COMBINATORICA
Bolyai Society – Springer-Verlag

COMBINATORICA **17** (3) (1997) 393–400

COVERING ODD CYCLES

JÁNOS KOMLÓS

Received December 31, 1996

To the memory of Pál Erdős

We estimate the number of vertices/edges necessary to cover all odd cycles in graphs of given order and odd girth.

1. Introduction

1.1. Notation

Given a graph G, we write $V(G)$ and $E(G)$ for the vertex-set and the edge-set of G, $e(G)$ for the number of edges in G, $\delta(G)$ for the minimum degree of G, $og(G)$ for the odd girth of G (length of shortest odd cycles), and $\mathcal{O}(G)$ for the set of all odd cycles in G. We also write $\kappa(G)$ and $\lambda(G)$ for the minimum size of a vertex-cover and edge-cover of $\mathcal{O}$, that is,

$$\kappa(G) = \min\{|U| : U \subset V(G), \text{ and } U \cap V(C) \neq \emptyset \text{ for all } C \in \mathcal{O}(G)\},$$
$$\lambda(G) = \min\{|U| : U \subset E(G), \text{ and } U \cap E(C) \neq \emptyset \text{ for all } C \in \mathcal{O}(G)\}.$$

(If G is bipartite, we write $og=+\infty$ and $\kappa=\lambda=0$.) $\mathcal{G}_n$ is the set of all graphs of order at most n. $c_1, c_2, c_3, \ldots$ are absolute constants.

1.2. Covering odd cycles with vertices

Let $f_\kappa(g,n) = \max\{\kappa(G) : G \in \mathcal{G}_n, og(G) \geq g\}$ — the minimum number of *vertices* whose removal would make any n-graph with odd-girth g (or more) bipartite. One's first instinct is to guess that the extreme example for covering all odd cycles with vertices is the vertex-disjoint union of cycles, that is, $f_\kappa(n,g) \sim n/g$. However, a random graph argument shows that there are graphs of order n and

Mathematics Subject Classification (1991): 05C35, 05C38

odd girth as large as $c\log n$ for which more than half (in fact almost all of) the vertices are needed to cover all odd cycles. That is, $f_\kappa(n,g)>n/2$ for all $g<c\log n$. Subdividing the edges of such graphs would give lower bounds for larger values of g, and these turn out to be optimal examples within a constant factor as the following theorem shows.

Theorem 1.

$$\min\left\{c_1 n,\, c_2\frac{n}{g}\log\frac{2n}{g}\right\} \le f_\kappa(g,n) \le \min\left\{n,\, c_3\frac{n}{g}\log\frac{2n}{g}\right\}.$$

1.3. Covering odd cycles with edges

Let $f_\lambda(g,n)=\max\{\lambda(G): G\in\mathcal{G}_n,\, og(G)\ge g\}$ — the minimum number of *edges* whose removal would make any n-graph with odd-girth g (or more) bipartite. Béla Bollobás, Pál Erdős, Miklós Simonovits and Endre Szemerédi [1] established the inequalities $c_4n^2/g^2<f_\lambda(g,n)\le c_5n^2/g$, and they conjectured that the lower bound is the true order of magnitude. Here we prove the Bollobás–Erdős–Simonovits–Szemerédi conjecture.

Theorem 2.

$$f_\lambda(g,n)\le c_6n^2/g^2.$$

Remark. The lower bound in [1] ($\lfloor(n/g\rfloor^2$) was achieved by the graph $C_g(n/g)$ which is obtained from a cycle of length g by replacing each vertex with n/g vertices and replacing the edges by complete bipartite graphs.

The upper bound with $c_5=2$ is an immediate consequence of the fact that if G is a non-bipartite n-graph with $og(G)=g>3$, then $\delta(G)\le 2n/g$, which, in turn, follows from the simple fact that if C is a shortest odd cycle in such a G, then the sum of degrees of all vertices of C is at most $2n$ (in fact, $deg(x,C)\le 2$ for all $x\notin C$)— see [1] as well as the paper of Roland Häggkvist [2].

