The 2-Intersection Number of Paths and Bounded-Degree Trees

Michael S. Jacobson
André E. Kézdy
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF LOUISVILLE
LOUISVILLE, KENTUCKY

Douglas B. West
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF ILLINOIS
URBANA, ILLINOIS

ABSTRACT

We represent a graph by assigning each vertex a finite set such that vertices are adjacent if and only if the corresponding sets have at least two common elements. The *2-intersection number* $\theta_2(G)$ of a graph G is the minimum size of the union of sets in such a representation. We prove that the maximum order of a path that can be represented in this way using t elements is between $(t^2 - 19t + 4)/4$ and $(t^2 - t + 6)/4$, making $\theta_2(P_n)$ asymptotic to $2\sqrt{n}$. We also show the existence of a constant c depending on ϵ such that, for any tree T with maximum degree at most d, $\theta_2(T) \le c(\sqrt{n})^{1+\epsilon}$. When the maximum degree is not bounded, there is an n-vertex tree T with $\theta_2(T) > .945n^{2/3}$. © 1995 John Wiley & Sons, Inc.

1. INTRODUCTION

We study representations of graphs using intersections of finite sets. We use G to denote a finite simple graph with vertex set $V = V(G)$ and edge set $E = E(G)$. Given a positive integer p, a p-*intersection representation* of G is a collection of finite sets $\{S_v : v \in V\}$ such that $uv \in E$ if and only if $|S_u \cap S_v| \ge p$. The p-*intersection number* of G is the smallest integer t such that G has a p-intersection representation using subsets of a t-element set. Following the notation in [2], we denote the p-intersection number of G by $\theta_p(G)$. The concept of p-intersection representations for $p = 1$ first

Journal of Graph Theory, Vol. 19, No. 4, 461–469 (1995)
© 1995 John Wiley & Sons, Inc. CCC 0364-9024/95/040461-09

appeared in [3]. The general case $p > 1$ was introduced in [8] and studied in [2,6,5,9], the latter two of which explore its connection with competition graphs.

In this paper we investigate the 2-intersection number of paths and of trees with bounded degree. Let P_n denote the path with n vertices, and let $[t]$ denote $\{0, 1, \ldots, t-1\}$ (a k-*set* is a set with k elements). By adapting methods used to construct cyclic Steiner triple systems, we provide a 2-intersection representation consisting of 3-sets from $[t]$ for the path with $n = \lfloor (t-19)(t)/4 \rfloor + 1$ vertices. A simple counting argument shows that this is asymptotically best possible. This proves $\theta_2(P_n) \sim 2\sqrt{n}$.

Let $\mathcal{T}_d(n)$ be the collection of trees with at most n vertices and maximum vertex degree at most d. Given a positive integer d, define $t_d(n) = \max\{\theta_2(T): T \in \mathcal{T}_d(n)\}$. We prove that for any $\epsilon > 0$, there exists a constant $c = c(\epsilon, d)$ such that $t_d(n) < c(\sqrt{n})^{1+\epsilon}$.

Finally, it was thought for some time that $\theta_2(T) \in O(n^{1/2})$ for all n-vertex trees, but this is not true. We provide a sequence of trees for which the 2-intersection number grows as $n^{2/3}$, where n is the number of vertices. This demonstrates the special role of the trees of bounded degree.

2. PATHS AND CYCLES

In this section, we compute bounds on the 2-intersection number of paths and cycles.

Lemma 1. A path having a 2-intersection representation using subsets of $[t]$ has at most $(\binom{t}{2} + 3)/2$ vertices.

Proof. Let $m = \max\{n: \theta_2(P_n) \le t\}$, and consider such a representation of P_m. For each edge e of P_m, we can choose a 2-set assigned to both end points of e, and the $m-1$ sets thus selected must be distinct. Since each of the $m-2$ interior vertices belongs to two such edges, the sets representing these vertices have size at least 3. Hence the set for each interior vertex contains at least three 2-sets. Since these vertices have only two neighbors, one of those 2-sets is not contained in the set representing any other vertex. We have associated distinct 2-sets with each edge and each interior vertex; this requires $\binom{t}{2} \ge 2m - 3$, from which the bound follows. ∎

We conjecture that the bound of Lemma 1 is best possible in all cases except $t = 6$. This has been verified by computer for $t \le 16$.

