The size of the largest bipartite subgraphs

Paul Erdős [1]

Mathematical Institute, Hungarian Academy of Sciences, P.O.B. 127, Budapest, H–1364 Hungary

András Gyárfás [2]

Computer and Automation Institute, Hungarian Academy of Sciences, Kende u. 13–17, Budapest, H–1111 Hungary

Yoshiharu Kohayakawa [3]

Instituto de Matemática e Estatística, Universidade de São Paulo, Cidade Universitária, Rua do Matão 1010, 05508–900 São Paulo, SP, Brazil

The last two authors dedicate this note to the memory of Professor Paul Erdős.

Abstract. Simple proofs are given for results of Edwards concerning the size of the largest bipartite subgraphs of a graph.

A widely applied remark of Paul Erdős [4] is that a graph with e edges always contains a bipartite subgraph of at least $e/2$ edges. The importance of this remark justifies the search for improvements. Let $f(e)$ be the largest integer such that any multigraph with e edges must contain a bipartite subgraph with $f(e)$ edges. Edwards [2] proved that

$$f(e) \geq \left\lceil \frac{e}{2} + \frac{1}{8}\left(\sqrt{8e+1} - 1\right) \right\rceil . \tag{1}$$

We shall refer to this lower bound for $f(e)$ as *Edwards's formula*. The purpose of this note is to give a simple proof of (1) which seems to be much more transparent and shorter than the original proof in [2]. Our approach also

[1] Part of this work was done while this author was visiting the University of São Paulo, supported by FAPESP (Proc. 94/2813–6).
[2] Supported by OTKA grant 7309.
[3] Partially supported by FAPESP (Proc. 93/0603–1) and by CNPq (Proc. 300334/93–1 and ProTeM-CC-II Project ProComb).

provides some other best possible lower bounds for the size of the largest bipartite subgraphs in terms of the size and order of the underlying graph (see Theorem 4).

It is easy to see that Edwards's formula gives $f(e)$ exactly when $e = \binom{m}{2}$ for some integer m, and one may check that $e = 19$ is the first case in which (1) is not tight: $12 = f(19) > 11$, the value given by Edwards's formula. This raises the question whether the difference between $f(e)$ and the value given by Edwards's formula is bounded. This question was circulated through [3] and has already been answered. Alon [1] proved that if n is even and $e = n^2/2$, then

$$f(e) - \left\lceil \frac{e}{2} + \frac{1}{8}\left(\sqrt{8e+1} - 1\right)\right\rceil \geq ce^{1/4} \tag{2}$$

for some absolute constant $c > 0$. On the other hand, a construction given in [1] shows that, for any positive integer e, the left-hand side of (2) is bounded from above by $Ce^{1/4}$ for some constant C. A slight improvement of (1) was proved by Hofmeister and Lefmann (cf. [6], Corollary 2.4). Alon [1] and Hofmeister and Lefmann [6] have independently found a short, nonconstructive (in today's terminology, probabilistic) proof of (1). Our methods are constructive.

We note that the asymptotics of $f(e)$ can be determined not only for graphs but for hypergraphs as well, even in a more general form. This was done by Erdős and Kleitman [5] with the probabilistic method. In this note, however, we restrict ourselves to graphs.

Assume that G is a multigraph with e edges. We shall use $E(G)$ and $V(G)$ to denote the set of edges and vertices of G; their respective cardinalities are called the *size* and the *order* of the graph G. The subgraph induced by a subset S of vertices of G is denoted by $G[S]$. If $G[S]$ is bipartite, connected and has at least two vertices, we call it a *bipartite block*. Since a bipartite block is an induced bipartite subgraph, its partite classes are independent sets in G. We refer to these classes as the *classes of the block*. A partition P of $V(G)$ into pairwise disjoint sets $I, S_1, S_2, \ldots, S_t$ is called a *partition of* G if I is an independent set in G (called the *independent block*) and each $G[S_i]$ is a bipartite block. The sum of the orders (respectively, sizes) of the $G[S_i]$ $(1 \leq i \leq t)$ are referred to as the *order* (respectively, *size*) of P. Let us also define the *pseudosize* of the partition P to be the total number of parallel classes of edges in the bipartite blocks of P. Thus the pseudosize of P is its 'size' when the multiplicity of the edges are disregarded. The rôle of partitions will be clear from the following simple lemma.

