Covering the edges of a random graph by cliques

Alan Frieze^{*} Bruce Reed[†]

July 8, 2021

Introduction

The clique cover number $\theta_1(G)$ of a graph G is the minimum number of cliques required to cover the edges of graph G. In this paper we consider $\theta_1(G_{n,p})$, for p constant. (Recall that in the random graph $G_{n,p}$, each of the $\binom{n}{2}$ edges occurs independently with probability p). Bollobás, Erdős, Spencer and West [1] proved that **whp** (i.e. with probability 1-o(1) as $n \to \infty$)

$$\frac{(1-o(1))n^2}{4(\log_2 n)^2} \le \theta_1(G_{n,.5}) \le \frac{cn^2 \ln \ln n}{(\ln n)^2}$$

They implicitly conjecture that the $\ln \ln n$ factor in the upper bound is unnecessary and in this paper we prove

Theorem 1. There exist constants $c_i = c_i(p) > 0, i = 1, 2$ such that whp

$$\frac{c_1 n^2}{(\ln n)^2} \le \theta_1(G_{n,p}) \le \frac{c_2 n^2}{(\ln n)^2}.$$

Remark 1: a simple use of a martingale tail inequality shows that θ_1 is close to its mean with very high probability.

^{*}Department of Mathematics, Carnegie Mellon University. Supported in part by NSF grants CCR9024935 and CCR9225008.

[†]Department of Mathematics, Carnegie Mellon University and Equipe Combinatoire, CNRS, Université Pierre et Marie Curie, Paris

2 Proof of Theorem 1

We write $a_n \approx b_n$ if $a_n/b_n \to 1$ as $n \to \infty$.

The lower bound is simple as the number of edges m of $G_{n,p}$ whp satisfies

$$m \approx \frac{np^2}{2}$$

and the size of the largest clique $\omega = \omega(G_{n,p})$ whp satisfies

$$\omega \approx 2 \log_b n$$

where b = 1/p. We may thus choose $c_1 \approx (\ln b)^2 p/2$.

The upper bound requires more work. Our method does not seem to yield the correct value for c_2 and so we will not work hard to keep c_2 small. Let α be some small constant and let

$$k = \lfloor \alpha \log_b n \rfloor.$$

We consider an algorithm for randomly selecting cliques to cover the edges of $G = G_{n,p}$. It bears some relation to part of the algorithm described in Pippenger and Spencer [2]. At iteration *i* we randomly select cliques of size $k_i = \lfloor k/i \rfloor$ none of whose edges are covered by previously chosen cliques. Our idea is to choose these cliques so that at the start of iteration *i* the graph G_i formed by the set E_i of edges which have not been covered behaves, for our purposes, similarly to G_{n,p_i} , $p_i = pe^{1-i}$. That is it will contain about $m_i = \binom{n}{2}p_i$ edges, it will have about $N_i = \binom{n}{k_i}p_i^{\binom{k_i}{2}}$ cliques of size k_i and the intersection of these cliques will be similar to that for the k_i -cliques in G_{n,p_i} . In particular, in both G_{n,p_i} and G_i almost all of the edges are in about $\zeta_i = N_i \binom{k_i}{2}/m_i k_i$ -cliques.

Now in iteration *i* we choose a set C_i of k_i -cliques from G_i to add to our cover. The available cliques are chosen independently with probability about $1/\zeta_i$. By our assumptions on G_i , an edge is left uncovered with probability about e^{-1} . With a bit of care we can show that our assumptions continue to hold for G_{i+1} as well.

We do this for $i_0 = \lceil 4 \ln \ln n \rceil$ iterations. After this there are about $\binom{n}{2} pe(\ln n)^{-4}$ uncovered edges and we can add these as cliques of size two to the cover. In iteration *i* we choose about

 $m_i/{\binom{k_i}{2}} \approx n^2 i^2 p e^{1-i} k^{-2}$ cliques and so the total number of cliques used is $O(n^2/(\ln n)^2)$ as required.

