Combinatorics, Probability and Computing (2011) 20, 837–853. © Cambridge University Press 2011
doi:10.1017/S0963548311000460

Turán Numbers of Multiple Paths and Equibipartite Forests

NEAL BUSHAW[1] and NATHAN KETTLE[2]

[1]Department of Mathematical Sciences, University of Memphis, Memphis, TN 38152, USA
(e-mail: nobushaw@memphis.edu)

[2]Department of Pure Mathematics and Mathematical Statistics,
University of Cambridge, Cambridge CB3 0WB, UK
(e-mail: n.kettle@dpmms.cam.ac.uk)

Received 29 June 2011; revised 8 September 2011; first published online 12 October 2011

The *Turán number* of a graph H, $\mathrm{ex}(n, H)$, is the maximum number of edges in any graph on n vertices which does not contain H as a subgraph. Let P_l denote a path on l vertices, and let $k \cdot P_l$ denote k vertex-disjoint copies of P_l. We determine $\mathrm{ex}(n, k \cdot P_3)$ for n appropriately large, answering in the positive a conjecture of Gorgol. Further, we determine $\mathrm{ex}(n, k \cdot P_l)$ for arbitrary l, and n appropriately large relative to k and l. We provide some background on the famous Erdős–Sós conjecture, and conditional on its truth we determine $\mathrm{ex}(n, H)$ when H is an equibipartite forest, for appropriately large n.

1. Introduction

Our notation in this paper is standard (see, *e.g.*, [3]). Thus $G \cup H$ denotes the disjoint union of graphs G and H, and $k \cdot G$ denotes k disjoint copies of G. We write $G + H$ for the join of G and H, the graph obtained from $G \cup H$ by adding edges between all vertices of G and all vertices of H, K_t for the complete graph on t vertices, E_t for the empty graph on t vertices, and M_t for a maximal matching on t vertices; that is, the graph on t vertices consisting of $\lfloor \frac{t}{2} \rfloor$ independent edges. We will use $N(S)$ to denote the neighbourhood of the vertex set S, those vertices which are adjacent to some element of S. We also take this opportunity to point out that unless explicitly stated, any graph named G is assumed to be on vertex set $V = [n]$ and edge set E; we also make no requirement that the subgraphs we find be induced.

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 for assorted graphs traces its history back to 1907, when Mantel (see, *e.g.*, [3]) proved that the maximum number of edges in an n-vertex triangle-free graph is $\lfloor \frac{n^2}{4} \rfloor$. In 1940, Pál Turán [13, 14] proved that the extremal graph avoiding K_r as a subgraph is the complete $(r-1)$-partite graph on n vertices which is 'as balanced as possible': this is

https://doi.org/10.1017/S0963548311000460 Published online by Cambridge University Press

the Turán graph $T_{r-1}(n)$. Later, Simonovits [12] showed that for large n, the extremal graph forbidding $p \cdot K_r$ is $K_{p-1} + T_{r-1}(n-p+1)$. For notation, $H_{\text{ex}}(n, G)$ will be used to represent a graph on n vertices with no copy of G as a subgraph, and exactly $\text{ex}(n, G)$ edges. We note that, in general, the extremal graph(s) may not be unique.

Recently, Gorgol [6] proved upper and lower bounds on the extremal number for forbidding several vertex-disjoint copies of an arbitrary connected graph. We determine this number for paths of length 3 in Section 2.1, longer paths in Section 2.2, and for forests of equibipartite trees in Section 3.2. We also provide some background on the Erdős–Sós conjecture in Section 3.1, as our result for trees is conditional on its validity.

2. Extremal numbers for disjoint paths

We start by looking at graphs with no disjoint paths of length three. The extremal case here is slightly different than for longer paths, but the proof introduces the main ideas we shall use in proving the result for all paths, as well as the general tools needed for our results on forests.

2.1. Paths of length 3

As a starting point, we state the following trivial lemma.

Lemma 2.1. *If G is a graph on n vertices which contains no P_3, then G contains at most $\lfloor \frac{n}{2} \rfloor$ edges; that is,* $\text{ex}(n, P_3) = \lfloor \frac{n}{2} \rfloor$.

Proof. If G contains no P_3, then no vertex can have degree $\geqslant 2$, and so G consists of independent edges, giving $\text{ex}(n, P_3) \leqslant \lfloor \frac{n}{2} \rfloor$. Clearly this maximum number of edges is obtained by a perfect matching when n is even and a matching leaving one vertex uncovered when n is odd, and thus the lemma holds. □

Gorgol [6] gave constructions giving the following lower bound regarding the extremal number forbidding several paths of length three:

$$\text{ex}(n, k \cdot P_3) \geqslant \begin{cases} \binom{3k-1}{2} + \lfloor \frac{n-3k+1}{2} \rfloor & \text{for } \; 3k \leqslant n < 5k-1, \\ \binom{k-1}{2} + (n-k+1)(k-1) + \lfloor \frac{n-k+1}{2} \rfloor & \text{for } \; n \geqslant 5k-1. \end{cases}$$

Remark. This bound is obtained by noting that, for any connected graph G on v vertices, and for any positive integers n, k such that $n \geqslant kv$, the graphs $H_{\text{ex}}(n-kv+1, G) \cup K_{kv-1}$ and $H_{\text{ex}}(n-k+1, G) + K_{k-1}$ do not contain k vertex-disjoint copies of G. Applying this to forbidding copies of K_3 and counting the edges in these graphs gives the above bound.

Gorgol conjectured that this is the correct value of $\text{ex}(n, k \cdot P_3)$, and proved that this is indeed true for $k = 2, 3$. Our first result shows that the second construction is best possible for any k and large enough n.

Theorem 2.2. $\text{ex}(n, k \cdot P_3) = \binom{k-1}{2} + (n-k+1)(k-1) + \lfloor \frac{n-k+1}{2} \rfloor$, *for $n \geqslant 7k$.*

https://doi.org/10.1017/S0963548311000460 Published online by Cambridge University Press

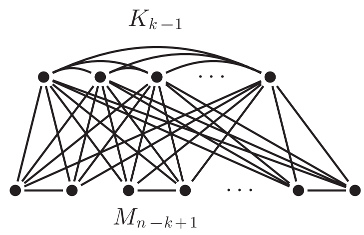

Figure 1. $H_{\text{ex}}(n, k \cdot P_3)$.

There is a unique graph for which this bound is attained, namely $K_{k-1} + M_{n-k+1}$, as in Figure 1. This graph does not contain k disjoint copies of P_3, since each P_3 must contain at least one vertex from the $(k-1)$-clique.

