On the Cover Time for Random Walks on Random Graphs

JOHAN JONASSON
Chalmers University of Technology

March 26, 1997

Abstract

The cover time, C, for simple random walk on a realization, G_N, of $\mathcal{G}(N,p)$, the random graph of N vertices where all two vertices have an edge between them with probability p independently, is studied. The parameter p is allowed to decrease with N and p is written on the form $f(N)/N$. It is shown that if $f(N)$ is of higher order than $\log N$, then with probability $1-o(1)$, $(1-\epsilon)N\log N \leq \mathbf{E}[C|G_N] \leq (1+\epsilon)N\log N$ for any fixed $\epsilon > 0$ whereas if $f(N) = O(\log N)$ there exists a constant $a > 0$ such that with probability $1-o(1)$, $\mathbf{E}[C|G_N] \geq (1+a)N\log N$. It is furthermore shown that if $f(N)$ is of higher order than $(\log N)^3$ then $\mathbf{Var}(C|G_N) = o((N\log N)^2)$ so that with probability $1-o(1)$ the stronger statement that $(1-\epsilon)N\log N \leq C \leq (1+\epsilon)N\log N$ holds.

1 Introduction

Let $G = (V,E)$ be a simple connected graph on N vertices and consider a simple random walk on G. Let C denote the *cover time*, i.e. the time taken to visit all vertices of G and let $\mathbf{E}_v C$ denote the expectation of C for a walk starting from v. Much work has been done on finding this quantity for different graphs and on finding bounds for it. The interested reader is urged to look into the reference list.

It is today known that $(1-o(1))N\log N$ is a general lower bound for $\min_v \mathbf{E}_v C$ and that a general upper bound for $\max_v \mathbf{E}_v C$ is given by $(1+o(1))\frac{4}{27}N^3$. These bounds were until quite recently just conjectures but were proved by Feige ([8] and [7]). It is fairly easy to come up with examples to show that these bounds are also tight. In both cases one can use a so called lollipop graph where a path of suitable length extends from a clique. By starting at the end of the path, the cover time becomes extremely low, whereas starting from the clique produces an extremely high cover time.

The lollipop graph is clearly an example of a graph which seems "artificial" in the sense that it looks constructed by a human mind rather than something that would apper just by chance. As a matter of fact most of the common special cases of graphs are artificial in this sense. This is of course quite natural since such graphs are much simpler to think of. However, the mere fact that a graph is constructed in this way of course affects its properties. Among other things, as for the lollipop graph, the cover time might be affected. Heuristically the usually "simple" structure of a constructed graph often tends to make some vertices harder to reach so that the cover time becomes high (at least if the walk starts from a suitably chosen vertex). Examples are the graph $\mathbf{Z}$ mod N for which $\mathbf{E}_v C = \frac{1}{2}N^2$ for any v and the binary tree for which $\mathbf{E}_{root} C = N(\log N)^2$ (see [16]). Therefore a natural question, which is the principal question of this paper, is

How long does it take to cover a *typical* graph?

One natural interpretation of the term "typical graph" is a graph which is likely to be produced from a random mechanism. In short, we might replace the word "typical" by the word "random". To make the question more precise we need to decide on what kind of random graphs we wish to consider. We will work with the well known random graph model $\mathcal{G}(N,p)$, i.e. the graph one gets by letting the edges of a complete graph be present with probability p independently. Thus the question above becomes

How long does it take to cover a random graph with distribution $\mathcal{G}(N,p)$?

We need to introduce a mathematical framework, so let $(\Omega, \mathcal{F}, P)$ be a probability space and on this space define the random graph G_N with distribution $\mathcal{G}(N,p)$ and, given G_N, a simple random walk on it. The construction of such a probability space is just straightforward standard construction.

For the trivial case $p = 1$, G_N becomes the complete graph and it is well known that for this graph $\mathbf{E}_v C = N \log N$ and $\mathbf{Var}_v(C) = O(N^2)$ for any v so that

$$P((1-\epsilon)N\log N \leq C \leq (1+\epsilon)N \log N) = 1 - o(1)$$

for any fixed $\epsilon > 0$. Is this threshold at $N \log N$ still there if $p = p(N) < 1$? What if $p(N) \to 0$ as $N \to \infty$? We shall see that if we write $p(N)$ on the form $f(N)/N$ we get the following result.

THEOREM 1.1 *Let* $p(N) = \frac{f(N)}{N}$ *and consider a simple random walk on* G_N.