We now need to describe our clique choosing process a little more formally: let $C_{j,i}$ denote the set of *j*-cliques all of whose edges are in E_i . If

$$c_{s,j,i} = {\binom{n-s}{j-s}} (be^i)^{\binom{s}{2} - \binom{j}{2}},$$

then $c_{s,j,i}$ is close to the expected number of cliques in $\mathcal{C}_{j,i}$ which contain a particular fixed clique in $\mathcal{C}_{s,i}$.

For a clique $S \in \mathcal{C}_{s,i}$ we let

$$X_{S,j,i} = |\{C \in \mathcal{C}_{j,i} : C \supseteq S\}|$$

and for integer $s \ge 0$,

$$X_{s,j,i}^* = \max\{X_{S,j,i} : S \in \mathcal{C}_{s,i}\}.$$

Algorithm COVER

begin

$$E_1 := E(G_{n,p}); \ \mathcal{C}_{COVER} := \emptyset;$$

for i = 1 to i_0 do

begin

A: independently place each $C \in \mathcal{C}_{\lfloor k/i \rfloor, i}$ into \mathcal{C}_{COVER} with probability

$$X_{2,\lfloor k/i\rfloor,i}^*^{-1};$$

B: for each $u \in E_i$ which is not covered by a clique in Step A, add u

(as a clique of size 2) to \mathcal{C}_{COVER} with probability ρ_u where

$$e^{-1} - X_2^{*-1} = \left(1 - \frac{1}{X_2^*}\right)^{X_u} (1 - \rho_u),$$

 $X_2^* = X_{2,\lfloor k/i \rfloor,i}^*$ and $X_u = X_{u,\lfloor k/i \rfloor,i}$.

end

$$\mathcal{C}_{COVER} := \mathcal{C}_{COVER} \cup E_{i_0+1}.$$

end

Observe first that the definition of ρ_u assumes that X_2^* is large (which it is **whp**) and so

$$\left(1 - \frac{1}{X_2^*} \right)^{X_u} \geq \left(1 - \frac{1}{X_2^*} \right)^{X_2^*} \\ \geq e^{-1} - X_2^{*-1},$$

and ρ_u is properly defined.

The following lemma contains the main core of the proof:

Lemma 1. Let \mathcal{E}_i refer to the following two conditions:

(a)

$$X_{S,j,i} \le (1+\beta_i)c_{s,j,i}, \qquad 0 \le s \le j \le k/i \text{ and } S \in \mathcal{C}_{s,i},$$

where $\beta_i = i n^{-1/4}$,

(b)

$$X_{u,j,i} \ge (1 - \gamma_i)c_{2,j,i}, \qquad e \in E_i \text{ and } 2 \le j \le k/i$$

for all but at most $in^{31/16}$ edges, where $\gamma_i = in^{-1/16}$.

Then

$$\mathbf{Pr}(\mathcal{E}_1) = 1 - o(n^{-1}), \tag{1}$$

$$\mathbf{Pr}(\mathcal{E}_{i+1} \mid \mathcal{E}_i) \geq 1 - O(n^{-1/16} \log n).$$
(2)

We defer the proof of the lemma to the next section and show how to use it to prove Thereom 1. Observe first that

$$\frac{c_{s+1,j,i}}{c_{s,j,i}} = \left(\frac{j-s}{n-s}\right) (be^i)^s,\tag{3}$$

and

$$c_{s,j,i} \ge n^{7/8} \tag{4}$$

when α is small and $0 \le s < j \le k/i$.

Next let Y_i and Z_i denote the number of $\lfloor k/i \rfloor$ -cliques and edges respectively added to C_{COVER} in iteration *i*.

$$\mathbf{E}(Y_i \mid \mathcal{E}_i) \leq \mathbf{E}\left(\frac{X_{0,\lfloor k/i \rfloor,i}^*}{X_{2,\lfloor k/i \rfloor,i}^*} \middle| \mathcal{E}_i\right) \\
\leq (1+o(1)) \frac{c_{0,\lfloor k/i \rfloor,i}}{c_{2,\lfloor k/i \rfloor,i}} \\
\approx \frac{n^2 i^2}{bk^2 e^i},$$
(5)

on using (3)

Since Y_i is binomially distributed, we see using standard bounds on the tails of the binomial, that

$$\mathbf{Pr}\left(Y_i \ge \frac{2n^2 i^2}{bk^2 e^i} \middle| \mathcal{E}_i\right) \le n^{-1}.$$

