Separating systems and oriented graphs of diameter two

Béla Bollobás *   Alex Scott †

April 24, 2006

Abstract

We prove results on the size of weakly and strongly separating set systems and matrices, and on cross-intersecting systems. As a consequence, we improve on a result of Katona and Szemerédi [6], who proved that the minimal number of edges in an oriented graph of order n with diameter 2 is at least $(n/2)\log_2(n/2)$. We show that the minimum is $(1+o(1))n\log_2 n$.

1 Introduction

The diameter of a strongly connected digraph G is defined as $\text{diam}(G) = \max\{d(x,y) : x \in V(G), y \in V(G) \setminus x\}$, where $d(x,y)$ is the minimal length of a directed path from x to y. It is easily seen that a digraph of order n and diameter 2 can have as few as $2n-2$ edges: just pick a vertex x and take all edges into and out of x.

For *oriented* graphs the situation is surprisingly different. Katona and Szemerédi [6] showed that the minimal number of edges in an oriented graph of order n and diameter 2 is at least $(n/2)\log_2(n/2)$. They also suggested (but did not prove) an upper bound $n\log_2 n$. In this paper, we will show

*Trinity College, Cambridge CB2 1TQ, UK *and* Department of Mathematical Sciences, University of Memphis, Memphis TN38152, USA; email: bollobas@msci.memphis.edu Research supported in part by NSF grants CCR-0225610 and DMS-0505550

†Mathematical Institute, University of Oxford, 24-29 St Giles, Oxford OX1 3LB; email: scott@maths.ox.ac.uk.

that every oriented graph with diameter 2 has at least $(1 + o(1))n \log_2 n$ edges (an upper bound matching this to within a $1 + o(1)$ factor is given by a construction at the end of Section 4; in fact, the two bounds differ by $O(n \log_2 \log_2 n)$).

We will also be concerned with the size of weakly and strongly separating set systems. Indeed, our result on oriented graphs rests on a defect result for strongly separating systems. A sequence $(S_1, T_1), \ldots, (S_k, T_k)$ of pairs of disjoint subsets of a ground set X is called a *weakly separating system*[1] if for every $x, y \in X$ with $x \neq y$ there is an i such that either $x \in S_i$ and $y \in T_i$ or $x \in T_i$ and $y \in S_i$. The sequence is said to be a *strongly separating system* if for every $x, y \in X$ with $x \neq y$ there is an i such that $x \in S_i$ and $y \in T_i$. Equivalently, the sequence is weakly separating if the complete bipartite graphs with vertex classes S_i, T_i cover the edges of the complete graph with vertex set X; the sequence is strongly separating if the complete bipartite oriented graphs with vertex classes S_i, T_i and all edges oriented from S_i to T_i cover the complete digraph with vertex set X.

Strongly separating set systems are closely related to cross-intersecting systems. A sequence $(A_1, B_1), \ldots, (A_k, B_k)$ of pairs of sets is said to be *cross-intersecting* if $A_i \cap B_i = \emptyset$ for all i, and the intersection $A_i \cap B_j$ is nonempty for all $i \neq j$. (Many authors use the term *cross-intersecting* for the different setup of two set systems $\mathcal{A}$, $\mathcal{B}$ such that every $A \in \mathcal{A}$ meets every $B \in \mathcal{B}$.) Note that, in contrast to weakly or strongly separating systems, the ground set of a cross-intersecting system need not be specified. Strongly intersecting systems are dual to cross-intersecting systems, and it will be convenient for us to work in terms of the latter.

We begin the paper in Section 2 by giving an account of weakly separating systems, presenting a defect result of Katona and Szemerédi [6] and noting a couple of variants. In Section 3, we turn to strongly separating and cross-intersecting systems, proving analogous results to the weakly separating case. Finally, in Section 4, we prove a result on the number of nonzero entries in a strongly separating matrix. As an application of this result, we prove our bound on the size of oriented graphs with diameter at most 2.

[1] Weakly separating systems are usually referred to just as *separating systems*. We have adopted the weak/strong terminology to emphasize the difference.

2 Weakly separating systems

In this section we give some background on weakly separating systems. The investigation of weakly separating systems was started by Rényi [11], and continued by a number of other authors (see, for instance, [4, 5, 6, 14, 8, 9, 13, 10, 7, 2]). It is easily seen that $\lceil \log_2 n \rceil$ sets are necessary and sufficient for weak separation of a set of size n. If we are concerned only with the number of sets, then we may clearly asssume that $T_i = {S_i}^c$ for each pair (S_i, T_i) in a weakly separating system. However, Hansel proved the following stronger result that takes into account the sizes of the sets (a slight sharpening is given in [2]).