(i) If $f(N)$ *is of higher order than* $\log N$ *then, for any fixed* $\epsilon > 0$,

$$P((1-\epsilon)N\log N \leq \min_v \mathbf{E}_v[C|G_N] \leq \max_v \mathbf{E}_v[C|G_N] \leq (1+\epsilon)N\log N)$$

$$= 1 - o(1).$$

Moreover, if $f(N)$ *is of higher order than* $(\log N)^3$ *then*

$$P(\max_v \mathbf{Var}_v(C|G_N) = o((N\log N)^2)) = 1 - o(1)$$

so that

$$P((1-\epsilon)N\log N \leq C \leq (1+\epsilon)N\log N) = 1 - o(1)$$

regardless of the choice of the starting vertex, i.e. there is indeed a threshold at $N \log N$.

(ii) *If $f(N) = (c + o(1)) \log N$ for some constant $c > 1$, then there is a constant $a > 0$ such that*

$$P(\min_v \mathbf{E}_v[C|G_N] \geq (1+a) N \log N) = 1 - o(1).$$

The reason for restricting to $c > 1$ in part (ii) of the above theorem is that it is known that in this case $P(G_N$ is connected $) \to 1$ as $N \to \infty$, whereas if $c < 1$ this probability tends to 0 (see e.g. [3, Theorem VII.3]). As a matter of fact the result at (i) also applies for $f(N)$ of smaller order if we consider a random walk on the *giant component* of G_N; it can be proved that if $f(N) > 1 + \delta$ for some $\delta > 0$, then with a probability tending to 1 there will be a unique giant component of order $O(N)$. For details we refer to [3, Chapter VI]. The validity of this extension will be obvious later on as it is easily seen that the same techniques as those used to prove Theorem 1.1 apply to this case as well. In Section 2 (ii) is proved. Part (i), being the longer part, is saved for Section 3.

Random walks on various kinds of random graphs have been studied earlier. Grimmett, Kesten and Zhang [9] look at random walk on the infinite cluster of a supercritical bond percolation model on $\mathbf{Z}^d$ for $d \geq 3$ and prove that the walk is a.s. transient. Lyons, Pemantle and Peres study random walk on a supercritical Galton-Watson tree. In [11] the simple random walk case is studied and in [12] they consider random walks biased towards or away from the root. Their results concern the speed of the random walk and its harmonic measure.

Before going into the main section we need some technical beckground.

DEFINITIONS. Let G be a finite connected graph and let (u, v) be a pair of vertices of G. We define $T(u, v)$ as the random time taken for a simple random walk starting from u to reach v. The *hitting time* is defined as the expectation $\mathbf{E}T(u, v)$ and is denoted $H(u, v)$. The *commute time*, $C(u, v)$, is defined as $H(u, v) + H(v, u)$ and the *difference time*, $D(u, v)$, is given by $H(u, v) - H(v, u)$. The total number of edges of G is denoted by m.

LEMMA 1.2 *The difference time is additive, i.e. for any three vertices, u, v and w,*

$$D(u, w) = D(u, v) + D(v, w).$$

A proof can be found in [5].

A useful technique in the analysis of random walks on graphs is to regard the graph as an electrical network (see e.g. [6] and [15]). The edges are then regarded as resistors with unit resistance and it can be shown that the effective resistances, $R(u, v)$, between vertices relate to the corresponding hitting times. Tetali [15] shows

LEMMA 1.3 *For any two vertices, u and v,*

$$H(u, v) = \frac{1}{2} \sum_{w \in V} d_w (R(u, v) + R(u, w) - R(v, w))$$

where d_w is the degree of the vertex w.

By adding $H(u, v)$ and $H(v, u)$ and using Lemma 1.3 we get

LEMMA 1.4 *For any two vertices, u and v,*

$$C(u, v) = 2mR(u, v).$$

The following lower bound on the effective resistance between two vertices, u and v, follows from the monotonicity property of electrical resistance by shortcutting all other vertices and, if u and v are neighbours, regard the adjoining edge as two parallel two ohm resitors.

LEMMA 1.5 $$R(u,v) \geq \frac{1}{d_u+1} + \frac{1}{d_v+1}.$$

An important technique for bounding the cover time was introduced by Matthews [13]. It is given by the lemma below. For later purposes we supply a proof.

