2013

On the Turán Number of Forests

Bernard Lidicky, *University of Illinois at Urbana-Champaign*
Hong Liu, *University of Illinois at Urbana-Champaign*
Cory Palmer, *University of Illinois at Urbana-Champaign*

SELECTEDWORKS™

On the Turán number of forests

Bernard Lidický   Hong Liu   Cory Palmer*

Department of Mathematical Sciences,
University of Illinois at Urbana-Champaign,
Urbana, Illinois 61801, USA

{lidicky,hliu,ctpalmer}@illinois.edu

Abstract

The Turán number of a graph H, $\mathrm{ex}(n, H)$, is the maximum number of edges in a graph on n vertices which does not have H as a subgraph. We determine the Turán number and find the unique extremal graph for forests consisting of paths when n is sufficiently large. This generalizes a result of Bushaw and Kettle [Combinatorics, Probability and Computing 20:837–853, 2011]. We also determine the Turán number and extremal graphs for forests consisting of stars of arbitrary order.

1 Introduction

Notation in this paper is standard. For a graph G let $E(G)$ be the set of edges and $V(G)$ be the set of vertices. The *order* of a graph is the number of vertices. The number of edges of G is denoted by $e(G) = |E(G)|$. For a graph G with subgraph H, the graph $G - H$ is the induced subgraph on vertex set $V(G) \setminus V(H)$ i.e. $G[V(G) \setminus V(H)]$. For $U \subset V(G)$ we define $N(U)$ to be the set of vertices in $V(G) \setminus U$ that have a neighbor in U, while the *common neighborhood* of $U \subset V(G)$ is the set of vertices in $V(G) \setminus U$ that are adjacent to every vertex in U. We denote the degree of a vertex v by $d(v)$ and the minimum degree in a graph by $\delta(G)$. A *universal vertex* in G is a vertex that is adjacent to all other vertices in G. A *star forest* is a forest whose connected components are stars and a *linear forest* is a forest whose connected components are paths. A path on k vertices is denoted P_k and a star with $k + 1$ vertices is denoted S_k. Superscript is used to denote the index of a particular graph in a set of graphs. Let $k \cdot H$ denote the graph of the disjoint union of k copies of the graph H.

The *Turán number*, $\mathrm{ex}(n,H)$, of a graph H is the maximum number of edges in a graph on n vertices which does not contain H as a subgraph. The problem of determining Turán numbers is one of the cornerstones of graph theory. A traditional starting point of extremal graph theory is a theorem of Mantel (see e.g. [1]) that the maximum number of edges in a triangle-free graph on n vertices is $\lfloor \frac{n^2}{4} \rfloor$. Turán [17, 18] generalized this result to find the extremal graph of any complete graph. In particular he showed that $\mathrm{ex}(n,K_r) = \left(\frac{r-2}{r-1} \right) \frac{n^2}{2}$. The Erdős-Stone theorem [9] states that asymptotically Turán's construction is best-possible for any r-chromatic graph H (as long as $r > 2$). More precisely $\mathrm{ex}(n,H) = \left(\frac{r-2}{r-1} \right) \frac{n^2}{2} + o(n^2)$. Thus when H is bipartite Erdős-Stone only states that $\mathrm{ex}(n,H) = o(n^2)$. There are many classical theorems on determining the Turán number of bipartite graphs. For example for complete bipartite graphs the Kővári-Sós-Turán theorem [15] gives $\mathrm{ex}(n,K_{s,t}) = O(n^{2-1/s})$ (also see e.g. [8, 4, 11]) and for even cycles Bondy and Simonovits [2] have $\mathrm{ex}(n,C_{2k}) \leqslant n^{1+1/k}$.

In 1959, Erdős and Gallai [7] determined the Turán number for paths. We state the theorem here since it is an important tool in our proofs.

Theorem 1 ([7]). *For any $k, n > 1$, $\mathrm{ex}(n, P_k) \leqslant \frac{k-2}{2} n$, where equality holds for the graph of disjoint copies of K_{k-1}.*

A well-known conjecture of Erdős and Sós [6] states that the Turán number for paths is enough for any tree i.e. a graph G on n vertices and more than $\frac{k-2}{2} n$ edges contains any tree on k vertices. A proof of the Erdős-Sós conjecture for large trees was announced by Ajtai, Komlós, Simonovits and Szemerédi.

A natural extension of the problem is the determination of the Turán number of forests. Erdős and Gallai [7] considered the graph H consisting of k independent edges (note that H is a linear forest) and found $\mathrm{ex}(n,H) = \max\{\binom{k-1}{2} + (k-1)(n-k+1), \binom{2k-1}{k}\}$. When n is large enough compared to k, the first bound in the maximum is larger, and the extremal graph attaining this bound is obtained by adding $k-1$ universal vertices to an independent set of $n - k + 1$ vertices. This construction clearly does not contain k independent edges as every such edge must include at least one of the universal vertices. This construction forms a model for the constructions presented throughout the paper. Brandt [3] generalized the above result by proving that any graph G with $e(G) > \max\{\binom{k-1}{2} + (k-1)(n-k+1), \binom{2k-1}{k}\}$ contains every forest on k edges without isolated vertices. An analogue of Erdős-Gallai theorem for linear paths in hypergraphs was considered by Füredi, Jiang and Seiver [12].

Recently, Bushaw and Kettle [5] found (among other things) the Turán number and extremal graph for linear forests where all components have even order as well as for linear forests where all components have the same odd order.

Motivated by these results we determined the Turán number and extremal graph for arbitrary linear forests.