Lemma 1. *[4] Let $(S_1, T_1), \ldots, (S_k, T_k)$ be a weakly separating system on a ground set of size n. Then $\sum_{i=1}^{k}(|S_i| + |T_i|) \geq n \log_2 n$.*

Katona and Szemerédi [6] independently proved Lemma 1, and gave a defect version of the result as follows. Let G be a graph with vertex set V. A sequence $(S_1, T_1), \ldots, (S_k, T_k)$ of pairs of disjoint subsets of V is *weakly separating off G* if for every pair of distinct vertices $x, y \in V$ with $xy \notin E(G)$ there is i such that either $x \in S_i, y \in T_i$ or $x \in T_i, y \in S_i$.

Lemma 2. *[6] Let G be a graph with vertex set V. If $(S_1, T_1), \ldots, (S_k, T_k)$ is weakly separating off G then*

$$\sum_{i=1}^{k}(|S_i| + |T_i|) \geq \sum_{v \in V} \log_2 \left(\frac{n}{d(v) + 1} \right), \tag{1}$$

where $d(v)$ denotes the degree of v.

For completeness, we give a short proof of Lemma 2.

Proof. Define a weight function on $V = V(G)$ by setting $w(v) = 1/(d(v)+1)$ for each $v \in V$. If $X \subset V$ induces a complete subgraph, then each vertex in X has degree at least $|X| - 1$ and so $\sum_{v \in X} w(v) \leq 1$. Thus any subset of weight greater than 1 contains a pair of vertices that are not joined.

Now consider the random subset of V obtained by deleting, for each i, either S_i or T_i (independently, and with probability 1/2 each). Since the system is separating, the surviving vertices induce a clique and so have total weight at most 1. Thus the expected weight of vertices that survive is at most 1.

For each vertex v, let $f(v) = |\{i : v \in S_i \cup T_i\}|$. Then v survives with probability $2^{-f(v)}$, and calculating the expected weight of surviving vertices gives

$$\sum_{v \in V} \frac{1}{d(v)+1} 2^{-f(v)} \leq 1.$$

Defining $a(v)$ by $f(v) = \log_2 n - \log_2(d(v)+1) - a(v)$, this becomes

$$\frac{1}{n} \sum_{v \in V} 2^{a(v)} \leq 1.$$

By Jensen's inequality, $\sum_{v \in V} a(v) \leq 0$, and so

$$\sum_{v \in V} f(v) \geq n \log_2 n - \sum_{v \in V} \log_2(d(v)+1).$$

Since $\sum_{v \in V} f(v) = \sum_{i=1}^{k} |S_i \cup T_i|$, we are done. □

If G has maximal degree Δ then (1) trivially implies that $\sum(|S_i|+|T_i|) \geq n \log_2 n - n \log_2(\Delta+1)$. However, it is easy to obtain a bound in terms of average degree: if d is the average degree of G then by (1) and convexity we have

$$\begin{aligned} \sum_{i=1}^{k}(|S_i| + |T_i|) &\geq n \log_2 n - \sum_{i=1}^{n} \log_2(d_i+1) \\ &\geq n \log_2 n - n \log_2(d+1). \end{aligned} \tag{2}$$

We note that the proof of Lemma 2 also gives a bound in terms of clique number.

Lemma 3. *Let G be a graph with n vertices. If $(S_1, T_1), \ldots, (S_k, T_k)$ is weakly separating off G then*

$$\sum_{i=1}^{k}(|S_i| + |T_i|) \geq n \log_2 n - n \log_2(\mathrm{cl}(G)). \tag{3}$$

Proof. Follow the proof of Lemma 2, except with weight function $w(v) = 1/\mathrm{cl}(G)$. □

Both (2) and (3) are sharp when $n/(d+1)$ is a power of 2. Let $n = 2^k(d+1)$ and let G be the union of 2^k pairwise disjoint copies of K_{d+1}. We can separate a set of 2^k vertices with a system of pairs of sets with sizes summing to $k2^k$ (arrange the 2^k vertices in a cube, and take the k pairs of opposite faces). Replacing each vertex of the cube by the vertices from one copy of K_{d+1}, we obtain a system of pairs whose sizes sum to $(d+1)\cdot k2^k = n\log_2(n/(d+1))$.

