Optimal and Random Partitions of Random Graphs

JOSEPH L. GANLEY* AND LENWOOD S. HEATH†

*Department of Computer Science, University of Virginia, Charlottesville, VA 22903, USA

†Department of Computer Science, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061-0106, USA

The behavior of random graphs with respect to graph partitioning is considered. It is shown that, for a random graph with *n* vertices and with expected degree exceeding a constant times ln *n*, the graph cannot be partitioned well, i.e. a random partition is likely to be almost as good as an optimal partition.

Received April 1994; revised July 1994

Graph algorithms are often tested using random graphs as input (see, e.g. Ganley and Heath, 1994). In the simplest and most common model (Bollobás, 1985), $R_{n,p}$ denotes the class of random graphs with n labeled vertices in which each edge is present with probability p. A natural question is whether such random graphs are useful for testing algorithms for various graph optimization problems. Turner (1986) considers the problem of minimizing the *bandwidth* of a linear ordering a graph, which is the length of the longest edge under that ordering. He shows (Turner, 1986, theorem 2.2) that the bandwidth of a random graph from $R_{n,p}$ almost certainly exceeds $(1-\epsilon)n$ for any constant $\epsilon > 0$; that is, with respect to bandwidth minimization, a *random ordering* of a large random graph is almost as good as an *optimal ordering*. Thus, random graphs are not useful in testing bandwidth-minimization algorithms. McDiarmid and Miller (1991) show analogous results for an extension of bandwidth to multi-dimensional lattices. In addition, a plethora of results exist regarding the behavior of random graphs with respect to colorability, expected clique sizes, and many other combinatorial features. These results can be used similarly to determine the applicability of random graphs for testing algorithms for the respective problems. The reader is referred to Bollobás (1985) for details on these results.

In this paper, we prove that a similar phenomenon occurs with respect to partitioning a random graph. The problem of partitioning a graph consists of dividing the vertices of the graph into subsets of cardinality not exceeding some bound K, such that the number of edges whose endpoints lie in the different subsets is minimized. More precisely, the problem of *graph partitioning* is the following:

Given an undirected graph $G = (V, E)$ and a set size K, find a partition of V into disjoint subsets $V_1, V_2, , V_m$ such that $|\{(u, v) : (u, v) \in E,\ u \in V_i,\ v \in V_j,\ i \neq j\}|$ is minimized, subject to $|V_i| \leqslant K$ for all i, $1 \leqslant i \leqslant m$.

The corresponding decision problem is NP-complete (see Garey and Johnson, 1979, p. 209). In this paper, we prove that a random graph cannot be partitioned well, in the sense that a random partition is almost as good as an optimal partition. A similar result with respect to graph bisection appears in Bui (1986).

For simplicity, assume that the number of vertices in the graph is an integer multiple of K. Suppose $G = (V, E)$ is a random graph from $R_{n,p}$. Let $\Pi = \{V_1, V_2, \ldots, V_{n/K}\}$ be a random partition of V in which every subset V_i contains exactly K vertices. An edge whose endpoints lie in different subsets is *external*, Let $\phi(\Pi)$ be the number of external edges in the partition Π. Let $\phi(G)$ be the number of external edges in an optimal partition of G.

The number of non-external edges (both of whose endpoints lie in the same subset) in Π falls between 0 and $n/K(K/2)$. The number of external edges falls between 0 and

$$\begin{aligned} N &= \binom{n}{2} - \frac{n}{K}\binom{K}{2} \\ &= n(n-K)/2. \end{aligned}$$

The expected number of external edges in a random partition Π is

$$E[\phi(\Pi)] = pN.$$

We proceed to derive conditions under which a random partition is almost as good as an optimal partition of G (Corollaries 3 and 4). The following lemma bounds the probabiilty that $\phi(\Pi)$ is much below its expected value.