Theorem 2. *Let F be a linear forest with components of order $v_1, v_2, \ldots, v_k$. If at least one v_i is not 3, then for n sufficiently large,*

$$\mathrm{ex}(n, F) = \left(\sum_{i=1}^{k} \left\lfloor \frac{v_i}{2} \right\rfloor - 1 \right) \left(n - \sum_{i=1}^{k} \left\lfloor \frac{v_i}{2} \right\rfloor + 1 \right) + \binom{\sum_{i=1}^{k} \left\lfloor \frac{v_i}{2} \right\rfloor - 1}{2} + c,$$

where $c = 1$ if all v_i are odd and $c = 0$ otherwise. Moreover, the extremal graph is unique.

Notice that the theorem avoids the case of linear forest with every component of order three. This case was solved by Bushaw and Kettle [5] and settled a conjecture of Gorgol [14]. We prove Theorem 2 in Section 2 and describe the unique F-free graph on n vertices with $\mathrm{ex}(n, F)$ edges.

Another motivation is the following conjecture of Goldberg and Magdon-Ismail [13]:

> Let F be a forest with k components, then every graph G with at least $e(F) + k$ vertices and average degree $> e(F) - 1$ contains F as a subgraph.

This is a natural generalization of the Erdős-Sós conjecture, however it is false. From Theorem 2, some simple calculation shows that the conjecture fails for linear forests with at least two components of even order.

We also investigate the other extreme case when each component is a star and we determine the Turán number and also find all the extremal graphs.

Theorem 3. *Let $F = \bigcup_{i=1}^{k} S^i$ be a star forest where d_i is the maximum degree of S^i and $d_1 \geqslant d_2 \geqslant \cdots \geqslant d_k$. For n sufficiently large,*

$$\mathrm{ex}(n, F) = \max_{1 \leqslant i \leqslant k} \left\{ (i-1)(n-i+1) + \binom{i-1}{2} + \left\lfloor \frac{d_i - 1}{2}(n-i+1) \right\rfloor \right\}.$$

Note that for a single component, Theorem 3 describes what is technically the Turán number of a star i.e. the maximum number of edges in a graph of fixed maximum degree. Furthermore, Theorem 3 includes the forest where every component is of order 3 (proved in [5]). In other words a forest of components each of order 3 behaves like a star forest rather than a linear forest. We prove Theorem 3 and characterize all extremal graphs in Section 3. In the proofs of Theorems 2 and 3 we make no attempt to minimize the bound on n. The last section contains a result about forests with small components.

2 Linear forests

We first consider the Turán problem for a linear forest. Throughout this section, unless otherwise specified, F is a linear forest i.e. $F = \bigcup_{i=1}^{k} P^i$, such that P^i is a path on v_i vertices and $v_1 \geqslant v_2 \geqslant \ldots \geqslant v_k \geqslant 2$.

Let F be a linear forest where at least one v_i is not equal to 3. Define $G_F(n)$ to be the graph on n vertices with a set, U, of $\left(\sum_{i=1}^{k} \lfloor v_i/2 \rfloor \right) - 1$ universal vertices together with a single edge in $G_F(n) - U$ if each v_i is odd or $n - |U|$ independent vertices otherwise
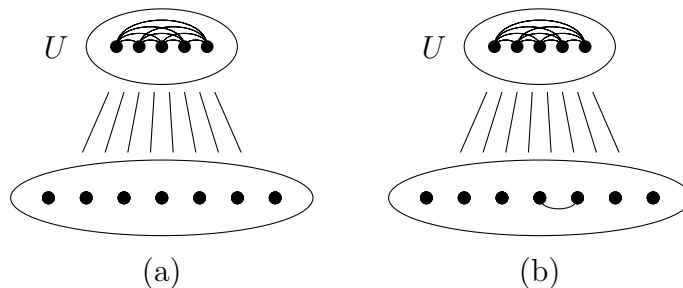
Figure 1: $G_F(n)$, (a) is the case where at least one path F is of even order, (b) is the case where all paths in F are of odd order.

(see Figure 1). Observe that $G_F(n)$ is F-free. Indeed, any path P^i in $G_F(n)$ must use at least $\lfloor v_i/2 \rfloor$ vertices from U but $|U| < \sum_{i=1}^{k} \lfloor v_i/2 \rfloor$. We show that $G_F(n)$ is the unique extremal graph for the linear forest F. We begin with a lemma that is a special case of Lemma 2.3 in [5]. We include a proof for the sake of completeness.

Lemma 4. *Let $p, v \geqslant 2$ be given and let n be at least $4p(p+v)\binom{p}{\lfloor p/2 \rfloor}$. If G is a graph on n vertices with a set P of p vertices such that $e(G) - e(G - P) - \binom{p}{2} \geqslant (\lfloor p/2 \rfloor - \frac{1}{2})n$, then P contains a subset of $\lfloor p/2 \rfloor$ vertices with a common neighborhood of $v + p$ vertices in $G - P$.*

Proof. Let n_0 be the number of vertices in $G - P$ with at least $\lfloor p/2 \rfloor$ neighbors in P. Therefore the number of edges between P and $G-P$ is at most $n_0 p + (n - p - n_0)(\lfloor p/2 \rfloor - 1)$. So

$$ e(G) - e(G - P) - \binom{p}{2} \leqslant n_0 p + (n - p - n_0) \left(\lfloor p/2 \rfloor - 1 \right). $$

Rearranging terms gives $n_0 \geqslant \frac{n/2}{\lceil p/2 \rceil + 1}$. There are $\binom{p}{\lfloor p/2 \rfloor}$ sets of $\lfloor p/2 \rfloor$ vertices in P. Thus there is some set with a common neighborhood of order at least

$$ \frac{n_0}{\binom{p}{\lfloor p/2 \rfloor}} \geqslant v + p. $$

$\square$

First, let us consider the base case when F consists of only two paths. The case when the two paths have the same order is resolved in [5], so let us assume they have different order.