We next construct a 2-intersection representation of a path. The main tool is difference systems, a complete development of which can be found in Anderson's book [1]. This construction actually produces a 2-intersection representation for a cycle. By splitting a vertex of the cycle we generate a 2-intersection representation for a path of the same length. In particular, if three consecutive vertices u, v, w on the cycle are represented by $\{a,b,c\},\{b,c,d\},\{c,d,e\}$, respectively, then we delete v and add ends of the path represented by $\{b,c\}$ and $\{c,d\}$.

Theorem 1. If $t = 16k - 1$, for some integer $k > 2$, then there is a 2-intersection representation of a cycle with $t(t-19)/4$ vertices using 3-sets from $[t]$.

Proof. Fix $k > 2$, and let $t = 16k - 1$; we construct such a 2-intersection representation. Consider the following two systems of ordered triples from $\{1,2,\dots,8k-1\}$:

System 1.

$$(5k+r, 2k-2r+1, 7k-r+1) \qquad 1 \le r \le k-1$$
$$(4k+r, 4k-2r, 8k-r) \qquad 1 \le r \le k-2$$

System 2.

$$(5k+r, 2k-2r, 7k-r) \qquad 1 \le r \le k-2$$
$$(6k-1, 2k, 8k-1)$$
$$(4k+r, 4k-2r-1, 8k-r-1) \qquad 1 \le r \le k-2$$
$$(1, 7k, 7k+1)$$

Observe that all these ordered triples have the form $(a,b,a+b)$ and that any pair of triples from the same system are disjoint, except that the last two triples are $(5k-2, 2k+3, 7k+1)$ and $(1,7k,7k+1)$ in System 2 share one element. Define the set $D = \{\pm 1, \pm 2, \dots, \pm 4k\}$. Every triple contains a (positive) element of D, and no element of D appears in both system 1 and system 2.

Let $S_1', S_2', \dots, S_{2k-3}'$ be the triples from the first system in the order listed. Similarly, let $T_1', T_2', \dots, T_{2k-2}'$ be the triples from the second system in the order listed. For $S_i' = (a,b,a+b)$, define $S_i = \{0,a,a+b\}$. Similarly, for $T_i' = (a,b,a+b)$, define $T_i = \{0,a,a+b\}$. The differences arising from the 3-set $\{0,a,a+b\}$ are $\pm a, \pm b, \pm(a+b)$. In particular, the S_i's determine $12k - 18$ distinct differences from $\{-8k+1,\dots,-1,+1,\dots, 8k-1\}$; we view these as congruence classes mod t. Similarly, the T_i's determine $12k - 11$ distinct differences from $\{-8k-1,\dots, -1,+1,\dots,8k-1\}$ (the difference $7k$ appears in T_{2k-3} and T_{2k-2}).

Let $C(0)$ be the column vector of 3-sets that is the transpose of

$$(S_1, T_1, S_2, T_2, \dots, S_{2k-3}, T_{2k-3}, T_{2k-2}).$$

Here

$$S_r = \{0, 5k + r, 7k - r + 1\} \qquad 1 \le r \le k - 1$$
$$S_r = \{0, 3k + r + 1, 9k - r - 1\} \qquad k \le r \le 2k - 3$$
$$T_r = \{0, 5k + r, 7k - r\} \qquad 1 \le r \le k - 2$$
$$T_{k-1} = \{0, 6k - 1, 8k - 1\}$$
$$T_r = \{0, 3k + r + 1, 9k - r - 2\} \qquad k \le r \le 2k - 3$$
$$T_{2k-2} = \{0, 1, 7k + 1\}$$

For $j = 1, \ldots, t - 1$, define $C(j)$ as the vector of 3-sets obtained from $C(0)$ by adding j to every element of every 3-set (modulo t). For convenience, we use the notation $S_i + j$ to denote the translation of the triple S_i by j units. Similarly, $T_i + j$ is the translation of the triple T_i. See Table 1 for an example.

