Journal of Combinatorics
Volume 5, Number 3, 335–354, 2014

Separating path systems

Victor Falgas-Ravry, Teeradej Kittipassorn, Dániel Korándi, Shoham Letzter, and Bhargav P. Narayanan

We study separating systems of the edges of a graph where each member of the separating system is a path. We conjecture that every n-vertex graph admits a separating path system of size linear in n and we prove this in certain interesting special cases. In particular, we establish this conjecture for random graphs and graphs with linear minimum degree. We also obtain tight bounds on the size of a minimal separating path system in the case of trees.

1. Introduction

Given a set S, we say that a family $\mathcal{F}$ of subsets of S *separates* a pair of distinct elements $x, y \in S$ if there exists a set $A \in \mathcal{F}$ which contains exactly one of x and y. If $\mathcal{F}$ separates all pairs of distinct elements of S, we say that $\mathcal{F}$ is a *separating system* of S.

The study of separating systems was initiated by Rényi [19] in 1961. It is essentially trivial that the minimal size of a separating system of an n-element set is $\lceil \log_2 n \rceil$. However, the question of finding the minimal size of a separating system becomes much more interesting when one imposes restrictions on the elements of the separating system. For example, separating systems with restrictions on the cardinalities of their members have been studied by Katona [14], Wegener [24], Ramsay and Roberts [18] and Kündgen, Mubayi and Tetali [15], amongst others. Stronger notions of separation as well as other extremal questions about separating systems have also been studied; see, for example, the papers of Spencer [20], Hansel [12], and Bollobás and Scott [5].

Another interesting direction involves imposing a graph structure on the underlying ground set and imposing graph theoretic restrictions on the separating family (see, for instance, [9, 6]). In this paper, we investigate the question of separating the edges of a graph using paths. Given a graph $G = (V, E)$, we say that a family $\mathcal{P}$ of subsets of the edge set $E(G)$ is a *separating path system of* G if $\mathcal{P}$ separates $E(G)$ and every element of $\mathcal{P}$ is

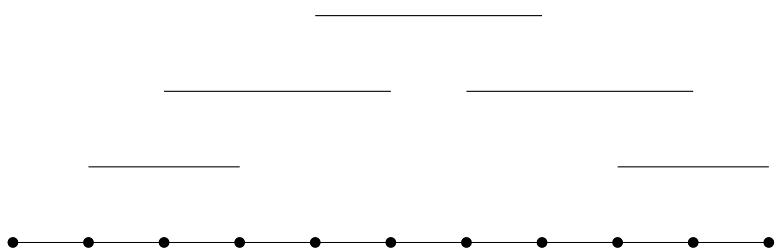

Figure 1: A path on 11 vertices and a separating path system with 5 paths.

a path of G. The analogous question of separating the vertices of a graph with paths has also been studied; we refer the reader to [11] for details.

Separating path systems arise naturally in the context of network design (see, for instance, [1, 13, 21]). We are presented with a communication network with one (and at most one) defective link and our goal is to identify this link. Of course, one could test every link, but this is not very efficient; can we do better? A natural test to perform is to send a message between a pair of nodes along a predetermined path; if the message does not reach its intended destination, we conclude that the defective link lies on this path. If we model the communication network as a graph, a fixed set of such tests succeeds in locating any defective link if and only if the corresponding family of paths is a separating path system of the underlying graph. We are naturally led to the following question: what is the size of a minimal separating path system of a given graph?

For a graph G, let $f(G)$ be the size of a minimal separating path system of G. As a separating path system of G is also a separating system of $E(G)$, it follows that $f(G) \geq \lceil \log_2 |E(G)| \rceil$. In particular, for any connected n-vertex graph G, $f(G) = \Omega(\log n)$. With a little work, we can construct graphs that come close to matching this bound. Let L_n be the *ladder* of order $2n$, that is, the Cartesian product of a path of length $n-1$ with a single edge. Given any subset A of $[n-1]$, there is (see Figure 2) a natural way of mapping A to a path P_A in L_n. With this, it is an easy exercise to establish that $f(L_n) = O(\log n)$; indeed, one can show that $f(L_n) \leq 3 \log_2 n + 1$.

A more interesting problem is to determine $f(n)$, the *maximum* of $f(G)$ taken over all n-vertex graphs; this question was raised by Gyula Katona in August 2013 at the 5$^{\text{th}}$ Emléktábla Workshop in Budapest.

Clearly, at most one edge of a graph can be left uncovered by the paths of a separating path system of the graph; it is thus unsurprising that the question of building small separating path systems is closely related to the well-studied question of covering a graph with paths. It would be remiss

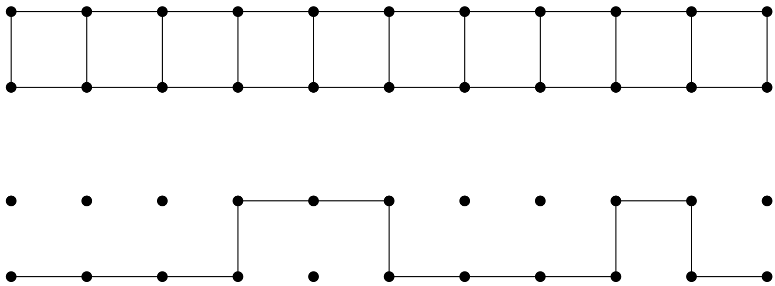

Figure 2: The graph L_{11} and the path P_A corresponding to the subset $A = \{4, 5, 9\}$.

not to remind the reader of a beautiful conjecture of Gallai which asserts that every connected graph on n vertices can be decomposed into $\lfloor (n+1)/2 \rfloor$ paths. The following fundamental result of Lovász [17], which provides support for Gallai's conjecture, will prove useful. Here and elsewhere, by a decomposition of a graph we mean a covering of its edges with edge disjoint subgraphs.

Theorem 1.1 (Lovász). *Every n-vertex graph can be decomposed into at most $n/2$ paths and cycles. Consequently, every n-vertex graph can be decomposed into at most n paths.* □

Let G be any graph on n vertices and let $E_1, E_2, \ldots, E_k$ be a separating system of the edge set $E(G)$ where $k = \lceil \log_2 |E(G)| \rceil \leq 2 \log_2 n$. Let G_i be the subgraph of G induced by the edges of E_i. By Theorem 1.1, each G_i may be decomposed into at most n paths. Putting these together, we get a separating path system of G of cardinality at most kn. Consequently, we note that $f(n) \leq 2n \log_2 n$.