Theorem 5. *Suppose $F = P_a \cup P_b$, with $a > b \geqslant 2$. Let G be any F-free n-vertex graph with $n \geqslant 8a^2(a+b)^2\binom{a}{\lfloor a/2 \rfloor}^2$. Then $e(G) \leqslant e(G_F(n))$, with equality only when $G \simeq G_F(n)$.*

We use a standard trick to reduce the problem to graphs with large minimum degree.

Lemma 6. *Suppose $F = P_a \cup P_b$, with $a > b \geqslant 2$. Let G be any F-free n-vertex graph with $n \geqslant 4a(a+b)\binom{a}{\lfloor a/2 \rfloor}$ and $\delta(G) \geqslant \lfloor a/2 \rfloor + \lfloor b/2 \rfloor - 1$. Then $e(G) \leqslant e(G_F(n))$, with equality only when $G \simeq G_F(n)$.*

We first show how Lemma 6 implies Theorem 5 and we give the proof of the lemma afterwards.

Proof of Theorem 5 using Lemma 6. Suppose G is an extremal graph for F on $n \geqslant 8a^2(a+b)^2\binom{a}{\lfloor a/2 \rfloor}^2$ vertices and $e(G) \geqslant e(G_F(n))$. We start by removing vertices of small degree from G. Suppose that there exists a vertex v in G with $d(v) < \delta(G_F(n)) = \lfloor a/2 \rfloor + \lfloor b/2 \rfloor - 1$. Let $G^n = G$. We then define $G^{n-1} = G^n - \{v\}$. We construct each successive G^{i-1} from G^i by removing a vertex of degree less than $\delta(G_F(i))$. The process continues while $\delta(G^i) < \delta(G_F(i))$. Notice that

$$e(G^{i-1}) - e(G_F(i-1)) \geqslant e(G^i) - e(G_F(i)) + 1.$$

Hence the process terminates after $n - \ell$ steps. We get G^ℓ with $\delta(G^\ell) \geqslant \delta(G_F(\ell))$ and

$$\binom{\ell}{2} \geqslant e(G^\ell) \geqslant e(G_F(\ell)) + n - \ell = (\lfloor a/2 \rfloor + \lfloor b/2 \rfloor - 1)\ell + n - \ell + O(a^2b^2) \geqslant n,$$

this implies $\ell > \sqrt{2n} \geqslant 4a(a+b)\binom{a}{\lfloor a/2 \rfloor}$. Since $e(G^\ell) > e(G_F(\ell))$, Lemma 6 then implies $F \subseteq G^\ell \subseteq G$, a contradiction. □

We give two useful lemmas before proving Lemma 6. Let H be a graph and let $U \subset V(H)$ such that U has at least $|U| + 3$ common neighbors in $V(H) - U$. Furthermore, let $1 \leqslant c \leqslant |U| + 1$.

Lemma 7. *Let $v \in N(U)$ be an endpoint of P_t in $H - U$ where $1 \leqslant t \leqslant 3$. Then H contains a path P of order $2c - 2 + t$ such that $|P \cap U| = c - 1$.*

Proof. Let $u \in U$ be a neighbor of v. Let P' be a path in H of order $2c - 2$ starting at u and avoiding vertices of P_t and having at most $c - 1$ vertices of U. It can be obtained from U and $N(U) \setminus V(P_t)$ by alternating between U and $N(U) \setminus V(P_t)$ since U has at least $|U| + 3$ common neighbors and $|U| + 3 \geqslant |U| + t$. The path P_t can be used to extend P' and finding a path of order $2c - 2 + t$. See Figure 2(a) for case $t = 3$. □

Lemma 8. *If there exist nonadjacent $u, v \in N(U)$ both of degree at least one in $H - U$ then H contains a path P of order $2c + 1$ such that $|P \cap U| = c - 1$.*

Proof. If there is a common neighbor of u and v, then the claim follows from Lemma 7 because u is an endpoint of a path of order three. Let u' and v' be distinct neighbors of u and v, respectively, in $H - U$. A greedily obtainable path with end vertices u and v avoiding u' and v' of order $2c - 1$ can be extended by u' and v' to a path of order $2c + 1$. See Figure 2(b). □
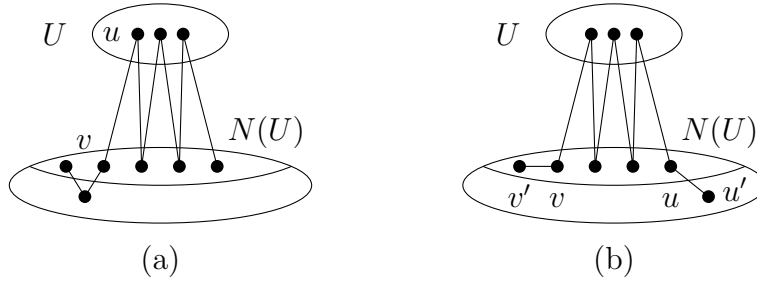
U $\quad$ u $\bullet\bullet\bullet$

N(U)

v

(a)

U

N(U)

v' $\quad$ v $\qquad$ u $\quad$ u'

(b)

Figure 2: Finding a long path.

Proof of Lemma 6. Let G be an extremal graph for F with $\delta(G) \geqslant \lfloor a/2 \rfloor + \lfloor b/2 \rfloor - 1$. First we show that there is a set U_a of $\lfloor a/2 \rfloor$ vertices with a common neighborhood of more than $a + b$ vertices. Such a set U_a can be easily extended to a copy of P_a, which implies $G - U_a$ must be P_b-free. However, $G - U_a$ is not P_{b-2}-free and hence we can find a set U_b of $\lfloor b/2 \rfloor - 1$ vertices that share a large common neighborhood. Then we show that $G - U_a - U_b$ has at most one edge, which then implies $G \subseteq G_F(n)$.

