ELSEVIER

Discrete Mathematics 163 (1997) 279–284

DISCRETE MATHEMATICS

Note

Graphs without short odd cycles are nearly bipartite

Ervin Györi[a,*], Alexandr V. Kostochka[b], Tomasz Łuczak[c,1]

[a] Mathematical Institute of the Hungarian Academy of Sciences, P.O.B. 127, Budapest H-1364, Hungary
[b] Novosibirsk State University, Novosibirsk, Russia 630090
[c] Mathematical Institute of the Polish Academy of Sciences, Poznań, Poland

Received 17 May 1995

Abstract

It is proved that for every constant $\varepsilon > 0$ and every graph G on n vertices which contains no odd cycles of length smaller than εn, G can be made bipartite by removing $(15/\varepsilon)\ln(10/\varepsilon))$ vertices. This result is best possible except for a constant factor. Moreover, it is shown that one can destroy all odd cycles in such a graph G also by omitting not more than $(200/\varepsilon^2)(\ln(10/\varepsilon))^2$ edges.

For a graph G, let $X_v(G)$ $[X_e(G)]$ denote the minimum number of vertices [edges] which must be removed from G in order to make it bipartite. Furthermore, let $\mathscr{G}_{n,\varepsilon}$ be the family of all graphs with the vertex set $\{1, \dots, n\}$ which contain no odd cycles of length smaller than εn. Erdős and Sós asked whether, for a given ε, $X_v(G)$ and $X_e(G)$ are uniformly bounded for every n and $G \in \mathscr{G}_{n,\varepsilon}$, i.e. whether functions

$$f_v(\varepsilon) = \max\{X_v(G)\colon G \in \mathscr{G}_{n,\varepsilon} \text{ and } n \geqslant 1/\varepsilon\}$$

and

$$f_e(\varepsilon) = \max\{X_e(G)\colon G \in \mathscr{G}_{n,\varepsilon} \text{ and } n \geqslant 1/\varepsilon\}$$

are well defined. We show that this is indeed the case and estimate f_v up to a constant and f_e up to a logarithmic factor.

Theorem. *For every $\varepsilon > 0$ we have*

$$\frac{1}{600\varepsilon}\ln\frac{10}{\varepsilon} \leqslant f_v(\varepsilon) \leqslant \frac{15}{\varepsilon}\ln\frac{10}{\varepsilon} \tag{1}$$

* Corresponding author.
[1] On leave from Department of Discrete Mathematics, Adam Mickiewicz University.

SSDI 0012-365X(95)00321-5

and

$$\left\lfloor \frac{1}{\varepsilon} \right\rfloor^2 \leqslant f_e(\varepsilon) \leqslant \frac{200}{\varepsilon^2}\left(\ln \frac{10}{\varepsilon}\right)^2. \tag{2}$$

Remark. In the note we make no attempts to choose the optimal values of constants, trying to simplify calculations whenever possible.

Let us start with the following elementary observation.

Claim 1. *Let C be a shortest odd cycle of a graph G and let v, w be vertices of C. Then the distance between v and w in G is the same as the distance between v and w on the cycle C, i.e. no path joining v and w in G is shorter than the path between these two vertices contained in C.*

Our proof of Theorem is based on the following result.

Lemma. *Let $\varepsilon n > 33$ and $G \in \mathscr{G}_{n,\varepsilon}$. Then there exist subsets V', V'' of vertices of G and a constant $0 < \alpha \leqslant 1$ such that*

(i) $|V'| + |V''| \leqslant (28\alpha/\varepsilon)\ln(10/\varepsilon)$,

(ii) *removing edges between V' and V'' from G leads to a graph G^*, in which at least αn vertices belong to bipartite components.*

Proof. Let G be a graph on n vertices without odd cycles of length shorter than $\varepsilon n > 33$ and let C be a shortest odd cycle in G of length $k_0 \geqslant \varepsilon n$. Moreover, let v be any vertex which belongs to C, and by $N_i(v)$ we denote the ith neighbourhood of v, i.e. $N_0(v) = \{v\}$ and $N_i(v)$ is the set of all vertices of G which lie at the distance i from v. Claim 1 implies that for every i, $1 \leqslant i \leqslant \lfloor k_0/2 \rfloor$, $N_i(v)$ contains two vertices which lie at the distance i from v on the cycle C, so $|N_i(v)| \geqslant 2$ for $1 \leqslant i \leqslant \lfloor k_0/2 \rfloor$.

