Discrete Mathematics 114 (1993) 3–7
North-Holland

Bisection of trees and sequences

N. Alon, Y. Caro and I. Krasikov

School of Mathematical Sciences, Sackler Faculty of Exact Sciences, Tel-Aviv University, Israel

Received 11 December 1988
Revised 6 May 1990

Abstract

Alon, N., Y. Caro and I. Krasikov, Bisection of trees and sequences, Discrete Mathematics 114 (1993) 3–7.

A graph G is called *bisectable* if it is an edge-disjoint union of two isomorphic subgraphs. We show that any tree T with e edges contains a bisectable subgraph with at least $e-O(e/\log\log e)$ edges. We also show that every forest of size e, each component of which is a star, contains a bisectable subgraph of size at least $e-O(\log^2 e)$.

1. Introduction

Let G be a graph with $n=n(G)$ vertices and $e=e(G)$ edges. The number of edges e of G is called the size of G. G is *bisectable* if it is an edge-disjoint union of two isomorphic subgraphs.

Let $B(G)$ be a bisectable subgraph of maximum size of G.

The function $R(G)=e(G)-e(B(G))$ for general graphs G has been studied by Erdős et al. [1] and independently by Alon and Krasikov (unpublished). It was shown that any graph of size e contains a bisectable subgraph with at least $\Omega(e^{2/3})$ edges, and that there are graphs of size e containing no bisectable subgraphs of size more than $O(e^{2/3}\log e/\log\log e)$.

Here we consider the function $R(G)$ in two special cases; when G is a tree and when G is a forest, each connected component of which is a star. Some other results dealing with decompositions of trees into isomorphic subgraphs appear in [4] and in some of its references.

2. Trees

The first class of graphs we consider is the class of trees.

We start with the following result of Otter [5].

Correspondence to: N. Alon, Dept. of Mathematics, Tel Aviv University, Ramat Aviv, Israel.

0012-365X/93/$06.00

Lemma 2.1. *The number of isomorphism types of rooted trees with n vertices is at most* $\frac{1}{n}\binom{2n-2}{n-1}$.

Lemma 2.2. *Let F be a forest and let q be the maximum size of a connected component of it. Let F′ be a tree obtained from F by inserting one additional vertex adjacent to each component of F. Then* $R(F)\leqslant 4^q$ *and* $R(F')\leqslant 4^q$.

Proof. Split all components of F into classes, according to their isomorphism type. Drop out one tree from each class containing an odd number of components, and bisect the classes obtained in the obvious way. Since at most one tree for each isomorphism type is omitted, we have, by the previous lemma,

$$R(F)\leqslant \sum_{i=1}^{q}\binom{2i-2}{i-1}<4^q.$$

The proof that $R(F')\leqslant 4^q$ is similar. □

Theorem 2.3. *For any tree T with n vertices,*

$$R(T)\leqslant O\left(\frac{n}{\log\log n}\right).$$

On the other hand there are trees T on n vertices such that $R(T)\geqslant\Omega(\log n)$.

Proof. To prove the lower bound, consider the forest F consisting of the stars $F_i\cong K_{1,5^i}$, $i=1,2,\ldots,t$. Consider the tree T obtained from F by inserting a new vertex v adjacent to the centers of all stars. It is not too difficult to check that $R(T)>\Omega(\log n)$. We omit the details.

We now prove the upper bound. Fix two numbers $f=(\log n)^{1/4}\log\log n$ and $g=\frac{1}{2}\log\log n$, where all logarithms above are in base 4. Given a tree T, produce a forest in it according to the following three steps:

(i) Delete from T the minimum number of edges such that in the resulting forest F_1 each component has at most one vertex of degree more than f.

(ii) Consider vertices having degree more than f in F_1. For each such vertex v consider the branches at v having more than g edges and delete the edges joining these branches to v. Denote the forest so obtained by F_2.

(iii) Consider those components of F_2 which do not contain a vertex of degree more than f whose size is more than f^2. Delete the minimum possible number of edges from F_2 to obtain a forest F_3 having no such components.

Observe that the total number of deleted edges does not exceed $3n/f+n/g$. Indeed, the number of edges deleted from T in step (i) is at most $2n/f$. (To see this, choose, arbitrarily, a root v of T. For each vertex $u\neq v$ of degree greater than f, delete the edge joining it to its parent.) In step (ii) at most n/g additional edges are deleted. Also, in any tree T of size $|T|$ and maximum degree Δ, there is a branch of size s, where $c \leqslant s \leqslant c\Delta$ for any integer c satisfying $1 < c \leqslant |T|/\Delta$; hence, at most n/f edges are deleted in step (iii).

Now we can build the required bisection. To this end, split all components of F_3 into two classes F_4 and F_5 such that F_4 consists of all components of size less than f^2.