Proof. We proceed by induction on k. The case $k = 1$ is covered by Lemma 2.1. For the induction step, suppose G is a graph on n vertices, with $m > \binom{k-1}{2} + (n-k+1)(k-1) + \lfloor \frac{n-k+1}{2} \rfloor$ edges, and containing no $k \cdot P_3$. The number of edges incident to any P_3 in G must be at least

$$\begin{aligned} m - \text{ex}(n-3, (k-1) \cdot P_3) \geqslant \binom{k-1}{2} + (n-k+1)(k-1) + \left\lfloor \frac{n-k+1}{2} \right\rfloor + 1 \\ - \binom{k-2}{2} - (n-k-1)(k-2) - \left\lfloor \frac{n-k-1}{2} \right\rfloor \\ = n + 2k - 3. \end{aligned}$$

Otherwise, the graph induced by the vertices not on this P_3 contains $(k-1) \cdot P_3$ by induction, showing that G does contain $k \cdot P_3$.

By the induction hypothesis we can find $k-1$ vertex-disjoint copies of P_3 in our graph, and each of these must contain a vertex of degree at least $(n + 2k - 3)/3$. Otherwise, the total number of edges with an endpoint on this P_3 is smaller than $n + 2k - 3$. Taking such a high-degree vertex from each P_3 gives us a set U of $k-1$ vertices, each of degree at least $(n + 2k - 3)/3$.

Assume that $G[V \setminus U]$ contains P_3. Then, we can still construct another $k-1$ copies of P_3, each centred on a vertex from U, as long as each vertex in U has degree large enough to ensure it is connected to at least two vertices not contained on any of the other $k-1$ copies of P_3, *i.e.*, if $(n + 2k - 3)/3 \geqslant 3k - 1$, and this is the case when $n \geqslant 7k$. Therefore $G[V \setminus U]$ consists of independent edges and isolated vertices, and so G has at most $\binom{k-1}{2} + (n-k+1)(k-1) + \lfloor \frac{n-k+1}{2} \rfloor$ edges, a contradiction. □

The above proof gives the extremal graph for $n \geqslant 7k$. No construction is known giving a better bound for $n \geqslant 5k - 1$, and we conjecture that the above example is optimal in this range.

https://doi.org/10.1017/S0963548311000460 Published online by Cambridge University Press

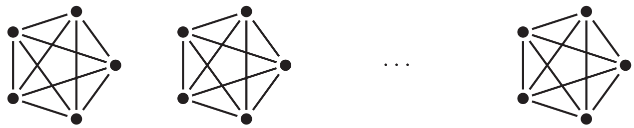

Figure 2. Extremal graph forbidding P_6.

2.2. Longer paths

We note at this point that in the proof of Theorem 2.2, in order to find a P_3 it was enough to find a vertex of degree two; to find subsequent copies of P_3, it sufficed to find vertices of large degree. To adapt this idea to longer paths, we will look for sets of vertices with large common neighbourhood. This notion will continue to be an integral part of our proofs, and thus we formalize it here.

Lemma 2.3. *Let G be a graph on n vertices with m edges, $t \in \mathbb{N}$, and let F_1, F_2 be arbitrary graphs. Then if $F_1 \cup F_2 \nsubseteq G$, any F_1 in G contains t vertices with shared neighbourhood of size at least $n' \geqslant \frac{m'-(n-r)(t-1)}{r-t+1} / \binom{r}{t}$, where $m' = m - \mathrm{ex}(n-r, F_2) - \binom{r}{2}$, and $r = |V(F_1)|$.*

Proof. Assume $F_1 \subseteq G$, say on vertex set U. Since G contains no $F_1 \cup F_2$, $G[V \setminus U]$ contains no F_2. Thus $G[V \setminus U]$ contains at most $\mathrm{ex}(n-r, F_2)$ edges, and so U must have at least $m - \mathrm{ex}(n-r, F_2) - \binom{r}{2} = m'$ edges to $V \setminus U$. Let n_0 be the number of vertices in $V \setminus U$ with neighbourhood of size at least t in U; that is, $n_0 = |\{v \in V \setminus U : |N_U(v)| \geqslant t\}|$.

Since U has at most $n_0 r + (n-r-n_0)(t-1)$ edges to $V \setminus U$, $n_0 r + (n-r-n_0)(t-1) \geqslant m'$, and so $n_0 \geqslant \frac{m'-(n-r)(t-1)}{r-t+1}$. Trivially, there are only $\binom{r}{t}$ subsets of size t in F_1, and so some subset has shared neighbourhood of size $n' \geqslant \frac{m'-(n-r)(t-1)}{r-t+1} / \binom{r}{t}$ as claimed. □

The proof of Lemma 2.2 also required the value of $\mathrm{ex}(n, P_3)$; for longer paths, we shall use the following result due to Erdős and Gallai [5].

Theorem 2.4. *For any $n, l \in \mathbb{N}$, $\mathrm{ex}(n, P_l) \leqslant \frac{l-2}{2} n$.*

We note that the bound in Theorem 2.4 is attained by taking disjoint copies of K_{l-1} as in Figure 2; this gives a tight result whenever n is divisible by $l-1$.

We are now ready to prove the main result of this section.

Theorem 2.5. *For $k \geqslant 2$, $l \geqslant 4$, and $n \geqslant 2l + 2kl\left(\lceil \frac{l}{2} \rceil + 1\right)\binom{l}{\lfloor \frac{l}{2} \rfloor}$,*

$$\mathrm{ex}(n, k \cdot P_l) = \binom{k\lfloor \frac{l}{2} \rfloor - 1}{2} + \left(k\left\lfloor \frac{l}{2} \right\rfloor - 1\right)\left(n - k\left\lfloor \frac{l}{2} \right\rfloor + 1\right) + c_l,$$

where $c_l = 1$ if l is odd, and $c_l = 0$ if l is even.

Note that the result above for $k \cdot P_l$ for $l \geqslant 4$ does not match the earlier result for $k \cdot P_3$ in Theorem 2.2.

https://doi.org/10.1017/S0963548311000460 Published online by Cambridge University Press

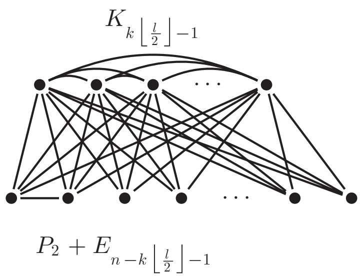

Figure 3. $H_{\text{ex}}(n, k \cdot P_l)$, for l odd.

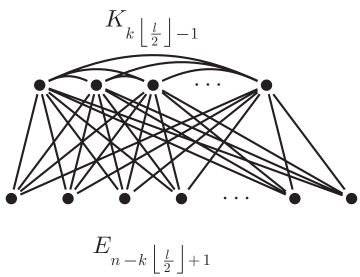

Figure 4. $H_{\text{ex}}(n, k \cdot P_l)$, for l even.