To bound $f(n)$ from below, let us consider K_n, the complete graph on n vertices. Suppose that we have a separating path system $\mathcal{P}$ of K_n with k paths. Note that at most one edge of K_n goes uncovered by the paths of $\mathcal{P}$ and further at most k edges of K_n belong to exactly one path of $\mathcal{P}$. Since any path of K_n has length at most $n-1$, we deduce that

$$k(n-1) \geq 1 + k + 2\left(\binom{n}{2} - k - 1\right)$$

or equivalently, $k \geq n - 1 - 1/n$. Thus, we note that $f(n) \geq n - 1$. We believe that the lower bound, rather than the upper bound, is closer to the truth, and we make the following conjecture.

Conjecture 1.2. *There exists an absolute constant C such that for every graph G on n vertices, $f(G) \leq Cn$.*

Let us remark that it is not inconceivable that $f(n) = (1 + o(1))n$ and Conjecture 1.2 is true for every $C > 1$. In this paper, we shall prove Conjecture 1.2 in certain special cases. Our first result establishes the conjecture for graphs of linear minimum degree.

Theorem 1.3. *Let $c > 0$ be fixed. Every graph G on n vertices with minimum degree at least cn has a separating path system of cardinality at most $122n/c^2$ for all sufficiently large n.*

Building upon the ideas used to prove Theorem 1.3, we shall prove Conjecture 1.2 for the Erdős-Rényi random graphs using the fact that these graphs have good connectivity properties.

Theorem 1.4. *For any probability $p = p(n)$, with high probability, the random graph $G(n, p)$ has a separating path system of size at most $48n$.*

Note in particular that Theorem 1.4 implies that Conjecture 1.2 is true for almost all n-vertex graphs with $C = 48$. Using Theorem 1.3, we shall also establish the conjecture for a class of dense graphs, which includes quasi-random graphs (in the sense of Chung, Graham and Wilson [10] and Thomason [23]) as a subclass.

Theorem 1.5. *Let $c > 0$ be fixed and let G be a graph on n vertices such that every subset $U \subseteq V(G)$ of size at least $\sqrt{n}$ spans at least $c|U|^2$ edges. Then $f(G) \leq 638n/c^3$ for all sufficiently large n.*

The above results are far from best possible but we make no attempt to optimise our bounds since it seems unlikely that our methods will yield the best possible constants. In the case of trees, however, we are able to obtain tight bounds.

Theorem 1.6. *Let T be a tree on $n \geq 4$ vertices. Then*

$$\left\lceil \frac{n+1}{3} \right\rceil \leq f(T) \leq \left\lfloor \frac{2(n-1)}{3} \right\rfloor .$$

Furthermore, these bounds are best possible.

We use standard graph theoretic notions and notation and refer the reader to [3] for terms and notation not defined here. We shall also make use of some well known results about random graphs without proof, see [4] for details.

In the next section, we describe a general strategy that we adopt to prove Theorems 1.3 and 1.4. We then prove Theorems 1.3 and 1.4 in Sections 3 and 4 respectively. Section 5 is devoted to the proof of Theorem 1.5. For the sake of clarity, we shall systematically omit ceilings and floors in Sections 3, 4 and 5. We then prove Theorem 1.6 in Section 6. We conclude the paper in Section 7 with a discussion of related questions and problems.

2. A general strategy

Theorems 1.3 and 1.4 are proved similarly, using the following strategy. Let G_1 and G_2 be subgraphs of G which partition the edge set of G. First, we decompose the edges of G_1 into at most $3n$ matchings $M_1, \ldots, M_{3n}$ as follows. Initially, each M_i is empty; we add the edges of G_1 one by one to a suitably chosen matching M_i. By Theorem 1.1, there exists a path decomposition of G_1 into (at most) n paths $P_1, \ldots, P_n$. Given an edge $e = xy \in E(G_1)$, let j be such that $e \in P_j$. We add e to a matching M_i which contains no edge of P_j and no edge incident to x or y. As the length of P_j is at most $n-1$ and there are at most $2n$ edges incident to either x or y, this process is well defined; indeed, we can always find a matching M_i satisfying the required conditions. Note that we have ensured that $|M_i \cap P_j| \leq 1$ for each $1 \leq i \leq 3n$ and $1 \leq j \leq n$.

Next, for each $1 \leq i \leq 3n$, we find a covering of $E(M_i)$ with paths using edges from $E(G_2) \cup E(M_i)$. These covering paths together with the paths $P_1, \ldots, P_n$ separate the edges of G_1 from each other and from the edges of G_2. To check this, consider an edge $e \in E(G_1)$ such that $e \in P_j$ and $e \in M_i$ and note that P_j separates e from every edge of G_2 as well as each edge in $E(G_1) \backslash P_j$, while the path covering M_i separates e from every other edge of P_j since $|M_i \cap P_j| \leq 1$. Repeating this process with the roles of G_1 and G_2 reversed, we obtain a separating path system of G.

In order to prove the existence of a small separating path system, we shall partition the graph G into G_1 and G_2 in a way that will enable us to keep the cardinalities of the above coverings small.

3. Graphs of linear minimum degree

Proof of Theorem 1.3. Let G be a graph on n vertices with minimum degree at least cn, for some fixed $0 < c < 1$. It is easy to decompose G into two disjoint subgraphs G_1 and G_2 in such a way that both subgraphs have minimum degree at least $cn/3$. Indeed, one way to do this is to define G_1 by randomly selecting each edge of G with probability $1/2$ and to take G_2 to be the complement of G_1 in G, i.e., $V(G_2) = V(G)$ and $E(G_2) = E(G)\backslash E(G_1)$; the minimum degree conditions follow from the standard estimates for the tail of the binomial distribution.

Following the strategy described in Section 2, let $P_1, \ldots, P_n$ be a path decomposition of G_1 and let $M_1, \ldots, M_{3n}$ be a decomposition of G_1 into matchings such that the intersection $M_i \cap P_j$ contains at most one edge for each i and j.

Define a graph H on $V(G)$ as follows: two distinct vertices $x, y \in V(G)$ are adjacent in H if they have at least $c^2n/24$ common neighbours in G_2. Note that H has no independent set of size $4/c$. Indeed, if $A \subseteq V(G)$ is an independent set in H of size $k = 4/c$, then writing $\Gamma(.)$ to denote vertex neighbourhoods, we have

$$\begin{aligned} n = |V(G)| &\geq \sum_{x\in A} |\Gamma(x, G_2)| - \sum_{x\neq y\in A} |\Gamma(x, G_2) \cap \Gamma(y, G_2)| \\ &> \frac{kcn}{3} - \frac{k^2c^2n}{48} = (4/3 - 1/3)n = n \end{aligned}$$

which is a contradiction.