Lemma 1 *If a graph G of size e admits a partition P of size s, then G has a* *bipartite subgraph of at least $\frac{1}{2}(e+s)$ edges.*

PROOF. Starting from the bipartition $(A,B) = (I,\emptyset)$, take the bipartite blocks of P in turn, in any order, and at each step add one of its classes to A and the other to B, favouring the choice which brings as many edges as possible to (A,B). $\square$

We shall apply the above lemma with suitable partitions. As customary, let $\nu(G)$ denote the maximum number of pairwise disjoint edges in G. In what follows, we only consider *ν-partitions* of G, namely, partitions of G with $t = \nu(G)$ bipartite blocks. Note that any maximum matching of G gives such a partition. Moreover, observe that the bipartite blocks of any ν-partition are induced stars (possibly having multiple edges), since any two edges of any fixed block must meet. Among all ν-partitions of G, consider the ones with largest possible pseudosize. Amongst those, select one, say P_1, with smallest possible independent block I. Let s be the size of P_1. The following properties are immediate from the definition of P_1.

Property 1 *The order of P_1 is at most $2s$.*

Property 2 *Each bipartite block of P_1 that has at least 3 vertices sends no edge into I.*

Property 3 *If a bipartite block of P_1 sends an edge into I, then it sends exactly two, one from each of its two vertices. These two edges, which may have multiplicity larger than 1, are incident to the same vertex in I, thus forming a triangle with the bipartite block.*

The next observation is a little less obvious.

Lemma 2 *If G is connected, the partition P_1 may be chosen so that its independent block I contains at most one vertex.*

PROOF. Suppose to the contrary that $|I| \geq 2$ for all possible choices for P_1. Among all these choices, let P_1 be a partition of G with smallest possible *separation* within I, that is, such that $\min_{u,v} d(u,v)$ is smallest, where the minimum is taken over all pairs of distinct vertices u, $v \in I$ and $d(u,v)$ denotes the distance between u and v in G. Let this minimum be attained by the pair u_0, $v_0 \in I$. Let $u_0, x_1, x_2, \ldots, v_0$ be a minimum length u_0–v_0 path in G, and, for each $i = 1, 2, \ldots$, let y_i be a neighbour of x_i within the bipartite block of x_i.

Property 3 immediately gives that $d(u_0, v_0) \geq 3$. Suppose $d(u_0, v_0) = 3$. Again by Property 3, we have that u_0y_1 and y_2v_0 are edges of G. We now obtain a contradiction by observing that $u_0, y_1, x_1, x_2, y_2, v_0$ is an augmenting path that contradicts the fact that P_1 has $\nu(G)$ bipartite blocks.

Let us now assume that $d(u_0, v_0) \geq 4$. Replace the block $\{x_1, y_1\}$ of the partition P_1 by the block $\{u_0, y_1\}$ to define a new partition P_1' of G. Note that P_1' is a ν-partition, it has the same pseudosize as P_1, and, furthermore, the independent blocks of P_1 and P_1' have the same cardinality. However, P_1' has separation no greater than $d(x_1, v_0) < d(u_0, v_0)$, which contradicts the choice of P_1. This contradiction shows that indeed $|I| \leq 1$, as required. $\square$

Before we proceed, let us state the following consequence of the above lemma, which may be of independent interest.

Corollary 3 *Any connected graph G contains a forest F, all components of which are induced stars, with F covering all but possibly one vertex of G.* $\square$

Properties 1, 2, and 3 combined with Lemmas 1 and 2 give lower bounds for the size of the largest bipartite subgraphs of a graph in terms of its size and order. The bounds are sharp in the sense that there are infinitely many graphs where these bounds are attained. The second assertion in Theorem 4 below is also due to Edwards (see Theorem 6 in [2]). Here we give a shorter, more transparent proof.