The extremal graph here is $G(n,k,l) := K_t + E_{n-t}$, with a single edge added to the empty class when l is odd, and $t = k\lfloor \frac{l}{2} \rfloor - 1$, as seen in Figures 3 and 4 respectively.

Remark. We note that for paths of even lengths, the above bound can be proved, and the extremal structure determined, via a paper of Balister, Győri, Lehel and Schelp [2] as a consequence of a theorem regarding the maximal number of edges in a connected graph containing no path of some fixed length. One can divide a long path into many short even paths, and this allows one to deduce our Theorem 2.5 from their Theorem 1.3; for odd length paths this result gives a non-optimal number of edges due to parity issues. This extremal number within connected graphs was also determined earlier by Kopylov in 1977 [7], but the approach in the proof given there did not give the extremal structure.

Proof. We proceed by induction on k, starting with the base case, $k = 2$. Let G be a graph with

$$|V| = n \geqslant 2l + 4l\left(\left\lceil \frac{l}{2} \right\rceil + 1\right)\binom{l}{\lfloor \frac{l}{2} \rfloor},$$
$$|E(G)| \geqslant \binom{2\lfloor \frac{l}{2} \rfloor - 1}{2} + \left(2\left\lfloor \frac{l}{2} \right\rfloor - 1\right)\left(n - 2\left\lfloor \frac{l}{2} \right\rfloor + 1\right) + c_l,$$

https://doi.org/10.1017/S0963548311000460 Published online by Cambridge University Press

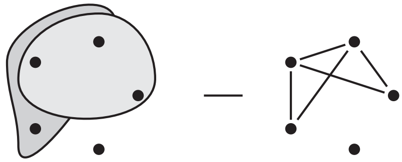

Figure 5. Flattening a hypergraph.

and which contains no $2 \cdot P_l$. As $n \geqslant l^2$, we have that $|E(G)| > \mathrm{ex}(n, P_l)$, and so G contains a P_l on vertex set U, say. Using Lemma 2.3 with $F_1 = P_l$, $F_2 = P_l$, and

$$m = \binom{2\lfloor \frac{l}{2} \rfloor - 1}{2} + \left(2\left\lfloor \frac{l}{2} \right\rfloor - 1\right)\left(n - 2\left\lfloor \frac{l}{2} \right\rfloor + 1\right) + c_l,$$

some elementary simplification shows that any P_l contained in G must have at least $\lfloor \frac{l}{2} \rfloor$ vertices sharing a neighbourhood of size at least

$$\begin{aligned} n' &= \frac{m - \mathrm{ex}(n-l, P_l) - \binom{l}{2} - (n-l)\left(\lfloor \frac{l}{2} \rfloor - 1\right)}{\left(\lceil \frac{l}{2} \rceil + 1\right)\binom{l}{\lfloor \frac{l}{2} \rfloor}} \\ &\geqslant \frac{\binom{2\lfloor \frac{l}{2} \rfloor - 1}{2} + \left(2\lfloor \frac{l}{2} \rfloor - 1\right)\left(n - 2\lfloor \frac{l}{2} \rfloor + 1\right)}{\left(\lceil \frac{l}{2} \rceil + 1\right)\binom{l}{\lfloor \frac{l}{2} \rfloor}} \\ &\quad + \frac{c_l - (n-l)\left(\frac{l}{2} - 1\right) - \binom{l}{2} - (n-l)\left(\lfloor \frac{l}{2} \rfloor - 1\right)}{\left(\lceil \frac{l}{2} \rceil + 1\right)\binom{l}{\lfloor \frac{l}{2} \rfloor}} \\ &\geqslant \frac{\left(1 - \frac{c_l}{2}\right)(n-l)}{\left(\lceil \frac{l}{2} \rceil + 1\right)\binom{l}{\lfloor \frac{l}{2} \rfloor}}. \end{aligned} \tag{2.1}$$

By our assumption on n, (2.1) is at least $2l$.

We now create an $\lfloor \frac{l}{2} \rfloor$-uniform hypergraph $\mathcal{H}$ with $V(\mathcal{H}) = V(G)$ as follows. For any $P_l \subseteq G$, we find a subset U' of $\lfloor \frac{l}{2} \rfloor$ vertices with a large common neighbourhood, as above, and add U' as an edge in $\mathcal{H}$.

We now flatten this hypergraph to form a simple graph G' on the same vertex set, with $uv \in E(G')$ whenever u and v are contained in the same hyperedge.

Since vertices adjacent in G' have large common neighbourhood in G, a path of length $\lfloor \frac{l}{2} \rfloor$ in G' lets us find a path of length l in G. More formally, as $n' \geqslant 2l$, if G' contains $2 \cdot P_{\lfloor \frac{l}{2} \rfloor}$, we can choose distinct common neighbours for each pair of consecutive vertices in these paths, and distinct neighbours for the end vertices, giving us $2 \cdot P_l$ in G. Thus G' cannot contain $2 \cdot P_{\lfloor \frac{l}{2} \rfloor}$.

We further note that certainly two disjoint hyperedges in $\mathcal{H}$ give rise to two such disjoint paths. Thus every pair of edges in $\mathcal{H}$ intersect; such a hypergraph is called *intersecting*. We will further call a hypergraph *k-intersecting* if every pair of edges intersect in at least k vertices.

We now claim that if there exists $X \subseteq V(\mathcal{H})$, with $|X| = t < \lfloor \frac{l}{2} \rfloor$, and such that X contains some vertex from each edge in $\mathcal{H}$, then $|E(G)| < |E(G(n, 2, l))|$. We will later refer to this as Claim A, for clarity.

https://doi.org/10.1017/S0963548311000460 Published online by Cambridge University Press

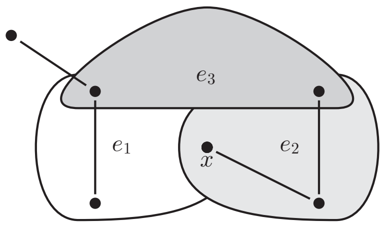

Figure 6. Case 1.

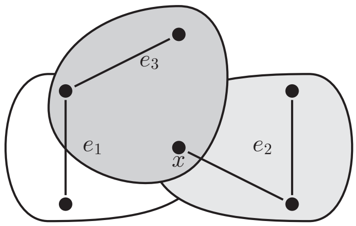

Figure 7. Case 2.