For each $1 \leq i \leq 3n$, define a sequence of paths in $E(M_i) \cup E(H)$ as follows. Colour the edges of M_i blue and the edges of H red — note that there may be edges coloured both red and blue. Let $Q_{i,1}$ be a longest path alternating between blue and red edges and starting with a blue edge. Having defined $Q_{i,1}, \ldots, Q_{i,j-1}$, we set

$$E_{i,j} = E(M_i)\backslash(E(Q_{i,1}) \cup \cdots \cup E(Q_{i,j-1})).$$

If $E_{i,j} = \varnothing$, we stop. If not, let $Q_{i,j}$ be a longest path alternating between blue edges from $E_{i,j}$ and red edges, starting with a blue edge. Note that we might reuse red edges in this process, but not blue edges.

Since each $Q_{i,j}$ is a longest path, the starting vertices of the paths $Q_{i,j}$ form an independent set in H. Thus for each $1 \leq i \leq 3n$, we have at most $4/c$ such paths $Q_{i,j}$ and consequently at most $12n/c$ paths in total. Note that every edge of G_1 appears exactly once in one of these $12n/c$ paths as a blue edge. Thus the sum of the lengths of these paths $Q_{i,j}$ is at most $2|E(G_1)| \leq n^2$. We split each of the paths $Q_{i,j}$ into paths of length $c^2n/48$, where we allow one of the subpaths to have length less than $c^2n/48$. We thus obtain at most $n^2/(c^2n/48) + 12n/c \leq 60n/c^2$ red-blue paths. Note that for every red edge xy, the vertices x and y have at least $c^2n/24$ common neighbours in G_2. Consequently, we can transform all the red-blue paths into simple paths in G: we replace every red edge with a path of length two in G_2 with the same endpoints. We can do this because the number of common neighbours in G_2 of the ends of a red edge is at least twice the length of the original red-blue path. The family consisting of these paths and the paths $P_1, \dots, P_n$ separates the edges of G_1 and has size at most $60n/c^2 + n \leq 61n/c^2$.

By repeating the above process with the roles of G_1 and G_2 reversed, we obtain a separating path system of G of size at most $122n/c^2$. □

4. Random graphs

Proof of Theorem 1.4. We use different arguments for different ranges of the edge probability.

Case 1: $p \geq 10 \log n/n$. Let G be a copy of $G(n, 2p)$, where $p \geq 5 \log n/n$. We define graphs G_1 and G_2 on the vertex set of G as follows. We construct G_1 by randomly selecting each edge of G with probability $1/2$ and we take G_2 to be the complement of G_1 in G; clearly, G_1 and G_2 are edge-disjoint copies of $G(n, p)$.

The following lemma is easily proved using the standard estimates for the tail of the binomial distribution.

Lemma 4.1. *Let $p \geq 5 \log n/n$. Then with high probability, the following assertions hold.*

1. $n^2p/4 \leq |E(G(n, p))| \leq n^2p$.
2. $G(n, p)$ *has minimum degree at least* $np/5$.
3. $G(n, p)$ *is* $np/10$*-connected.*

We shall also need the notion of a *k-linked graph.* A graph is said to be k-linked if it has at least $2k$ vertices and for every sequence of $2k$ distinct vertices $u_1, \dots, u_k, v_1, \dots, v_k$, there exist vertex disjoint paths $P_1, \dots, P_k$ such that the endpoints of P_i are u_i and v_i. Bollobás and Thomason [7] showed that every $22k$-connected graph is k-linked. This was later improved by Thomas and Wollan [22], who proved that every $2k$-connected graph on n vertices with at least $5kn$ edges is k-linked. From the latter result and Lemma 4.1, we conclude that with high probability, both G_1 and G_2 are $np/20$-linked.

Following the strategy described in Section 2, we find a decomposition of G_1 into paths $P_1, \dots, P_n$ and a decomposition of G_1 into matchings $M_1, \dots, M_{3n}$ such that the intersection $M_i \cap P_j$ contains at most one edge for each i and j.

We decompose each matching M_i into submatchings of size at most $np/20$. Since G_1 has at most n^2p edges, we thus obtain at most $23n$ different matchings $M'_1, \ldots, M'_{23n}$. Now, since G_2 is $np/20$-linked, we can complete each such matching M'_i into a path using the edges of G_2. These paths along with $P_1, \ldots, P_n$ constitute a separating path system of G_1 of size at most $24n$. Reversing the roles of G_1 and G_2, we obtain a set of $24n$ paths separating the edges of G_2. The union of these two families of paths is a separating path system of G of cardinality at most $48n$.

Case 2: $p \le 10/n$**.** In this case, with high probability, $G(n,p)$ has at most $20n$ edges and so the edges of G constitute a separating path system of size at most $20n$.

Case 3: $10/n \le p \le 10 \log n/n$**.** We begin by collecting together some useful properties of sparse random graphs. We will need some notation: Given a graph G, write $B_i(v) = B_i(v, G)$ for the set of vertices at (graph) distance at most i from v and let $\Gamma_i(v) = \Gamma_i(v, G) = B_i(v) \backslash B_{i-1}(v)$. The following lemma is somewhat technical; we defer its proof to the end of the section.

Lemma 4.2. *Let $10 \le d \le 10 \log n$. Then with high probability, the following assertions hold for $G = G(n, d/n)$.*

1. *G has at most dn edges.*
2. *$|\Gamma_i(x)| \le (2d)^i \log n$ for every $x \in V(G)$ and $i \le n$.*
3. *Every set of $i \le \sqrt{n}$ vertices spans at most $2i$ edges. Furthermore, every set of $i \le 10 \log\log n$ vertices spans at most i edges.*
4. *Let G' be a subgraph of G with minimum degree at least 10. Then $|\Gamma_i(x, G')| \ge 2^i$ for every $x \in V(G')$ and every $1 \le i \le 10 \log\log n$.*
5. *Let G' be a subgraph of G obtained by deleting at most $20d \log n$ vertices and edges and let $l = 3 \log\log n$. For every pair of vertices $x, y \in V(G')$ such that $|B_l(x, G')|, |B_l(y, G')| \ge (\log n)^3$, there is a path of length at most $2 \log n$ between x and y in G'.*

The *k-core* of a graph is its largest induced subgraph with minimum degree at least k. Let H be the 15-core of $G = G(n,p)$ and let $d = np$. By Theorem 1.1, we can decompose H into n paths. Since by Lemma 4.2(1) there are at most dn edges in G, we can decompose these n paths into at most $2n$ subpaths $Q_1, \ldots, Q_{2n}$, each of which has length at most d.