But more is proved in [1]: the use of the Regularity Lemma is used there to show that if the odd cycles of an n-graph G cannot be covered by $n^2/(2k+1)$ edges then not only does G have an odd cycle of size at most $2k+1$ but it has a whole blown-up version of such a cycle.

2. Proof of Theorem 1

2.1. Proof of the upper bound

We will actually prove the upper bound $f_\kappa(g,n)\le((4n)/(g-1))\log((4n)/(g+1))$. Let n and g be given, and write $t=(g-1)/2$ and

$$\alpha=\frac{4}{g-1}\log\frac{4n}{g+1}. \tag{1}$$

Let G be an arbitrary graph of order at most n with finite odd girth at least g. Here is an algorithm to select a set S of at most αn vertices that cover all odd cycles.

If $\alpha \geq 1$ then $S = V(G)$ (all vertices). Otherwise, pick an arbitrary vertex x, and for $k=0,1,\ldots$, let the k-th level Γ_k be defined as the set of vertices at distance k from x, and let $N_k = \cup_{0\leq i\leq k}\Gamma_i$ be the set of vertices at distance at most k from x. Let $a_k = |\Gamma_k|$, $s_k = |N_k|$.

Now, choose a k, $1\leq k\leq t$, such that

$$\frac{a_k}{s_k} \leq \alpha \tag{2}$$

and put all vertices of the level Γ_k into S. (They certainly cover all odd cycles which have at least one vertex in N_k.) Delete N_k and proceed in the leftover graph $G - N_k$ similarly (pick a new vertex x, build the neighbourhoods, etc; always use the original α defined by (1)).

If no such index k exists then stop — the current graph is already bipartite.

It remains to show that indeed if no $k \leq t$ satisfying (2) exists then the graph is bipartite. (This is sufficient, for α is an increasing function of n and a decreasing function of g, and in the leftover graph the order is smaller than, and the odd girth is at least as large as, in the original graph.) Assume that it is not, and use the following simple lemma. (The lemma is a discrete analogue of the fact that the solution to $f'/f = \alpha$, $f(0)=1$, is $f(x)=e^{\alpha x}$.)

Lemma 1. *Let t be a positive integer, let $a_0, a_1, \ldots, a_t$ be positive real numbers, and write $s_k = \sum_{0\leq i\leq k} a_i$. If for some $\alpha > 0$ the sequence a_k satisfies*

$$a_k > \alpha s_k, \quad k = 1, 2, \ldots, t,$$

then

$$s_k > s_0 e^{\alpha k}, \quad k = 1, 2, \ldots, t.$$

Now let $\ell = \lfloor (g-1)/4 \rfloor$, and apply the lemma for the numbers $a_\ell, \ldots, a_t$. We have $s_\ell \geq \ell + 1$ and hence

$$n \geq s_t > s_\ell e^{\alpha(t-\ell)} \geq (\ell+1)e^{\alpha(t-\ell)} \geq \frac{g+1}{4}\, e^{\alpha(g-1)/4} = n$$

— a contradiction. ∎

Remark. By choosing $\ell = \lceil \varepsilon g \rceil$ we get the following better bound:
For every $\varepsilon > 0$ there is a $K > 0$ such that $f_\kappa(g,n) \leq (2+\varepsilon)(n/g)\log(Kn/g)$.

Proof of Lemma 1. It easily follows by induction using the inequality $1/(1-x) > e^x$ for all $0 < x < 1$. For an alternative proof see Section 4. ∎

2.2. Proof of the lower bound

We will use the following existence theorem (proved in the Appendix by standard random graph methods).

Lemma 2. *There are positive constants γ_1, z such that for every n there is a graph H_n with the following properties:*
(a) the order of H_n is at most n,
(b) the size of H_n is between zn and $2zn$,
(c) the odd girth of H_n is greater than $\gamma_1 \log n$,
(d) $\kappa(H_n) > n/2$.

We will assume that n is large, and also that $g < \gamma_2 n$, where γ_2 is very small. (If $g > \gamma_2 n$ then the lower bound is less than 1, so it is trivially valid if c_2 is small enough.)