3 Strongly separating systems

We now turn to considering strongly separating systems. The investigation of strongly separating systems was started by Dickson [3], who showed that every strongly separating system on a ground set of size n has at least $(1 + o(1))\log_2 n$ pairs of sets. The exact minimum was found by Spencer [12], who showed that the minimum number of pairs is $t(n)$, where $t = t(n)$ is the smallest positive integer such that $\binom{t}{\lfloor t/2\rfloor} \geq n$. Thus

$$t(n) = \log_2 n + \frac{1}{2}\log_2\log_2 n + O(1),$$

as compared to a minimum of $\lceil \log_2 n\rceil$ pairs in a weakly separating system.

An analogue for strongly separating systems of Hansel's result (Lemma 1) was proved in [2].

Theorem 4. *[2] Suppose that $(S_1, T_1), \ldots, (S_k, T_k)$ is a strongly separating system on a ground set of size n. Let $t^*(n)$ be the largest integer such that $\binom{t^*}{\lfloor t^*/2\rfloor} \leq n$. Then*

$$\sum_{i=1}^{k}(|S_i| + |T_i|) \geq nt^*(n) = n\log_2(n) + \frac{n}{2}\log_2\log_2 n + O(n), \qquad (4)$$

with equality if and only if $n = \binom{t^*}{\lfloor t^*/2\rfloor}$.

Our aim in this section is to prove a defect version (Corollary 7) of this result.

It is convenient to work in terms of dual systems. If $(S_1, T_1), \ldots, (S_k, T_k)$ is a sequence of pairs of sets on ground set V then the *dual set system* (A_v, B_v), $v \in V$, has ground set $[k] = \{1, \ldots, k\}$ and is defined by $A_v = \{j : v \in S_j\}$ and $B_v = \{j : v \in T_j\}$. The system $(S_1, T_1), \ldots, (S_k, T_k)$ is strongly separating if and only if the dual system (A_v, B_v), $v \in V$, is cross-intersecting. Furthermore, it is clear that

$$\sum_{i=1}^{k}(|S_i| + |T_i|) = \sum_{v\in V}(|A_v| + |B_v|), \tag{5}$$

so a bound on $\sum_{v\in V}(|A_v| + |B_v|)$ immediately provides a bound of the form (4)

Cross-intersecting systems were investigated by Bollobás [1], who proved the following result.

Lemma 5. *[1] Every cross-intersecting system (A_v, B_v), $v \in V$, satisfies*

$$\sum_{v\in V}\binom{|A_v| + |B_v|}{|A_v|}^{-1} \le 1. \tag{6}$$

Let G be a graph with vertex set V. A sequence $(S_1, T_1), \dots, (S_k, T_k)$ of pairs of disjoint subsets of V is *strongly separating off* G if for every pair of distinct vertices $x, y \in X$ with $xy \notin E(G)$ there is i such that $x \in S_i$, $y \in T_i$. A sequence (A_v, B_v), $v \in V$, of pairs of sets is *cross-intersecting off* G if $A_v \cap B_v$ is empty for all $v \in V$, and $A_v \cap B_w$ is nonempty whenever $v \neq w$ and $vw \notin E(G)$. Note that $V(G)$ appears as the ground set in a strongly separating system, but as the index set in a cross-interesting system: the edges of G correspond to pairs (of vertices or of sets respectively) where we do not insist on the strong separation or cross-intersection condition. Importantly, a system $(S_1, T_1), \dots, (S_k, T_k)$ with ground set $V(G)$ is strongly separating off G if and only if the dual system (with sets indexed by $V(G)$) is cross-intersecting off G. We can therefore prove results on systems that are cross-intersecting off G and then dualise (using (5)) to obtain results on systems that are strongly intersecting off G.

Let us prove a defect version of Lemma 5.

Theorem 6. *Let G be a graph with vertex set V and suppose that (A_v, B_v), $v \in V$, is cross-intersecting off G. Then*

$$\sum_{v\in V}\frac{1}{d(v)+1}\binom{|A_v| + |B_v|}{|A_v|}^{-1} \le 1 \tag{7}$$

and

$$\sum_{v\in V}\binom{|A_v| + |B_v|}{|A_v|}^{-1} \le \mathrm{cl}(G). \tag{8}$$

Proof. We prove (7) and (8) by induction on $|\bigcup_{v\in V} B_v|$.