Theorem 4 *Let G be a graph of order n and size e. If G has no isolated vertices then it has a bipartite subgraph of size at least $\frac{1}{2}\left(e + \frac{1}{3}n\right)$. If G is connected, then it has a bipartite subgraph of size at least $\frac{1}{2}\left(e + \frac{1}{2}(n-1)\right)$.*

PROOF. Let us first assume that G has no isolated vertices. In view of Lemma 1, to prove the first assertion in our theorem it suffices to prove that $n \leq 3s$. Let $E(P_1)$ denote the set of edges that belong to the bipartite blocks of P_1. Thus $s = |E(P_1)|$. Define a function $\varphi\colon V(G) \to E(P_1)$ as follows. If v is a vertex in a bipartite block, let $\varphi(v)$ be any edge of this bipartite block that is incident to v. Now suppose v belongs to the independent block I of P_1. Since G has no isolated vertices, our vertex v and some bipartite block of P_1 form a triangle (cf. Property 3). Let $\varphi(v)$ be any edge in this bipartite block. It is now easy to see that any edge in $E(P_1)$ is the image of at most 3 vertices of G. Thus $n \leq 3|E(P_1)| = 3s$, as required.

Let us now assume that G is connected. By Lemma 2, we may assume that P_1 has order at least $n-1$. Combined with Property 1, this gives that $n-1 \leq 2s$. The second assertion of our theorem now follows from Lemma 1. $\square$

We now prove (1). Amongst all ν-partitions of G, let P_2 have the maximal possible size. Then Properties 1, 2, and 3 hold for P_2. Moreover, because of the maximality of the size of P_2, the following extra property holds.

Property 4 *Let T be a triangle of G induced by a vertex in the independent block of P_2 and a bipartite block of P_2. The side of T determined by the bipartite block of P_2 has multiplicity at least as large as the multiplicity of the other two sides.*

Let $H = G - I$ be the subgraph of G obtained from G by removing all the vertices in the independent block I of P_2, and let h be its size. By Property 1, the graph H has order at most $2s$, where s is the size of P_2. Therefore, considering a factorization of K_{2s}, one sees that there is a partition P_3 of G with size at least $h/(2s-1)$, all of its bipartite blocks are (possibly multiple) edges. Property 4 implies that $h \geq e - 2s$. Applying Lemma 1 to the partitions P_2 and P_3, we obtain the following result.

Proposition 5 *Assume that G is a graph of size e. Then G has a bipartite subgraph with at least*

$$\frac{e}{2} + \frac{1}{2} \min \left\{ \max \left\{ s, \frac{e-2s}{2s-1} \right\} \right\}$$

edges, where the minimum is taken over $1 \leq s \leq e$. $\square$

Edwards's lower bound (1) follows from the above proposition on solving the equation $s = (e-2s)/(2s-1)$ for s.

Acknowledgement. The authors are grateful to a referee for pointing out a small imprecision in their original manuscript.

References

[1] N. Alon. Bipartite subgraphs. *Combinatorica.* To appear; manuscript of 1995.

[2] C.S. Edwards. Some extremal properties of bipartite subgraphs. *Canadian Journal of Mathematics*, 25:475–485, 1973.

[3] P. Erdős. Some recent problems in combinatorics and graph theory. Manuscript (1995), and lecture at the *26th Southeastern International Conference on Graph Theory, Combinatorics, and Computing*, Boca Raton, 1995.

[4] P. Erdős. On some extremal problems in graph theory. *Israel Journal of Mathematics*, 3:113–116, 1965.

[5] P. Erdős and D.J. Kleitman. On coloring graphs to maximize the proportion of multicolored k-edges. *Journal of Combinatorial Theory*, 5:164–169, 1968.

[6] T. Hofmeister and H. Lefmann. On k-partite subgraphs. *Ars Combinatoria*. To appear; manuscript of 1995.