LEMMA 1.6 *Let μ_- and μ_+ be real numbers and asssume that $\mu_- \leq H(u,v) \leq \mu_+$ for every pair, (u,v), of vertices. Then*

$$\mu_- \sum_{i=1}^{N} \frac{1}{i} \leq \mathbf{E}C \leq \mu_+ \sum_{i=1}^{N} \frac{1}{i}$$

for any starting vertex.

Proof. Let σ be a uniformly chosen random permutation, $\{v_1, v_2, \ldots, v_N\}$ of the vertices, defined on the same probability space as the random walk. Let, for $i = 1, \ldots, N$, T_i be the first time all the vertices $1, \ldots, i$ have been visited. (As a convention we do not regard the starting vertex as having been visited at time 0.) Then, with $R_i = T_i - T_{i-1}$, where of course $T_0 \equiv 0$,

$$C = T_N = \sum_{i=1}^{N} R_i$$

so that, since $P(R_i > 0) = 1/i$,

$$\mathbf{E}C = \sum_{i=1}^{N} \mathbf{E}R_i = \sum_{i=1}^{N} \frac{1}{i}\mathbf{E}[R_i | R_i > 0].$$

Under the condition $R_i > 0$, let S_i be the vertex $u \in \{v_1, \ldots, v_{i-1}\}$ which is the last one of these vertices to be reached by the random walk. Then

$$\mathbf{E}[R_i|R_i > 0] = \sum_{u,u'} \mathbf{E}[T(u,v_i)|R_i > 0, S_i = u, v_i = u']P(S_i = u, v_i = u'|R_i > 0)$$

and since the walk from u to v_i above is independent of the event $\{R_i > 0, S_i = u, v_i = u'\}$, we get

$$\mu_- \leq \mathbf{E}[R_i|R_i > 0] \leq \mu_+$$

and the result follows on summation. □

Since the starting vertex for the random walk will not make any difference in the sequel, it will be omitted from the notation.

Much of our interest is focused on making highly probable statements about conditional expectations and variances given G_N. For convenience such conditional entities will be denoted $\mathbf{E}^*$ and $\mathbf{Var}^*$. Consequently $H^*(u,v)$, $C^*(u,v)$ and $D^*(u,v)$ will denote the corresponding conditional entities. For example $H^*(u,v)$ is the random variable $\mathbf{E}^*T(u,v) = \mathbf{E}[T(u,v)|G_N]$.

Also for convenience, we make the following definition.

DEFINITION. An event, A_N, depending on N, is said to occur *with high probability*, abbrievated "whp", if $P(A_N) \to 1$ as $N \to \infty$.

Consequently the phrase "with probability" is sometimes abbrievated "wp".

2 The case $f(N) = (c + o(1)) \log N$

Before proving the main theorem of this section, we prove the following elementary lemma.

LEMMA 2.1 *Let $\mathcal{B}(m, \alpha)$ denote the law of a binomially distributed random variable with parameters m and α. Let X have law $\mathcal{B}(n, p)$ and let X' have law $\mathcal{B}(n-1, p)$ for some $p \in (0, 1)$. Then*

$$Pr(X' \leq k) - Pr(X \leq k) \leq \frac{p}{1-p} Pr(X \leq k).$$

Proof. This is just straightforward calculation for

$$Pr(X' \leq k) - Pr(X \leq k) = \sum_{j=0}^{k} [\binom{n-1}{j} p^j (1-p)^{n-1-j} - \binom{n}{j} p^j (1-p)^{n-j}]$$

$$= \sum_{j=0}^{n} [\frac{n-j}{n} \frac{1}{1-p} - 1] Pr(X = j) \leq (\frac{1}{1-p} - 1) Pr(X \leq k)$$

$$= \frac{p}{1-p} Pr(X \leq k).$$

□

THEOREM 2.2 *Let G_N be a random graph with distribution $\mathcal{G}(N, f(N)/N)$ with $f(N) = (1 + o(1)) c \log N$ for some $c > 1$ and let C be the cover time for a simple random walk on G_N. Then there is a constant $a > 0$ such that whp*

$$\mathbf{E}^* C \geq (1 + a) N \log N.$$

Proof. The idea is to show that there whp will be a large number of vertices with substantially lower degree than $c \log N$. (This can also be derived as a consequence of [2, Theorem 5.E].) Using Lemma 1.4 we will then show that the commute times for pairs of such vertices are high and by using the additivity of difference times we will show that this necessarily means that a large number of the hitting times are also high. Finally Lemma 1.6 will yield the result.

