COMBINATORICA
Akadémiai Kiadó – Springer-Verlag

COMBINATORICA 13 (1) (1993) 1–5

CLIQUE COVERINGS OF THE EDGES OF A RANDOM GRAPH

BÉLA BOLLOBÁS, PAUL ERDŐS, JOEL SPENCER, and DOUGLAS B. WEST[1]

Received November 29, 1989

The edges of the random graph (with the edge probability $p = 1/2$) can be covered using $O(n^2 \ln\ln n/(\ln n)^2)$ cliques. Hence this is an upper bound on the intersection number (also called clique cover number) of the random graph. A lower bound, obtained by counting arguments, is $(1-\varepsilon)n^2/(2\lg n)^2$.

1. Introduction

We consider several parameters related to clique coverings of graphs. The number of cliques required to cover the *vertices* of a graph G equals the chromatic number of its complement $\bar{G}$; for this reason the usual definition of the *clique cover number* $\theta_1(G)$ is the number of cliques required to cover the *edges* of G. Erdős, Goodman, and Pósa [2] proved that the clique cover number of G always equals its *intersection number*, which is the minimum k such that G is the intersection graph of a collection of subsets of a k-element set. (The *intersection graph* of a family of sets is the graph formed by creating vertices corresponding to the sets and placing an edge between two vertices if and only if the corresponding sets intersect.) They also showed that the maximum of $\theta_1(G)$ for an n-vertex graph is $\lfloor n^2/4 \rfloor$.

For a random graph generated by independent edge probability $1/2$, studying the clique coverings of the vertices is equivalent to studying vertex colorings. Using the technique of martingales, Bollobás [1] determined the asymptotic value of the chromatic number for almost all graphs: $n/(2\lg n)$ (lg denotes $\log_2$). In this note, we obtain bounds for the number of cliques needed to cover the edges.

An easy lower bound follows from the fact that almost every graph has at least $(1-\varepsilon)n^2/4$ edges and has fewer than $2\lg n$ vertices in every clique [6]; this implies that $\theta_1(G) \geq (1-\varepsilon)n^2/[8(\lg n)^2]$ for almost every graph. By a more careful

AMS subject classification code (1991): 05 C 80

[1] Research supported in part by ONR Grant N00014–85K0570 and by NSA/MSP Grant MDA904–90–H–4011.

examination of intersection representations, we will improve this bound by a factor of 2.

An upper bound of $O(n^2 \ln\ln n/(\ln n)^2)$ follows from a related clique covering problem we considered at the First China-USA International Conference in Graph Theory, in Jinan, China (1986). In this problem, we seek the minimum number of cliques needed to cover almost all the edges of the random graph with no vertex used very often. The main result of this note is that a subquadratic number of cliques almost always suffices. Specifically, we prove that almost every graph has $O(n^2 \ln\ln n/(\ln n)^2)$ cliques that cover all but $O(n^2/(\ln n)^{3(1-\varepsilon)})$ of its edges in such a way that each vertex appears in at most $O(n \ln\ln n/\ln n)$ of the cliques. Since the remaining edges can be added individually as cliques, we obtain an upper bound on the intersection number: $\theta_1(G) \le O(n^2/(\ln n)^{3(1-\varepsilon)}) + O(n^2 \ln\ln n/(\ln n)^2) = O(n^2 \ln\ln n/(\ln n)^2)$ for almost every graph.

Our original motivation for this problem was to improve the upper bound then known for the interval number of a random graph. Letting a *t-interval* be a union of t intervals in R, the *interval number* $i(G)$ of a graph G is the minimum t such that G can be expressed as the intersection graph of a family of t-intervals. This application is the source of the requirement above that no vertex appear in many cliques. Counting the possible representations of n-vertex graphs by t-intervals (we may assume the endpoints are the integers $\{1, \ldots, 2n\}$) yields $i(G) \ge (1-\varepsilon)n/(4 \lg n)$ for almost every graph [3].