Now set $k_1 = \lfloor (\lfloor k_0/2 \rfloor - 1)/2 \rfloor$ and observe that the graph H induced in G by set $\bigcup_{i=0}^{k_1} N_i(v)$ is bipartite. Indeed, suppose that this is not the case. Let C_H be a shortest odd cycle contained in H of length $k' \geqslant k_0$ and let v', v'' be vertices of C_H which are joined by a path of length $\lfloor k'/2 \rfloor \geqslant \lfloor k_0/2 \rfloor$ contained in C_H. Since both v' and v'' lie within the distance k_1 from v, they are joined by a walk of length $2k_1 < \lfloor k_0/2 \rfloor$ in H, contradicting Claim 1 and the choice of C_H. Thus, H must be bipartite.

Set $k_2 = \lfloor k_1/2 \rfloor$. Note that for $33 < k_0 < 41$ we have $k_2 = 4 \geqslant 0.1\varepsilon n$ and for $k_0 \geqslant 41$ we have $k_2 \geqslant (k_0 - 9)/8 \geqslant 4\varepsilon n/41$. We show that for some i_0, $2 \leqslant i_0 \leqslant k_2$, and

$$\alpha = \sum_{j=1}^{2i_0-2} |N_j(v)|/n, \tag{3}$$

we have

$$|N_{2i_0-1}(v)| + |N_{2i_0}(v)| \leqslant (28\alpha/\varepsilon)\ln(10/\varepsilon). \tag{4}$$

In order to verify this fact set

$$u_i = |N_{2i-1}(v)| + |N_{2i}(v)| \geqslant 4$$

for $i = 1, \dots, k_2$, and let $\beta = \exp[(20.5/\varepsilon n)\ln(10/\varepsilon)]$. Note that for every $x \geqslant 0$ we have $1 - e^{-x} \leqslant x$, so

$$\frac{\beta - 1}{\beta} = 1 - \frac{1}{\beta} \leqslant \frac{20.5}{\varepsilon n}\ln\frac{10}{\varepsilon}.$$

Observe now that for at least one i, where $1 \leqslant i \leqslant k_2$, we must have $u_i \leqslant \beta^i$, since otherwise

$$\sum_{j=1}^{k_2} u_i > \sum_{j=1}^{k_2} \beta^i \geqslant \beta\frac{\beta^{k_2+1} - 1}{\beta - 1}$$

$$\geqslant \left(\frac{100}{\varepsilon^2} - 1\right)\frac{\beta}{\beta - 1} \geqslant \frac{99}{\varepsilon^2}\,\frac{\varepsilon n}{20.5\ln(10/\varepsilon)}$$

$$\geqslant \frac{99n}{20.5\varepsilon\ln(10/\varepsilon)} \geqslant \frac{99n}{20.5\ln 10} > n,$$

which, of course, is impossible.

Thus, let i_0 be the smallest index such that $1 \leqslant i_0 \leqslant k_2$ and $\beta^{i_0} \geqslant u_{i_0} \geqslant 4$. Then, for α defined as in (3), we get

$$\alpha > \sum_{j=1}^{i_0-1} \beta^j/n \geqslant \frac{\beta^{i_0} - 1}{n}\,\frac{\beta}{\beta - 1}$$

$$\geqslant \frac{3\beta^{i_0}}{4n}\,\frac{\varepsilon n}{20.5\ln(10/\varepsilon)} > \frac{u_{i_0}\varepsilon}{28\ln(10/\varepsilon)},$$

and, consequently, (4) follows. Hence the assertion holds with $V' = N_{2i_0-1}(v)$ and $V'' = N_{2i_0}(v)$. □

