COLLOQUIA MATHEMATICA SOCIETATIS JÁNOS BOLYAI
60. SETS, GRAPHS AND NUMBERS, BUDAPEST (HUNGARY), 1991

The Maximum Size of 4- and 6-cycle Free Bipartite Graphs on m, n Vertices

D. DE CAEN* and L. A. SZÉKELY

1. Introduction

Let $f(n,m)$ denote the maximum size of a bipartite graph, whose vertex sets are of sizes n, m, not containing any 4- or 6-cycle. It is well-known [3] that $f(n,n) = \Theta(n^{4/3})$, and it is folklore that $f(n,m) = \Theta(n)$ for $m = O(\sqrt{n})$. In 1979, Erdős [5] conjectured that $f(n,m) = O(n)$ for $m = O(n^{2/3})$, since in this range the probabilistic method fails to give a superlinear lower bound. Actually, the conjecture is even older, since Erdős referred to it as an "old and nearly forgotten conjecture of mine".

In this paper we prove $f(n,m) = O(n^{2/3}m^{2/3})$. Faudree and Simonovits [7] have also obtained this result. Our proof is motivated by Mantel's argument [9], [8, 10.30]; we also sketch a second proof that uses an interesting matrix inequality of Atkinson & al. [1]. The remarks above show that this upper bound on $f(n,m)$ is asymptotically tight when $n = m$, $m = O(\sqrt{n})$. Using generalized quadrangles, we show that $f(n,m) = \Omega(n^{2/3}m^{2/3})$ also when $m \sim n^{4/5}$ and $m \sim n^{7/8}$.

* The first author's research is supported financially by NSERC grant OGP0093041.

We disprove Erdős' conjecture by constructing a bipartite graph without 4- and 6-cycles, on n, $m = \sqrt{n}e^{f(n)\sqrt{\log n}}$ vertices with $g(n)n$ edges, where $f(n) \to \infty$ arbitrarily slowly and $g(n) \to \infty$. For n, $m \sim n^{1/2+\delta}$ $(0 < \delta < 1/4)$ our construction gives a 4- and 6-cycle–free bipartite graph with $n^{1+\varepsilon}$ $(\varepsilon = \varepsilon(\delta) > 0)$ edges. In particular, for n, $m \sim n^{2/3}$, it gives $n^{58/57+o(1)}$ edges; this disproves Erdős' conjecture, but falls short of our upper bound, $O(n^{10/9})$.

The second author is indebted to Professors Erdős, Faudree and Simonovits for conversations on the topic of the present paper.

2. Upper bounds

Theorem 1. *Suppose G is a bipartite graph with vertex sets of size n, m $(\sqrt{n} \leq m \leq n)$ without 4- and 6-cycles. Then the number of edges is $O(n^{2/3}m^{2/3})$.*

Proof. First, we prove the theorem for $m = O(\sqrt{n})$. This is the only case that we settle in terms of the original graph, all the other cases will be settled in terms of a set system. Let the degrees of the bipartite graph on the $n-$side be $d_1, ..., d_n$. Since every pair of vertices on the $m-$side belongs to at most one neighbourhood, we have

$$\sum_{i=1}^{n} (d_i - 1) \leq \sum_{i=1}^{n} \binom{d_i}{2} \leq \binom{m}{2} = O(n),$$

and $\sum d_i = O(n)$, which is equivalent to the theorem for $\sqrt{n} \leq m = O(\sqrt{n})$.

Let X denote the $n-$element vertex set. Identify the graph G with a collection $\mathcal{F}$ of m subsets of X, whose members are the neighbourhoods of G contained in X.

$$\text{For } E, F \in \mathcal{F},\ |E \cap F| \leq 1, \tag{1}$$

since G has no 4-cycle, and

$$\text{for } D,E,F \in \mathcal{F},\ |D \cap E| = |D \cap F| = |E \cap F| = 1 \text{ implies } |D \cap E \cap F| = 1, \tag{2}$$

since G has no 6-cycle. Hence, it is sufficient to prove, that for a collection of sets $\mathcal{F}$ satisfying (1) and (2),

$$\sum_{F \in \mathcal{F}} |F| = O(n^{2/3}m^{2/3}). \tag{3}$$

We may assume without loss of generality that $\mathcal{F}$ contains no singletons.