Thus

$$\mathbf{Pr}\left(\sum_{i=1}^{i_0} Y_i \ge \sum_{i=1}^{i_0} \frac{2n^2 i^2}{bk^2 e^i} \middle| \mathcal{E}_0\right) = O\left(\frac{i_0 \log n}{n^{1/16}}\right),$$

and so

$$\mathbf{Pr}\left(\sum_{i=1}^{i_0} Y_i \ge \sum_{i=1}^{i_0} \frac{2n^2 i^2}{bk^2 e^i}\right) = o(1).$$
(6)

Now a simple calculation gives

$$\rho_u = O\left(\frac{X_2^* - X_u}{X_2^*}\right) \tag{7}$$

and so

$$\mathbf{E}(Z_i \mid \mathcal{E}_i) = O(in^{31/16} + \beta_i |E_i|) = O(n^{31/16} \ln n).$$

Thus

$$\mathbf{Pr}(Z_i \ge n^{63/32} \mid \mathcal{E}_i) = O(n^{-1/32} \ln n)$$

and so

$$\mathbf{Pr}(\exists 1 \le i \le i_0 : Z_i \ge n^{63/32} \mid \mathcal{E}_0) = O(n^{-1/32} (\ln n)^2)$$

and

$$\mathbf{Pr}(\sum_{i=1}^{i_0} Z_i \ge i_0 n^{63/32}) = o(1).$$
(8)

Also

$$\mathbf{Pr}(u \in E_{i+1} \mid u \in E_i) = \left(1 - \frac{1}{X_2^*}\right)^{X_u} (1 - \rho_u) < e^{-1}.$$

Thus

$$\mathbf{E}(|E_{i_0+1}|) = O\left(\frac{n^2}{(\ln n)^4}\right)$$

and

$$\mathbf{Pr}\left(|E_{i_0+1}| \ge \frac{n^2}{(\ln n)^3}\right) = o(1).$$
(9)

Theorem 1 follows from (6), (8) and (9) and

$$|\mathcal{C}_{COVER}| = \sum_{i=1}^{i_0} Y_i + \sum_{i=1}^{i_0} Z_i + |E_{i_0+1}|$$

As we only use estimates for $X_{0,\lfloor k/i\rfloor,i}^*$ and $X_{2,\lfloor k/i\rfloor,i}^*$ the reader may wonder why it is necessary to prove Lemma 1(a) for $0 \le s \le j \le k/i$. The reason is simply that the lemma is proved by induction and we use a stronger induction hypothesis than the needed outcome.

3 Proof of Lemma 1

If s = j then $X_{S,j,i} = c_{s,j,i} = 1$ and so we can assume s < j from now on.

Let us first consider \mathcal{E}_1 . Fix a set S of size $s, 0 \leq s \leq k$. Assume it forms a clique in G. This does not condition any edges not contained in S. For a set T let $N_c(T)$ denote the set of common neighbours of T in G. We can enumerate the set of j-cliques containing S as follows: choose $x_1 \in N_c(S), x_2 \in N_c(S \cup \{x_1\}), \ldots, x_{j-s} \in N_c(S \cup \{x_1, x_2, \ldots, x_{j-s-1}\})$. The number of choices ν_t for x_t given $x_1, x_2, \ldots, x_{t-1}$ is distributed as $Bin(n - (s - t + 1), p^{s+t-1})$. Thus for $0 \le \epsilon \le 1$

$$\begin{aligned} &\mathbf{Pr}\left(\left|\frac{\nu_t}{(n-s-t+1)p^{s+t-1}} - 1\right| \ge \epsilon\right) &\le & 2\exp\left\{-\frac{\epsilon^2(n-s-t+1)p^{s+t-1}}{3}\right\} \\ &\le & 2\exp\{-\epsilon^2 n^{1-\alpha}/4\}. \end{aligned}$$

Putting $\epsilon = n^{-1/3}$ we see that since there are $n^{O(\ln n)}$ choices for $x_1, x_2, \ldots, x_{j-s}$,

$$\mathbf{Pr}\left(\left|\frac{X_{S,j,0}}{c_{s,j,0}} - 1\right| \ge n^{-1/3 + o(1)}\right) \le \exp\{-n^{1/4}\}.$$

There are $n^{O(\ln n)}$ choices for S and (1) follows.