If all B_v are empty then G must be complete and so both (7) and (8) are satisfied. Otherwise, for each $x \in \bigcup_{v\in V}(A_v \cup B_v)$, let $I(x) = \{v : x \in A_v\}$ and $J(x) = \{v : x \in B_v\}$. Let $V_x = V \setminus I(x)$ and let $G_x = G \setminus I(x)$. Consider the sequence (A'_v, B'_v), $v \in V_x$, obtained from (A_v, B_v), $v \in V$, as follows. If $x \in A_v$ then we delete the pair (A_v, B_v); if $x \in B_v$ we replace it with the pair $(A_v, B_v \setminus x)$; otherwise, we leave it unchanged. Note that the new system (indexed by V_x) is cross-intersecting off G_x.

Let $a_v = |A_v|$ and $b_v = |B_v|$ for each $v \in V$, and let λ_v, $v \in V$, be any sequence of positive reals. Consider the weight function assigning weight $\lambda_v\binom{|A_v|+|B_v|}{|A_v|}^{-1}$ to the pair (A_v, B_v). If (A_v, B_v) survives as (A'_v, B'_v), it will then have weight $\lambda_v\binom{|A'_v|+|B'_v|}{|A'_v|}^{-1}$. The increase Δ_x in weight between the new system and the old system is

$$\begin{aligned}\Delta_x &= \sum_{v\in V_x} \lambda_v \binom{|A'_v \cup B'_v|}{|A'_v|}^{-1} - \sum_{v\in V} \lambda_v \binom{|A_v \cup B_v|}{|A_v|}^{-1} \\ &= \sum_{v\in J(x)} \lambda_v \left[\binom{|A_v \cup B_v| - 1}{|A_v|}^{-1} - \binom{|A_v \cup B_v|}{|A_v|}^{-1}\right] \\ &\qquad - \sum_{v\in I(x)} \lambda_v \binom{|A_v \cup B_v|}{|A_v|}^{-1}.\end{aligned}$$

But

$$\binom{|A_v \cup B_v| - 1}{|A_v|}^{-1} - \binom{|A_v \cup B_v|}{|A_v|}^{-1} = \frac{a_v}{b_v}\binom{a_v + b_v}{a_v}^{-1}.$$

So

$$\begin{aligned}\sum_{x\in V} \Delta_x &= \sum_x \sum_{v\in J(x)} \lambda_v \frac{a_v}{b_v}\binom{a_v + b_v}{a_v}^{-1} - \sum_x \sum_{v\in I(x)} \lambda_v \binom{a_v + b_v}{a_v}^{-1} \\ &= \sum_v \lambda_v \sum_{x:v\in J(x)} \frac{a_v}{b_v}\binom{a_v + b_v}{a_v}^{-1} - \sum_v \lambda_v \sum_{x:v\in I(x)} \binom{a_v + b_v}{a_v}^{-1} \\ &= \sum_v \lambda_v a_v \binom{a_v + b_v}{a_v}^{-1} - \sum_v \lambda_v a_v \binom{a_v + b_v}{a_v}^{-1} \\ &= 0.\end{aligned}$$

Here we have used the fact that $|\{x : v \in I(x)\}| = |A_v|$ and $|\{x : v \in J(x)\}| = |B_v|$.

It follows that $\Delta_x \geq 0$ for some choice of x. But now to prove (8), take the weighting $\lambda_v = 1$ for all v. Since (8) holds for the reduced system on G_x and $\mathrm{cl}(G_x) \leq \mathrm{cl}(G)$, we are done. Similarly, for (7), taking $\lambda_v = 1/(d_G(v)+1)$, we obtain a system on G_x with weights λ_v and larger total weight. We know by induction that (7) holds on G_x with the larger weights $\lambda'_v = 1/(d_{G_x}(v)+1)$, and so the inequality holds on G_x with the weights λ_v. Therefore (7) also holds on G. □

Note that the bounds are sharp in some cases. If all the d_i are equal to d, and $n = (d+1)\binom{a+b}{a}$ then both bounds in the theorem are sharp for the graph consisting of n copies of K_{d+1}. Consider the system consisting of one pair $(A, [a+b] \setminus A)$ for each subset $A \subset [a+b] = \{1, \ldots, a+b\}$ with $|A| = a$. This has $\binom{a+b}{a}$ pairs, and satisfies (6) with equality. Taking $d+1$ copies of each of these pairs (i.e. taking each pair of sets $d+1$ times), we obtain a system satisfying (7) and (8) with equality.