Indeed, assume X is such a set. By the construction of $\mathcal{H}$, since $\mathcal{H}[V(\mathcal{H}) \setminus X]$ contains no hyperedges, $G[V(G) \setminus X]$ contains no P_l, and so Theorem 2.4 tells us that

$$|E(G)| \leqslant \binom{t}{2} + t(n-t) + \frac{l-2}{2}(n-t) \leqslant \left(2\left\lfloor\frac{l}{2}\right\rfloor - \frac{3}{2}\right)n.$$

Recall that

$$\begin{aligned}|E(G(n,2,l))| &= \binom{2\lfloor\frac{l}{2}\rfloor - 1}{2} + \left(2\left\lfloor\frac{l}{2}\right\rfloor - 1\right)\left(n - 2\left\lfloor\frac{l}{2}\right\rfloor + 1\right) + c_l \\ &\geqslant \left(2\left\lfloor\frac{l}{2}\right\rfloor - 1\right)n - l^2,\end{aligned}$$

and so as $n > 2l^2$, $|E(G)| < |E(G(n,2,l))|$. Thus Claim A holds.

Now, assume we have at least $2\lfloor\frac{l}{2}\rfloor$ vertices contained in edges of $\mathcal{H}$, but without $2 \cdot P_{\lfloor\frac{l}{2}\rfloor}$ in G'. We will now show that no two hyperedges can intersect in only a single vertex.

If $E_1, E_2 \in E(\mathcal{H})$ with $E_1 \cap E_2 = \{x\}$, then $|E_1 \cup E_2| = 2\lfloor\frac{l}{2}\rfloor - 1$ vertices, and so $\mathcal{H}$ contains an edge E_3 not contained in their union. We may assume that this edge intersects $E_1 \cup E_2$ outside $\{x\}$, as if no such edge exists, we are done by Claim A applied to the set $\{x\}$. Without loss of generality, $E_3 \cap E_1 \nsubseteq E_2$.

Let us consider two cases.

Case 1: $E_3 \cap (E_2 \setminus E_1) \neq \emptyset$. Then we can find a cycle in G' through all the vertices in $E_1 \cup E_2$. Since we have at least $2\lfloor\frac{l}{2}\rfloor$ vertices in edges of G', there is another vertex adjacent to this cycle. This gives us a path of length $2\lfloor\frac{l}{2}\rfloor$, and so G' contains $2 \cdot P_{\lfloor\frac{l}{2}\rfloor}$.

Case 2: $E_3 \cap (E_2 \setminus E_1) = \emptyset$. Then there is some $y \in E_3 \setminus (E_1 \cup E_2)$, and so we can form one $P_{\lfloor\frac{l}{2}\rfloor}$ in $(E_1 \setminus \{x\}) \cup \{y\}$ and a disjoint $P_{\lfloor\frac{l}{2}\rfloor}$ entirely inside E_2.

https://doi.org/10.1017/S0963548311000460 Published online by Cambridge University Press

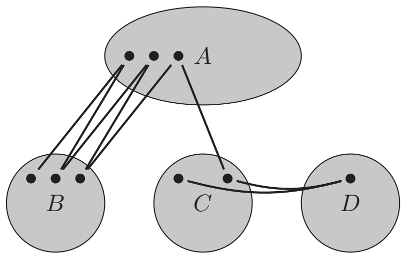

Figure 8. A vertex in D of degree 2.

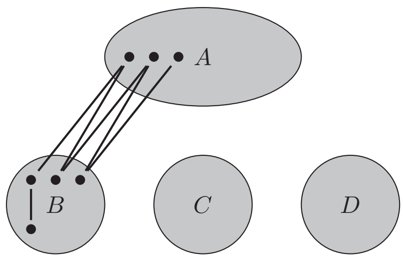

Figure 9. An edge inside B.

We now have that $\mathcal{H}$ is an intersecting hypergraph, with at least $2\lfloor \frac{l}{2} \rfloor$ vertices contained in its edges, and no two edges can intersect in a single vertex, and so $\mathcal{H}$ is 2-intersecting. The edge set of $\mathcal{H}$ is non-empty, so pick an edge E, and any vertex in $x \in E$. Each edge in $\mathcal{H}$ intersects E in at least two vertices, so any edge in $\mathcal{H}$ intersects $E \setminus \{x\}$, a set of size $\lfloor \frac{l}{2} \rfloor - 1$. We have already ruled out such a set of vertices in Claim A.

We now know that all edges of $\mathcal{H}$ are contained in a set A of vertices with $|A| \leqslant 2\lfloor \frac{l}{2} \rfloor - 1$, and hence any P_l in G contains at least $\lfloor \frac{l}{2} \rfloor$ vertices from A. We define three more sets of vertices as follows:

$$B = \left\{ x \in G \setminus A \,|\, d_A(x) \geqslant \left\lfloor \frac{l}{2} \right\rfloor \right\},$$
$$C = \left\{ x \in G \setminus A \,|\, \left\lfloor \frac{l}{2} \right\rfloor > d_A(x) > 0 \right\},$$
$$D = \{ x \in G \setminus A \,|\, d_A(x) = 0 \}.$$

Certainly D can contain no P_l, since every P_l meets A. Thus the number of edges entirely within D is at most $\frac{l-2}{2}|D|$ by Theorem 2.4.

We now claim that every vertex $x \in B \cup C$ is the end vertex of a P_l in G, with alternate vertices in A, which also misses any given $y_1, y_2 \in B \cup C$. Since x is adjacent to some $y \in A$, and y is contained in some hyperedge E, as long as $n' > |A| + \lfloor \frac{l}{2} \rfloor + 2$, we can find $\lfloor \frac{l}{2} \rfloor$ vertices in $(B \cup C) \setminus \{x, y_1, y_2\}$ adjacent to all vertices in E, allowing us to find such a P_l.

Further, no vertex in D can have degree more than 1 to $B \cup C$; assume uv, uw are both edges with $u \in D$, and $v, w \in B \cup C$. We can find a P_l leaving v, that misses w, with alternate vertices in A. This gives a P_l starting at w with only $\lfloor \frac{l}{2} \rfloor - 1$ vertices from A, as

https://doi.org/10.1017/S0963548311000460 Published online by Cambridge University Press

in Figure 8. A vertex in $B \cup C$ with degree 2 to $B \cup C$ allows us to create a path in the same way, so our graph contains none of these.

Similarly, if l is even, an edge inside B allows us to create a P_l using only $\lfloor \frac{l}{2} \rfloor - 1$ vertices from A, as in Figure 9, so in this case B must be empty. If l is odd, then since every vertex in B is adjacent to vertices in every edge of $\mathcal{H}$, then the existence of two disjoint edges in B allows us to create a P_l with only $\lfloor \frac{l}{2} \rfloor - 1$ vertices from A. A single edge does not create this problem, however: this is where the c_l in the theorem arises.