Fix a small $\alpha > 0$. Since d_u is a random variable with law $\mathcal{B}(N-1, \frac{(c+o(1)) \log N}{N})$, it follows from noticing that

$$P(d_u \leq c(1-\epsilon) \log N) \geq P(d_u = [c(1-\epsilon) \log N])$$

$$= \binom{N}{[c(1-\epsilon) \log N]} (\frac{c \log N}{N})^{[c(1-\epsilon) \log N]} \sim e^{-(\frac{1}{2}\epsilon^2 + o(\epsilon^2))(c+o(1)) \log N}$$

that

$$P(d_u - c \log N < -\epsilon c \log N) > e^{-\frac{2}{3}\epsilon^2 (c+o(1)) \log N}$$

$$= \frac{1}{N^{\frac{2}{3}\epsilon^2 (c+o(1))}} > \frac{1}{N^\alpha}$$

for large N as soon as

$$\epsilon < \sqrt{\frac{3\alpha}{2c}}.$$

Now, letting $X = \sum_u I_{\{d_u \leq (1-\epsilon)c\log N\}} \equiv \sum_u I_u$ for such an ϵ, we have that

$$\mathbf{E}X \geq N^{1-\alpha}$$

and

$$\mathbf{Var}(X) \leq N^{1-\alpha} + \sum\sum_{u\neq v}\mathbf{Cov}(I_u, I_v).$$

Since I_u and I_v only depend through the absence or presence of the edge between u and v, Lemma 2.1, with $p = P(I_u = 1)$ yields

$$\mathbf{Cov}(I_u, I_v) \leq (1 + \frac{(c+o(1))\log N}{N - (c+o(1))\log N})p^2 - p^2 \leq 2\frac{(c+o(1))\log N}{N}p^2$$

for large N. If ϵ is chosen to be larger than $\sqrt{\frac{\alpha}{c}}$, then $p \leq N^{-\alpha/2}$ so that

$$\mathbf{Cov}(I_u, I_v) \leq 2(c+o(1))N^{-1-\alpha}\log N.$$

This means that

$$\mathbf{Var}(X) \leq N^{1-\alpha} + N^{1-\alpha}(c+o(1))\log N$$

and since this is clearly $o(N^{2(1-\alpha)})$, Chebyshev's inequality yields that whp we have that $X \geq N^{1-2\alpha}$. For these at least $N^{1-2\alpha}$ vertices

$$R(u,v) \geq \frac{1}{d_u+1} + \frac{1}{d_v+1} \geq \frac{2}{(1-\epsilon)c\log N}$$

by Lemma 1.5. Order these vertices in an order $\{u_1, u_2, \ldots, u_X\}$ such that if $i < j$ then $D^*(u_i, u_j) > 0$. That this is possible is a consequence of Lemma 1.2, which also says that

$$D^*(u_1, u_X) = \sum_{i=1}^{X-1} D^*(u_i, u_{i+1}) \leq C^*(u_1, u_X).$$

It is safe to assume that $C^*(u_1, u_X) \leq 3N\log N$, for if not, there is nothing to prove. We can then find, for any fixed $K < \infty$, indices i and j such that $|i-j| \geq N^{1-2\alpha}/(K\log N)$ and

$$D^*(u_i, u_j) \leq \frac{3}{K}N$$

which for $K \geq 9/\epsilon$ is smaller than $\epsilon N/3$. However, since the Central Limit Theorem implies that whp

$$m \geq (1 - \frac{\epsilon}{2})\frac{N^2}{2}\frac{c\log N}{N} = (1-\frac{\epsilon}{2})\frac{cN\log N}{2},$$

Lemma 1.4 implies that whp

$$C^*(u_i, u_j) \geq \frac{1-\frac{\epsilon}{2}}{1-\epsilon}2N \geq 2(1+\frac{\epsilon}{3})N$$

for all i and j, so that we must whp have for a set of vertices of size at least $N^{1-2\alpha}/(K \log N)$ that

$$H^*(u,v) \geq (1+\frac{\epsilon}{3})N - \frac{\epsilon}{6}N = (1+\frac{\epsilon}{6})N.$$