LEMMA 1 Let G be a random graph from $R_{n,p}$. Constrain partitions of G to n/K subsets. Let $\zeta = \zeta(n)$ be a real-valued function of n such that $\zeta(n) < 1$. Then

$$\ln \Pr[\phi(G) \leqslant (1-\zeta)pN] \leqslant n \ln n - n \ln K - \frac{\zeta^2 pN}{4(1-p)} + \frac{\ln n}{2} - \frac{n \ln K}{2K} + O(1).$$

Proof Let Π be a random partition of G into n/K subsets. The number of external edges $\phi(\Pi)$ follows a binomial distribution with parameters N and p, i.e.

$$\Pr[\phi(\Pi) = k] = \binom{N}{k} p^k (1-p)^{N-k}.$$

The probability of interest is a tail of this distribution, to wit,

$$\Pr[\phi(\Pi) \leqslant (1-\zeta)pN] = \sum_{i=0}^{(1-\zeta)pN} \binom{N}{i} p^i (1-p)^{N-i}.$$

An upper bound on this probability can be obtained from Hoeffding's inequality (see Cormen *et al.*, 1990, p. 126), which asserts

$$\Pr[E[\phi(\Pi)] - \phi(\Pi) \geqslant r] \leqslant e^{-r^2/4Np(1-p)}$$

for arbitrary r. Substituting $E[\phi(p)] = pN$ and selecting $r = \zeta pN$, we have

$$\Pr[\phi(\Pi) \leqslant (1-\zeta)pN] \leqslant e^{-\zeta^2 pN/4(1-p)}.$$

There are $n!/(K!)^{n/K}$ possible partitions Π of G into n/K subsets of size K. For the optimal layout of G to have at most $(1-\zeta)pN$ external edges, at least one of these $n!/(K!)^{n/K}$ partitions must have $(1-\zeta)pN$ or fewer external edges. Hence,

$$\Pr[\phi(G) \leqslant (1-\zeta)pN] \leqslant \frac{n! e^{-\zeta^2 pN/4(1-p)}}{(K!)^{n/K}}.$$

Taking natural logarithms, we obtain

$$\ln \Pr[\phi(G) \leqslant (1-\zeta)pN] \leqslant \ln n! - \frac{n}{K} \ln K! - \frac{\zeta^2 pN}{4(1-p)}.$$

Applying Stirling's approximation for the factorial function (Graham *et al.*, 1989, p. 467), we have

$$\begin{aligned}\ln \Pr[\phi(G) \leqslant (1-\zeta)pN] &\leqslant n \ln n - n + \frac{\ln n}{2} \\ &\quad - \frac{n}{K}\left(K \ln K - K + \frac{\ln K}{2} + O(1)\right) \\ &\quad - \frac{\zeta^2 pN}{4(1-p)} + O(1). \\ &\leqslant n \ln n - n \ln K - \frac{\zeta^2 pN}{4(1-p)} + \frac{\ln n}{2} \\ &\quad - \frac{n \ln K}{2K} + O(1),\end{aligned}$$

since the $O(1)$ term in the approximation is positive.* The lemma follows. □

Our central result is Theorem 2, which identifies conditions under which $\Pr[\phi(G) \leqslant (1-\zeta)pN]$ is asymptotically 0.

*We use the standard asymptotic notation $O(f(n))$ to denote a function that is bounded above by a positive constant factor times $f(n)$ for n sufficiently large. Similarly, $\Omega(f(n))$ denotes a function that is bounded below by a positive constant factor times $f(n)$ for n sufficiently large. The notation $g(n) = \omega(f(n))$ means $\lim_{n\to\infty} g(n)/f(n) = \infty$.

THEOREM 2 Let G be a random graph from $R_{n,p}$. Constrain partitions of G to n/K subsets, where $K \geqslant 2$. Let $\zeta = \zeta(n)$ be a real-valued function of n such that $\zeta(n) < 1$. Suppose

$$1 > p > \frac{8(\ln n - \ln K) + c}{\zeta^2 n},$$

for some positive constant c and n sufficiently large. Then

$$\Pr[\phi(G) \leqslant (1-\zeta)pN] \to 0$$

as $n \to \infty$.