By Lemma 2.2, F_4 can be bisected after omitting at most 4^{f^2} edges. Similarly, each component of F_5 can be bisected after omitting at most 4^g edges. But, since each component in F_5 has size at least f^2, the total number of components in F_5 does not exceed n/f^2. Thus, $R(F_5) < n4^g/f^2$.

Altogether, we get

$$R(T) < \frac{3n}{f} + \frac{n}{g} + 4^{f^2} + \frac{n4^g}{f^2} < \mathrm{O}\left(\frac{n}{\log\log n}\right). \qquad \square$$

Corollary 2.4. *If G is a forest with e edges then* $R(G) < \mathrm{O}\left(\dfrac{e}{\log\log e}\right)$.

Remark 2.5. The last proof showing that, for any tree T with e edges, $R(T) \leqslant \mathrm{O}(e/\log\log e)$ easily supplies a polynomial-time algorithm for producing, for any such T, a bisectable subgraph H of size at least $e - \mathrm{O}(e/\log\log e)$ in it (together with an actual bisection of H). This is in contrast to the result in [3], which asserts that the decision problem 'given a tree T, decide if $R(T)=0$, i.e., if T is bisectable' is NP-complete.

3. Star forests and sequences

In this section we estimate $R(G)$ for forests each component of which is a star. In this case an isomorphic decomposition has a natural interpretation as a decomposition of a sequence of positive integers. Let us first introduce some relevant definitions. All sequences we consider are finite sequences of nonnegative integers. We use capital letters to denote our sequences, and the corresponding small letters for their elements.

Let $n = n(A)$ be the number of elements in a sequence A, $S = S(A) = \sum a_i$. We write $A \leqslant B$ if, after an appropriate ordering, $n(A) = n(B)$ and $a_i \leqslant b_i$ for all i.

A sequence A is called *bisectable* if there is a sequence B and a permutation π such that $a_i = b_i + b_{\pi(i)}$, $1 \leqslant i \leqslant n$. A sequence A is called *irreducible* if it does not contain a proper bisectable subsequence.

Observe that a forest F consisting of vertex-disjoint stars is bisectable iff the sequence A, whose elements are the sizes of the corresponding components of F, is bisectable. Moreover, $R(F) = \min(S(A) - S(B))$, where B ranges over all bisectable sequences satisfying $B \leqslant A$. The right-hand side of this equality can be used as a definition of $R(A)$ for sequences.

To estimate $R(A)$, we need the following well-known result which we state without its simple proof.

Lemma 3.1. *If δ_i are reals, $1 \leqslant i \leqslant n$, $|\delta_i| \leqslant c$, and $\sum \delta_i = 0$ then there is a permutation $\delta_{\pi(1)}, \ldots, \delta_{\pi(n)}$ such that $\pi(1) = 1$ and $|\sum_{i=1}^{j} \delta_{\pi(i)}| \leqslant c$ for all $1 \leqslant j \leqslant n$.*

Lemma 3.2. *Let A be an irreducible sequence satisfying* $\max a_i / \min a_i < 1.5$. *If A can be partitioned into two pairwise disjoint subsequences B and C, such that $n(B) = n(C) = \frac{1}{2}n(A)$, $S(B) = S(C) = \frac{1}{2}S(A)$, then A is bisectable.*

Proof. Put $n = n(A)$ and let us index the elements of A by $a_0, a_1, \ldots, a_{n-1}$ such that $\min a_i = a_0$ and $\sum a_{2i+1} = \sum a_{2i}$. By the previous lemma, applied to the numbers $(a_0 - a_1)$, $(a_2 - a_3), \ldots, (a_{n-2} - a_{n-1})$, we can rearrange the sequence such that $\min a_i = a_0$ and $|(a_0 - a_1) + \cdots + (a_{2i-2} - a_{2i-1})| \leqslant \lfloor a_0/2 \rfloor$ for all $1 \leqslant 2i - 1 < n$.

Define

$$d_0 = \left\lceil \frac{a_0}{2} \right\rceil, \qquad d_i = (-1)^i d_0 + \sum_{j=1}^{i} (-1)^{i-j} a_j, \quad i \geqslant 1.$$

Observe that $a_i = d_i + d_{i-1}$, where $d_{-1} = d_{n-1}$. Indeed, this is obvious for all $i > 0$, and, for $i = 0$, $d_{n-1} = d_{-1} = a_0 - d_0$ since $\sum a_{2i+1} = \sum a_{2i}$. Moreover, for every $1 \leqslant 2i < n$,

$$d_{2i} = (a_0 - a_1) + \cdots + (a_{2i-2} - a_{2i-1}) + \left(a_{2i} - \left\lceil \frac{a_0}{2} \right\rceil \right) \geqslant -\left\lfloor \frac{a_0}{2} \right\rfloor + \left\lfloor \frac{a_0}{2} \right\rfloor = 0.$$

Similarly, for every $1 \leqslant 2i - 1 < n$,

$$d_{2i-1} = -(a_0 - a_1) - \cdots - (a_{2i-2} - a_{2i-1}) + \left\lfloor \frac{a_0}{2} \right\rfloor \geqslant -\left\lfloor \frac{a_0}{2} \right\rfloor + \left\lfloor \frac{a_0}{2} \right\rfloor = 0.$$

Thus, each d_i is nonnegative and A is bisectable, as needed. □

Theorem 3.3. *Let A be a sequence whose sum of elements is S. Then $R(A) \leqslant O(\log^2 S)$. On the other hand, there are sequences A with sum $S(A) = S$ satisfying $R(A) \geqslant \Omega(\log S)$.*

Proof. The lower bound can be easily proved for the sequence defined by $a_i = 3^i$, $0 \leqslant i \leqslant n-1$.