The relationship between our problem and $i(G)$ is as follows. It is known that $i(G) \le \lceil (n+1)/4 \rceil$ for every n-vertex graph [4] and that $i(G) \le \lceil 1 + \sqrt{e/2} \rceil$ for every e-edge graph [8]. If in almost every graph we could find cliques covering all but $O(n^2/(\log n)^2)$ edges and using each vertex at most $O(n/\log n)$ times, we would have $i(G) \le O(n/\log n)$ for almost every graph. If the number of cliques in this "almost covering" were $O(n^2/(\log n)^2)$, then we would have bounded $\theta_1(G)$ between two constant multiples of $(n/\log n)^2$.

Subsequently, Scheinerman [7] established the desired upper bound on $i(G)$ by another method. He proved $i(G) \le (n+\varepsilon)/(2 \lg n)$ by using the lemma of Bollobás [1] (see also [7]) that in almost every graph every set of at least $n/(\log n)^2$ vertices contains a clique of size at least $2 \lg n - 6 \lg\lg n$. He also raised the easy lower bound on $i(G)$ to $(n-\varepsilon)/(2 \lg n)$ by more carefully counting intersection representations for almost all graphs, much like the method discussed in the next section. In light of this success for $i(G)$, the "loglog" factors in our result and its application seem to be excessive. Indeed, we conjecture that our lower bound $\theta_1(G) \ge (1-o(1))\,(n/(2\lg n))^2)$ is optimal.

2. Lower Bound on Intersection Number

We will use the notation of asymptotics, writing $f(n) \sim g(n)$ when $\lim_{n\to\infty} f(n)/g(n) = 1$, i.e. when $f(n) - g(n) \in o(g(n))$.

Let K_t denote a labeled clique on t vertices, and let $[k] = \{1, \ldots, k\}$. If we form graphs on n vertices by assigning each vertex v a set $X_v \subseteq [k]$ and taking the intersection graph, there are 2^{nk} possible representations of labeled graphs on n

vertices. However, almost all the resulting intersection graphs equal K_n. That is, if we generate random representations by assigning each $j \in [k]$ to X_v with probability $1/2$, then the probability that edge uv does not appear in the intersection graph is $(3/4)^k$. With $k > (1-\varepsilon)n/2$, the expected number of missing edges goes rapidly to zero, so almost every representation yields K_n. However, very few graphs have large cliques. As mentioned earlier, almost every graph has maximum clique size $(2+o(1))\lg n$ [6]. If we restrict our attention to such graphs, we keep almost all the graphs and lose almost all the representations, so we will need a larger value of k to have enough representations.

Theorem 1. *$\theta_1(G) \geq (1-\varepsilon)n^2(2\lg n)^2$ for almost every graph G.*

Proof. Let $\mathbf{G}_n$ denote the set of graphs on n vertices having no clique with more than r vertices, where $r = (2+\varepsilon)\lg n$. Asymptotically, $|\mathbf{G}_n| \sim 2^{n^2/2}$. Any intersection representation of $G \in \mathbf{G}_n$ by subsets of $[k]$ uses each element of $[k]$ in at most r sets. Hence we consider only representations obtained by choosing for each element of $[k]$ a collection of at most r vertices to which it is assigned. The number of ways to specify such a representation is asymptotic to $\binom{n}{r}^k < (en/r)^{rk}$. Taking logarithms, we must have $n^2/2 \leq kr(\lg n - \lg r + o(\lg\lg n))$, else almost all graphs in $\mathbf{G}_n$ go unrepresented. Hence we require $k \geq (1-\varepsilon)n^2(2\lg n)^2$ to represent almost all graphs. (Note: we can divide $\binom{n}{r}^k$ by $k!$ to bound the asymptotic number of distinct labeled graphs represented, because the order of the sets does not affect the resulting graph, but this does not change the asymptotic behavior of the lower bound.) ∎

3. A Subquadratic Clique Covering

The random graph has a quadratic number of edges. To cover a quadratic number of edges with a subquadratic number of cliques, we must use a large number of large cliques. The largest clique has about $2\lg n$ vertices, but there won't be many of that size, so we use slightly smaller cliques.