Let $l = 3 \log\log n$. We shall define a collection of at most $2n$ matchings in H of size d each using the paths $Q_1, \ldots, Q_{2n}$. Each of these matchings will consist of d edges $e_1, \ldots, e_d$ chosen from some d distinct paths $Q_{i_1}, \ldots, Q_{i_d}$ which have the additional property that for every $j \ne j'$ and every $x \in V(Q_{i_j})$ and $x' \in V(Q_{i_{j'}})$ we have $B_l(x) \cap B_l(x') = \varnothing$.

We begin with a collection of paths $R_1, \ldots, R_{2n}$ which we modify as we go along. Initially we set $R_i = Q_i$ for every i. We define our collection of matchings in H in a sequence of rounds.

At the beginning of a round, if we have fewer than $2\sqrt{n}$ non-empty paths R_i, we stop. Otherwise, we select d of the R_i (in a way we specify below), remove the initial edge from each of these paths and use these d removed edges to form a matching of size d. To choose our d paths $R_{i_1}, \ldots, R_{i_d}$ we proceed as follows. Let R_{i_1} be any non-empty path. Now, assume that we have chosen $R_{i_1}, \ldots, R_{i_{t-1}}$, where $t \leq d$. Let $N_t = \bigcup_x B_{2l+1}(x)$, where the union is taken over all $x \in V(Q_{i_1}) \cup \cdots \cup V(Q_{i_{t-1}})$. From Lemma 4.2(2), we see that

$$|N_t| < (t-1)d(2d)^{2l+2} \log n < (2d)^{2l+4} \log n < \sqrt{n}.$$

Thus by Lemma 4.2(3), N_t spans at most $2\sqrt{n}$ edges. Since we started the round with more than $2\sqrt{n}$ non-empty paths, there is a path which contains no edge induced by the vertex set N_t. Let R_{i_t} be any such a path. We repeat the procedure until the d paths $R_{i_1}, R_{i_2}, \ldots, R_{i_d}$ have been obtained. Clearly the matchings defined by this process are disjoint and of size d, so there are at most n of them, which we can denote by $M_1, \ldots, M_n$.

In Lemma 4.3 (stated below), we show that for each such matching M_i, there is a path containing $E(M_i)$ and avoiding, for every $e \in E(M_i)$, the other edges of the path $Q \in \{Q_1, Q_2 \ldots, Q_{2n}\}$ containing e.

We leave it to the reader to verify that we then obtain a separating system of size at most $19n$ by taking the union of

1. the edges $E(G) \backslash E(H)$ of which there are at most $15n$,
2. the paths $Q_1, \ldots, Q_{2n}$,
3. the edges of H which are not covered by the matchings $M_1, \ldots, M_n$ of which there are at most $2d\sqrt{n} \leq n$, and
4. the set of n paths promised by Lemma 4.3.

We now state and prove Lemma 4.3.

Lemma 4.3. *Let $G = G(n,p)$ be a graph satisfying Lemma 4.2 and let $S_1, \ldots, S_d$ be vertex-disjoint paths of length at most d in the 15-core H of G. Set $l = 3 \log \log n$ and assume that $B_l(x) \cap B_l(y) = \varnothing$ for every $x \in V(S_i)$ and $y \in V(S_j)$ with $i \neq j$. For each i, select an edge $e_i = x_i y_i$ from S_i, and set $M = \{e_1, e_2, \ldots, e_d\}$. Then there exists a path in G containing all the edges of M and no other edge from $\bigcup_{1 \leq i \leq d} E(S_i)$.*

Proof. Write $E' = (\bigcup_{1\le i\le d} E(S_i))\setminus E(M)$, let G_0 be the graph on $V(G)$ with edge set $E(G)\setminus E'$, and let G_1 be the graph obtained from G_0 by deleting x_1. Consider the graph H_1 on the vertex set $V(H)\setminus\{x_1\}$ with edge set $E(H)\cap E(G_1)$. Note that H_1 has minimum degree at least 12, since by removing the vertex-disjoint paths $S_1,\dots,S_d$ and the vertex x_1 we decrease vertex degrees in H by at most 3. Thus by Lemma 4.2(4), $|B_l(v,G_1)|\ge(\log n)^3$ for every $v\in V(M)$.

We define vertex-disjoint paths $P_1,\dots,P_{d-1}$ of size at most $2\log n$ as follows. Suppose that we have already defined the paths $P_1,P_2,\dots,P_{i-1}$ for some $i<d$. Set $G_i = G_1\setminus\bigcup_{1\le j<i} V(P_j)$ and let P_i be a shortest path in G_i connecting y_i to a vertex from $\bigcup_{i+1\le j\le d}\{x_j,y_j\}$. Relabelling the remaining vertices and edges if necessary, assume that this path connects y_i to x_{i+1}.

We shall show by induction that P_i has length at most $2\log n$. Assume that we have defined $P_1,\dots,P_{i-1}$. By the inductive hypothesis, note that we may assume that G_i is obtained by removing at most $2d\log n$ vertices and at most $d^2\le 10d\log n$ edges from G. Consequently, the bound on the length of P_i would follow from Lemma 4.2(5) by showing that both $|B_l(y_i,G_i)|, |B_l(x_{i+1},G_i)|\ge(\log n)^3$.

First, we claim that $B_l(x_{i+1},G_i)=B_l(x_{i+1},G_1)$. To see this, first note that for every $j\le i-1$, the sets $B_l(y_j,G)$ and $B_l(x_{i+1},G)$ are disjoint and, consequently, so are $B_l(y_j,G_j)$ and $B_l(x_{i+1},G_j)$. Since P_j is a shortest path from y_j to $\bigcup_{j+1\le k\le d}\{x_k,y_k\}$, it follows that $V(P_j)\cap B_l(x_{i+1},G_j)=\varnothing$. Hence, $B_l(x_{i+1},G_{j+1})=B_l(x_{i+1},G_j)$ for every $j\le i-1$. Therefore $|B_l(x_{i+1},G_i)|\ge(\log n)^3$ and notice that $B_l(y_i,G_{i-1})=B_l(y_i,G_1)$ by the same argument.