Assume now that \mathcal{E}_i holds. We first prove

Lemma 2. Suppose $e_1, e_2, \ldots, e_t \in E_i$. Then

$$\mathbf{Pr}(e_t \in E_{i+1} \mid e_1, e_2, \dots, e_{t-1} \in E_{i+1}) = e^{-1} \left(1 + O\left(\frac{t \ln n}{n}\right) \right)$$

uniformly for $1 \leq t \leq n^{1/2}$.

Proof

$$\mathbf{Pr}(e_t \in E_{i+1} \mid e_1, e_2, \dots, e_{t-1} \in E_{i+1}) \geq \mathbf{Pr}(e_t \in E_{i+1})$$

$$= \left(1 - \frac{1}{X_2^*}\right)^{X_u} (1 - \rho_u)$$

$$= e^{-1} - X_2^{*-1}.$$
(10)

Here $u = e_t$, $X_u = X_{u,\lfloor k/i \rfloor,i}$ and $X_2^* = X_{2,\lfloor k/i \rfloor,i}^*$ and inequality (10) follows from the fact that knowing $e_1, e_2, \ldots e_{t-1} \in E_{i+1}$ tells us that certain cliques (and edges) were not chosen for \mathcal{C}_{COVER} . On the other hand

$$\mathbf{Pr}(e_t \in E_{i+1} \mid e_1, e_2, \dots, e_{t-1} \in E_{i+1}) \leq \left(1 - \frac{1}{X_2^*}\right)^{X_u - tX_3^*} (1 - \rho_u)$$
(11)
$$= \left(e^{-1} - X_2^{*-1}\right) \left(1 - \frac{1}{X_2^*}\right)^{tX_3^*}$$

$$= e^{-1} \left(1 + O\left(\frac{tX_3^*}{X_2^*}\right)\right),$$

where $X_3^* = X_{3,\lfloor k/i \rfloor,i}^*$. If \mathcal{E}_i holds then $X_3^*/X_2^* = O(\ln n/n)$.

Inequality (11) follows from the fact that $e_t = u$ lies in at least $X_u - (t-1)X_3^*$ cliques which contain none of $e_1, e_2, \ldots, e_{t-1}$. This in turn arises from a two term inclusion-exclusion inequality and the fact that e_t and e_i together lie in at most X_3^* cliques, for $1 \le i \le t - 1$. \Box

Now fix a set $S \in \mathcal{C}_{s,i}$ and let $X = X_{S,j,i+1}$ for some $j \leq k/(i+1)$. Condition on $S \in \mathcal{C}_{s,i+1}$. Let $\mathcal{C}_{S,j,i} = \{C \in \mathcal{C}_{j,i} : C \supseteq S\}$. Then on using Lemma 2, we have

$$\mathbf{E}(X) = \sum_{C \in \mathcal{C}_{S,j,i}} \mathbf{Pr}(C \in \mathcal{C}_{j,i+1} \mid S \in \mathcal{C}_{s,i+1}) \\
= X_{S,j,i} \exp\left\{ \binom{s}{2} - \binom{j}{2} \right\} \left(1 + O\left(\frac{j^4 \ln n}{n}\right) \right), \quad (12) \\
= \mathbf{E}(X_{S,j,0}) \exp\left\{ (i+1) \left(\binom{s}{2} - \binom{j}{2} \right) \right\} \left(1 + O\left(\frac{j^4 \ln n}{n}\right) \right),$$

by induction on i

$$= c_{s,j,0} \exp\left\{ (i+1) \left(\binom{s}{2} - \binom{j}{2} \right) \right\} \left(1 + O\left(\frac{j^4 \ln n}{n} \right) \right),$$

$$= c_{s,j,i+1} \left(1 + O\left(\frac{j^4 \ln n}{n} \right) \right).$$
(13)

We are going to use the Markov inequality

$$\mathbf{Pr}(X \ge x) \le \frac{\mathbf{E}((X)_r)}{(x)_r} \tag{14}$$

where $(x)_r = x(x-1)(x-2)\dots(x-r+1)$ and $r = \lfloor n^{3/8} \rfloor$.