Now split $\mathcal{F}$ into $\mathcal{F}_1$ and $\mathcal{F}_2$ by $\mathcal{F}_1 = \{F \in \mathcal{F} : |F| \leq 2\sqrt{n}\}$, $\mathcal{F}_2 = \mathcal{F}\backslash\mathcal{F}_1$. We have to prove the theorem for both families. Set $m_2 = |\mathcal{F}_2|$. The reader may easily prove, or may find it in Székely [12], that $m_2 = O(\sqrt{n})$, and hence the theorem is proved for $\mathcal{F}_2$. From now on we simply assume that the size of the largest set in $\mathcal{F}$ has size at most $2\sqrt{n}$, and $|\mathcal{F}| = m$.

The "main case" of the proof will be to show that (3) holds for $\mathcal{F}$ if

$$\sum_{F\in\mathcal{F}} |F|^3 \leq \left(\sum_{F\in\mathcal{F}} |F|^2\right)^2 /2n. \tag{4}$$

We postpone the proof of the main case.

Either we have

$$\max_{E\in\mathcal{F}} |E| \leq \frac{\sum_{F\in\mathcal{F}} |F|^2}{2n} \tag{5}$$

or the opposite inequality. If (5) holds, then we are in the main case:

$$\sum_{F\in\mathcal{F}} |F|^3 \leq \sum_{F\in\mathcal{F}} |F|^2 \max_{E\in\mathcal{F}} |E| \leq \left(\sum_{F\in\mathcal{F}} |F|^2\right)^2 /2n.$$

Hence we may assume that the opposite of (5) holds:

$$\sum_{F\in\mathcal{F}} |F|^2/2n \leq \max_{F\in\mathcal{F}} |F| \leq 2\sqrt{n},$$

hence by the inequality of arithmetic and quadratic means, $\sum_{F\in\mathcal{F}} |F| \leq 2m^{1/2}n^{3/4} = O(n^{2/3}m^{2/3})$ in the range that we consider.

Finally, we prove the theorem in the main case. Let $d(p)$ denote the degree of the vertex p for the collection $\mathcal{F}$. Fix any $E \in \mathcal{F}$ and note that

$$\sum_{F\in\mathcal{F}:|E\cap F|=1} |F| \leq n + \sum_{p\in E} [d(p) - 2], \tag{6}$$

since the sets $F \setminus E : F \in \mathcal{F}$ are pairwise disjoint. Apply the operator $\sum_{E\in\mathcal{F}} |E|\cdot$ to (6). Using the convention of underlining the summation variable in dubious cases, we have

$$\sum_{p\in X}\left(\sum_{p\in \underline{E}\in\mathcal{F}}|E|\right)^2-\sum_p\sum_{p\in \underline{E}\in\mathcal{F}}|E|^2\le n\sum_{E\in\mathcal{F}}|E|+\sum_{E\in\mathcal{F}}|E|\left(\sum_{\underline{p}\in E}d(p)-2\right)$$

$$\le (n+m)\sum_{E\in\mathcal{F}}|E|\le 2n\sum_{E\in\mathcal{F}}|E|.$$

On the other hand, (4) and two applications of the inequality of arithmetic and quadratic means yields

$$\sum_{p\in X}\left(\sum_{p\in \underline{E}\in\mathcal{F}}|E|\right)^2-\sum_p\sum_{p\in \underline{E}\in\mathcal{F}}|E|^2\ge\frac{\left(\sum_p\sum_{p\in\underline{E}\in\mathcal{F}}|E|\right)^2}{n}-\sum_{E\in\mathcal{F}}|E|^3$$