Because $e(G) \geqslant e(G_F(n)) > \mathrm{ex}(n, P_a)$, we have $P_b \subseteq P_a \subseteq G$. The graph G does not contain the forest $P_a \cup P_b$, so the graph $G - P_a$ is P_b-free. Thus $e(G - P_a) \leqslant \frac{b-2}{2}(n-a)$ by Theorem 1 (Erdős-Gallai). Because $e(G) \geqslant e(G_F(n))$, the number of edges between P_a and $G - P_a$ is at least $e(G_F(n)) - \frac{b-2}{2}(n-a) - \binom{a}{2}$. Thus, by Lemma 4, there is a set $U_a \subset P_a$ such that $|U_a| = \lfloor a/2 \rfloor$ and U_a has at least $a + b$ many common neighbors.

Now observe that $G - U_a$ is P_b-free. Indeed, suppose there is a copy of P_b in $G - U_a$, then U_a can be extended to P_a avoiding vertices of P_b, since U_a has more than $a + b$ common neighbors. Thus $F \subseteq G$ which is a contradiction.

We now distinguish three cases based on the value of b. If $b = 2$, then $G - U_a$ is P_2-free, namely $G - U_a$ is an independent set. Thus $G \subseteq G_F(n)$.

If $b = 3$, then $G - U_a$ is P_3-free. Thus $G - U$ contains only isolated edges and vertices. If uv is an isolated edge in $G - U_a$ then both u and v have a neighbor in U_a as the minimum degree of G is $\delta(G) \geqslant \lfloor a/2 \rfloor \geqslant 2$. Furthermore, at most one of u and v is not adjacent to all vertices of U_a, otherwise the graph G_0 obtained from G by removing the edge uv and adding all edges between $\{u, v\}$ and U_a has more edges and remains F-free, which contradicts the extremality of G. The graph G_0 is indeed F-free. If not, then a copy of F in G_0 contains at least one of u, v. However, u and v are isolated vertices in $G_0 - U_a$. Hence they could be replaced by some of the $a + b$ common neighbors of U_a, which would give a copy of F in G, a contradiction. Let z be a vertex in U_a. If a is even, then $U_a \setminus z$ together with an edge in $G - U_a$ complete a copy of P_a, and $N(z) \cup \{z\}$ contains a $P_3 = P_b$ (see Figure 3(a)), thus $F \subseteq G$, contradiction. So $G - U_a$ is an independent set if a is even. When a is odd, then two edges in $G - U_a$ together with $U_a \setminus z$ completes a copy of P_a, and again $N(z) \cup \{z\}$ contains a P_b (see Figure 3(b)), thus $G - U_a$ has at most one edge. Therefore, $G \subseteq G_F(n)$.

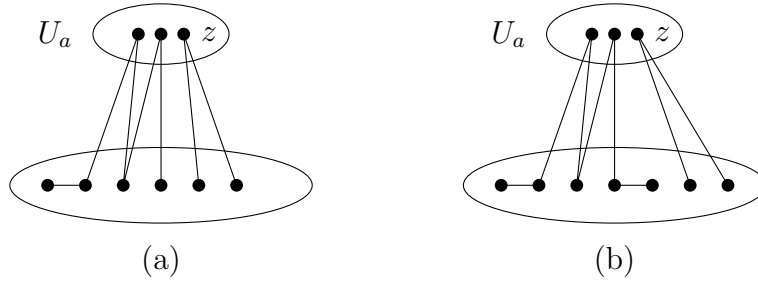Now we may assume $b \geqslant 4$. Let $|V(G - U_a)| = n'$. Denote by G' the graph obtained

Figure 3: Cases for $b = 3$.

by deleting $\lfloor a/2 \rfloor$ universal vertices from $G_F(n)$, clearly $|V(G')| = n'$.

If $G - U_a$ is P_{b-2}-free, then $e(G - U_a) \leqslant (\frac{b}{2} - 2)n' < (\lfloor \frac{b}{2} \rfloor - 1)n' - \binom{\lfloor \frac{b}{2} \rfloor}{2} = e(G')$, implying $e(G) < e(G_F(n))$ which is a contradiction. Thus we may assume $G - U_a$ is not P_{b-2}-free. Take a maximum path, P, in $G - U_a$ and let u be an end vertex of P. Since $G - U_a$ is P_b-free, P has at most $b - 1$ vertices. Then

$$d(u) - (|V(P)| - 1) \geqslant \delta(G) - b + 2 \geqslant \lfloor a/2 \rfloor + \lfloor b/2 \rfloor - 1 - b + 2 = \lfloor a/2 \rfloor - \lceil b/2 \rceil + 1 \geqslant 1.$$

Thus u has a neighbor, say w, not in P. Furthermore w must be in U_a, since otherwise we could extend P to a longer path in $G - U_a$. Recall that U_a has more than $a + b$ common neighbors. Then P with w and a neighbor of w in $G - U_a - P$ form a P_b. By Lemma 4 there is a set of $\lfloor b/2 \rfloor$ vertices of P_b with a common neighborhood of order at least $a + b$ (by the same argument as before). Note that all vertices in this P_b except w are in $G - U_a$. Thus we find a set of vertices in $G - U_a$ of order $\lfloor b/2 \rfloor - 1$, call it U_b, with a large common neighborhood (U_b is nonempty since $b \geqslant 4$). We are done if we show $G - U_a - U_b$ is empty if one of a and b is even or has at most one edge when they are both odd, as this implies $G \subseteq G_F(n)$.