Theorem 2. *Almost every graph has a collection of $O(n^2 \ln\ln n/(\ln n)^2)$ cliques that cover all but $O(n^2/(\ln n)^{3(1-\varepsilon)})$ edges and include each vertex at most $O(n\ln\ln n/\ln n)$ times.*

Proof. The idea of the proof is quite simple. We choose "large" (not quite maximum-sized) cliques at random in the random graph, choosing them independently with probability designed so that the number of cliques chosen will almost always be about the desired number. We then show that it almost always holds that almost all edges are covered by these cliques.

Let $t = c\lg n$, where $c < 2$, and let X be a random variable counting t-cliques in the random graph. Since there are $\binom{n}{t}$ possible t-cliques, we have $E(X) = \binom{n}{t}2^{-t(t-1)/2}$. When $c < 2$, we have $E(X) \to \infty$ and $E(X^2) \sim E(X)^2$, and then Tchebysheff's Inequality implies that almost every graph has $(1+o(1))\binom{n}{t}2^{-t(t-1)/2}$ t-cliques. Let $s = E(X) = \binom{n}{t}2^{-t(t-1)/2}$.

Given a fixed edge known to appear in a random graph, let Y be the number of t-cliques that contain it. We have $E(Y) = \binom{n-2}{t-2}2^{-t(t-1)/2+1} = 2s\binom{t}{2}/\binom{n}{2} \sim 2st^2/n^2$. Let $r = E(Y)$. For any fixed $\varepsilon > 0$, we say that an edge is *regular* if it appears in at least $(1-\varepsilon)r$ t-cliques. As above, it is easy to show $E(Y) \to \infty$ and $E(Y^2) \sim E(Y)^2$. Tchebysheff's Inequality then implies that $\mathrm{Prob}(|Y - r| \geq \varepsilon r) \leq o(1)$, and that therefore the expected number of irregular edges is $o(n^2)$. Indeed, almost every graph has $o(n^2)$ irregular edges. To obtain the desired result, we must show that in fact almost every graph has $O(n^2/(\ln n)^{3(1-\varepsilon)})$ irregular edges. We postpone this computation, which is the most difficult of the proof.

Let $q = x/r$, where $x = 3\ln\ln n$. Choose a set of t-cliques in the graph by choosing independently with probability q whether to include each t-clique that occurs in the graph. Since there are $(1+o(1))s$ such cliques, these Bernoulli trials produce a collection of $(1+o(1))qs = (1+o(1))n^2x/(2t^2) = c'n^2\ln\ln n/(\ln n)^2$ t-cliques. Furthermore, it is almost always true that each vertex appears in at most $c'c(1+\varepsilon)n\ln\ln n/\ln n$ of the cliques.

We claim that these cliques cover almost all the edges. For any regular edge, the probability that it is not covered is bounded by $(1-q)^{(1-\varepsilon)r} < e^{-q(1-\varepsilon)r} = (\ln n)^{-3(1-\varepsilon)}$. The expected number of regular edges missed is bounded by $n^2/(\ln n)^{3(1-\varepsilon)}$. Another use of Tchebysheff's Inequality implies that in almost every graph, the number of regular edges missed is $O(n^2/(\ln n)^{3(1-\varepsilon)})$. If in addition there are almost always at most $O(n^2/(\ln n)^{3(1-\varepsilon)})$ irregular edges, which may also be missed, then the total number of edges missed by these cliques is $O(n^2/(\ln n)^{3(1-\varepsilon)})$, which yields the desired covering.

To bound the number of irregular edges, we need a more precise computation of $E(Y)$ and $E(Y^2)$. We have $Y = \Sigma Y_i$, where the Y_i's are 0,1-variables indicating the presence of each of the $\binom{n-2}{t-2}$ t-cliques that can contain the specified edge, and $E(Y^2) = \Sigma_{i,j}E(Y_iY_j)$. Here the summand equals the probability that the ith and jth cliques are both present, which is $2^{-t(t-1)}2^{\binom{l+2}{2}+1}$, where l is the number of vertices the two cliques share outside the specified edge. For fixed l, the number of choices for the pair i,j is $\binom{n-2}{t-2}\binom{t-2}{l}\binom{n-t}{t-2-l}$. Grouping the sum by the value of l and writing $(m)_l$ for $m(m-1)\ldots(m-l+1)$, we have

$$E(Y^2) = \sum_{l=0}^{t-2}\binom{n-2}{t-2}\binom{t-2}{l}\binom{n-t}{t-2-l}2^{-t(t-1)}2^{\binom{l+2}{2}+1} =$$

$$= E(Y)^2\frac{\binom{n-t}{t-2}}{\binom{n-2}{t-2}}\sum_{l=0}^{t-2}\frac{(t-2)_l^2}{l!(n-2t+l+2)_l}2^{\binom{l+2}{2}-1}.$$