Proof of the Theorem. The upper bounds for $f_v(\varepsilon)$ and $f_e(\varepsilon)$ follow straightforwardly from the Lemma and an elementary induction with respect to the number of vertices in a graph. Clearly, it holds for every ε and $n \leqslant 15$ (since then we may delete either all except two vertices or all except two edges of the graph). Thus, let us assume that the assertion is valid for every $\varepsilon > 0$ and every n' such that $n' < n$. Moreover, let G be a graph on n vertices without odd cycles of length less than εn. Note now that, again, when $\varepsilon n \leqslant 33$ there is nothing to prove, because for such ε and n

$$\frac{15}{\varepsilon}\ln\frac{10}{\varepsilon} \geqslant \frac{15n}{33}\ln 10 > n$$

and

$$\frac{200}{\varepsilon^2}\left(\ln\frac{10}{\varepsilon}\right)^2 \geqslant \frac{n^2}{2}\,\frac{400}{33^2}(\ln 10)^2 > \binom{n}{2},$$

so we can remove from G either $n-2$ vertices or all, except two, edges. Hence, we may assume that $\varepsilon n > 33$ and apply the Lemma. Then, there exists $\alpha > 0$ such that deleting from G at most $(14\alpha/\varepsilon)\ln(10\varepsilon)$ vertices (the smallest set of V' and V'') splits it into a bipartite part and a graph G' with at most $(1-\alpha)n$ vertices. According to the inductional assumption, G' can be made bipartite by removing not more than $(15(1-\alpha)/\varepsilon)\ln(10(1-\alpha)/\varepsilon)$ vertices, so to destroy all odd cycles in G it is enough to delete from it

$$\frac{14\alpha}{\varepsilon}\ln\frac{10}{\varepsilon}+\frac{15(1-\alpha)}{\varepsilon}\ln\frac{10(1-\alpha)}{\varepsilon} \leqslant \frac{15\alpha}{\varepsilon}\ln\frac{10}{\varepsilon}+\frac{15(1-\alpha)}{\varepsilon}\ln\frac{10}{\varepsilon} = \frac{15}{\varepsilon}\ln\frac{10}{\varepsilon}$$

vertices.

Similarly, to show the upper bound for $f_e(\varepsilon)$ we apply the Lemma and remove $(14\alpha/\varepsilon)^2(\ln(10/\varepsilon))^2$ edges from G disconnecting it into a bipartite part of size αn and the remaining part with $(1-\alpha)n$ vertices. Then, from the inductional assumption, we deduce that to make G bipartite one needs to omit not more than

$$\left(\frac{14\alpha}{\varepsilon}\ln\frac{10}{\varepsilon}\right)^2+\frac{200(1-\alpha)^2}{\varepsilon^2}\left(\ln\frac{10(1-\alpha)^2}{\varepsilon}\right)^2 \leqslant \frac{200}{\varepsilon^2}\left(\ln\frac{10}{\varepsilon}\right)^2$$

edges of G.

Now note that if we replace vertices of an odd cycle C of length εn by sets of size either $\lfloor 1/\varepsilon\rfloor$ or $\lceil 1/\varepsilon\rceil$, and the edges of C by complete bipartite graphs, we get a graph $G\in\mathscr{G}_{n,\varepsilon}$ such that to destroy all odd cycles in G we must delete from it either at least $\lfloor 1/\varepsilon\rfloor$ vertices or at least $\lfloor 1/\varepsilon\rfloor^2$ edges. This simple construction gives the promised lower bound for $f_e(\varepsilon)$. However, in order to get an additional logarithmic factor in the lower bound for $f_v(\varepsilon)$, we need the following fact.