Suppose that G is not connected. Then there exists a connected component C in G not containing U_a. The minimum degree $\lfloor a/2 \rfloor + \lfloor b/2 \rfloor - 1 \geqslant b - 1$ implies that it is possible to find P_b in C (for example by a greedy algorithm). It contradicts the claim that $G - U_a$ is P_b-free. Hence we assume that G is connected.

Notice that if a vertex is not adjacent to all vertices of $U_a \cup U_b$ then it has degree at least 1 in $G - U_a - U_b$ because of the minimum degree condition. If there are at least three such vertices, Lemma 7 or Lemma 8 applies with $H = G - U_a$, $U = U_b$ and $c = \lfloor \frac{b}{2} \rfloor$. It leads to the existence of a path of order $2\lfloor \frac{b}{2} \rfloor + 1$ in $G - U_a$ which is a contradiction. Hence there are at most two vertices in $N(U_b) \setminus U_a$ that are not adjacent to all vertices of U_a.

Put $U = U_a \cup U_b$. The previous observation together with the fact that vertices of U_b have many common neighbors implies that there is a common neighborhood of U of order at least $\frac{n}{b\binom{b}{b/2}} - 2$. Hence whenever we find in G a path of order a using at most $\lfloor \frac{a}{2} \rfloor - 1$ vertices of U or a path of order b using at most $\lfloor \frac{b}{2} \rfloor - 1$ vertices of U then we can easily find the other path of F.

If a vertex v in $G - U$ is not adjacent to any vertex of U then by connectivity and minimum degree condition of G it is easy to find a vertex $u \in N(U)$ such that u is an end vertex of a path of order 3. Hence Lemma 7 applies and gives a path of order b using $\lfloor \frac{b}{2} \rfloor - 1$ vertices of U which is a contradiction. Thus each vertex in $G - U$ is adjacent to a vertex of U.

Therefore $N(U)$ is exactly $G - U$. Lemma 7 implies that $G - U$ is P_3-free and together with Lemma 8 this implies that $G - U$ contains at most one edge. Moreover, if a or b is even, Lemma 7 implies that $G - U$ is P_2-free. Hence $G - U$ contains no edges if a or b is even and contains at most one edge if both a and b are odd. Therefore, G is a subgraph of $G_F(n)$. This completes the proof of Lemma 6. $\qquad\square$

Now we are ready to prove Theorem 2.

Proof of Theorem 2. We proceed by induction on k, the number of components of F. The base case, $k = 2$, is given by Theorem 5. Assume the theorem is true for $2 \leqslant \ell \leqslant k - 1$. Pick j such that $1 \leqslant j \leqslant k$ and $F' = F - P^j$ is not $(k-1) \cdot P_3$. Let G be an extremal graph for F. By Theorem 1 (Erdős-Gallai) G must contain a P^j. Clearly $G - P^j$ is F'-free, thus $e(G - P^j) \leqslant e(G_{F'}(n - v_j))$. Simple algebra and applying induction to $G - P^j$ gives $e(G) - e(G - P^j) - \binom{v_j}{2} \geqslant \lfloor v_j/2 \rfloor n$ i.e. the number edges between P^j and $G - P^j$ is at least $(\lfloor v_j/2 \rfloor - \frac{1}{2})n$. Thus, by Lemma 4, P^j contains a set U_j of $\lfloor v_j/2 \rfloor$ vertices who have at least $\sum_{i=1}^{k} v_i$ neighbors in $G - P^j$ (for n large enough). Therefore $G - U_j$ must be F'-free and by the induction hypothesis $G - U_j$ is $G_{F'}$, so $G \subseteq G_F(n)$. Hence $G_F(n)$ is the unique extremal graph. $\qquad\square$

3   Star forests

In this section we give a proof of Theorem 3. Without going into all the details we remark that our proof gives that the size of n for which the result holds depends on $\sum_{i=1}^{k} d_i$.

Let $F = \bigcup_{i=1}^{k} S^i$ be a star forest as in the statement of Theorem 3. Let d_i be the maximum degree of S^i. Recall that $d_1 \geqslant d_2 \geqslant \cdots \geqslant d_k$.

We begin by describing the extremal graph for F. Let $F(n, i)$ be a graph obtained by adding a set U of $i - 1$ universal vertices to an extremal graph, H, for S^i on $n - i + 1$ vertices (see Figure 4). Observe that H is a $(d_i - 1)$-regular graph if one of $(d_i - 1)$ and $n - i + 1$ is even and n large enough (see [16]). If both are odd, then H has exactly one vertex of degree $(d_i - 2)$ and the remaining vertices have degree $(d_i - 1)$. Therefore, we have $e(H) = \lfloor \frac{d_i - 1}{2}(n - i + 1) \rfloor$.

Observe that for all $1 \leqslant i \leqslant k$, $F(n, i)$ is F-free. Indeed, each star $S^1, \ldots, S^{i-1}$ must have at least one vertex from U and S^i is not a subgraph of $F(n, i) - U$.

Throughout the proof, unless otherwise specified, i is always the index maximizing the number of edges. Notice that $e(F(n, i))$ is equal to the Turán number claimed by the theorem, i.e.

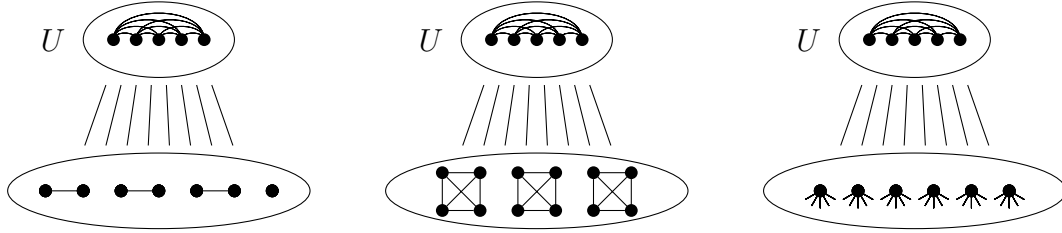$$e(F(n, i)) = \left\lfloor \left(i - 1 + \frac{d_i - 1}{2}\right) n - \frac{i - 1}{2}(i + d_i - 1) \right\rfloor.$$

Figure 4: The extremal graph for a star-forest.