We have now counted edges between B and C and between $B \cup C$ and D respectively, and counted the edges inside each of B, C and D. We can use the degree conditions in their definitions to count edges from A to B, C and D. Putting these together, we see that

$$\begin{aligned} |E(G)| \leqslant \binom{|A|}{2} + (n - |A| - |C| - |D|)|A| \qquad (2.2) \\ + \left(1 + \left\lfloor \frac{l}{2} \right\rfloor - 1\right)|C| + \left(1 + \frac{l-2}{2}\right)|D| + c_l, \\ = \binom{|A|}{2} + (n - |A|)|A| + \left(\left\lfloor \frac{l}{2} \right\rfloor - |A|\right)|C| + \left(\frac{l}{2} - |A|\right)|D|. \qquad (2.3) \end{aligned}$$

As any P_l in G contains at least $\lfloor \frac{l}{2} \rfloor$ vertices of A, $|A| \geqslant \lfloor \frac{l}{2} \rfloor$. If $|A| = \lfloor \frac{l}{2} \rfloor$, then $|D| \leqslant n - |A|$, and so, whenever $n > |A| + 2$,

$$\begin{aligned} |E(G)| &\leqslant \binom{|A|}{2} + (n - |A|)|A| + \frac{c_l}{2}(n - |A|) + c_l \\ &= \binom{|A|+1}{2} + (n - |A| - 1)(|A| + 1) + \left(\frac{c_l}{2} - 1\right)(n - |A| - 1) + \frac{3}{2}c_l \\ &< \binom{|A|+1}{2} + (n - |A| - 1)(|A| + 1) + c_l. \end{aligned}$$

In fact, $|A| + 1 \leqslant 2\lfloor \frac{l}{2} \rfloor - 1$, and so $|E(G)| < |G(n, 2, k)|$.

If $|A| > \lfloor \frac{l}{2} \rfloor$, the coefficients of $|C|$ and $|D|$ in (2.3) are negative, and so $|E(G)|$ is maximized when C and D are empty. This gives our bound on $|E(G)|$ as claimed. Further, since C and D must be empty to attain this bound, it also shows that the extremal graph is $G(n, 2, l) = K_{2\lfloor \frac{l}{2} \rfloor - 1} + E_{n - 2\lfloor \frac{l}{2} \rfloor + 1}$ with an extra edge in the empty class for odd l, as in the statement of the theorem.

We have now established the base case $k = 2$. Somewhat surprisingly, the inductive step is easy to show.

Let G be a graph on n vertices with

$$m \geqslant \binom{k\lfloor \frac{l}{2} \rfloor - 1}{2} + \left(k\left\lfloor \frac{l}{2} \right\rfloor - 1\right)\left(n - k\left\lfloor \frac{l}{2} \right\rfloor + 1\right) + c_l$$

https://doi.org/10.1017/S0963548311000460 Published online by Cambridge University Press

edges, not containing $k \cdot P_l$. This graph does contain a P_l, and from Lemma 2.3, we can find $\lfloor \frac{l}{2} \rfloor$ vertices with shared neighbourhood of size at least

$$\begin{aligned}
n' &= \frac{\binom{k\lfloor \frac{l}{2}\rfloor - 1}{2} + \left(k\lfloor \frac{l}{2}\rfloor - 1\right)\left(n - k\lfloor \frac{l}{2}\rfloor + 1\right)}{\left(\lceil \frac{l}{2}\rceil + 1\right)\binom{l}{\lfloor \frac{l}{2}\rfloor}} \\
&\quad + \frac{c_l - \mathrm{ex}(n-l,(k-1)\cdot P_l) - \binom{l}{2} - (n-l)\left(\lfloor \frac{l}{2}\rfloor - 1\right)}{\left(\lceil \frac{l}{2}\rceil + 1\right)\binom{l}{\lfloor \frac{l}{2}\rfloor}} \\
&= \frac{\binom{k\lfloor \frac{l}{2}\rfloor - 1}{2} + \left(k\lfloor \frac{l}{2}\rfloor - 1\right)\left(n - k\lfloor \frac{l}{2}\rfloor + 1\right)}{\left(\lceil \frac{l}{2}\rceil + 1\right)\binom{l}{\lfloor \frac{l}{2}\rfloor}} \\
&\quad + \frac{c_l - \binom{(k-1)\lfloor \frac{l}{2}\rfloor - 1}{2} - \left((k-1)\lfloor \frac{l}{2}\rfloor - 1\right)\left(n - l - (k-1)\lfloor \frac{l}{2}\rfloor + 1\right)}{\left(\lceil \frac{l}{2}\rceil + 1\right)\binom{l}{\lfloor \frac{l}{2}\rfloor}} \\
&\quad + \frac{-c_l - \binom{l}{2} - (n-l)\left(\lfloor \frac{l}{2}\rfloor - 1\right)}{\left(\lceil \frac{l}{2}\rceil + 1\right)\binom{l}{\lfloor \frac{l}{2}\rfloor}} \\
&= \frac{n + k\lfloor \frac{l}{2}\rfloor^2 - \frac{3}{2}\lfloor \frac{l}{2}\rfloor^2 + c_l k\lfloor \frac{l}{2}\rfloor - \frac{5+4c_l}{2}\lfloor \frac{l}{2}\rfloor - 2c_l}{\left(\lceil \frac{l}{2}\rceil + 1\right)\binom{l}{\lfloor \frac{l}{2}\rfloor}} \\
&\geqslant \frac{n-l}{\left(\lceil \frac{l}{2}\rceil + 1\right)\binom{l}{\lfloor \frac{l}{2}\rfloor}}.
\end{aligned}$$

The second equality is valid since

$$n - l \geqslant 2l + 2l(k-1)\left(\left\lceil \frac{l}{2}\right\rceil + 1\right)\binom{l}{\lfloor \frac{l}{2}\rfloor}.$$

Note that by our assumption on n, $n' \geqslant 2kl$. Write U for the set of vertices given by Lemma 2.3. Then $G[V \setminus U]$ is a graph on $n - \lfloor \frac{l}{2}\rfloor$ vertices and at least $\mathrm{ex}\big(n - \lfloor \frac{l}{2}\rfloor, (k-1)\cdot P_l\big)$ edges. If we can find $(k-1)\cdot P_l$, then since $n' \geqslant kl$, we can find another P_l in G disjoint from these $k-1$. Therefore there cannot be $k-1$ disjoint copies of P_l in $G[V \setminus U]$, so by the inductive hypothesis, $G[V \setminus U] = G\big(n - \lfloor \frac{l}{2}\rfloor, k-1, l\big)$. Thus $G = G(n,k,l)$. □

The above proof shows that our construction is optimal for $n = \Omega(kl^{\frac{3}{2}}2^l)$. We conjecture that this construction is optimal for $n = \Omega(kl)$. We also note a comparison between Theorem 2.5 for even paths and Theorem 2.4: certainly if one forbids $k \cdot P_{2l}$, then one is also forbidding P_{2kl}. Thus an easy upper bound on $\mathrm{ex}(n, k \cdot P_{2l})$ is $\mathrm{ex}(n, P_{2kl})$. The difference between this bound and the precise result established above is relatively small, $(kl-1)\binom{kl}{2}$. In particular, it is not dependent on n for fixed k and l, despite the significant difference between the extremal graphs.