We claim that the successive 3-sets in $C(0), \ldots, C(t - 1)$ form a 2-intersection representation of a cycle with $t(4k - 5)$ vertices. First, we verify that the 2-intersection graph contains the desired edges. By construction, successive 3-sets of $C(0)$ intersect twice, and hence this holds for all $C(j)$. Furthermore, $T_{2k-2} + j$ and $S_1 + (j + 1)$ share the elements $j + 1$ and $7k + j + 1$. Hence it suffices to show that no nonconsecutive pair of 3-sets in the list have two common elements. For a pair of 3-sets arising from the same system of triples, this is quite easy. Because the differences generated by the 3-sets in $\{S_i\}_{i=1}^{2k-3}$ or the 3-sets in $\{T_i\}_{i=1}^{2k-2}$ are all distinct (except for the common difference in T_{2k-3} and T_{2k-2}), there is no way to translate two of the 3-sets in one triple system to have an undesired common 2-set.

Suppose two 3-sets $X = \{x, y, z\}$ and $Y = \{x, y, w\}$ arising from distinct triple systems intersect twice, with X arising from system 1. One of the differences $(x - y), (y - z), (x - z)$ belongs to the set $D = \{\pm 1, \pm 2, \ldots, \pm 4k\}$, since the differences in our 3-sets are the elements of some triple in the systems. Furthermore, $x - y \notin D$, since 3-sets from different systems do not share a common difference in D. Therefore, either $x - z$ or $y - z$ is in D. Now let us consider the two neighbors of X in the

TABLE 1. Example of Theorem 1: $k = 3$; $t = 47$.

Label	Column C(0)	Column C(1)	...	Column C(46)
S_1	{0, 16, 21}	{1, 17, 22}	...	{46, 15, 20}
T_1	{0, 16, 20}	{1, 17, 21}	...	{46, 15, 19}
S_2	{0, 17, 20}	{1, 18, 21}	...	{46, 16, 19}
T_2	{0, 17, 23}	{1, 18, 24}	...	{46, 16, 22}
S_3	{0, 13, 23}	{1, 14, 24}	...	{46, 12, 22}
T_3	{0, 13, 22}	{1, 14, 23}	...	{46, 12, 21}
T_4	{0, 1, 22}	{1, 2, 23}	...	{46, 0, 21}

cycle. If either of these two neighbors contain both x and y, then it must be the 3-set Y itself, since it would be a 3-set from the same system as Y that shares two elements with Y. The only triples arising from System 2 that share two elements arise from T_{2k-3} and T_{2k-2}, but the difference between the two common elements of such a pair does not occur in System 1, and hence x, y cannot be such a pair.

Hence we may assume that the two neighboring 3-sets of X in the cycle do not contain both x and y. Therefore they both contain the element z: suppose the neighbors are $\{z, x, p\}$ and $\{q, y, z\}$. Let d be either $x - z$ or $y - z$, whichever is in D. Now d is a difference in D appearing in both systems, and this contradicts our earlier observation. Therefore the only 3-sets intersecting X twice are the desired neighbors of X. ∎

3. BOUNDED-DEGREE TREES

In this section we investigate the 2-intersection number of trees with bounded maximum degree. We wish to estimate the function $t_d(n)$ defined in the introduction. We begin with lemmas about decomposing trees of bounded degree into pieces of comparable size. These techniques have been used before for trees of bounded degree; we include the proofs to make the presentation self-contained. If G is a graph, we let $|G| = |V(G)|$.

Lemma 2. If T is a tree with maximum degree d, then there is an edge $e \in E(T)$ such that, if T_1, T_2 are the components of $T - e$, then $|T_i| \leq (d - 1)|T_j| + 1$ for $i \neq j$.