Define $f_i = i - 1 + \frac{d_i - 1}{2}$, namely f_i is the coefficient of n, the leading term of $e(F(n, i))$. We first show that for any $j < i$, $f_j < f_i$. Indeed, if $f_j > f_i$, then j would be the index maximizing the number of edges, a contradiction. Thus we may assume $f_j = f_i$, namely by definition, $i - 1 + \frac{d_i - 1}{2} = j - 1 + \frac{d_j - 1}{2}$ and

$$d_j = d_i + 2(i - j). \tag{1}$$

Thus we have:

$$
\begin{aligned}
e(F(n, j)) - e(F(n, i)) &\geqslant \frac{i - 1}{2}(i + d_i - 1) - \frac{1}{2} - \frac{j - 1}{2}(j + d_j - 1) \\
&= \frac{i - 1}{2}(i + d_i - 1) - \frac{1}{2} - \frac{j - 1}{2}(j + d_i + 2(i - j) - 1) \quad \text{by (1)} \\
&= \frac{i}{2}(d_i + i - 1) - \frac{j}{2}(d_i + 2i - j - 1) + \frac{1}{2}(i - j) - \frac{1}{2} \\
&= \frac{i}{2}(d_i + i) - \frac{j}{2}(d_i + i - j) - \frac{ji}{2} - \frac{1}{2} \\
&= \frac{i - j}{2}(d_i + i - j) - \frac{1}{2} \\
&> 0
\end{aligned}
$$

But then $e(F(n, j)) > e(F(n, i))$, contradicting the choice of i.

We use induction on k, the number of components of F. We prove the theorem in three cases distinguished by the index i that maximizes the number of edges: (1) $i = k$, (2) $1 < i < k$, and (3) $i = 1$. Throughout this section we say a vertex has *linear degree* if the degree of the vertex is linear in n.

Case 1: $i = k$

Let $F' = F - S^k$. Let G be an extremal graph for F with n vertices. By the induction hypothesis, we have $\mathrm{ex}(n, F') = e(F(n, i'))$, where i' is the index maximizing the number of edges. Since $i = k$, we have $i' < i$. Then as shown above, we have

$$f_{i'} < f_i = f_k. \tag{2}$$

Note that $F(n, i)$ is F-free, we have

$$\mathrm{ex}(n, F') = f_{i'} n + O(d_{i'}) < f_k n + O(d_k) = e(F(n, k)) \leqslant e(G).$$

Thus $F' = S^1 \cup S^2 \cup \cdots \cup S^{k-1} \subseteq G$ by the induction hypothesis.

It suffices to prove that there exists a vertex subset $U \subseteq V(G)$ of order $k-1$, such that every vertex in U has linear degree, that is, $d(v) = \Omega(n)$. Indeed, if such a U exists, then $G - U$ must be S^k-free. Otherwise, say there is a S^k in $G - U$, then we can get F' using vertices in U as centers and their neighbors in $G - U - S^k$ as leaves, which gives a copy of $F' \cup S^k = F$, which is a contradiction. Hence $G \subseteq F(n, k)$ as desired.

Now we prove that such a U exists. We know $F' \subseteq G$, namely there are $k-1$ disjoint stars in G. Take any one of them, say S^j, $1 \leqslant j \leqslant k-1$. Notice that $G - S^j$ has to be F'-free, since otherwise a copy of F' in $G - S^j$ together with S^j yields a copy of $S^1 \cup S^2 \cup \cdots \cup S^{k-1} \cup S^j$. Since $S^k \subseteq S^j$ we get that $F \subseteq G$, which is a contradiction. Note that $e(G[S^j]) \leqslant \binom{d_j+1}{2}$ and $e(G) \geqslant e(F(n,k))$. Let e_0 be the number of edges between S^j and $G - S^j$. Then we have

$$
\begin{aligned}
e_0 &= e(G) - e(G - S^j) - e(G[S^j]) \\
&\geqslant e(F(n,k)) - \mathrm{ex}(n, F') - \binom{d_j + 1}{2} \\
&\sim f_k n - f_{i'} n = \Omega(n) \qquad\qquad\qquad\qquad \text{by (2).}
\end{aligned}
$$

Thus there is a vertex in S^j with linear degree. This is true for every j with $1 \leqslant j \leqslant k-1$, so consider a set containing a vertex of linear degree in each star. These $k-1$ vertices form the desired set U.

This finishes the proof of Case 1.

Case 2: $1 < i < k$

Let $F^* = S^i \cup S^{i+1} \cup \cdots \cup S^k$ and $F' = F - F^*$. As before, if i' is the index maximizing the number of edges for extremal graph of F', then $f_{i'} < f_i$. A crucial observation about F^* is the following.

$$
\mathrm{ex}(n, F^*) = \left\lfloor \frac{d_i - 1}{2} n \right\rfloor. \tag{3}
$$