Using Lemma 1.6 and setting, for instance, $\epsilon = \sqrt{\frac{36\alpha}{25c}}$, we get

$$\mathbf{E}^* C \geq (1+\frac{\sqrt{\alpha/c}}{5})N \log(\frac{N^{1-2\alpha}}{K \log N})$$

$$= (1+\frac{\sqrt{\alpha/c}}{5})(1-2\alpha)N\log N - (1+\frac{\sqrt{\alpha/c}}{5})(\log K + \log\log N)$$

$$\geq (1+\frac{\sqrt{\alpha/c}}{5})(1-3\alpha)N\log N$$

whp for large N. Since

$$(1+\frac{\sqrt{\alpha/c}}{5})(1-3\alpha) > 1$$

for a suitably chosen α close to 0, the proof is complete on letting a be the value of this expression for such a choice of α. □

3 The case $f(N) = \Omega(\log N)$

We start by observing the elementary fact that, as opposed to in Section 2, the graph will in this case become "almost regular", i.e. all vertices will have about the same degree.

LEMMA 3.1 *Let the random graph G_N have distribution $\mathcal{G}(N, f(N)/N)$ with $f(N)$ being of higher order than $\log N$ and fix $\epsilon > 0$. Then, wp $1 - o(1/N^k)$, all vertices, u, of G_N, will satisfy*

$$(1-\epsilon)f(N) \leq d_u \leq (1+\epsilon)f(N)$$

for any fixed $k > 0$.

Proof. This is a matter of upper bounding the tail in the binomial distribution as

$$d_u \stackrel{D}{=} \mathcal{B}(N-1, \frac{f(N)}{N}).$$

Use e.g. [3, Theorem I.7] for such a bound. Since $f(N)$ is of higher order than $\log N$, we have for large N that

$$P(|d_u - f(N)| \geq \epsilon f(N)) \leq e^{-\frac{1}{3}\epsilon^2 f(N)} = \frac{1}{N^{\frac{1}{3}\epsilon^2 f(N)/\log N}} = o(\frac{1}{N^k})$$

for any fixed k. □

The following Proposition and its proof provide, along with Lemma 1.6, the main tools of this section. The proof is, for later purposes, made in more detail than necessary. Note that the fact that for all two vertices there whp are not too long paths between them is an immediate consequence of [3, Theorem X.10]. Note also that the first part of the proof is very similar to the proof of [3, Lemma X.7]. Since our problem, however, does not require as much care with the details, we can without too much effort make the proof selfcontained.

Proposition 3.2 *Fix an $\epsilon > 0$. If $f(N) = o(N^{1/2})$, then G_N will whp satisfy*

$$R(u,v) \le \frac{2}{(1-\epsilon)N}$$

for all (u,v).

Proof. The idea is to look at the random graph as being built out from u like a branching process originating from u. We will thereby show that there is whp a quite tight lower bound on the number of paths of length $[\log N/\log f(N)] + 3$ between u and v. This will be done in such a way that the resulting paths between the second neighbours of u and the second neighbours of v become disjoint. Using this we give a bound for the effective resistance.

Fix $\delta > 0$. By Lemma 3.1, whp all vertices have degree between $(1-\delta)f(N)$ and $(1+\delta)f(N)$, so we can safely condition on this being the case. (Strictly speaking this means that most of the probabilities below are rather conditional probabilities.) Let $\phi(N) = (1-\delta)f(N) - 1$. For each vertex of our branching process we will only consider $\phi(N)$ edges. Since this waste only serves to decrease the number of paths between u and v it is no restriction. It is now clear that $\phi(N)+1$ vertices are reached in generation 1. Now, suppose we are given that x vertices ($x \ge \phi(N)$) are reached by only one edge in generation k and assume that $x\phi(N) \le O(N/f(N))$. (Since $f(N) = o(1/N^2)$ this is always at least true for $k = 1$.) Letting X be the number of edges stemming out from these x vertices towards the next generation reaching a vertex which has already been reached or is reached by another one of these edges, we have that

$$X \overset{D}{\le} Y$$

where Y has law

$$\mathcal{B}(x\phi(N), \frac{b}{\phi(N)})$$

for some constant $b < \infty$. By bounding the tail in this distribution using [3, Theorem I.7] we get, since $x \ge \phi(N)$ and $\phi(N)$ is of higher order than $\log N$,

$$P(X \ge 2bx) \le e^{-\frac{1}{3}bx} = o(\frac{1}{N^3}).$$

This means that wp $1 - o(1/N^3)$ the number of vertices reached by only one edge at generation $k+1$ will be at least

$$x\phi(N) - 2bx = (1 - \frac{2b}{\phi(N)})x\phi(N) \ge (1 - \frac{3b}{f(N)})x\phi(N).$$