We shall apply the following consequence of Theorem 6, which provides a defect result complementing Lemma 4.

Corollary 7. *Let G be a graph with vertex set V of size n and average degree $d \leq n/630 - 1$. Let t^* be the largest positive integer such that $\binom{t^*}{\lfloor t^*/2 \rfloor} \leq n/(d+1)$. Suppose that (A_v, B_v), $v \in V$, is cross-intersecting off G. Then*

$$\sum_{v \in V} (|A_v| + |B_v|) \geq nt^*. \tag{9}$$

This bound is sharp when $n = (d+1)\binom{t}{\lfloor t/2 \rfloor}$: let $a = \lceil t/2 \rceil$ and $b = \lfloor t/2 \rfloor$, and consider the example after Theorem 6. This has $n = (d+1)\binom{t}{\lfloor t/2 \rfloor}$ pairs (A_i, B_i), each of which satisfies $|A_i| + |B_i| = t$, so (9) holds with equality.

We will need the following lemma.

Lemma 8. *If a and b are integers with $5 \leq a \leq b-2$ then*

$$\binom{a}{\lfloor a/2 \rfloor}^{-1/2} + \binom{b}{\lfloor b/2 \rfloor}^{-1/2} \geq \binom{a+1}{\lfloor (a+1)/2 \rfloor}^{-1/2} + \binom{b-1}{\lfloor (b-1)/2 \rfloor}^{-1/2}.$$

Also, if $0 \leq a \leq b-5$ then

$$\binom{a}{\lfloor a/2 \rfloor}^{-1/2} + \binom{b}{\lfloor b/2 \rfloor}^{-1/2} \geq \binom{a+2}{\lfloor a/2+1 \rfloor}^{-1/2} + \binom{b-2}{\lfloor b/2-1 \rfloor}^{-1/2}.$$

Proof. Define $f(a) = \binom{a}{\lfloor a/2 \rfloor}$. Note that $f(a+1) = 2f(a)$ if a is odd and $f(a+1) = 2(a+1)/(a+2)$ if a is even. Thus $(2f(a))^{-1/2} \le f(a+1)^{-1/2} \le (2(a+1)f(a)/(a+2))^{-1/2}$. It follows from a short calculation that $f(a)^{-1/2} - f(a+1)^{-1/2}$ is monotone decreasing for $a \ge 5$. The first part of the result follows immediately.

The second part of the result is implied by the first unless $a \le 5$; the remaining cases are easily checked. □

Proof of Corollary 7. We know from Theorem 6 that, given $(d(v))_{v\in V}$, the minimal value of $\sum_{v\in V}(|A_v| + |B_v|)$ is at least

$$\min\Big\{\sum_{v\in V}(a_v + b_v) : \sum_{v\in V}\frac{1}{d(v)+1}\binom{a_v+b_v}{b_v}^{-1} \le 1\Big\},$$

which is clearly at least

$$\min\Big\{\sum_{v\in V} c_v : \sum_{v\in V}\frac{1}{d(v)+1}\binom{c_v}{\lfloor c_v/2\rfloor}^{-1} \le 1\Big\},$$

where the minimum is taken over nonnegative integers $c_v, v \in V$. Now this is at least

$$\min\Big\{\sum_{v\in V} c_v : \sum_{v\in V}\frac{1}{e_v}\binom{c_v}{\lfloor c_v/2\rfloor}^{-1} \le 1, \frac{1}{n}\sum_{v\in V} e(v) = d+1\Big\}, \tag{10}$$

where d is the average degree, the e_v range over nonnegative reals, and the c_v range over nonnegative integers.

For fixed nonnegative reals C_v, and nonnegative reals e_v summing to D, Cauchy-Schwarz gives $\sum_{v\in V} e_v \sum_{v\in V}(C_v/e_v) \ge (\sum_{v\in V}\sqrt{C_v})^2$, and so $\sum_{v\in V} C_v/e_v \ge \big(\sum_{v\in V}\sqrt{C_v}\big)^2/D$. Setting $C_v = \binom{c_v}{\lfloor c_v/2\rfloor}^{-1}$ and $D = n(d+1)$, we see that the first condition in (10) implies

$$1 \ge \sum_{v\in V}\frac{C_v}{e_v} \ge \frac{(\sum_{v\in V}\binom{c_v}{\lfloor c_v/2\rfloor}^{-1/2})^2}{n(d+1)}.$$

Thus

$$\sum_{v\in V}\binom{c_v}{\lfloor c_v/2\rfloor}^{-1/2} \le \sqrt{n(d+1)}. \tag{11}$$