To show (3), notice that i is the index maximizing the number of edges for F, thus for any $\ell > i$, $f_\ell \leqslant f_i$ or $\ell - 1 + \frac{d_\ell - 1}{2} \leqslant i - 1 + \frac{d_i - 1}{2}$. This implies

$$
\ell - i + \frac{d_\ell - 1}{2} \leqslant \frac{d_i - 1}{2}. \tag{4}
$$

By the induction hypothesis, if i^* with $i \leqslant i^* \leqslant k$ is the index maximizing the size of the extremal graph for $F^* = S^i \cup S^{i+1} \cup \cdots \cup S^k$, we have,

$$
\begin{aligned}
\mathrm{ex}(n, F^*) &= \max_{i \leqslant \ell \leqslant k} \left\{ \left[\ell - 1 + \frac{d_\ell - 1}{2} \right] n - \left\lfloor \frac{\ell - 1}{2}(\ell + d_\ell - 1) \right\rfloor \right\} \\
&= \left[i^* - 1 + \frac{d_{i^*} - 1}{2} \right] n - \left\lfloor \frac{i^* - 1}{2}(i^* + d_{i^*} - 1) \right\rfloor \\
&\leqslant \left[i^* - 1 + \frac{d_{i^*} - 1}{2} \right] n \\
&\leqslant \frac{d_i - 1}{2} n \qquad\qquad\qquad\qquad\qquad\qquad \text{by (4) and } i^* \geqslant i.
\end{aligned}
$$

On the other hand, a S^i-free graph is F^*-free, thus $\mathrm{ex}(n, F^*) \geqslant \left\lfloor \frac{d_i-1}{2}n \right\rfloor$.

Let G be an extremal graph for F on n vertices. Recall $F' = S^1 \cup \cdots \cup S^{i-1}$, the choice of i and the extremality of G implies $e(G) \geqslant e(F(n,i)) > \mathrm{ex}(n, F')$. Thus $F' \subseteq G$. As before if we can show there is a vertex with linear degree in each star in $F' \subseteq G$, then we have a set U of order $i-1$, each vertex of which has linear degree and $G - U$ is F^*-free. Then by (3) we get $G \subseteq F(n,i)$.

Take any star in a copy of F' in G, say S^j, note that $j < i$. We take the same approach as before to prove that the number of edges between S^j and $G - S^j$ is linear, namely $e_0 = \Omega(n)$, which implies the existence of a vertex with linear degree as desired.

Note that $G - S^j$ must be $(F - S^j)$-free. We now give an upper bound on $e(G - S^j)$. Let $F'' = F - S^j$ and let i'' be the index maximizing the number of edges for F''.

If $i'' < j < i$, then $f_{i''} < f_i$ and thus

$$e(G - S^j) \leqslant \mathrm{ex}(n, F'') \leqslant f_{i''}n.$$

Therefore, $e_0 = e(G) - e(G[S^j]) - e(G - S^j) \geqslant f_i n - f_{i''}n - \binom{d_j+1}{2} = \Omega(n)$ as desired.

Notice that in F'', all indices after j were shifted to the left by one, that is $F'' = S^1 \cup \cdots \cup S^{j-1} \cup S^{j+1} \cup \cdots \cup S^k$. Thus if $i'' \geqslant j$, by the definition of f_i, then it is the same index as F that maximizes the number of edges for F'', that is i in F and $i-1$ in F''.

Thus $i'' = i - 1$. In this case $e(G - S^j) \leqslant [(i-1) - 1] + \frac{d_i-1}{2} = f_i - 1$. Thus $e_0 \geqslant f_i n - f_{i''}n - \binom{d_j+1}{2} = n - \binom{d_j+1}{2} = \Omega(n)$ as desired.

This finishes the proof of Case 2.

Case 3: $i = 1$.

Let G an extremal graph for F. We want to show that $e(G) = \mathrm{ex}(n, F) = \left\lfloor \frac{d_1-1}{2}n \right\rfloor$. Since an S^1-free graph is F-free, $e(G) \geqslant \left\lfloor \frac{d_1-1}{2}n \right\rfloor$.

We may assume $\Delta(G) \geqslant d_1$, since otherwise $e(G) \leqslant \frac{d_1-1}{2}n$. Let v be a vertex of degree $\Delta(G)$, so $d(v) \geqslant d_1$. Thus we can get a S^1 from $N(v) \cup \{v\}$ with v as its center. Note that since $i = 1$, we have for any $j > 1$, $f_j \leqslant \frac{d_1-1}{2}$. Let $F^* = F - S^1 = S^2 \cup \cdots \cup S^k$. Observe that in F^*, all indices were shifted to the left by one. Hence if i^* is the index maximizing the number of edges for F^*, then it was $j = i^* + 1 \geqslant 2$ in F. Thus $f_{i^*} = [(j-1) - 1] + \frac{d_j-1}{2} = f_j - 1 \leqslant \frac{d_1-1}{2} - 1$.

Let e_0 be the number of edges between S^1 and $G - S^1$, then $e_0 = e(G) - e([S^1]) - e(G - S^1) \geqslant \frac{d_1-1}{2}n - \binom{d_1+1}{2} - f_{i^*}n \geqslant n - \binom{d_1+1}{2} = \Omega(n)$. Thus there is a vertex of linear degree in S^1, let it be u. This implies $G - \{u\}$ is F^*-free. Thus

$$e(G) = d(u) + e(G - u) \leqslant n - 1 + \mathrm{ex}(n, F^*) \leqslant n - 1 + f_{i^*}n \leqslant \frac{d_1 - 1}{2}n - 1 < e(F(n,1)),$$

which is a contradiction.