3. Trees

Throughout the following section, we need an analogue of Lemma 2.1 as a starting point. For longer paths, we used the Erdős–Gallai result, Lemma 2.4. The analogous result for trees is known as the Erdős–Sós conjecture.

https://doi.org/10.1017/S0963548311000460 Published online by Cambridge University Press

3.1. The Erdős–Sós conjecture

We note that a path can be viewed as an extreme kind of tree: $l-2$ vertices have degree two, and the two leaves of course have degree one. The opposite extreme is the star: one central vertex of degree $l-1$, and the other $k-1$ vertices are leaves. For both examples, it is easily seen that $\mathrm{ex}(n, G) = \frac{l-2}{2}n$. Legend has it that Vera T. Sós presented the proofs of these two results to her graph theory class in Budapest in 1962, and left the following conjecture as a homework problem; by now, this is known as the notoriously difficult Erdős–Sós conjecture.

Conjecture 3.1 (Erdős–Sós conjecture). *For any tree T on l vertices,* $\mathrm{ex}(n, T) = \frac{l-2}{2}n$.

In 2008, a proof of the conjecture was announced for very large trees by Ajtai, Komlós, Simonovits and Szemerédi. For small trees, however, the conjecture is mostly open. There is a sequence of results in the direction of the full theorem for smaller trees. We present a representative sample of these results here, which is certainly only the tip of the iceberg. Many more partial results related to the Erdős–Sós conjecture exist; see, for example, [1] and [15]. The first result here establishes the conjecture for graphs of large girth and is due to Dobson [4].

Theorem 3.2. *If T is a tree on l vertices, and G is a graph with girth at least five and minimum degree $\delta \geqslant \frac{l}{2}$, then G contains T. Thus Conjecture 3.1 holds if the maximum number edges is taken over all T-free graphs of girth at least five.*

Similarly, Saclé and Woźniak [10] proved that whenever G is a graph with at least $\frac{l-2}{2}n$ edges and no C_4, G contains any tree on l vertices. In 2005, McLennan [8] proved the Erdős–Sós bound for trees of diameter at most four.

The Erdős–Sós conjecture has also been proved for caterpillars; this result is attributed to Perles in [9]. Later, Sidorenko [11] showed that the Erdős–Sós conjecture holds for trees of order l containing a vertex which is the parent of at least $\frac{l-1}{2}$ leaves.

3.2. Forests of equibipartite trees

Our proof of Theorem 2.5 can be adapted to work on a significantly larger class of graphs. A key element of our proof was finding a set of vertices which intersected every long path in at least half its vertices. This continues to be an essential idea, and thus we restrict ourselves to trees which have the same number of vertices in each vertex class, when viewed as a bipartite graph. We call such trees *equibipartite*, and a forest in which each component is an equibipartite tree is called an equibipartite forest. Clearly any equibipartite tree or equibipartite forest has an even number of vertices.

If we allow ourselves the considerable benefit of assuming that Erdős–Sós holds for all equibipartite trees, we can determine the extremal number for any equibipartite forest, for large n. There is a slight difference in the extremal number and the structure of the extremal graph depending on whether the forest admits a perfect matching.

https://doi.org/10.1017/S0963548311000460 Published online by Cambridge University Press

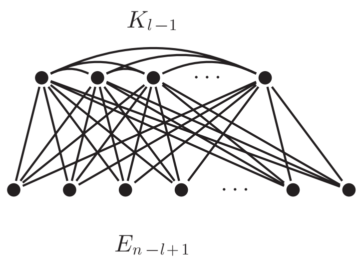

Figure 10. $H_{\text{ex}}(n, H)$, for H containing a perfect matching.

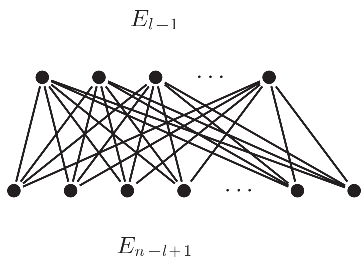

Figure 11. $H_{\text{ex}}(n, H)$, for H containing no perfect matching.

Theorem 3.3. *Let H be an equibipartite forest on $2l$ vertices which comprises at least two trees. If the Erdős–Sós conjecture holds for each component tree in H, then for $n \geqslant 3l^2 + 32l^5\binom{2l}{l}$,*

$$\text{ex}(n, H) = \begin{cases} \binom{l-1}{2} + (l-1)(n-l+1) & \textit{if } H \textit{ admits a perfect matching,} \\ (l-1)(n-l+1) & \textit{otherwise.} \end{cases}$$

Remark. The extremal graphs here are $K_{l-1} + E_{n-l+1}$ for any forest with a perfect matching, and $E_{l-1} + E_{n-l+1}$ for any forest with no perfect matching, as in Figures 10 and 11. To prove the eventual extremal number for equibipartite trees as in Theorem 3.3, we do not need the full strength of the Erdős–Sós conjecture; in fact, it suffices to know that $\text{ex}(n, T) = \frac{|T|-2}{2}n + o(n)$ for any of the equibipartite trees $T \subseteq H$. In this case, however, the bound on n for which the result holds is worse. We also note that in order to avoid many lower-order terms, the bound on n in the statement of the theorem has not been optimized.

Lemma 3.4. *Let H be a equibipartite tree on $2l$ vertices. If H contains a perfect matching, then every partition of $V(H)$ into two classes of different sizes is such that the larger class induces at least one edge.*

https://doi.org/10.1017/S0963548311000460 Published online by Cambridge University Press

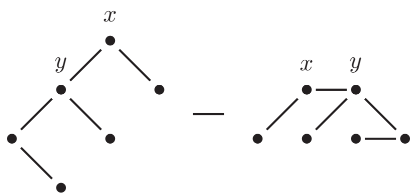

Figure 12. Example partition for Lemma 3.4.

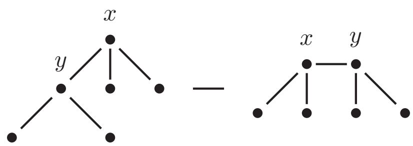

Figure 13. Example partition for Lemma 3.5.