We must be somewhat careful in bounding the summation because of the exponent on 2, which causes difficulty when l is large and the cliques overlap a lot. Since $\binom{l+2}{2} - 1 = l(l+3)/2$, the summand is bounded by $t^{2l}(2^{l/2})^l(2^{3/2})^l/(n-2t)^l$. Furthermore, $2^{l/2} < 2^{(c/2)\lg n} = n^{c/2}$. Since $c < 2$, for sufficiently large n any tail of the summation is bounded by a geometric series with ratio less than a negative power of n. For $l<k$ we need a better bound and use the geometric series only for $l \geq k$ (where k is defined so that $c<2-4/k$). Combining this with the fact that $\binom{n-t}{t-2}/\binom{n-2}{t-2} \to e^{-(t-2)^2/(n-2)}$ when $t=c\lg n$, we have

$$\begin{aligned} E(Y^2) &= E(Y)^2\left[1-\frac{(t-2)^2}{n-2}+O\left(\frac{(\ln n)^4}{n^2}\right)\right]\left[1+\frac{4(t-2)^2}{n-2t+3}+O\left(\frac{(\ln n)^4}{n^2}\right)\right] \\ &= E(Y)^2\left[1+\frac{3(c\ln n)^2}{n}+O\left(\frac{(\ln n)^4}{n^2}\right)\right] \end{aligned}$$

Let Z be the number of irregular edges, so that $E(Z) = \frac{1}{2}\binom{n}{2}\cdot \mathrm{Prob}(E(Y)-Y > \varepsilon r)$. By Tchebysheff's Inequality, this probability is bounded by $[E(Y^2)-E(Y)^2]/(\varepsilon r)^2 = [3(c\ln n)^2/(\varepsilon^2 n)](1+O((\ln n)^2/n))$. Hence we conclude that $E(Z) \sim 3n(c\ln n)^2/(2\varepsilon^2) = O(n(\ln n)^2)$. This is actually quite small, and it is easy to show that the number of irregular edges is almost always within the desired bound. ∎

References

[1] B. Bollobás: The chromatic number of the random graph, *Combinatorica* **8** (1988).

[2] P. Erdős, A. Goodman, and L. Pósa: The representation of a graph by set intersections. *Canad. J. Math.* **18** (1966), 106–112.

[3] P. Erdős and D.B. West: A note on the interval number of a graph. *Disc. Math.* **55** (1985), 129–133.

[4] J.R. Griggs: Extremal values of the interval number of a graph, II. *Disc. Math.* **28** (1979), 37–47.

[5] J.R. Griggs and D.B. West: Extremal values of the interval number of a graph, I. *SIAM J. Alg. Disc. Meth.* **1** (1980), 1–7.

[6] D.W. Matula: The largest clique size in a random graph, Tech. Rep. Dept. Comp. Sci. Southern Methodist Univ., Dallas [1976].

[7] E.R. Scheinerman: On the interval number of random graphs, *Discrete Math.* **82** (1990), 105–109.

[8] J. Spinrad, G. Vijayan, and D.B. West: An improved edge bound on the interval number of a graph. *J. Graph Th.* **11** (1987), 447–449.

Béla Bollobás

Louisiana State University
Baton Rouge LA 70803, and
Cambridge Univ. England

Paul Erdős

Hungarian Academy of Sciences,
Budapest, Hungary

Joel Spencer

Courant Institute,
New York, New York
`spencer@cs.nyu.edu`

Douglas B. West

University of Illinois,
Urbana, Illinois 61801
`west@math.uiuc.edu`