Next, by the minimality of P_{i-1}, it is clear that the set $V(P_{i-1})\cap B_l(y_i,G_{i-1})$ is contained in the set V'_{i-1} of the last $l+1$ vertices of P_{i-1}. Let H_i be the subgraph of H_1 induced by the vertex subset $V(H_1)\setminus V'_{i-1}$. We deduce from Lemma 4.2(3) that no vertex of G_1 has more than two neighbours in V'_{i-1} and so H_i has minimum degree at least 10. By Lemma 4.2(4) we then have $|B_l(y_i,H_i)|\ge(\log n)^3$. Since $V(H_i)\cap B_l(y_i,G_1)\subseteq V(G_i)\cap B_l(y_i,G_1)$, it follows that $B_l(y_i,H_i)\subseteq B_l(y_i,G_i)$. Hence, $|B_l(y_i,G_i)|\ge(\log n)^3$ and Lemma 4.3 follows by Lemma 4.2(5). □

We now complete the proof of Theorem 1.4 by proving Lemma 4.2.

Proof of Lemma 4.2. Parts (1) and (2) of Lemma 4.2 follow easily from the standard Chernoff-type bounds for the tails of binomial random variables. Part (3) is also routine: the probability that a given set of i vertices induces k or more edges is at most $\binom{i^2}{k}(d/n)^k$ and a straightforward union bound over all sets of i vertices establishes both the claimed statements.

To prove part (4), we assume that G satisfies parts (2) and (3). Let G' be a subgraph of G with minimum degree at least 10. Let $x \in V(G')$ and write $\Gamma_i = \Gamma_i(x, G')$ and $B_i = B_i(x, G')$.

Claim 4.4. $|\Gamma_i| \geq 2|B_{i-1}|$ *for* $1 \leq i \leq 10 \log\log n$.

Proof. By part (2), $|\Gamma_i| \leq (2d)^i \log n$ for $1 \leq i \leq 10 \log\log n$ and so, $|B_i| \leq 2(2d)^{i+1} \log n \leq \sqrt{n}$. So, by part (3), B_i spans at most $2|B_i|$ edges for every $1 \leq i \leq 10 \log\log n$. Since every vertex in B_{i-1} has degree at least 10 in G' and B_{i-1} spans at most $2|B_{i-1}|$ edges, there are at least $6|B_{i-1}|$ edges from B_{i-1} to Γ_i. As B_{i-1} is connected, B_i must span at least $7|B_{i-1}| - 1$ edges. Since B_i spans at most $2|B_i|$ edges, this implies that $|B_i| \geq (7|B_{i-1}|-1)/2 \geq 3|B_{i-1}|$, i.e., $|\Gamma_i| \geq 2|B_{i-1}|$. □

In particular, Claim 4.4 implies that $|\Gamma_i| \geq 2^i$ for $i \leq 10 \log\log n$, proving part (4). In order to prove part (5), we shall need the following.

Claim 4.5. *Let* $l = 3 \log\log n$*. Let* G' *be a graph obtained from* G *by removing at most* $20d \log n$ *vertices and edges and let* x *be a vertex in* G' *satisfying* $|B_l(x, G')| \geq (\log n)^3$*. Then with high probability for every such* G' *and* x*, there exists an* $i < \log n$ *such that* $|\Gamma_i(x, G')| \geq n/2d$*.*

Proof. Write $\Gamma_i = \Gamma_i(x, G')$ and $B_i = B_i(x, G')$, and let V' and E' be the set of vertices and edges removed from G to obtain G'. In particular, note that the assumption on x implies that there exists a $k \leq l$ such that $|\Gamma_k| \geq (\log n)^{5/2}$. By part (2), with high probability we also have $|B_k| = o(n/d)$.

We show that with high probability for every G' and x as above and $i \geq k$, either $|\Gamma_{i+1}| \geq (d/2)|\Gamma_i|$ or $|\Gamma_i| \geq (n/2d)$. Note that this would prove Claim 4.5.

Conditional on $|\Gamma_i| \leq n/2d$ and on $|\Gamma_{j+1}| \geq (d/2)|\Gamma_j|$ for $k \leq j < i$, we shall bound from above the probability that $|\Gamma_{i+1}| \leq (d/2)|\Gamma_i|$. Write $A_i = V(G') \backslash (\Gamma_i \cup V' \cup V(E'))$, and let A_i' be the set of those vertices in A_i which are adjacent to some vertex in Γ_i. It follows that since $|\Gamma_i| \leq n/2d$ and $|V'|, |E'| \leq 20d \log n$, $|A_i| \geq 9n/10$ for all sufficiently large n. We shall estimate the probability that $|A_i'| \leq (d/2)|\Gamma_i|$, conditional on $|\Gamma_i| \geq (\log n)^{5/2}$ and $|A_i| \geq 9n/10$. The probability that a particular vertex in A_i is adjacent to some vertex in Γ_i is

$$\left(1 - (1 - d/n)^{|\Gamma_i|}\right) \geq \frac{d|\Gamma_i|}{n} - \frac{1}{2}\left(\frac{d|\Gamma_i|}{n}\right)^2 \geq \frac{3d|\Gamma_i|}{4n}$$

since no edge adjacent to any vertex in A_i is deleted in passing to G' from G. Thus the expected size of A'_i is at least $27d|\Gamma_i|/40$. By appealing to the standard bounds for the tail of a binomial random variable, we see that

$$\begin{aligned}\mathbb{P}\big[\,|A'_i| \leq \frac{d}{2}|\Gamma_i|\,\big] &\leq \mathbb{P}\Big[\,|A'_i| \leq \frac{3}{4}\mathbb{E}|A'_i|\,\Big]\\ &\leq \exp\left(-\mathbb{E}|A'_i|/32\right)\\ &\leq \exp\left(-d(\log n)^{5/2}/100\right).\end{aligned}$$

Since we have $2^{O(d(\log n)^2)}$ choices for G', x and i, this implies that $|A'_i| \geq (d/2)|\Gamma_i|$ with high probability, as required. □

We now complete the proof of part (5) of Lemma 4.2. Using Claim 4.5, we can find $s, t < \log n$ such that $|\Gamma_s(x, G')|, |\Gamma_t(y, G')| \geq n/2d$. If $B_s(x, G') \cap B_t(y, G') \neq \varnothing$, the assertion of part (5) follows. Otherwise, note that the probability that there are no edges between $\Gamma_s(x, G')$ and $\Gamma_t(y, G')$ is at most $(1 - d/n)^{n^2/4d^2 - 20d \log n} \leq e^{-n/5d}$. Since we have $2^{O(d(\log n)^2)}$ choices for x, y and G', this implies that with high probability the assertion of part (5) holds. □

We have established Lemma 4.2, thus completing the proof of Theorem 1.4. □

5. Dense graphs

Proof of Theorem 1.5. Let $c > 0$ and let G be a graph on n vertices such that for every $k \geq \sqrt{n}$, every set of k vertices spans at least ck^2 edges.