Proof. Choose an edge $e \in E(T)$ such that the components T_1, T_2 of $T - e$ differ as little as possible in size; we may assume $|T_1| \geq |T_2|$. Suppose $|T_1| > (d - 1)|T_2| + 1$. Let x be the end point of e in T_1. Consider the trees $Q_1, Q_2, \ldots, Q_p$ of $T_1 - x$. Because $d_T(x) \leq d$, we have $p < d$, and hence the pigeonhole principle guarantees that some Q_i is bigger than T_2. Replacing e by the edge from x to that Q_i contradicts the choice of e. ∎

The bound in this lemma is achieved for any $d > 1$ by any tree consisting of d subtrees of the same size attached to a single central vertex of degree d.

Lemma 3. Given a positive integer k and any tree T with maximum degree d, there exists a vertex partition of T into k subtrees $T_1, \ldots, T_k$ such that $\max_{1 \leq i \leq k}\{|T_i|\} \leq 1 + d \min_{1 \leq j \leq k}\{|T_j|\}$.

Proof. Induction on k. For $k = 1$, the claim is trivial. For $k > 1$, let $T_1, T_2, \ldots, T_{k-1}$ be a partition of T into $k - 1$ trees as guaranteed by the induction hypothesis, with $|T_1| = \max_{1 \leq i \leq k-1}\{|T_i|\}$. Choose e from among the edges of T_1 that satisfy the conclusions of the previous lemma. Let Q_1, Q_2 be the components of $T_1 - e$, with $|Q_1| \geq |Q_2|$. In order to show that $Q_1, Q_2, T_2, T_3, \ldots, T_{k-1}$ is a suitable partition into k trees, it suffices to prove $|T_i| \le d|Q_2| + 1$. This follows from

$$|T_i| \le |T_1| = |Q_1| + |Q_2| \le |Q_2| + (d-1)|Q_2| + 1 = d|Q_2| + 1. \quad \blacksquare$$

Lemma 4. Given a positive integer k and a tree T with maximum degree d, there exists a vertex partition of T into k subtrees each with at most $(d|T| + k - 1)/(k + d - 1)$ vertices.

Proof. Use the partition into k subtrees guaranteed by the previous lemma, and let r be the order of the largest subtree. The smallest of the subtrees has order at most $(|T| - r)/(k-1)$, by the pigeonhole principle. By Lemma 3, $r \le 1 + d(|T| - r)/(k-1)$, from which the result follows. ∎

It is well known that for all $x > 1$ there is a prime p between x and $2x$. We seek primes in a particular range because we need projective planes of order in that range, and there are projective planes of prime order. A *projective plane* of order p is a set of $p^2 + p + 1$ points and $p^2 + p + 1$ subsets of that set called lines, such that each line consists of $p + 1$ points, each point belongs to $p + 1$ lines, each pair of lines has one common point, and each pair of points belongs to one common line. Interested readers can find a discussion of projective planes in Anderson's book [1]; projective planes can be obtained from finite fields of the same order.

We use these lemmas and projective planes to inductively construct a 2-representation for a tree of bounded maximum degree d. Our approach is as follows, where "roughly $f(n)$" means "$f(n)$ times a constant determined by d and ϵ". We use Lemma 4 to partition T into roughly $\sqrt{n}$ subtrees, each of size bounded by roughly $\sqrt{n}$. Given a projective plane with roughly $(\sqrt{n})^{1+\epsilon}$ lines and points, we use distinct lines to build a 2-representation of each subtree, inductively. Adding 2 times roughly $\sqrt{n}$ elements to represent the edges between the subtrees completes the construction. The details of the proof concerns the meaning of each "roughly." We have not attempted to minimize the choice of the constant c. For example, Huxley [4] proved that, for any fixed $\epsilon > 0$, there exists a constant β such that for all $n > 2$ there is a prime p with $n \le p \le n + \beta n^{\alpha}$, where $\alpha = (7/12) + \epsilon$; use of this result instead of the bound $n \le p \le 2n$ would improve our constant.