Let $\mathcal{B}(\ell_2, \ell_3, \dots, \ell_r) = \{ (C_1, C_2, \dots, C_r) : (i) \ C_t \neq C_{t'} \text{ for } t \neq t', (ii) \ C_t \in \mathcal{C}_{S,j,i}, (iii) \ |\mathcal{C}_t \cap (C_1 \cup C_2 \cup \dots \cup C_{t-1})| = s + \ell_t, \text{ for } t, t' = 2, 3, \dots, r \}.$ Then

$$\mathbf{E}((X)_r) = \sum_{\ell_2, \ell_3, \dots, \ell_r} \sum_{\mathcal{B}(\ell_2, \ell_3, \dots, \ell_r)} \mathbf{Pr}(C_1, C_2, \dots, C_r \in \mathcal{C}_{j, i+1} \mid S \in \mathcal{C}_{s, i+1}).$$

From (12)

$$\mathbf{Pr}(C_1 \in \mathcal{C}_{j,i+1} | S \in \mathcal{C}_{s,i+1}) = \exp\left\{\binom{s}{2} - \binom{j}{2}\right\} \left(1 + O\left(\frac{j^4 \ln n}{n}\right)\right)$$

and

$$\mathbf{Pr}(C_t \in \mathcal{C}_{j,i+1} \mid C_1, C_2, \dots, C_{t-1} \in \mathcal{C}_{j,i+1}) = \exp\left\{ \binom{s+\ell_t}{2} - \binom{j}{2} \right\} \left(1 + O\left(\frac{j^4 \ln n}{n}\right) \right)$$
$$= \exp\left\{ \binom{s+\ell_t}{2} - \binom{s}{2} \right\} \frac{c_{s,j,i+1}}{c_{s,j,i}} \left(1 + O\left(\frac{j^4 \ln n}{n}\right) \right)$$

Also,

$$\begin{aligned} |\mathcal{B}(\ell_2, \ell_3, \dots, \ell_r)| &\leq \prod_{t=1}^r \left(\binom{(t-1)j-s}{\ell_t} X^*_{s+\ell_t, j, i} \right) \\ &\leq \prod_{t=1}^r (rj)^{\ell_t} (1+\beta_i) \left(\frac{b^{s+\ell_t} j e^{i(s+\ell_t)}}{n} \right)^{\ell_t} c_{s, j, i}. \end{aligned}$$

Hence,

$$\frac{\mathbf{E}((X)_{r})}{c_{s,j,i+1}^{r}} \leq \left(1 + O\left(\frac{(\ln n)^{4}r}{n}\right)\right) \sum_{\ell_{2},\ell_{3},\dots,\ell_{r}} \prod_{t=1}^{r} (1+\beta_{i}) \left(\frac{e^{(\ell_{t}+2s-1)/2}rj^{2}(be^{i})^{s+\ell_{t}}}{n}\right)^{\ell_{t}} \\
\leq \left(1 + O\left(\frac{(\ln n)^{4}r}{n}\right)\right) (1+\beta_{i})^{r} \sum_{\ell_{2},\ell_{3},\dots,\ell_{r}} \left(\frac{rk^{2}e^{3k}b^{2k}}{n}\right)^{\ell_{2}+\dots+\ell_{t}} \\
\leq (1+rn^{-3/4})(1+\beta_{i})^{r},$$
(15)

for α sufficiently small.

Hence, using (14),

$$\begin{aligned} \mathbf{Pr}(X \ge (1+\beta_{i+1})c_{s,j,i+1}) &\leq \frac{2(1+\beta_i)^r c_{s,j,i+1}^r}{((1+\beta_{i+1})c_{s,j,i+1})_r}, \qquad \text{by (16)} \\ &\leq 3\left(\frac{1+\beta_i}{1+\beta_{i+1}}\right)^r, \qquad \text{using (4)} \\ &\leq 3\exp\left\{-\frac{r(\beta_{i+1}-\beta_i)}{1+\beta_{i+1}}\right\} \\ &= \exp\{-n^{1/8-o(1)}\}. \end{aligned}$$

There are $n^{O(\ln n)}$ choices for S and j and so part (a) of the lemma is proven.

