017.

L. Gishboliner and A. Shapira. A generalized turán problem and its applications. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 760–772. ACM, 2018.

E. Győri and H. Li. The maximum number of triangles in C_{2k+1}-free graphs. *Combinatorics, Probability and Computing*, 21(1-2):187–191, 2012.

G. Kopylov. On maximal paths and cycles in a graph. In *Doklady Akademii Nauk*, volume 234, pages 19–21. Russian Academy of Sciences, 1977.

S. Letzter. Many H-copies in graphs with a forbidden tree. *arXiv preprint arXiv:1811.04287*, 2018.

R. Luo. The maximum number of cliques in graphs without long cycles. *Journal of Combinatorial Theory, Series B*, 128:219–226, 2018.

V. Nikiforov. The spectral radius of graphs without paths and cycles of specified length. *Linear Algebra and its Applications*, 432(9):2243–2256, 2010.

B. Ning and X. Peng. Extensions of Erdős-Gallai theorem and Luo's theorem with applications. *arXiv preprint arXiv:1801.09981*, 2018.