$$\ge\frac{\left(\sum_{E\in\mathcal{F}}|E|^2\right)^2}{n}-\sum_{E\in\mathcal{F}}|E|^3\ge\frac{\left(\sum_{E\in\mathcal{F}}|E|^2\right)^2}{2n}\ge\frac{\left(\sum_{E\in\mathcal{F}}|E|\right)^4}{2nm^2},$$

and from here we immediately get (3). ■

An alternative proof of Theorem 1 can be given using an inequality of Atkinson & al. [1]:

$$\sigma(AA^tA)\ge\frac{\sigma(A)^3}{mn} \tag{7}$$

for any $m\times n$ non-negative matrix A, where $\sigma(M)$ denotes the sum of entries of M.

To see the connection with $f(n,m)$, let A be the bipartite adjacency matrix of a graph G satisfying the conditions of Theorem 1. It is straightforward to see that

$$\frac{e^3}{nm}\le\sigma(AA^tA)\le nm-2e+\sum_{i\in V(G)}d_i^2, \tag{8}$$

where $e=\sigma(A)$ is the number of edges of G. One can show that $\sum_{i\in V(G)}d_i^2=O(mn)$ for the family $\mathcal{F}_1$ defined after (3). Then (8) implies $e=O(n^{2/3}m^{2/3})$ for $\mathcal{F}_1$ again.

3. Lower bounds

Lemma. *Let $r(k,n)$ denote the maximum number of natural numbers $1 \le a_1 < \ldots < a_r \le n$, such that the simultaneous equations $a_i x + a_j y = a_l z$, $x + y = z$ have only trivial solution $(a_i = a_j = a_l)$ for $1 \le x, y, z \le k = O(n^{1/4-\varepsilon})$ $(\varepsilon > 0)$. Then*

$$r(k,n) > n^{1 - 3(\frac{\log 2k}{\log n})^{1/2} + 2\frac{\log 2k}{\log n} + O\left(\frac{\log \log n}{\log n}\right)}.$$

Proof. Use the base-t representation of integers in $[1,n]$. Let $(g)_i$ denote the i^{th} digit of g in base-t representation. Let A denote the set of those numbers in $[1,n]$, whose base-t digits are less than $t/(2k)$. Assume t be a multiple of $2k$. Let N denote the most frequent sum of squares of the digits of the members of A; and let B denote the subset of A whose members have sum of squares of digits equal to N. We claim that no $a, b, c \in B$ solves the forbidden simultaneous equations. Indeed, assume on the contrary that there is a solution:

$$N = \sum_i (a)_i^2 = \sum_i \left(\frac{(b)_i x + (c)_i y}{x+y}\right)^2 = \frac{x^2+y^2}{(x+y)^2} N + \frac{2xy}{(x+y)^2} \sum_i (b)_i (c)_i,$$

and hence the Cauchy-Schwartz inequality holds with equality, and the digits of b are proportional to that of c. Since the digits have the same square sum, $b = c$, the solution is trivial. Clearly $N \le 2(\frac{t}{2k})^2 \frac{\log n}{\log t}$, $|A| \ge (\frac{t}{2k})^{\frac{\log n}{\log t}}$ and $|B| \ge (\frac{t}{2k})^{\frac{\log n}{\log t} - 2} \frac{\log t}{2 \log n}$. In order to optimize our lower bound for $|B|$, select t to be the first multiple of $2k$ exceeding $e^{\sqrt{\log 2k \log n}}$. It yields the required lower bound. (We need the upper bound for k to make sure that the lower bound for $|B|$ is nontrivial.) ■