We define a sequence of subgraphs $G = G_0 \supseteq G_1 \supseteq \cdots \supseteq G_{l-1}$ and a related sequence of graphs $H_1, H_2, \ldots, H_l$ as follows. Start by setting $G_0 = G$. If $|V(G_{i-1})| \leq \sqrt{n}$, we stop and take $H_i = G_{i-1}$. Otherwise, we take H_i to be the $(c|G_{i-1}|/2)$-core of G_{i-1} and define G_i to be the graph induced by $V(G_{i-1}) \backslash V(H_i)$. Note that the sets $V(H_i)$ form a partition of $V(G)$.

Let us write g_i and h_i for the number of vertices of G_i and H_i, respectively. It is well known that the k-core of a graph can be found by removing vertices of degree at most $k-1$, in arbitrary order, until no such vertices exist. So the number of edges removed from G_{i-1} to obtain its $(cg_{i-1}/2)$-core is at most $cg_{i-1}^2/2$. Thus, at least $cg_{i-1}^2/2$ edges remain, so $h_i \geq \sqrt{c}g_{i-1} \geq cg_{i-1} \geq cg_i$.

We first separate the internal edges of the graphs H_i. Note that H_i has minimum degree at least $ch_i/2$. So we conclude from Theorem 1.3 that H_i has a separating system of size at most $488h_i/c^2$ for every $1 \leq i < l$. Also, since $V(H_l) \leq \sqrt{n}$, we may separate the edges of H_l trivially (by adding each edge individually to our separating path system). This contributes at most n paths. Since the graphs H_i are pairwise vertex disjoint, we may separate the internal edges of the H_i using at most $488n/c^2$ paths.

It remains to separate the crossing edges between the H_i. For $1 \leq i < l$, let E_i be the set of edges of the form xy where $x \in V(H_i)$ and $y \in V(G_i)$, and let E'_i be the set of such edges xy where y has at least 3 neighbours in H_i. Note that every edge of G not contained in any of the H_i is contained in one of the E_i.

We define a g_i-edge-coloured multigraph F_i on the vertex set of H_i as follows. If $v \in G_i$ has at least three neighbours in H_i, say $x_1, \ldots, x_k$, we add the edges $x_1x_2, x_2x_3, \ldots, x_kx_1$ to F_i and colour these edges with the colour v; in other words, we add a v-coloured cycle through the neighbours of v. Note that the degree of every vertex in F_i (as a multigraph) is at most $2g_i$ and every colour class contains at most h_i edges. Since each edge has at most $4g_i + h_i \leq 5h_i/c$ edges which are either incident to it or from the same colour class, we can, as in Section 2, decompose F_i into at most $5h_i/c$ rainbow matchings $M_1, \ldots, M_{5h_i/c}$ (by a rainbow matching, we mean a matching containing at most one edge from each colour class).

We now construct another sequence of rainbow matchings decomposing F_i with the following property. Denote by $e_1, \ldots, e_m$ the edges in M_j and let $v_1, \ldots, v_m$ be their respective colour classes. Let α_k and β_k be the two neighbours of e_k in the cycle whose edges have colour v_k. In our second sequence of rainbow matchings, we would like the matching containing e_k to avoid e_t, α_t and β_t for every $t \neq k$. Since each edge has to avoid at most $4g_i + h_i + 3h_i \leq 8h_i/c$ other edges, we can find such a decomposition into at most $8h_i/c$ matchings, say, $M_{5h_i/c+1}, \ldots, M_{13h_i/c}$.

We now mimic the proof of Theorem 1.3. Let us define a graph H'_i on $V(H_i)$ where we join two vertices if they have more than $c^2h_i/24$ common neighbours in H_i. For each $1 \leq j \leq 13h_i/c$, we can find a collection of at most $4/c$ paths whose edges alternate between those of M_j and H'_i which cover each edge of M_j exactly once; we obtain $52h_i/c^2$ such paths in total. We divide these paths into subpaths of length at most $c^2h_i/48$ each, resulting in a collection of at most $96h_i/c^3 + 52h_i/c^2$ paths. Each such path can be transformed into a path in G by replacing every edge from H'_i with a suitably chosen path of length two in H_i and every coloured edge $e = xy$ from M_j with the path xvy where v is the colour of e. Since the matchings are rainbow matchings, these paths are guaranteed to be simple. It is easy to see that the collection of $96h_i/c^3 + 52h_i/c^2$ paths defined above separates E'_i.

It remains to separate edges in $E_i \backslash E_i'$ for $1 \leq i < l$. Note that there are at most $2(g_1 + \cdots + g_l) \leq 2(h_1 + \cdots + h_l)/c \leq 2n/c$ such edges; we add each such edge to our separating path system.

It is easy to check that we have constructed a separating path system of G of cardinality at most $488n/c^2 + 96n/c^3 + 52n/c^2 + 2n/c \leq 638n/c^3$. The result follows. □

6. Trees

We begin by collecting together a few simple observations into the following lemma.

Lemma 6.1. *Let T be a tree on $n \geq 3$ vertices, and let $\mathcal{P}$ be a separating path system of T. Then the following assertions hold.*

1. *With the exception of at most one leaf, every leaf of T is an endpoint of a path in $\mathcal{P}$.*
2. *If a path in $\mathcal{P}$ has two leaves u and v as its endpoints, then there must be at least one path in $\mathcal{P}$ which has exactly one of u and v as an endpoint.*
3. *Every vertex of degree two in T is an endpoint of a path in $\mathcal{P}$.*

Proof. Clearly, a leaf must be an endpoint of any path through it. Since $\mathcal{P}$ separates $E(T)$, there is at most one edge of T which is not covered by any path in $\mathcal{P}$. As $n \geq 3$, T does not consist of a single edge and thus at most one leaf of T is visited by no path in $\mathcal{P}$. This establishes part (1).

Suppose that we have a path $P \in \mathcal{P}$ having two leaves $u, v \in V(T)$ as its endpoints. Let e_u and e_v be the edges incident to u and v respectively. Since $\mathcal{P}$ separates $E(T)$, there must be some path $P' \in \mathcal{P}$ containing exactly one of e_u and e_v. This establishes part (2).

Suppose that v is a vertex of degree two in T; let e_1 and e_2 be the two edges of T incident to v. Since $\mathcal{P}$ separates $E(T)$, there must be some path $P \in \mathcal{P}$ containing exactly one of e_1 and e_2. Since v has degree two, it must be an endpoint of this path P. This establishes part (3). □

We split the proof of Theorem 1.6 into two parts.