Theorem 2. For all $d \ge 2$ and $0 < \epsilon < 1$, there is a constant $c = c(d, \epsilon)$ such that

$$t_d(n) \le c(\sqrt{n})^{1+\epsilon}.$$

Proof. Fix $1 > \epsilon > 0$ and $d > 1$. Pick a value c large enough to satisfy the following conditions:

$$(c/2)^{\epsilon/(1-\epsilon)} > (32d)^{2+\epsilon}c^{\epsilon} \quad (1)$$

$$(c/2)^{2/(1-\epsilon)} \geq (32c)^2. \quad (2)$$

In each inequality, the exponent on c is larger on the left side than on the right, so the inequalities are true for all sufficiently large c.

Given this choice of c, the proof is by induction on $n = |T|$. For the basis of the induction, suppose $n \leq (c/2)^{2/(1-\epsilon)}$. For such n, the conclusion holds because we can use disjoint 2-sets representing each edge to obtain $t_d(n) \leq 2(n-1) \leq 2n \leq c(\sqrt{n})^{1+\epsilon}$, where the last inequality is equivalent to $n \leq (c/2)^{2/(1-\epsilon)}$.

Hence we may assume $n > (c/2)^{2/(1-\epsilon)}$. Let T be a tree with maximum degree at most d and with at most n vertices; we follow the approach outlined above. By Lemma 4, there is a vertex partition of T into x trees $T_1, T_2, \ldots, T_x$ such that $\max\{|T_i|\} \leq (dn + x - 1)/(x + d - 1) < (d/x)n + 1$; we apply this statement with $x = \lceil 32cd\sqrt{n}\rceil$. Letting $y = \max\{|T_i|\}$, we have $y < \sqrt{n}/(32c) + 1$.

We want the lines of our projective plane to have enough points to represent trees of order y. Because $y < n$, we know by the inductive hypothesis that $t_d(y) \leq c(\sqrt{y})^{1+\epsilon}$; set $z = c(\sqrt{y})^{1+\epsilon}$. There is a prime number p such that $z \leq p \leq 2z$; we use a projective plane of order p. Its lines have size $p + 1$; by the choice of z they are big enough to represent the subtrees T_i.

We also need to verify that we have enough lines in the plane for all the subtrees. For this it suffices to prove $c^2y^{1+\epsilon} \geq x$, because then we have $p^2 + p + 1 > p^2 \geq z^2 = c^2y^{1+\epsilon} \geq x$. To obtain the desired inequality, note that from $n > (c/2)^{2/(1-\epsilon)}$, condition (1), and the definition of x, we have $n^{\epsilon/2} \geq (32d)^{2+\epsilon}c^{\epsilon} \geq (x/c\sqrt{n})^{2+\epsilon}c^{\epsilon}$. The outer inequality simplifies to $c^2(n/x)^{1+\epsilon} \geq x$. Substituting $y \geq n/x$ yields the desired inequality.

We have now constructed a 2-intersection representation of T using the $p^2 + p + 1$ points of the projective plane plus $2(x-1)$ additional points to represent the $x - 1$ edges connecting the trees $T_1, \ldots, T_x$. It remains only to show that this is at most $c(\sqrt{n})^{1+\epsilon}$. Since $x \leq p^2$ and $p \leq 2z$, we have $p^2 + p + 1 + 2(x-1) \leq 4p^2 \leq 16z^2 = 16c^2y^{1+\epsilon}$. The desired inequality $16c^2y^{1+\epsilon} \leq c(\sqrt{n})^{1+\epsilon}$ is equivalent to $y^{1+\epsilon} \leq (\sqrt{n})^{1+\epsilon}/(16c)$, which is implied by $y \leq \sqrt{n}/(16c)$ and $(16c)^{1+\epsilon} \geq 1$. The latter inequality follows from (2). Since we earlier observed that $y \leq \sqrt{n}/(32c) + 1$, it suffices to show $\sqrt{n}/(32c) + 1 \leq \sqrt{n}/(16c)$. This is equivalent to $n \geq (32c)^2$, which follows from $n > (c/2)^{2/(1-\epsilon)}$ and condition (2). ∎