Now if all c_v are at most 9 then $\sum_v \binom{c_v}{\lfloor c_v/2 \rfloor}^{-1/2} \geq n\binom{9}{4}^{-1/2} \geq n/\sqrt{630}$ and so by (11) we have $d+1 > n/630$, which contradicts our assumption on d. Otherwise, by repeated applications of Lemma 8, we may assume that the c_v are clustered on at most two values. (Note that this does not change $\sum_v c_v$.) Then let $c = \min c_v$: we have $\sum_{v \in V} \binom{c_v}{\lfloor c_v/2 \rfloor}^{-1/2} > n\binom{c+1}{\lfloor (c+1)/2 \rfloor}^{-1/2}$ and so $\binom{c+1}{\lfloor (c+1)/2 \rfloor} > n/(d+1)$. The corollary follows immediately. □

A straightforward calculation shows that Corollary 7 implies the following bound.

Corollary 9. *Let G be a graph with vertex set V and average degree $d \leq n/630-1$, where $n = |V|$. Suppose that (A_v, B_v), $v \in V$, is cross-intersecting off G. Then*

$$\sum_{v \in V} |A_v| + |B_v| \geq n \log_2 \left(\frac{n}{d+1} \right) + \frac{1}{2} n \log_2 \log_2 \left(\frac{n}{d+1} \right) + O(n). \qquad (12)$$

Since strongly separating systems are dual to cross-intersecting systems, it follows that we obtain the same bound for systems that are strongly separating off G.

4 Separating matrices and oriented graphs with diameter 2

In this section, we consider strongly separating matrices, and oriented graphs of diameter 2. We shall prove a bound on the number of non-zero entries in an n by n strongly separating matrix, and then use this to show that every oriented graph with diameter 2 has at least $(1+o(1))n \log_2 n$ edges.

Given a matrix $M = (m_{ij})$ with n rows and columns, and entries in $\{0, 1, -1\}$, we can define a system $(A_1, B_1), \ldots, (A_n, B_n)$ of pairs of sets, where $A_i = \{j : m_{ij} = 1\}$ and $B_i = \{j : m_{ij} = -1\}$. We shall refer to this as the *row system* of M. The *column system* of M is defined similarly, and is the same as the row system of M^T. We say that M is *weakly separating* if its row and column systems are both weakly separating. M is *strongly separating* if its row and column systems are both strongly separating

How many nonzero entries must an n by n strongly separating matrix contain?

Theorem 10. *Let M be an n by n strongly separating matrix. Then M has at least $2n\log_2 n - 3n\log_2\log_2 n + O(n)$ nonzero entries.*

Proof. Suppose that M is an n by n strongly separating matrix with fewer than $2n\log_2 n$ nonzero entries. Let X be the set of rows and Y be the set of columns. For $X' \subset X$ and $Y' \subset Y$ we write $M[X', Y']$ for the submatrix of M induced by rows X' and columns Y'. Let $\alpha = 2(\log_2 n)^2$. Let X^+ be the set of rows with at least α nonzero entries and let Y^+ be the set of columns with at least α nonzero entries. Let $X^- = X \setminus X^+$ and $Y^- = Y \setminus Y^+$, and define $n_x = |X^-|$ and $n_y = |Y^-|$. We may assume that n is large, as smaller values are covered by the $O(n)$ term.

Since M has fewer than $2n\log_2 n$ entries, $|X^+| \leq 2n(\log_2 n)/\alpha$ and so $n_x \geq n - 2n(\log_2 n)/\alpha = n(1 - 1/\log_2 n)$. Now suppose there are m nonzero entries in $M[X^-, Y^-]$. Define a graph G_X on X^- by joining x to y if there is some $z \in Y^-$ such that $a_{xz} = 1$ and $a_{yz} = -1$ or $a_{xz} = -1$ and $a_{yz} = 1$ (so x and y are weakly separated by the columns of $M[X^-, Y^-]$). A column in $M[X^-, Y^-]$ with k entries weakly separates at most $k^2/4 \leq \alpha k/4$ pairs of rows. Since $M[X^-, Y^-]$ has m nonzero entries it therefore separates at most $\alpha m/4$ pairs of rows, and so the average degree of G_X is at most $d_x = \alpha m/2n_x$.