Claim 2. *For every $m \geqslant e^{16}$ there exists a graph G_m on $3m$ vertices, with at most $120m$ edges and girth at least $0.25\ln 3m$, which cannot be made bipartite by deleting less than m vertices.*

Proof. Let $\bar m = 4m$ and let G^{rand} be the random graph with vertex set $\{1,\dots,\bar m\}$ in which two vertices i, j are adjacent with probability $p = 30/\bar m$ independently for each pair $1 \leqslant i < j \leqslant \bar m$. Note that the expected number of edges of G^{rand} is $\binom{\bar m}{2}p < 15\bar m$, so due to Markov's inequality, the probability that G^{rand} has more than $2.15\bar m = 120m$ edges is less than $1/2$. Furthermore, the expected number of cycles of G^{rand} of length smaller than $k = \lceil 0.25\ln\bar m\rceil$ is bounded from above by

$$\sum_{i=3}^{k-1}\binom{\bar m}{i}\frac{(i-1)!}{2}p^i \leqslant \frac{1}{2}\sum_{i=3}^{k-1}\frac{30^i}{i} < \frac{1}{6}\sum_{i=0}^{k-1}30^i$$

$$\leqslant \frac{30^k}{6\cdot 29} \leqslant \frac{30}{174}\bar m^{0.25\ln 30} < \frac{\bar m}{12}.$$

Thus, again by Markov's inequality, the probability that more than $\bar{m}/4 = m$ vertices of G^{rand} belong to such short cycles is less than $\frac{1}{3}$. Finally, note that the probability that G^{rand} contains an independent set on m vertices can be bounded from above by

$$\binom{\bar{m}}{m}(1-p)^{\binom{m}{2}} \leqslant \left(\frac{e\bar{m}}{m}\exp\left(-\frac{p(m-1)}{2}\right)\right)^m$$
$$\leqslant (4e\exp(-13/4))^m$$
$$\leqslant (0.5)^m \leqslant 0.1.$$

Now let $\bar{G}^{\text{rand}}$ be a graph obtained from G^{rand} by deleting m vertices in such a way that, we remove vertices which cover as large as possible number of cycles of length smaller than $0.25 \ln \bar{m}$. Then, $\bar{G}^{\text{rand}}$ has $3m$ vertices and, with the probability at least $1 - \frac{1}{2} - \frac{1}{3} - \frac{1}{10} > 0$, has less than $120m$ edges, no cycles of length smaller than $0.25 \ln 3m < 0.25 \ln \bar{m}$ and no independent set of size bigger than m. Since this probability is positive, in particular, there must exist at least one graph which fulfills all these properties. □

Now we return to the proof of the Theorem. We have already observed that $f_v(\varepsilon) \geqslant \lfloor 1/\varepsilon \rfloor$, which is better than the lower bound required in (1) whenever $\varepsilon > e^{-100}$. Thus, assume that $\varepsilon \leqslant e^{-100}$, choose the smallest possible m for which

$$\varepsilon > \frac{1}{500}\,\frac{\ln 3m}{m},$$

and for such an m let G_m be a graph whose existence is guaranteed by Claim 2. Let $n > 200m$. Replace each edge of G_m by a path with either s or $s+1$ internal vertices, where s is such that a graph G obtained in this way has precisely n vertices (in particular, $s + 1 \geqslant (n - 3m)/(120m)$). Now one can easily check that the girth of G is at least

$$\frac{s+1}{4}\ln 3m \geqslant \frac{n(1-3m/n)}{480m}\,\varepsilon n$$

and one must take away from it at least

$$m > \frac{1}{500\varepsilon}\ln\left(\frac{3}{500\varepsilon}\right) > \frac{1}{600\varepsilon}\ln\left(\frac{10}{\varepsilon}\right)$$

vertices to make G bipartite. □

Remark. Notice that we did not use the fact that ε is a constant. The same proof works if ε is a function of n, however the strength of the result is different for different orders of magnitude of ε. It seems to be a nice problem to determine the orders of magnitude of ε when the obtained result is best possible apart from a constant factor.

Acknowledgements

Essential part of the work was made when the second author visited the Faculty of Mathematical and Computer Science at Adam Mickiewicz University in Poznań. He thanks the hosts for the hospitality. His work was supported by the grant 93-011-1486 of the Russian Foundation of Fundamental Research and the grant RPY000 of International Science Foundation. The work of the first author was supported by the Hungarian Research Grant T-7559 and partly it was done while he visited CWI (Amsterdam). He is very grateful for the hospitality of CWI.