Let $\varepsilon > 0$ be small enough and write

$$n_0 = \left\lceil \min\left\{ \varepsilon n,\ \frac{\gamma_1}{4z}\frac{n}{g}\log\frac{2n}{g} \right\} \right\rceil .$$

We start out with a graph $G_0 = H_{n_0}$ as in Lemma 2, and write $e_0 = |E(G_0)|$, $g_0 = og(G_0)$, $\kappa_0 = \kappa(G_0)$.

Let q be the smallest even integer greater than or equal to g/g_0, and subdivide each edge of G_0 with q new vertices. The new graph G is of order $n' = n_0 + qe_0$, odd girth $g' = (1+q)g_0$, and odd girth cover $\kappa' = \kappa_0$. We need to check the following inequalities all of which easily follow from our choices of parameters:

$$n' \le n, \qquad g' \ge g, \qquad \kappa' > \min\left\{ c_1 n,\ c_2\frac{n}{g}\log\frac{2n}{g} \right\}. \qquad \blacksquare$$

Remark. Note that the levels Γ_k (from any vertex x) grow exponentially in k with an exponent $1/q$ near 0, just as suggested by Lemma 1.

3. Proof of Theorem 2

We will in fact prove the upper bound $f_\lambda(g,n) \le 8en^2/(g-1)^2 < 22n^2/(g-1)^2$.

We start with a graph G of order at most n with finite odd girth at least g, and we let $t = (g-1)/2$ and $\beta = 2e/t^2 = 8e/(g-1)^2$. Here is an algorithm that selects a set T of at most βn^2 edges that cover all odd cycles.

Pick an arbitrary vertex x, and define Γ_k, N_k, a_k, s_k just as in Section 2.1.
Choose a k, $0 \le k < t$, such that

$$\frac{a_k}{s_k}\frac{a_{k+1}}{n} < \beta \tag{3}$$

and put all edges between the levels Γ_k and Γ_{k+1} into T. (They certainly cover all odd cycles which have at least one vertex in N_k.) Then proceed in the leftover graph $G-N_k$ similarly (for n use the order of the current graph, but always use the original β so that we don't have to recompute g).

Induction shows (bipartite graphs are the base cases) that the total number of edges needed is at most $a_k a_{k+1} + \beta(n-s_k)^2 < \beta[s_k n + n^2 - s_k(2n-s_k)] \le \beta n^2$ as claimed. It remains to show that if G is not bipartite then a $k<t$ satisfying (3) always exists, which is stated in the following lemma. (Here the corresponding equation is $(f')^2/f = \beta n$, $f(0)=f'(0)=0$ with solution $f(x)=(\beta n/4)x^2$. Thus the condition $f(t)\le n$ would lead to $\beta = 4/t^2$. We get a somewhat larger constant.)

Lemma 3. *Let $t\ge 1$ be an integer, let $a_0, a_1, \dots, a_t$ be arbitrary non-negative real numbers, and write $s_k = \sum_{0\le i\le k} a_i$. Then there is a $k\in[0,t-1]$ such that*

$$\frac{a_k a_{k+1}}{s_k s_t} < \frac{2e}{t^2}.$$

This lemma will be proved in the next section.

4. Proof of Lemma 3

An alternative — and more complicated — proof for Lemma 1 would be the following. (Just as before, $a_0, a_1, \dots, a_k$ are positive reals ($k\ge 1$), and we write $s_i = a_0 + \dots + a_i$.) We use the following extremal problem.

Lemma 4. *For fixed s_0 and s_k, the choice of the a_i for which the product*

$$P = \prod_{1\le i\le k} \frac{a_i}{s_i}$$

is maximal, is the one for which the numbers s_i form a geometric series:

$$s_i = s_0 q^i,\ i=0,1,\dots,k, \quad \textit{where } q = \left(\frac{s_k}{s_0}\right)^{1/k},$$

that is,

$$a_i = a_0(q-1)q^{i-1}, \quad i=1,\dots,t.$$

Hence, if a_i satisfy the conditions of Lemma 1 then

$$\prod_{1\le i\le k} \frac{a_i}{s_i} \le \prod_{1\le i\le k} \frac{s_0(q-1)q^{i-1}}{s_0 q^i} = \left(\frac{q-1}{q}\right)^k,$$

and thus

$$\alpha < \min_{1 \le i \le k} \frac{a_i}{s_i} \le \frac{q-1}{q} = 1 - \left(\frac{s_k}{s_0}\right)^{-1/k} = 1 - e^{-\log(s_k/s_0)/k} < \frac{\log(s_k/s_0)}{k}. \quad \blacksquare$$