Now consider $M[X^-, Y^+]$. The rows of Y^+ strongly separate the columns in X^- off G_x, and so by Corollary 9 the number of nonzero entries in the matrix $M[X^-, Y^+]$ is at least

$$n_x \log_2(n_x/(d_x+1)) + \frac{1}{2} n_x \log_2\log_2(n_x/(d_x+1)) + O(n_x).$$

Since $d_x = \alpha m/2n_x \leq 2(\log_2 n)^2 \cdot (2n\log_2 n)/n = (\log_2 n)^3$ for large enough n, we have $\log_2\log_2(n_x/(d_x+1)) = \log_2\log_2 n_x + O(1)$. Thus the number of nonzero entries in $M[X^-, Y^+]$ is at least

$$n_x\big(\log_2 n_x - \log_2(d_x+1) + \frac{1}{2}\log_2\log_2 n_x + O(1)\big),$$

which equals

$$n\big(\log_2 n - \log_2(d_x+1) + \frac{1}{2}\log_2\log_2 n + O(1)\big).$$

So the number of nonzero entries in $M[X^-, Y^+]$ and $M[X^+, Y^-]$ plus the m entries in $M[X^-, Y^-]$ totals at least

$$n\big(2\log_2 n - 2\log_2(d_x+1) + \log_2\log_2 n + m/n + O(1)\big).$$

If $d_x \leq 1$ then we are done immediately. Otherwise, note that

$$\begin{aligned} m/n - 2\log_2(d_x + 1) &= m/n - 2\log_2(\alpha m/n) + O(1) \\ &\geq -2\log_2 \alpha + O(1) \\ &\geq -4\log_2 \log_2 n + O(1). \end{aligned}$$

The result follows immediately. □

We can obtain a similar bound for weakly separating matrices by employing Lemma 3 instead of Corollary 9, at the cost of a slightly worse constant in the $n \log_2 \log_2 n$ term. In both the weakly and strongly separating cases, it would be interesting to know the best possible constant in this term.

We are now ready to prove our result on oriented graphs of diameter 2. However, let us begin by remarking that the analogous problem for *digraphs*, where we allow both edges xy and yx, is easily settled.

Lemma 11. *Let G be a digraph of order $n \geq 5$ and diameter at most 2. Then $e(G) \geq 2n-2$, with equality only if G is the digraph obtained by taking all $2n-2$ edges incident with a single vertex.*

We remark that, for $n = 4$, a copy $uvwx$ of C_4 with cyclic orientation, and additional edges uw and wu, gives another extremal graph. For $n = 3$, the triangle with cyclic orientation is extremal.

Proof. Suppose that G is a digraph of order n with diameter 2 and $2n-2$ edges. We shall show that G is the extremal digraph (which is edge-minimal). Note first that the diameter condition implies that every vertex has outdegree at least 1, so by summing outdegrees we see that some vertex v has exactly one outneighbour w. Since every vertex can be reached from v by a directed path of length at most 2, the outneighbourhood of w contains every vertex in $V(G) \setminus \{v, w\}$. There are now two cases.

If $\Gamma^+(w) = V(G) \setminus w$ then, since every vertex has at least one outedge, we have $e(G) = \sum_{v \in V(G)} d^+(v) = d^+(w) + \sum_{x \in V(G) \setminus w} d^+(x) \geq (n-1) + (n-1)$. Thus every vertex other than w must have outdegree exactly 1. If every vertex other than w has outneighbour w, we are done. Otherwise, some $x \neq w$ has outneighbour $y \neq w$. But then we can only reach two vertices (y and its outneighbour) from x with paths of length at most 2, which contradicts our assumption on diameter if $n \geq 5$.

In the remaining case, $\Gamma^+(w) = V(G) \setminus \{v, w\}$. Summing degrees, we see that there is some $x \neq w$ with outdegree 2, and every other vertex $y \neq x, w$

has outdegree 1. If some $y \neq v, x, w$ has outneighbour w, then there is no path of length at most 2 from y to v. But then the outneighbour of y has outdegree at most 2, so we can reach at most 3 vertices from y with paths of length at most 2. Thus this case cannot occur for $n \geq 5$. □

There is a striking difference between digraphs of diameter 2 and oriented graphs of diameter 2. We now concentrate on the oriented case.