Using induction we thus have that in generation $r \equiv [\log N/ \log f(N)] - 1$ at least

$$(1 - \frac{3b}{f(N)})^r \phi(N)^r$$

vertices are wp $1 - ro(1/N^3) = 1 - o(1/N^2)$ reached by only one edge. For the next step, assume that $\phi(N)^{r+1} = o(N)$. Then the number X' of edges reaching towards the next generation which fail to reach a vertex which is not reached by another one and which has not been reached earlier is for large N stochastically dominated by a random variable with law

$$\mathcal{B}(\phi(N)^{r+1}, \delta')$$

for an arbitrarily chosen but fixed $\delta' > 0$. Repeating the same argument as above yields

$$P(X' \geq 2\delta'\phi(N)^{r+1}) = o(\frac{1}{N^2})$$

so that after $r+1$ generations at least

$$(1 - 2\delta')(1 - \frac{3b}{f(N)})^r \phi(N)^{r+1}$$

vertices are wp $1 - o(1/N^2)$ reached by only one edge. For the last steps we go backwards from v assuming that v has not been reached yet. By the same arguments as above, v will wp $1 - o(1/N^2)$ have at least $(1 - 2\delta')\phi(N)^2$ second neighbours. Here we use the assumption that $f(N) = o(N^{1/2})$. Of these second neighbours a hypergeometric number will be one of the $r+1$'th neighbours of u and the expectation of this number is no less than

$$(1 - 2\delta')^2(1 - \frac{3b}{f(N)})^r \frac{\phi(N)^{r+3}}{N}.$$

Again by bounding the tail of the distribution noting that this tail is dominated by the tail of the corresponding binomial distribution we find, since $\phi(N)^{r+3} \geq O(\phi(N))$, that wp $1 - o(1/N^2)$ this number will be at least

$$(1 - 2\delta')^3(1 - \frac{3b}{f(N)})^r \frac{\phi(N)^{r+3}}{N}.$$

If the condition $\phi(N)^{r+1} = o(N)$ fails, then $\phi(N)^{r+1} = O(N)$ and we go backwards from v one step earlier and find analogously that there will wp $1 - o(1/N^2)$ be at least

$$(1 - 2\delta')^2(1 - \frac{3b}{f(N)})^r \frac{\phi(N)^{r+2}}{N}$$

paths of length $r+2$ between u and v. Since this procedure consists of a finite number of steps all having probability $1 - o(1/N^2)$ the whole procedure has probability $1 - o(1/N^2)$ and therefore holds wp $1 - o(1)$ for *all* pairs (u, v) of vertices. By the definition of $\phi(N)$ and the fact that δ is arbitrary, we can arrange things so that $r + 3 \leq [\log N/ \log f(N)] + 3$. (This means in particular that there are whp paths of length $[\log N/ \log f(N)] + 3$ between u and v for all (u, v).)

Now, in the above construction u and v had wp $1 - o(1/N^2)$ both at least $(1-\epsilon/4)f(N)$ neighbours. By a now familiar tail bounding argument it follows that wp $1 - o(1/N^2)$ no more than $\epsilon f(N)/4$ of these neighbours coincide or equal v or u respectively. This means that u as well as v have $(1-\epsilon/2)f(N)$ neighbours of which none are in common. Let us only consider the edges going to these neighbours and note that this waste can only serve to increase the effective resistance between u and v. Now, all these neighbours reached at least $(1-\epsilon/2)(1-\epsilon/4)f(N)^2$ second neighbours and by another tail bounding argument using that $f(N) = o(N^{1/2})$ it follows that wp $1 - o(1/N^2)$ both u and v have at least $(1-\epsilon/2)^2 f(N)^2$ second neighbours of which none are in common in the sense that none of the second neigbours of u is a first or second neighbour of v or v itself or vice versa. By what we proved above, these second neighbours are pairwise connected by disjoint paths of length $[\log N/\log f(N)] - 1 < \log N$. Since these last steps had probability $1 - o(1/N^2)$ for the particular pair (u,v) they hold whp for all (u,v). Therefore we have whp that

$$R(u,v) \leq \frac{2 + \frac{\log N + 2}{(1-\epsilon/2)f(N)}}{(1-\frac{\epsilon}{2})f(N)} \leq \frac{2}{(1-\epsilon)f(N)}$$

for N large enough for all (u,v). □

Theorem 3.3 *Let G_N be a random graph with distribution $\mathcal{G}(N, f(N)/N)$ and let C be the cover time for a simple random walk on G_N. If $f(N)$ is of higher order than $\log N$ but of lower order than $N^{1/2}$, then whp*

$$(1-\epsilon)N\log N \leq \mathbf{E}^*C \leq (1+\epsilon)N\log N$$

for any fixed $\epsilon > 0$.