Proof of Lemma 4. For $k=1$ there is nothing prove, so assume $k \ge 2$. Let a_i be optimal (compactness guarantees that it exists, since the maximum of the product is certainly not taken where any a_i is 0). For given i, $1 \le i < k$, we need to show $s_i^2 = s_{i-1}s_{i+1}$. Using a very small real number x, let us replace a_i and a_{i+1} by $a_i + x$ and $a_{i+1} - x$. The only terms that change are a_i, a_{i+1}, s_i. Since the original value was maximal, we get a new product $P' \le P$. That is,

$$\frac{a_i + x}{s_i + x} \frac{a_{i+1} - x}{s_{i+1}} \le \frac{a_i}{s_i} \frac{a_{i+1}}{s_{i+1}},$$

which is equivalent to $h(x) = x^2 - (a_i a_{i+1} + a_i s_i - a_{i+1} s_i)x \ge 0$ for all small x, that is,

$$h'(0) = a_i a_{i+1} + a_i s_i - a_{i+1} s_i = 0.$$

Hence $a_{i+1} = a_i s_i/(s_i - a_i) = a_i s_i / s_{i-1}$, which implies $s_i^2 = s_{i-1}s_{i+1}$ as required. $\blacksquare$

The advantage of this method (maximizing the product P, which is then easily done by using logarithmic derivatives) is that it generalizes to the harder Lemma 3.

Proof of Lemma 3. It is an immediate consequence of the following extremal problem.

Lemma 5. *For fixed s_0 and s_t, the choice of the a_i for which the product*

$$Q = \prod_{0 \le i < t} \frac{a_i a_{i+1}}{s_i s_t}$$

is maximal, is the one for which the numbers satisfy

$$\frac{s_{i+1}}{a_{i+1}} = \frac{s_i}{a_i} + \frac{1}{2}, \quad i = 2, \dots, t-2, \quad \text{and} \quad \frac{s_t}{a_t} = 2\,\frac{s_{t-1}}{a_{t-1}},$$

whence

$$Q < \left(\frac{2e}{t^2}\right)^t.$$

Proof. We rewrite Q as

$$Q = \frac{a_t}{s_t^t} \prod_{1 \le i \le t-1} \frac{a_i^2}{s_i}.$$

Just as in the proof of Lemma 4, we use a very small real number x and replace a_i and a_{i+1} (where $2 \le t-2$) by a_i+x and $a_{i+1}-x$. Again, the only terms that change are a_i, a_{i+1}, s_i, and we get a new product $Q' \le Q$. That is,

$$\frac{(a_i+x)^2}{(s_i+x)} \frac{(a_{i+1}-x)^2}{s_{i+1}} \le \frac{a_i^2}{s_i} \frac{a_{i+1}^2}{s_{i+1}},$$

which is equivalent to $h(x) = Ax^4 + Bx^3 + Cx^2 + (a_i^2 a_{i+1}^2 - 2s_i a_i a_{i+1}^2 - 2s_i a_{i+1} a_i^2)x \ge 0$ for all small x, that is,

$$(a_i^2 a_{i+1}^2 - 2s_i a_i a_{i+1}^2 - 2s_i a_{i+1} a_i^2) = 0.$$

Hence $a_{i+1} = 2s_i a_i/(2s_i - a_i)$, which implies $s_{i+1} = s_i(2s_i + a_i)/(2s_i - a_i)$, whence $s_{i+1}/a_{i+1} = (s_i/a_i) + 1/2$ as claimed.