Note that the construction above is just a slight generalization of Behrend's construction [2] for a large set without 3-term arithmetic progressions. We get back the special case by setting $k = 2$. As it turned out, Erdős & al. [6], proved a slightly weaker version of the Lemma in the following form: there are $n^{1-c\log k/\sqrt{\log n}}$ integers in $[1,n]$ with no triplet contained in a $(k+1)-$term arithmetic progression. We repeated the proof, since Erdős&al. did not optimize the lower bound as we did, and they did not have a logarithmic asymptotic expansion at all. The next construction is motivated by the papers of Ruzsa and Szemerédi [11] and Clark & al. [4]:

Theorem 2. *There exist a bipartite graph without 4- and 6-cycles, on*

$n,\ m = \sqrt{n}e^{f(n)\sqrt{\log n}}$ *vertices with* $g(n)n$ *edges, where* $f(n) \to \infty$ *arbitrarily slowly and* $g(n) \to \infty$;

$n^{1+\delta},\ n^{2+\delta-3\sqrt{\delta(1-\delta)}+o(1)}$ *vertices with* $n^{2+2\delta-3\sqrt{\delta(1-\delta)}+o(1)}$ *edges* $(0 < \delta < 1/4)$;

$n,\ m = n^{2/3}$ *vertices with* $n^{58/57+o(1)}$ *edges.*

Proof. Set $r' = r(\frac{n}{k\log k}, k)$, take $1 \le a_1 < \ldots < a_{r'} \le \frac{n}{k\log k}$ according to the Lemma. Define a hypergraph in the following way: the vertex set is $\{(i,j) : 1 \le i \le k,\ 1 \le j \le n\}$, which is to be considered as a grid in the plane. The edges of the $k-$uniform hypergraph are the points of the linear functions $y = a_l x + b$ which have k points in the grid, i.e. $0 \le b \le n - a_l k$. Let one vertex set of the bipartite graph be the grid (i.e. the vertex set of the hypergraph), the other vertex set of the bipartite graph be the edge set of the hypergraph, and define the edges of the bipartite graph by the incidences of edges and vertices in the hypergraph. Clearly (1) is satisfied, since the edges are geometrically represented by straight lines, and (2) is just the non-solvability of the simultaneous equations in the lemma. So the graph has no 4- and 6-cycles. Finally, the sizes of vertex sets are $n' = nk$, $m' = r'\Theta(n)$, and we have $m'k$ edges. We have

$$m' = n\left(\frac{n}{k\log k}\right)^{1-3\left(\frac{\log 2k}{\log(\frac{n}{k\log k})}\right)^{1/2}+2\frac{\log 2k}{\log(\frac{n}{k\log k})}+O\left(\frac{\log\log(\frac{n}{k\log k})}{\log(\frac{n}{k\log k})}\right)}.$$

Hence straightforward computation yields our first two claims. The third one comes from the second by setting $\delta = 1/37$. ■

The set system that Erdős & al. [6] constructed using the Lemma, is actually the hypergraph construction of Theorem 2. Erdős & al. [6] constructed it for a different purpose, considered only a fixed k and therefore missed the asymptotic analysis and overlooked the application for bipartite graph construction.

One can show that $f(n,m) = \Omega(n^{2/3}m^{2/3})$ in certain instances, using generalized quadrangles.

Theorem 3.

(i) *There exists an infinite sequence* (n,m) *with* $m \sim n^{4/5}$, *such that* $f(n,m) = \Omega(n^{2/3}m^{2/3})$.

(ii) Similarly there is such a sequence with $m \sim n^{7/8}$.

Proof. Consider generalized quadrangles $GQ(s,t)$, cf. Payne and Thas [10]. These may be defined as bipartite graphs $G(m,n)$, with $m = (1+s)(1+st)$, $n = (1+t)(1+st)$, every vertex on the left of degree $(1+t)$, every vertex on the right of degree $(1+s)$, and with girth 8, i.e. no 4- or 6-cycles. The number of edges of G is $e = (1+s)(1+t)(1+st)$.