Proof. If H contains a perfect matching, $M \subseteq E(H)$, then for any partition of $V(H)$ into non-equal classes, $|V_1| < |V_2|$, the number of edges in M which meet V_1 is at most $|V_1| < l$, and so some edge lies inside V_2. □

Lemma 3.5. *Let H be a equibipartite tree on $2l$ vertices. If H does not contain a perfect matching, then there exists a partition of $V(H)$ into two classes of different sizes such that the larger class induces no edges and the smaller class induces exactly one edge.*

Proof. Consider H as a bipartite graph with bipartition $V(H) = (A, B)$. Since H contains no perfect matching, there is a set $S \subseteq A$ for which Hall's condition (see, *e.g.*, [3]) fails, namely $|N(S)| < |S|$. If we take S minimal, then $H[S \cup N(S)]$ is connected, as otherwise one of its components would fail Hall's condition. Consider $H[(A \setminus S) \cup (B \setminus N(S))]$. Each component of this graph is joined to $N(S)$ by a single edge. Since the union of these components has larger intersection with B than with A, at least one of the components does. Let C be such a component, and let xy be the unique edge between C and $N(S)$, with $x \in C$ and $y \in N(S)$.

Consider the partition $(C, V(H) \setminus C)$. Then taking the set of vertices $V_{x,y}$ which are in the same bipartite class as x in C or in the same bipartite class as y in $V(H) \setminus C$ as one class of our new partition, and $V(H) \setminus V_{x,y}$ as the other, forms a partition of $V(H)$ with exactly one edge in $V_{x,y}$, and none in $V(H) \setminus V_{x,y}$.

Since our tree is equibipartite, $|V_{x,y} \cap (V(H) \setminus C)| + |(V(H) \setminus V_{x,y}) \cap C| = l$. By our definition of C, $|V_{x,y} \cap C| < |(V(H) \setminus V_{x,y}) \cap C|$. By construction, each of the sets $V_{x,y} \cap (V(H) \setminus C)$, $(V(H) \setminus V_{x,y}) \cap C$, and $V_{x,y} \cap C$ are non-empty. Then

$$|V_{x,y}| = |V_{x,y} \cap C| + |V_{x,y} \cap (V(H) \setminus C)| < |(V(H) \setminus V_{x,y}) \cap C| + |V_{x,y} \cap (V(H) \setminus C)| = l,$$

and so our partition is an unbalanced partition with no edges in the larger class and exactly one edge in the smaller class, as claimed. □

https://doi.org/10.1017/S0963548311000460 Published online by Cambridge University Press

See Figures 12 and 13 for an example of a partition of a tree with and without a perfect matching, respectively.

Proof of Theorem 3.3. Let H have components $H_1, H_2, \ldots, H_k$, each on $2l_1, 2l_2, \ldots, 2l_k$ vertices respectively, and G be a graph on n vertices with m edges which does not contain H, and with $m \geqslant (l-1)(n-l+1)$. Without loss of generality, $l_1 \leqslant l_i$, for each i. For notational ease, we also define $H' = H_2 \cup \cdots \cup H_k$ and $l' = \frac{1}{2}|H'| = l - l_1$.

As $n \geqslant l^2$, $m \geqslant \mathrm{ex}(n, H')$ by induction (or Erdős–Sós, if H' is a tree), and so we can find a copy of $H' \subseteq G$. As in the proof of Lemma 2.3, for any copy of H' we can bound from below the size of the set E' of edges between H' and $G \setminus H'$ by $m - \binom{2l'}{2} - \mathrm{ex}(n-2l', H_1)$. By the Erdős–Sós conjecture, this is at least $(l-1)(n-l+1) - \binom{2l'}{2} - (n-2l')(l_1-1) \geqslant l'n - 3l^2$.

Consider the set of vertices $X = \{v \in G \setminus H' : |N(v) \cap H'| \geqslant l'\}$. Then

$$2l'|X| + (l'-1)(n-2l'-|X|) \geqslant |E'| \geqslant l'n - 3l^2.$$

Thus $|X| \geqslant \frac{n-3l^2}{l'+1}$. As there are only $\binom{2l'}{l'}$ sets of l' vertices in H', we can find a set A of l' vertices in H' with at least

$$n' = \frac{n-3l^2}{(l'+1)\binom{2l'}{l'}}$$

common neighbours. By our assumption on n, $n' \geqslant 32l^3$.

Interchanging the roles of H_1 and H', for any H_1 we similarly bound from below the size of the set E_1 of edges between H_1 and $G \setminus H_1$ by $m - \binom{2l_1}{2} - \mathrm{ex}(n-2l_1, H')$. Note that $n - 2l_1$ is much larger than needed in the condition of the inductive hypothesis, and so

$$\begin{aligned} |E_1| &\geqslant (l-1)(n-l+1) - \binom{2l_1}{2} - (n-2l_1-l'+1)(l'-1) - \binom{l'-1}{2} \\ &\geqslant l_1 n - 3l^2. \end{aligned} \tag{3.1}$$

With this in mind, we define the following set of vertices which are not in A, but which are still of large degree:

$$B = \left\{ w \in G \mid w \notin A \text{ and } d_G(w) \geqslant \frac{n-3l^2}{l_1+1} \right\}.$$

Now, any copy of H_1 in G must contain at least l_1 vertices from $A \cup B$, as otherwise the sum of the degrees of vertices in H_1 is less than $(l_1+1)\frac{n-3l^2}{l_1+1} + (l_1-1)n$, contradicting (3.1) above.

As a rough bound on the number of edges in G, we note that if G contained more than $2ln$ edges, we can find a copy of H' by induction (or by the Erdős–Sós conjecture if H' is a single tree). Removing this copy of H' leaves a graph on $n - 2l'$ vertices with more than $2l_1 n \geqslant 2l_1(n-2l')$ edges, since each vertex is of course adjacent to at most n edges. Again by Conjecture 3.1, we can find a copy of H_1. Thus our graph can have at most $2ln$ edges.

https://doi.org/10.1017/S0963548311000460 Published online by Cambridge University Press

This means that for any $c > 0$, there are at most $\frac{4ln}{c}$ vertices of degree at least c. Choosing

$$c = \frac{8ln}{n'} = \frac{8l(l'+1)\binom{2l'}{l'}n}{n-3l^2},$$

there are at least $\frac{n'}{2}$ common neighbours of A with degree at most c. Since $n \geqslant 6l^2$, $c \leqslant 16l(l'+1)\binom{2l'}{l'}$. Then since $\frac{n'}{2} \geqslant l'$, we can find a copy of H' with l' vertices in A and the other l' vertices having degree at most c.

Since this copy of H' is incident to at least $l'n - 3l^2$ edges, any vertex in A has degree at least

$$\begin{aligned} &l'n - 3l^2 - l'c - (l'-1)(n-1) \\ &\geqslant n - 3l^2 - l'c \\ &= n - c'. \end{aligned} \tag{3.2}$$

There are at most $\frac{4ln}{c} = \frac{n'}{2}$ vertices of degree at least c, and at most $l'c'$ vertices not adjacent to all of A. Since

$$\frac{n-3l^2}{l_1+1} - \frac{n'}{2} \geqslant \frac{n-3l^2}{2(l_1+1)} \geqslant 16l^4\binom{2l}{l} > l'c',$$

by the definition of B each vertex $x \in B$ is adjacent to a vertex y which is adjacent to all of A and such that $d_G(y) \leqslant c$.