Proof. The lower bound on $\mathbf{E}^*C$ follows immediately from Feige's [8] general lower bound of $(1-o(1))N\log N$ on the expected cover time, so let us focus on the upper bound part. By Proposition 3.2, for any fixed $\epsilon' > 0$,

$$R(u,v) \leq \frac{2}{(1-\epsilon')f(N)}$$

whp for all (u,v) and by Lemma 3.1 combined with Lemma 1.5

$$R(u,v) \geq \frac{2}{(1+\epsilon')f(N)}.$$

whp for all (u,v). Therefore Lemma 1.3 implies that

$$H^*(u,v) = \frac{1}{2}\sum_{w\in E} d_w(R(u,v) + R(v,w) - R(u,w))$$

$$\leq \frac{1}{2}N(1+\epsilon')f(N)\left(\frac{4}{(1-\epsilon')f(N)} - \frac{2}{(1+\epsilon')f(N)}\right)$$

whp for all (u,v) and for ϵ' small enough this is smaller than $(1+\epsilon)N$. Thus, using Lemma 1.6, we have

$$\mathbf{E}^*C \leq (1+\epsilon)N\log N$$

whp. □

In the introduction we promised to prove that if $f(N)$ is of higher order than $(\log N)^3$ then whp $\mathbf{Var}^*(C) = o((N\log N)^2)$ so that $N\log N$ is in fact a threshold for C. The way in which this will be done is by noting that with $f(N)$ of such high order, the time $T(u,v)$ given G_N will with overwhelming probability be dominated by a random variable with expectation $(1+o(1))N$ and variance $(1+o(1))N^2$. By a modification of the proof of Lemma 1.6 it can then be shown that this will imply that $\mathbf{Var}^*(C) = o((N\log N)^2)$.

PROPOSITION 3.4 *Let G_N have distribution $\mathcal{G}(N, f(N)/N)$ and assume that $f(N)$ is of higher order than $(\log N)^3$. Then, whp given G_N,*

$$T(v,u) \overset{D}{\leq} Y$$

where Y is a random variable with expectation $(1+o(1))N$ and variance $(1+o(1))N^2$, for all (u,v).

Proof. Since $f(N)$ is of higher order than $(\log N)^3$ we can strengthen the statement of Lemma 3.1 to

$$(1 - \frac{1}{f(N)^{1/3}})f(N) \leq d_u \leq (1 + \frac{1}{f(N)^{1/3}})f(N)$$

whp for all (u,v). The argument is completely analogous. Therefore we can, if $f(N) = o(N^{1/2})$, replace the function $\phi(N)$ of the proof of Proposition 3.2 by

$$\phi(N) = (1 - \frac{1}{f(N)^{1/3}})f(N).$$

The conclusion of the first part of that proof then becomes that if $\phi(N)^{r+1} = o(N)$ then, whp for all (u,v), at least

$$(1-2\delta')^3(1 - \frac{3b}{f(N)})^r(1 - \frac{1}{f(N)^{1/3}})^{r+3}\frac{f(N)^{r+3}}{N} \tag{1}$$

of the second neighbours of v are $r+1$'th neighbours of u. Note that by the nature of the proof of Proposition 3.2 the paths of length $r+1$ joining the second neighbours of v to u are disjoint. If $\phi(N)^{r+1} = O(N)$ then, whp for all (u,v), at least

$$(1-2\delta')^2(1 - \frac{3b}{f(N)})^r(1 - \frac{1}{f(N)^{1/3}})^{r+2}\frac{f(N)^{r+2}}{N} \tag{2}$$

of the second neighbours of v are r'th neighbours of u. Since $f(N)$ is of higher order than $(\log N)^3$ and $r+3 < \log N$, the factors in front of $f(N)^{r+3}/N$ and $f(N)^{r+2}/N$ converge to $(1-2\delta')^3$ and $(1-2\delta')^2$ respectively as $N \to \infty$ so that, since δ' is arbitrary, the above numbers are