If we do the same for $i = t-1$, we get

$$\frac{(a_{t-1}+x)^2}{s_{t-1}+x} \frac{a_t - x}{s_t} \le \frac{a_{t-1}^2}{s_{t-1}} \frac{a_t}{s_t},$$

leading as above to $a_t = s_{t-1}a_{t-1}/(2s_{t-1} - a_{t-1})$, $s_t = 2s_{t-1}^2/(2s_{t-1} - a_{t-1})$, and finally, $s_t/a_t = 2s_{t-1}/a_{t-1}$ as claimed.

It remains to estimate this maximal Q. We got the explicit formulae

$$\frac{s_i}{a_i} = \frac{s_1}{a_1} + \frac{i-1}{2} > \frac{i+1}{2}, \quad i = 2, \ldots, t-1, \quad \text{and} \quad \frac{s_t}{a_t} = 2\frac{s_1}{a_1} + t.$$

Also,

$$\prod_{1 \le i \le t} a_i \le \left(\frac{1}{t} \sum_{1 \le i \le t} a_i \right)^t < \left(\frac{s_t}{t} \right)^t,$$

whence

$$Q < \frac{1}{t^t} \prod_{1 \le i \le t-1} \frac{2}{i+1} = \frac{1}{t^t} \frac{2^{t-1}}{t!} < \left(\frac{2e}{t^2} \right)^t. \quad \blacksquare$$

5. Appendix

Proof of Lemma 2. We will assume that n is large (otherwise the claim is vacuous). We start with a random graph $G_{n,p}$, where $p = \mu/n$ with $\mu = 3z$, and delete one vertex from each odd cycle of length less than $\gamma_1 \log n$. This is our graph H_n. We show that, with a positive probability, we have $zn < e(H_n) < 2zn$, as well as $\kappa(H_n) > n/2$, provided $\gamma_1 \log \mu < 1$ and n is large enough.

For this we show two things:

(a) D, the number of vertices dropped from $G_{n,p}$ to get H_n, is, with probability greater than 1/2, at most $D_0 := 2n^{\gamma_1 \log \mu}$

(b) Regardless how we drop $(n/2)+D_0$ vertices out of $G_{n,p}$ we still have, with probability $1-o(1)$, some odd cycles left.

Proof of (a). The expected number of cycles of length at most k is less than $\sum_{3\le i\le k} n^i p^i/(2i) < \mu^k$ if k is large in terms of z. Hence, by Markov's inequality, $P(D > 2\mu^k) < 1/2$. Apply this with $k = \gamma_1 \log n$.

Proof of (b). The probability that $G_{n,p}$ is bipartite is less than

$$\sum_i \binom{n}{i}(1-p)^{\binom{i}{2}+\binom{n-i}{2}} < 2^n e^{-n\mu/5}.$$

Hence the probability that dropping at most $0.6n$ vertices would make $G_{n,p}$ bipartite is less than $2^n 2^{0.4n} e^{-0.4n\mu/5} = o(1)$ e.g. for $\mu = 15$ (that is, $z = 5$). ∎

Remark. Miklós Simonovits remarked that Lemma 2 would also follow from some algebraic constructions like the one of Lubotzky, Phillips and Sarnak [3] (need to use Proposition 5.2 in [3] proved by Noga Alon and also the inequality $\kappa \ge n - 2\alpha$). I appreciate his thorough refereeing and helpful remarks.

References

[1] B. BOLLOBÁS, P. ERDŐS, M. SIMONOVITS, and E. SZEMERÉDI: Extremal graphs without large forbidden subgraphs, Advances in Graph Theory, B. Bollobás ed., *Annals of Discrete Mathematics*, **3** (1978), 29–41, also appeared as a problem in B. BOLLOBÁS: *Extremal Graph Theory*, Academic Press 1978, p. 367 Problem 47.

[2] R. HÄGGKVIST: Odd cycles of specified length in non-bipartite graphs, *Graph Theory*, B. Bollobás ed., *Annals of Discrete Mathematics*, **13** (1982), 89–100.

[3] A. LUBOTZKY, R. PHILLIPS, P. SARNAK: Ramanujan graphs, *Combinatorica*, **8** (1988), 261–277.

János Komlós

Department of Mathematics
Rutgers University
New Brunswick, NJ 08903
U.S.A.
komlos@math.rutgers.edu