Proof of the lower bound in Theorem 1.6. Let T be a tree on $n \geq 4$ vertices and let $\mathcal{P}$ be a separating system of T. We shall show that $|\mathcal{P}| \geq \lceil (n+1)/3 \rceil$.

First, suppose that there is a leaf v which is not the endpoint of any path in $\mathcal{P}$. Let e_v be the edge of T incident to v. Since $\mathcal{P}$ separates $E(T)$, e_v is the unique edge of T not covered by any path in $\mathcal{P}$. Delete v from T to obtain a tree T' on $n - 1 \geq 3$ vertices.

The family $\mathcal{P}$ both covers and separates $E(T')$. From Lemma 6.1, we note that every leaf and every vertex of degree two of T' is the endpoint of at least one path from $\mathcal{P}$. Furthermore, we know that if a path from $\mathcal{P}$ has a pair of leaves for its endpoints, then at least one of those leaves is the endpoint of at least one other path from $\mathcal{P}$.

Let d_1 and d_2 denote the number of leaves and degree two vertices in T'. We claim that $\mathcal{P}$ contains at least $(2d_1 + d_2)/3$ paths. To see this, start by placing a red token on every leaf and a blue token on every vertex of degree two in T'. We then iterate through the paths of $\mathcal{P}$ in some order and in each iteration, we remove whatever tokens there are at the endpoints of the current path. If both the tokens removed are red, then we know that both the endpoints, say u and v, of the current path are leaves and that at least one of them, say u, is the endpoint of a different path; we then place a blue token on u. Writing R and B respectively for the number of red and blue tokens remaining on the tree, we see that the quantity $2R+B$ does not decrease by more than three in any iteration. Since $\mathcal{P}$ is a separating path system, all the tokens must have been removed by the end of the procedure. It follows that

$$|\mathcal{P}| \geq \frac{2d_1 + d_2}{3}.$$

Now, note that

$$2e(T') = 2(n-2) \geq d_1 + 2d_2 + 3(n - 1 - d_1 - d_2),$$

which we can rearrange to get

$$\frac{2d_1 + d_2}{3} \geq \frac{n+1}{3}.$$

Taken together, these inequalities show that $|\mathcal{P}| \geq (n+1)/3$.

If on the other hand every leaf of T is the endpoint of some path from $\mathcal{P}$, then, by repeating the argument above with T instead of T', we find that $|\mathcal{P}| \geq (n+2)/3$. We know from Lemma 6.1 that these are the only two possibilities; consequently, we are done.

To see that this lower bound is best possible, consider the family of *hair combs*, where the hair comb of order $3n$ is obtained by starting with a *spine* consisting of a path of length $n-1$ and then attaching a path of length two to each vertex of the spine. It is an easy exercise to show that this lower bound is tight for hair combs. (See Figure 3 for an example of an optimal separating path system.) □

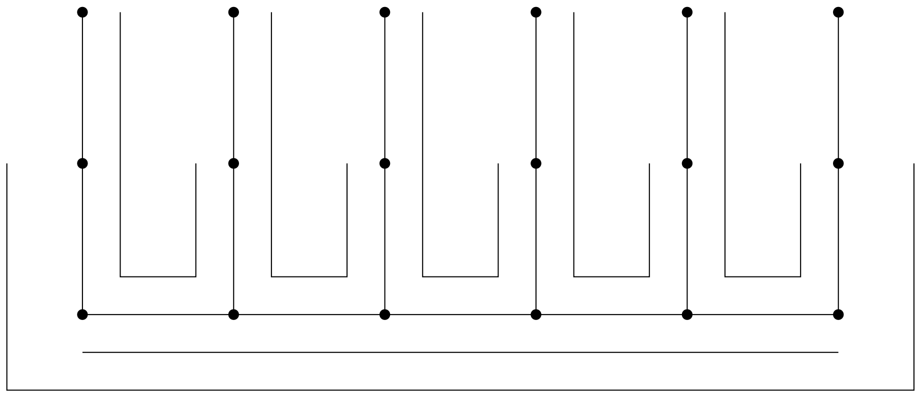

Figure 3: A hair-comb of order 18 and a separating system of 7 paths.

We now turn our attention to the second part of the proof of Theorem 1.6.

Proof of the upper bound in Theorem 1.6. We shall show by induction on $n = |V(T)|$ that $f(T) \leq \lfloor 2(n-1)/3 \rfloor$.

There is, up to isomorphism, only one tree of order n for each of $n = 1, 2, 3$, namely the path of length $n-1$. It is trivial to check that the claim holds for these trees.

Let T be a tree of order $n > 3$. If T is a path, then it is easy to show using Lemma 6.1 that $f(T) \leq \lceil (n-1)/2 \rceil$ and it is easy (see Figure 1) to construct a separating path system matching this bound. Since $\lceil (n-1)/2 \rceil \leq \lfloor 2(n-1)/3 \rfloor$ for all $n \geq 4$, we may suppose that T is not a path; hence T must contain at least one vertex u with three distinct neighbours, say v_1, v_2 and v_3. Contract the edges uv_1, uv_2 and uv_3 to obtain a new tree T' on $n-3$ vertices.

We find a separating path system $\mathcal{P}'$ of T' of size at most $2(n-4)/3$. We may think of $\mathcal{P}'$ as a family of paths of T since paths in T' map to paths in T in a natural way: a path in T' is lifted up to a path in T with the same endpoints (where we identify the vertex resulting from the contraction of u, v_1, v_2 and v_3 with u). Consider the family

$$\mathcal{P} = \mathcal{P}' \cup \{v_1uv_2, v_2uv_3\}.$$

Since $\mathcal{P}'$ separates $E(T')$, it readily follows that $\mathcal{P}'$, when viewed as a family of paths of T, separates $E(T) \setminus \{uv_1, uv_2, uv_3\}$. The two paths v_1uv_2 and

v_2uv_3 then separate uv_1, uv_2 and uv_3 from each other and from the rest of $E(T)$. Thus,

$$|\mathcal{P}| \le \frac{2(n-4)}{3} + 2 = \frac{2(n-1)}{3}.$$

We are done by induction.

To see that this upper bound is best possible, consider the family of *stars*, where the star of order n consists of a single internal vertex joined to $n-1$ leaves. By mimicking the proof of the lower bound using Lemma 6.1, it is an easy exercise to verify that the upper bound is tight for stars. □

7. Concluding remarks

There remain a number of interesting questions which merit investigation. While the main open problem is, of course, to establish that $f(n) = O(n)$, there are many other attractive related extremal questions. For instance, it would be interesting to determine the value of $f(K_n)$ exactly. One can also ask the same question for the the d-dimensional hypercube Q_d. It is easy to cover Q_d with $d-1$ ladders and so $f(Q_d) = O(d^2)$. On the other hand, we know from the information theoretic lower bound that $f(Q_d) = \Omega(d)$. It would be interesting to nail down the exact value of $f(Q_d)$.