Now there exists a $GQ(q,q^2)$ for every prime-power q. In this case we have $n \sim q^5$, $m \sim q^4 \sim n^{4/5}$ and $e \sim q^6 \sim (nm)^{2/3}$, as desired. Similarly there is a $GQ(q^2,q^3)$ yielding $m \sim n^{7/8}$. ∎

4. Open problems

1) If G is a bipartite graph without 4- and 6-cycles, on n, m vertices ($\sqrt{n} \le m \le \sqrt{n}e^{\sqrt{\log n}}$), does it imply that the edge number is $O(n)$?
2) We remark that in a $GQ(s,t)$ one has $s \le t^2$ (and dually $t \le s^2$). This suggests that our upper bound $f(n,m) = O(n^{2/3}m^{2/3})$ may not be tight when $m = o(n^{4/5})$.
3) For a fixed d and $n,k \to \infty$ find a version of the Lemma for the following problem: Give a large set F of integers in $[1,n]$, such that the simultaneous diophantine equations

$$\sum_{i=1}^{m} f_i x_i = \sum_{i=m+1}^{t} f_i x_i$$

$$\sum_{i=1}^{m} x_i = \sum_{i=m+1}^{t} x_i$$

has only trivial solutions for $t \le d$, $1 \le x_i \le k$, $f_i \in F$. (A solution is trivial if it repeats an f_i.) The above generalization of the Lemma would lead the way to similar constructions for n, m bipartite graphs with many edges and girth exceeding $2d$. When $d = 4$ the set F is a Sidon set, a well-studied structure.

References

[1] F. V. Atkinson, G. A. Watterson and P. A. D. Moran, A matrix inequality, *Quarterly J. of Math. Oxford Second Series* **11**(1960) 42, 137–140.

[2] F. Behrend, On sets of integers which contain no three terms in arithmetical progerssion, *Proc. Nat. Sci. U.S.A.* **32**(1946), 331–332.

[3] J. A. Bondy, M. Simonovits, Cycles of even length in graphs, *Journal of Combin. Theory* **16**(1974), 97–105.

[4] L. H. Clark, R. C. Entringer, J. E. McCanna, L. A. Székely, Extremal problems for local properties of graphs, *Australasian J. of Combinatorics* **4**(1991), 25–31.

[5] P. Erdős, Some old and new problems in various branches of combinatorics, Proc. of the Tenth Southeastern Conference on Combinatorics, Graph Theory and Computing, Boca Raton, 1979, Vol. 1, *Congressus Numerantium* **23**(1979), 19–38.

[6] P. Erdős, P. Frankl, V. Rödl, The asymptotic number of graphs not containing a fixed subgraph and a problem for hypergraphs having no exponents, *Graphs and Combinatorics* **2**(1986), 113–121.

[7] R. Faudree, M. Simonovits, On a class of degenerate extremal graph problems II, in preparation.

[8] L. Lovász, *Combinatorial Problems and Exercises*, North-Holland, Amsterdam, 1979.

[9] W. Mantel, *Wiskundige Opgaven* **10**(1906), 60–61.

[10] S. Payne and J. Thas, *Finite Generalized Quadrangles*, Pitman, 1984.

[11] I. Z. Ruzsa, E. Szemerédi, Triple systems with no six points carrying three triangles, in: *Infinite and Finite Sets* (eds.: A. Hajnal, R. Rado, V. T. Sós), Colloq. Math. Soc. János Bolyai, **11**, North-Holland, Amsterdam, 1975, 939–945.

[12] L. A. Székely, Inclusion-exclusion formulae without higher terms, *Ars Combinatoria* **23B**(1987), 7–20.

D. de Caen

Dept. Maths. and Stats.
Queen's University
Kingston, Ont. K7L 3N6
Canada

L. A. Székely

Department of Computer Science
Eötvös University
Budapest, Múzeum krt 6-8
1088, Hungary