4. TREES WITH LARGE 2-INTERSECTION NUMBER

Let T_k be the tree of diameter four having a central vertex x of degree k, plus $\binom{2k-1}{2}$ leaves appended to each neighbor of x. We compute a lower bound on θ_2 for trees in this sequence, so we technically prove a lower bound on the highest 2-intersection number of an n-vertex tree only for infinitely many values of n, but it is easy to interpolate the construction and argue for intermediate values.

Theorem 3. The tree T_k with $n = k\binom{2k-1}{2} + k + 1$ vertices has 2-intersection number at least $k(3k + 1)/2$.

Proof. Consider an optimal representation of T_k, using subsets of $[t]$. Let $y_1, \ldots, y_k$ be the neighbors of x, and let Y_i be the set assigned to y_i, for $1 \leq i \leq k$. Since the neighbors of y_i form an independent set, they must be assigned distinct 2-sets from Y_i. Since y_i has degree $1 + \binom{2k-1}{2}$, this forces $|Y_i| \geq 2k$.

Now consider $\{y_1, \ldots, y_k\}$. Because this is an independent set, no sets Y_i and Y_j can have more than one common element. Hence Y_j has at most $j - 1$ elements in $\cup_{i=1}^{j-1} Y_i$, for $j \geq 2$. This forces each Y_j to have at least $2k - j + 1$ elements not in $\cup_{i=1}^{j-1} Y_i$, for $j \geq 2$. Hence $|\cup_{j=1}^{k} Y_j| \geq \sum_{j=1}^{k} 2k + 1 - j = k(3k + 1)/2$. ∎

Corollary 1. For infinitely many values of n, there exists an n-vertex tree with 2-intersection number larger than $3/2(n/2)^{2/3}$.

Proof. Let n be the order of the tree T_k; we have $n \leq 2k^3$. Let $t = \theta_2(T_k)$; we have $t > 3k^2/2$. Hence $t > 1.5(n/2)^{2/3} \approx .945n^{2/3}$. ∎

It is also possible to find a 2-intersection representation of T_k with $k(3k + 3)/2$ elements; thus showing that $\theta_2(T_k) = k(3k + 3)/2$. This result has been generalized in [7] to show that max $\theta_p(T)$ over n-vertex trees is asymptotically as large as a constant times $n^{2/(p+1)}$.

ACKNOWLEDGMENT

The research of MSJ was supported in part by ONR Grant N00014-91-J-1098. The research of DBW was supported in part by NSA/MSP Grant MDA904-90-H-4011.

References

[1] I. Anderson, *Combinatorial Designs: Construction Methods,* Halsted Press, New York, New York (1990).

[2] M. S. Chung and D. B. West, The p-intersection number of a complete bipartite graph and orthogonal double coverings of a clique. Manuscript (1991).
[3] P. Erdös, A. Goodman, and L. Posa, The representation of a graph by set intersections. *Can. J. Math.* **18** (1966) 106–112.
[4] M. N. Huxley, On the difference between consecutive primes. *Invent. Math.* **15** (1972) 164–170.
[5] G. Isaak, S.-R. Kim, T. A. McKee, F. R. McMorris, and F. S. Roberts, DIMACS Technical Report 89-22 (1989).
[6] M. S. Jacobson, On the p-edge clique covering number of complete bipartite graphs. *SIAM J. Discrete Math.* **5** (1992) 539–544.
[7] M. S. Jacobson and A. E. Kézdy, Systems of p-intercycle sequences. Preprint.
[8] S.-R. Kim, Ph.D. thesis, Rutgers University (1988).
[9] S.-R. Kim, T. A. McKee, F. R. McMorris, and F. S. Roberts, DIMACS Technical Report 89-19 (to appear in *SIAM J. Discrete Math.*).

Received October 20, 1992