Theorem 12. *Let G be an oriented graph of diameter at most 2. Then $e(G) \geq n \log_2 n - \frac{3}{2} n \log_2 \log_2 n - O(n)$.*

Proof. Let G be an oriented graph of order n with diameter at most 2. Let $V(G) = \{v_1, \ldots, v_n\}$, and let $A = (a_{ij})$ be the adjacency matrix of G (so $a_{ij} = 1$ if $v_i v_j \in E(G)$, $a_{ij} = -1$ if $v_j v_i \in E(G)$ and $a_{ij} = 0$ otherwise). Define $M = A + I$, where I is the identity matrix.

We first show that M is a strongly separating matrix. Consider first the rows of M. It is enough to show that if i and j are distinct then there is k with $a_{ik} = 1$ and $a_{jk} = -1$. If $ij \in E(G)$ then we can take $k = i$, as we have $a_{ik} = a_{ii} = 1$ and $a_{jk} = a_{ji} = -a_{ij} = -1$. Otherwise there is k such that $v_i v_k \in E(G)$ and $v_k v_j \in E(G)$. Then $a_{ik} = 1$ and $a_{kj} = -a_{jk} = -1$. Now consider $M^T = A^T + I$. Since A^T is the adjacency matrix for the graph G' obtained from G by reversing the orientation of every edge, and G' also has diameter at most 2, it follows that the rows of M^T and therefore the columns of M form a strongly separating system.

By Theorem 10, M has at least $2n \log_2 n - 3n \log_2 \log_2 n + O(n)$ nonzero entries. Since A has two nonzero entries for each edge of G, and has n fewer nonzero entries than M, the theorem follows. □

The leading term in Theorem 12 is correct, as shown by the following example.

Example 13. Let k be an even positive integer and consider the complete bipartite graph with vertex classes V_1 and V_2, where $|V_1| = \binom{2k+1}{k}$ and $|V_2| = 2k+1$. We orient the graph such that every vertex in V_1 has outdegree k (and hence indegree $k+1$), and every vertex in V_1 has a distinct outneighbourhood. Every k-set of vertices from V_2 is the outneighbourhood of exactly one vertex in V_1 and the orientation is unique up to isomorphism. It is easy to check that there is a directed path of length at most 2 between any two vertices in V_1 and between any two vertices in V_2. Now add an edge between every pair of vertices in V_2, and orient them so that the oriented graph induced by

V_2 is k-out-regular. Pick $x \in V_1$ and $y \in V_2$. If there is no directed path of length at most 2 from x to y then the outneighbours of x must be precisely the inneighbours of y in V_1. Thus, for each y in V_2, there is at most one such x. (There must be a directed path of length at most two from y to x, as x has $k+1$ inneighbours in V_2, and these cannot be disjoint from y and its k outneighbours in V_2). Thus by adding an additional oriented edge inside V_1 for each $y \in V_2$ (there is plenty of room to do this) we obtain an oriented graph with diameter 2. We therefore have an oriented graph with $n = 2k+1+\binom{2k+1}{k}$ vertices and $(2k+1)\binom{2k+1}{k}+2k+1 = n(\log_2 n+(1/2)\log_2\log_2 n+O(1))$ edges, which is within $O(n\log_2\log_2 n)$ of the lower bound. The adjacency matrix of this digraph also shows that our bounds on the number of nonzero entries in a weakly or strongly separating matrix are also within $O(n\log_2\log_2 n)$ of optimal.

It would be interesting to have a lower bound that is sharp to within $o(n\log_2\log_2 n)$.

References

[1] B. Bollobás, On generalized graphs, *Acta Math. Acad. Sci. Hungar.* **16** (1965), 447–452

[2] B. Bollobás and A. Scott, On separating systems, *European Journal of Combinatorics*, to appear

[3] T.J. Dickson, On a problem concerning separating systems of a finite set, *J. Combinatorial Theory* **7** (1969), 191–196

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

[5] G. Katona, On separating systems of a finite set. *J. Combinatorial Theory* **1** (1966), 174–194

[6] G. Katona and E. Szemerédi, On a problem of graph theory, *Studia Sci. Math. Hungar.* **2** (1967), 23–28

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

[8] A. Nilli, Perfect hashing and probability, *Combin. Probab. Comput.* **3** (1994), 407–409

[9] J. Radhakrishnan, Entropy and Counting, *manuscript*

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

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

[12] J. Spencer, Minimal completely separating systems, *J. Combinatorial Theory* **8** (1970), 446–447

[13] Zs. Tuza, Applications of the set-pair method in extremal hypergraph theory, *in* Extremal problems for finite sets (Visegrád, 1991), 479–514, Bolyai Soc. Math. Stud. **3**, János Bolyai Math. Soc., Budapest, 1994.

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