This condition on the vertices in B enables us to find, for each $x \in B$, a copy of H' from which $l'-1$ of the vertices have small degree, and whose intersection with B contains x as a leaf. Further, we can find a set U of $l'-1$ vertices of degree at most c which are each adjacent to all of A, so for any $z \in A$, $G[(U \cup \{x\} \cup \{y\} \cup (A \setminus \{z\}))]$ is a graph on $2l'$ vertices which contains a copy of $K_{l',l'-1}$ with an extra vertex x adjacent to some vertex in the larger set. We can find a copy of H' in this by letting a leaf of H' correspond to x, and so as in (3.2), every vertex in B must have degree at least $n - c'$. If B contained at least l_1 vertices, they would have common neighbourhood of size at least $n - l_1c' \geqslant l$, allowing us to find H_1 in $G[V(G) \setminus A]$, and again as the common neighbourhood of A is of size at least $2l$, we can find a disjoint copy of H', giving a copy of H in G. Thus $|B| \leqslant l_1 - 1$, and so $|A \cup B| \leqslant l' + l_1 - 1 = l - 1$.

We now define two more sets of vertices as follows:

$$\begin{aligned} D &= \{x \in G \setminus (A \cup B) \mid d_{A\cup B}(x) \geqslant l_1\}, \\ E &= \{x \in G \setminus (A \cup B) \mid d_{A\cup B}(x) < l_1\}. \end{aligned}$$

We note that any vertex not in $A \cup B$ which is adjacent to all of A is in D, and thus $|E| \leqslant l'c'$. There can be no H_1 in E, so the number of edges in E is at most $(l_1 - 1)|E|$ by Erdős–Sós. We now claim that no vertex $v \in D$ can have a neighbour $y \in D \cup E$. Indeed, we can find a set U of $l_1 - 1$ vertices in $A \cup B$ adjacent to v since each vertex in $A \cup B$ has degree at least $n - c'$. Further, we can find $W \subseteq (D \cup E) \setminus \{v, y\}$ consisting of $l_1 - 1$ vertices adjacent to all of U. As before we can find a copy of H_1 on $U \cup W \cup \{v, y\}$ with only $l_1 - 1$ vertices from $A \cup B$, a contradiction. Thus all edges in $G[D \cup E]$ are in E.

https://doi.org/10.1017/S0963548311000460 Published online by Cambridge University Press

Letting $|A \cup B| = t$, we bound the number of edges in G by

$$\binom{t}{2} + t(n - t - |E|) + (l_1 - 1)|E| + (l_1 - 1)|E| \tag{3.3}$$
$$= \binom{t}{2} + t(n - t) + (2l_1 - 2 - t)|E|.$$

If $t < l - 1$, then since $|E| \leqslant l'c'$ the number of edges in G is at most $\binom{t}{2} + t(n-t) + 2l_1 l'c' < (l-1)(n-l+1)$, for $n \geqslant 2l^2c' + l^2$, and hence $|A \cup B| = l - 1$.

The common neighbourhood of $A \cup B$ has size at least $n - (l-1) - (l-1)c'$, as each vertex in $A \cup B$ is adjacent to all but c' vertices in G. Thus we can find a copy of $K_{l-1, n-(l-1)(c'+1)} \subseteq G$, where the smaller class is $A \cup B$. If H does not contain a perfect matching, then by Lemma 3.5 we can partition the vertices into unequal sets X, Y, the larger of which is empty, and the smaller of which contains one edge. This is clearly present in G if $A \cup B$ contains an internal edge.

Counting all edges in G, we see that by (3.3),

$$|E(G)| \leqslant (l-1)(n-l-1) - (l - 2l_1 + 1)|E| + C_H,$$

where $C_H = \binom{l-1}{2}$ if H admits a perfect matching, and $C_H = 0$ otherwise. As l_1 is minimal, $(l - 2l_1 + 1) > 0$, and so the number of edges is maximized when $|E| = 0$. □

It is unlikely that the bound on n in Theorem 3.3 is optimal. Determining the minimal value of n for which this construction is optimal remains an open question.

Acknowledgements

We would like to thank the anonymous referee for suggestions on improvements to the style, structure, and content of this paper. We would also like to thank Béla Bollobás for pointing us towards this interesting problem.

References

[1] Balasubramanian, S. and Dobson, E. (2007) On the Erdős–Sós conjecture for graphs with no $K_{2,s}$. *J. Graph Theory* **56** 301–310.
[2] Balister, P. N., Győri, E., Lehel, J. and Schelp, R. H. (2008) Connected graphs without long paths. *Discrete Math.* **308** 4487–4494.
[3] Bollobás, B. (2002) *Modern Graph Theory*, Springer.
[4] Brandt, S. and Dobson, E. (1996) The Erdős–Sós conjecture for graphs of girth 5. *Discrete Math.* **150** 411–414.
[5] Erdős, P. and Gallai, T. (1959) On maximal paths and circuits of graphs. *Acta Math. Acad. Sci. Hungar.* **10** 337–356.
[6] Gorgol, I. (2011) Turán numbers for disjoint copies of graphs. *Graphs Combin.* **27** 661–667.
[7] Kopylov, G. N. (1977) On maximal paths and cycles in a graph. *Soviet Math. Dokl.* **18** 593–596.
[8] McLennan, A. (2005) The Erdős–Sós conjecture for trees of diameter four. *J. Graph Theory* **49** 291–301.
[9] Moser, W. and Pach, J. (1993) Recent developments in combinatorial geometry. In *New Trends in Discrete and Computational Geometry*, Springer, pp. 281–302.

https://doi.org/10.1017/S0963548311000460 Published online by Cambridge University Press

[10] Saclé, J. F. and Woźniak, M. (1997) The Erdős–Sós conjecture for graphs without C_4. *J. Combin. Theory Ser. B* **70** 367–372.
[11] Sidorenko, A. (1989) Asymptotic solution for a new class of forbidden r-graphs. *Combinatorica* **9** 207–215.
[12] Simonovits, M. (1968) A method for solving extremal problems in extremal graph theory. In *Theory of Graphs* (P. Erdős and G. Katona, eds), Academic Press, pp. 279–319.
[13] Turán, P. (1941) Egy gráfelméleti szélsőértékfeladatról. *Mat. es Fiz. Lapok.* **48** 436–452.
[14] Turán, P. (1954) On the theory of graphs. *Colloquium Math.* **3** 19–30.
[15] Woźniak, M. (1996) On the Erdős–Sós conjecture. *J. Graph Theory* **21** 229–234.

https://doi.org/10.1017/S0963548311000460 Published online by Cambridge University Press