It remains only to deal with $X_{u,j,i+1}$ for an edge $u \in E_i$. It follows from (13) that if $X = X_{u,j,i+1}$ then

$$\mathbf{E}(X) = c_{2,j,i} \left(1 + O\left(\frac{j^4 \ln n}{n}\right) \right), \tag{17}$$

and from (16) that

$$\mathbf{E}(X(X-1)) \le \left(1 + \frac{3i}{n^{1/4}}\right) c_{2,j,i+1}^2.$$
 (18)

Suppose now that $X_{u,j,i} \ge (1 - \gamma_i)c_{2,j,i}$. Then (17) and (18) imply that

$$\begin{aligned} \mathbf{Pr}(X \leq (1 - \gamma_{i+1})c_{2,j,i+1}) &= \\ \mathbf{Pr}(\mathbf{E}(X) - X \geq \mathbf{E}(X) - (1 - \gamma_{i+1})c_{2,j,i+1}) \leq \\ \mathbf{Pr}\left(\mathbf{E}(X) - X \geq (1 - \gamma_{i})c_{2,j,i} \exp\left\{1 - \binom{j}{2}\right\} \left(1 + O\left(\frac{j^{4}\ln n}{n}\right)\right) - (1 - \gamma_{i+1})c_{2,j,i+1}\right) &= \\ \mathbf{Pr}\left(\mathbf{E}(X) - X \geq (1 - \gamma_{i})c_{2,j,i+1} \left(1 + O\left(\frac{j^{4}\ln n}{n}\right)\right) - (1 - \gamma_{i+1})c_{2,j,i+1}\right) = \\ \mathbf{Pr}\left(\mathbf{E}(X) - X \geq (1 - o(1))n^{-1/16}c_{2,j,i+1}\right) \leq \\ \frac{(\mathbf{E}(X) - X)^{2}}{(1 - o(1))n^{-1/8}c_{2,j,i+1}^{2}} &= \\ \frac{\mathbf{E}(X(X - 1)) + \mathbf{E}(X) - \mathbf{E}(X)^{2}}{(1 - o(1))n^{-1/8}c_{2,j,i+1}^{2}} \leq \\ \frac{(1 + \frac{3i}{n^{1/4}})c_{2,j,i+1}^{2} + \left(1 + O\left(\frac{j^{4}\ln n}{n}\right)\right)c_{2,j,i+1} - c_{2,j,i+1}^{2}\left(1 + O\left(\frac{j^{4}\ln n}{n}\right)\right)}{(1 - o(1))n^{-1/8}c_{2,j,i+1}^{2}} \leq \\ \leq 4in^{-1/8}. \end{aligned}$$
(19)

Now let Z_{i+1} denote the number of edges $u \in E_{i+1}$ for which $X_{u,j,i+1} \leq (1 - \gamma_{i+1})c_{2,j,i+1}$ and \hat{Z}_{i+1} those u counted in Z_{i+1} for which $X_{u,j,i} \geq (1 - \gamma_i)c_{2,j,i}$. Then

$$Z_{i+1} \le Z_i + \hat{Z}_{i+1}$$

and from (19)

$$\mathbf{E}(\hat{Z}_{i+1} \mid \mathcal{E}_i) \le 4i |E_i| n^{-1/8}.$$

So

$$\mathbf{Pr}(Z_{i+1} \ge (i+1)n^{31/16} \mid \mathcal{E}_i) \le \mathbf{Pr}(\hat{Z}_{i+1} \ge n^{31/16} \mid \mathcal{E}_i) \\ = O(n^{-1/16} \log n).$$

this completes the proof of Lemma 1.

Acknowledgement: We thank Reza Akhtar for pointing out an error in the paper.

References

- B.Bollobás, P.Erdős, J.Spencer and D.B.West, Clique coverings of the edges of a random graph, Combinatorica 13 (1993) 1-5.
- [2] N.Pippenger and J.Spencer, Asymptotic behaviour of the chromatic index for hypergraphs, Journal of Combinatorial Theory A (1989) 24-42.