A different question, though of a similar flavour, raised by Bondy [8] and answered by Li [16], is that of finding *perfect path double covers*, i.e., a set of paths of a graph such that each edge of the graph belongs to exactly two of the paths and each vertex of the graph is an endpoint of exactly two of the paths. We suspect that the tools developed to tackle this problem and its variants might prove useful in attempting to establish that $f(n) = O(n)$.

Acknowledgements

Some of the research in this paper was carried out while the authors were participating in the 5$^{\text{th}}$ Emléktábla Workshop in Budapest, Hungary. This research was continued while the fourth and fifth authors were visitors at the University of Memphis. The authors are grateful to the organisers of the Emléktábla Workshop for their hospitality, and the fourth and fifth authors are additionally grateful for the hospitality of the University of Memphis.

Note added in proof

Shortly after this article was submitted, Balogh, Csaba, Martin and Pluhár [2], working independently, announced some results on a similar problem. Given a graph G, they consider the problem of finding a family of paths $\mathcal{P}$ such that for every pair of edges $e, f \in E(G)$, there exist $P_e, P_f \in \mathcal{P}$ such that $e \in P_e, f \notin P_e$ and $f \in P_f, e \notin P_f$. The methods developed in this paper are applicable to their notion of separation in certain cases; it is easy to check that the separating path systems constructed in the proofs of Theorems 1.3, 1.4 and 1.5 satisfy their notion of separation as well. However, if we are interested in exact results (as we are in the case of trees and forests, for instance), then it would seem that the two notions of separation lead to different behaviours.

References

[1] S. S. Ahuja, S. Ramasubramanian, and M. M. Krunz (2009). Single-link failure detection in all-optical networks using monitoring cycles and paths. *IEEE/ACM Trans. Netw.* **17** 1080–1093.

[2] J. Balogh, B. Csaba, R. R. Martin, and A. Pluhár. On the path separation number of graphs. Preprint. arXiv: 1312.1724.

[3] B. Bollobás (1998). *Modern Graph Theory. Graduate Texts in Mathematics.* Springer-Verlag, New York. MR1633290

[4] B. Bollobás (2001). *Random graphs*, 2nd ed. *Cambridge Studies in Advanced Mathematics.* Cambridge University Press, Cambridge. MR1864966

[5] B. Bollobás and A. D. Scott (2007). On separating systems. *European J. Combin.* **28** 1068–1071. MR2305574

[6] B. Bollobás and A. D. Scott (2007). Separating systems and oriented graphs of diameter two. *J. Combin. Theory Ser. B* **97** 193–203. MR2290320

[7] B. Bollobás and A. Thomason (1996). Highly linked graphs. *Combinatorica* **16** 313–320. MR1417341

[8] J. A. Bondy (1990). Perfect path double covers of graphs. *J. Graph Theory* **14** 259–272. MR1053608

[9] M. Cai (1984). On separating systems of graphs. *Discrete Math.* **49** 15–20. MR0737613

[10] F. R. K. Chung, R. L. Graham, and R. M. Wilson (1989). Quasi-random graphs. *Combinatorica* **9** 345–362. MR1054011

[11] F. Foucaud and M. Kovše (2013). Identifying path covers in graphs. *J. Discrete Algorithms* **23** 21–34. MR3132877

[12] G. Hansel (1964). Nombre minimal de contacts de fermeture nécessaires pour réaliser une fonction booléenne symétrique de n variables. *C. R. Acad. Sci. Paris* **258** 6037–6040. MR0163851

[13] N. J. A. Harvey, M. Patrascu, W. Yonggang, S. Yekhanin, and V. W. S. Chan (2007). Non-Adaptive Fault Diagnosis for All-Optical Networks via Combinatorial Group Testing on Graphs. In *26th IEEE International Conference on Computer Communications* 697–705.

[14] G. O. H. Katona (1966). On separating systems of a finite set. *J. Combinatorial Theory* **1** 174–194. MR0195749

[15] A. Kündgen, D. Mubayi, and P. Tetali (2001). Minimal completely separating systems of k-sets. *J. Combin. Theory Ser. A* **93** 192–198. MR1807121

[16] H. Li (1990). Perfect path double covers in every simple graph. *J. Graph Theory* **14** 645–650. MR1079213

[17] L. Lovász (1968). On covering of graphs. In *Theory of Graphs (Proc. Colloq., Tihany, 1966).* Academic Press, New York. pp. 231–236. MR0233723

[18] C. Ramsay and I. T. Roberts (1996). Minimal completely separating systems of sets. *Australas. J. Combin.* **13** 129–150. MR1383413

[19] A. Rényi (1961). On random generating elements of a finite Boolean algebra. *Acta Sci. Math. Szeged* **22** 75–81. MR0126398

[20] J. H. Spencer (1970). Minimal completely separating systems. *J. Combinatorial Theory* **8** 446–447. MR0256890

[21] J. Tapolcai, L. Rónyai, and P. Ho (2013). Link fault localization using bi-directional M-trails in all-optical mesh networks. *IEEE Transactions on Communications* **61** 291–300.

[22] R. Thomas and P. Wollan (2005). An improved linear edge bound for graph linkages. *European J. Combin.* **26** 309–324. MR2116174

[23] A. Thomason (1987). Pseudorandom graphs. In *Random graphs '85 (Poznań, 1985).* North-Holland Math. Stud. **144** 307–331. North-Holland, Amsterdam. MR0930498

[24] I. Wegener (1979). On separating systems whose elements are sets of at most k elements. *Discrete Math.* **28** 219–222. MR0546657

Victor Falgas-Ravry
Vanderbilt University
USA
E-mail address: victor.falgas-ravry@vanderbilt.edu

Teeradej Kittipassorn
University of Memphis
USA
E-mail address: t.kittipassorn@memphis.edu

Dániel Korándi
ETH Zürich
Switzerland
E-mail address: daniel.korandi@math.ethz.ch

Shoham Letzter
University of Cambridge
UK
E-mail address: s.letzter@dpmms.cam.ac.uk

Bhargav P. Narayanan
University of Cambridge
UK
E-mail address: b.p.narayanan@dpmms.cam.ac.uk

Received November 12, 2013