Proof By Lemma 1,

$$\ln \Pr[\phi(G) \leqslant (1-\zeta)pN] \leqslant n \ln n - n \ln K - \frac{\zeta^2 pN}{4(1-p)} + \frac{\ln n}{2} - \frac{n \ln K}{2K} + O(1).$$

Using the fact that $N \geqslant n^2/4$ and the bounds on p, we obtain

$$\begin{aligned}\ln \Pr[\phi(G) \leqslant (1-\zeta)pN] &\leqslant n \ln n - n \ln K \\ &\quad - \frac{\zeta^2}{4(1-p)}\left(\frac{16(\ln n - \ln K) + c}{\zeta^2 n}\right)\frac{n^2}{2} \\ &\quad + \frac{\ln n}{2} - \frac{n \ln K}{2K} + O(1) \\ &\leqslant n \ln n - n \ln K - n(\ln n - \ln K) \\ &\quad - \frac{cn}{8} + \frac{\ln n}{2} - \frac{n \ln K}{2K} + O(1) \\ &= -\frac{cn}{8} + \frac{\ln n}{2} - \frac{n \ln K}{2K} + O(1),\end{aligned}$$

which approaches $-\infty$ as $n \to \infty$. Hence

$$\Pr[\phi(G) \leqslant (1-\zeta)pN] \to 0$$

as $n \to \infty$, as desired. □

As a corollary, we obtain this special case.

COROLLARY 3 Let G be a random graph from $R_{n,p}$ where $1 > p > 0$. Constrain partitions of G to n/K subsets, where $K \geqslant 2$. Suppose that ϵ is a positive constant and that the expected degree of G is $\omega(\ln n)$. Then

$$\Pr[\phi(G) \leqslant (1-\epsilon)pN] \to 0$$

as $n \to \infty$.

Proof The expected degree of G is $p(n-1) = \omega(\ln n)$. Hence we have that $p = \omega(\ln n/n)$. When n is sufficiently large, we have

$$p > \frac{8(\ln n - \ln K) + c}{\epsilon^2 n},$$

for any positive constants c and ϵ. Choosing $\zeta = \epsilon$ in Theorem 2, we conclude

$$\Pr[\phi(G) \leqslant (1-\epsilon)pN] \to 0$$

as $n \to \infty$, as desired. □

In other words, if the degree of a random graph is slightly greater than logarithmic, then the number of external edges in an optimal partition is asymptotically almost the same as the number of external edges in a random partition.

If we want the degree to be *exactly* logarithmic, we obtain a slightly weaker corollary.

COROLLARY 4 Let G be a random graph from $R_{n,p}$, where $p = \tau \ln n/n$ for some constant τ. Constrain partitions of G to n/K subsets, where $K \geqslant 2$. Suppose that ϵ is a positive constant. If $\tau > 8/\epsilon^2$, then

$$\Pr[\phi(G) \leqslant (1-\epsilon)pN] \to 0$$

as $n \to \infty$.

In particular, if $\tau > 32$, then the probability that an optimal partition of G has fewer than half the expected number of external edges in a random partition is asymptotically 0.

From Corollaries 3 and 4 we conclude that random graphs, in the standard model we study here, are not useful for testing heuristics for the graph partitioning problem. Experimental evidence supporting this conclusion is presented in Ganley and Heath (1994). In that work, some alternate random graph models based on geometry *do* prove useful in testing heuristics.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the support of National Science Foundation grant IRI-9116991.

REFERENCES

Bollobás, B. (1985) *Random Graphs*. Academic Press, Orlando, FL.

Bui, T. N. (1986) *Graph Bisection Algorithms*, PhD thesis, MIT.

Cormen, T. H., Leiserson, C. E. and Rivest, R. L. (1990) *Introduction to Algorithms*. MIT Press, Cambridge, MA.

Ganley, J. L. and Heath, L. S. (1994) Heuristics for laying out information graphs, *Computing*, **52**, 389–405.

Garey, M. R. and Johnson, D. S. (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman, San Francisco, CA.

Graham, R.L., Knuth, D. E. and Patashnik, O. (1989) *Concrete Mathematics*. Addison-Wesley, Reading, MA.

McDiarmid, C. and Miller, Z. (1991) Lattice bandwidth of random graphs. *Discrete Appl. Math.*, **30**, 221–227.

Turner, J. S. (1986) On the probable performance of heuristics for bandwidth minimization. *SIAM J. Comput.*, **15**, 561–580.