This finishes the proof of Case 3 and therefore Theorem 3. $\qquad \square$

4 Forests with components of order 4

Now we consider forests with components each of order 4. Notice that there are only two trees of order 4: the path P_4 and the star S_3. Let $F = a \cdot P_4 \cup b \cdot S_3$. Let $G_F^1(n)$ be the

n-vertex graph constructed as follows: assume $n - b = 3d + r$ with $r \leqslant 2$, $G_F^1(n)$ contains b universal vertices, the remaining graph is $K_r \cup d \cdot K_3$. Let $G_F^2(n)$ be the n-vertex graph containing $2a + b - 1$ universal vertices and the remaining graph is empty. See Figure 5. It is easy to check that both $G_F^1(n)$ and $G_F^2(n)$ are F-free as the set of universal vertices is too small to accommodate all components of F.
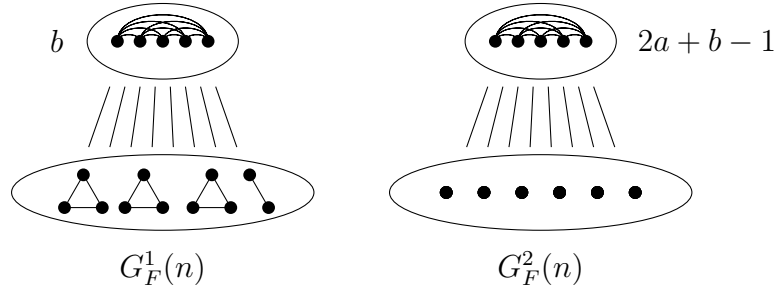

Figure 5: The extremal graphs for a forest with components of order 4.

Theorem 9. *Let $F = a \cdot P_4 \cup b \cdot S_3$, and $n = 3d + r$ be sufficiently large with $r \leqslant 2$, then*

(i) *If $a = 1$ and $r = 0$, then $G_F^1(n)$ is the unique extremal graph; if $a = 1$ and $r \neq 0$, then $G_F^1(n)$ and $G_F^2(n)$ are the only extremal graphs for F.*

(ii) *If $a > 1$, then $G_F^2(n)$ is the unique extremal graph for F.*

We only give a sketch of the proof. Use induction on b, the number of copies of S_3. Let G be an extremal graph with n vertices, then $e(G) \geqslant e(G_F^i(n))$, $i = 1, 2$. Thus G contains a copy of S_3, similar as the proof for star-forest, any copy of S_3 in G contains a vertex, say v, of degree $\Omega(n)$. Then $G - \{v\}$ must be F'-free where $F' = a \cdot P_4 \cup (b - 1) \cdot S_3$. Then by the inductive hypothesis, $G - \{v\}$ has to be $G_{F'}^i(n - 1)$, which then implies $G \subseteq G_F^i(n)$. The base case of the induction is when $b = 0$, then the graph is $a \cdot P_4$. This also explains why we have two different constructions for the extremal graphs. Because when $a = 1$, by a result of Faudree and Schelp [10], vertex-disjoint copies of triangles or a star are the extremal graphs (actually, they showed a combination of triangles and a smaller star is also an extremal case, but here, if any triangle appears, then the number of universal vertices will be fewer which yields a construction worse than G_F^1 and G_F^2); when $a > 1$, then Theorem 2 implies that $G_F^2(n)$ is the unique extremal graph.

The same technique can be applied to $a \cdot P_\ell \cup b \cdot S_t$, but the proof is very technical.

Acknowledgments

References

[1] B. Bollobás. *Modern Graph Theory.* Springer Science+Business Meda, Inc., USA, 3rd edition, 1998.

[2] J. A. Bondy and M. Simonovits. Cycles of even length in graphs. *J. Combinatorial Theory Ser. B*, 16:97–105, 1974.

[3] S. Brandt. Subtrees and subforests of graphs. *J. Comb. Theory, Ser. B*, 61(1):63–70, 1994.

[4] W. G. Brown. On graphs that do no contain a Thomsen graph. *Canad. Math. Bull.*, 9:281–289, 1966.

[5] N. Bushaw and N. Kettle. Turán numbers of multiple paths and equibipartite forests. *Combinatorics, Probability and Computing*, 20:837–853, 2011.

[6] P. Erdős. Extremal problems in graph theory. *Theory of Graphs and its Applications (Proc. Sympos. Smolenice, 1963)*, pages 29–36, 1964.

[7] P. Erdős and T. Gallai. On maximal paths and circuits of graphs. *Acta Math. Acad. Sci. Hungar.*, 10:337–356, 1959.

[8] P. Erdős, A. Rényi, and V. T. Sós. On a problem of graph theory. *Studia Sci. Math. Hungar.*, 1:215–235, 1966.

[9] P. Erdős and A. H. Stone. On the structure of linear graphs. *Bull. Amer. Math. Soc.*, 52:1087–1091, 1946.

[10] R. J. Faudree and R. H. Schelp. Path Ramsey numbers in multicolorings. *J. Combinatorial Theory Ser. B*, 19:150–160, 1975.

[11] Z. Füredi. New asymptotics for bipartite Turán numbers. *J. Combin. Theory Ser. A*, 75:141–144, 1996.

[12] Z. Füredi, T. Jiang, and R. Seiver. Exact solution of the hypergraph Turán problem for k-uniform linear paths, 2011. submitted. arXiv:1108.1247

[13] M. K. Goldberg and M. Magdon-Ismail. Embedding a forest in a graph. *Electr. J. Comb.*, 18(1):#P99, 2011.

[14] I. Gorgol. Turán numbers for disjoint copies of graphs. *Graphs and Combinatorics*, 27(5):661–667, 2011.

[15] T. Kővári, V. T. Sós, and P. Turán. On a problem of K. Zarankiewicz. *Colloquium Math.*, 3:50–57, 1954.

[16] M. Simonovits. A method for solving extremal problems in graph theory, stability problems. *Theory of Graphs (Proc. Colloq., Tihany, 1966)*, pages 279–319, 1968.

[17] P. Turán. Egy gráfelméleti szélsőértékfeladatról. *Mat. es Fiz. Lapok*, 48:436–452, 1941.

[18] P. Turán. On the theory of graphs. *Colloquium Math*, 3:19–30, 1954.