Given a sequence A with sum $S(A) = S$, let us prove the upper bound. Clearly, we may assume that A is irreducible. Order the members of A in a nondecreasing order $1 \leqslant a_0 \leqslant a_1 \leqslant \cdots \leqslant a_{n-1}$. Choose k to be the minimal number such that $a_{i+k} > 1.5 a_i$ for all $i < n - k$. (If there is no such $k < n$, choose $k = n$.)

Clearly,

$$S^k \geqslant \left(\frac{a_{n-1}}{a_1} \right)^k \geqslant \prod_{i=0}^{n-k-1} \frac{a_{i+k}}{a_i} \geqslant 1.5^{n-k}$$

and, hence, $k \geqslant \Omega(n / \log S)$. By the definition of k, there exists an l so that $a_{l+k-1} \leqslant 1.5 a_l$.

Observe now that any sequence X of t positive integers such that $1 + \Sigma_j (x_j - \min x_i) < \binom{t}{\lceil t/2 \rceil}$ contains two disjoint subsequences Y and Z such that

$S(Y)=S(Z)$ and $n(Y)=n(Z)$. (This is because, by trivial counting, there are two distinct subsequences of $\lceil t/2 \rceil$ terms each with equal sums, and, by omitting the elements in their intersection from both, we obtain Y and Z.) Since A was assumed to be irreducible, we must have, by Lemma 3.2,

$$S \geqslant \sum_{j=1}^{k-1} (a_{l+j}-a_l) \geqslant \binom{k}{\lceil k/2 \rceil} - 1.$$

Combining this with $k \geqslant \Omega(n/\log S)$ yields $n \leqslant O(\log^2 S(A))$. Since, trivially, $R(A) \leqslant n$, we conclude that $R(A) \leqslant O(\log^2 S(A))$, as needed. $\square$

Corollary 3.4. *Every forest of size e each connected component of which is a star contains a bisectable subgraph of size at least* $e - O(\log^2 e)$.

4. Concluding remarks and open problems

It would be interesting to close the gaps between our upper and lower bounds in Theorems 2.3 and 3.3. One can consider the following natural generalization of Theorem 3.3. For a sequence $A = a_0, \ldots, a_{n-1}$ and an integer k we say that A is *k-decomposable* if there is a sequence $B = b_0, \ldots, b_{n-1}$ and k permutations $\pi_1, \ldots, \pi_k$ such that $a_i = \Sigma_{j=1}^{k} \beta_{\pi_j(i)}$. A repeated application of Theorem 3.3 clearly implies that any sequence A whose sum of elements is S contains a 2^j-decomposable sequence C, with $C \leqslant A$ such that $S(A) - S(C) = O(2^j \log^2 S)$. When trying to study the case of k-decomposable sequences, one naturally obtains the problem of determining or estimating the function $f_k(n)$ defined as follows. For each $k \geqslant 2$, $f_k(n)$ is the maximum cardinality of a set A of integers not exceeding n, such that there are no k pairwise disjoint subsets $A_1, \ldots, A_k$ of A satisfying $|A_1| = \cdots = |A_k|$ and $S(A_1) = \cdots = S(A_k)$. Using the well-known results of Erdős and Rado [2] on sunflowers one can easily prove that for every fixed k, $f_k(n) \leqslant O(\log^2(n))$. Moreover, the validity of the Erdős–Rado conjecture would imply $f_k(n) \leqslant O(\log n)$. We have been informed that D. Coppersmith, motivated by a completely different problem, has considered recently a similar question and also noticed its connection to the Erdős–Rado results.

References

[1] P. Erdős, J. Pach and L. Pyber, Isomorphic subgraphs in a graph, in: Combinatorics, Proc. Colloq. Math. Soc. János Bolyai, Eger 1986, to appear.

[2] P. Erdős and R. Rado, Intersection theorems for systems of sets, J. London Math. Soc. 35 (1960) 85–90.

[3] R.L. Graham and R.W. Robinson, Isomorphic factorization IX, to appear.

[4] F. Harary and R.W. Robinson, Isomorphic factorization VIII: bisectable trees, Combinatorica 4 (1984) 169–179.

[5] R. Otter, The number of trees, Ann. of Math. 49 (1948) 583–599.