$$(1 - o(1))\frac{f(N)^{r+3}}{N}$$

and

$$(1 - o(1))\frac{f(N)^{r+2}}{N}$$

respectively. This means for the former case that at each time, k, the conditional probability given G_N that the random walk reaches u in $r+3$ steps is at least

$$\frac{(1-o(1))f(N)^{r+3}/N}{(1+1/f(N)^{1/3})^{r+3}f(N)^{r+3}} = (1-o(1))\frac{1}{N}$$

whp for all u. A completely analogous argument yields the same result for the latter case. Thus

$$\mathbf{E}^*T(v,u) \leq (1+o(1))N + r + 2 = (1+o(1))N$$

and

$$\mathbf{E}^*T(v,u)^2 \leq 2(1+o(1))N^2$$

whp for all (u,v) and the proof is complete for the case $f(N) = o(N^{1/2})$.

Assume next that $f(N)$ is of higher order than $(N \log N)^{1/2}$ and fix $\epsilon > 0$. Then the number of vertices which are neighbours to both u and v is binomial with expectation $f(N)^2/N$. Since $f(N)^2/N$ is of higher order than $\log N$ it is by [3, Theorem I.7] therefore whp true that this number deviates from its mean by no more than $\epsilon f(N)^2/N$ for any (u,v). Therefore the conditional probability given G_N that a simple random walk starting from v reaches u in two steps is by Lemma 3.1 always at least

$$(1-2\epsilon)\frac{1}{N}$$

whp for all (u,v) and the desired result follows.

Finally, in the case where $f(N) \leq O((N \log N)^{1/2})$ but $f(N)$ is not of lower order than $N^{1/2}$, note that wp $1-o(1/N^2)$ both u and v have at least $(1-\epsilon)f(N)$ neighbours. If none of these coincide and if u and v are not neighbours themselves, then the number of the $(1-\epsilon)^2 f(N)^2$ possible edges connecting neighbours of u to neighbours of v which are actually present in the graph is binomial with expectation

$$(1-\epsilon)^2\frac{f(N)^3}{N}$$

so that whp this number is no less than

$$(1-\epsilon)^3\frac{f(N)^3}{N}$$

for any (u,v). The rest of the proof is analogous to the previous cases. □

Now, let us take another look at the proof of Lemma 1.6. If we try to adjust this to give a bound for the variance of C in our special case, we observe that

$$\mathbf{Var}^*(C) = \sum_i \mathbf{Var}^*(R_i) + 2\sum\sum\nolimits_{i<j}\mathbf{Cov}^*(R_i, R_j).$$

By the same arguments as before

$$\mathbf{Var}^*(R_i) \leq \mathbf{E}^* R_i^2 \leq \frac{1}{i}2(1+o(1))N^2$$

whp for all i so that

$$\sum_i \mathbf{Var}^*(R_i) \leq N^2 \log N = o((N \log N)^2).$$

whp. Also, if $i < j$,

$$\mathbf{E}^* R_i R_j = \frac{1}{i}\frac{1}{j}\mathbf{E}^*[R_i R_j | R_i R_j > 0]$$

as the events $\{R_i > 0\}$ and $\{R_j > 0\}$ are independent. Now, we can write

$$\mathbf{E}^*[R_i R_j | R_i R_j > 0]$$

$$= \sum_{u,u',w,w'} \mathbf{E}^*[T(u,u')T(w,w') | R_i R_j > 0, S_i = u, S_j = w, v_i = u', v_j = w']$$

$$\cdot P^*(S_i = u, S_j = w, v_i = u', v_j = w' | R_i R_j > 0)$$

$$= \sum_{u,u',w,w'} \mathbf{E}^*[T(u,u') | R_i R_j > 0, S_i = u, S_j = w, v_i = u', v_j = w'] \mathbf{E}^*[T(w,w')]$$

as the walk from u to u' is independent of the walk from w to w' and as the conditions do not contain any information on the latter. The walk from u to u', however, is not independent of the conditions as they tell us that it will not pass w' and, unless $w = u'$, it will not pass w. This restriction might in certain cases seriously affect the distribution of $T(u, v_i)$, but in our case, however, this cannot happen. To see this, observe that conditioning on that the walk avoids w and w' only serves to exclude paths passing these vertices from the sample space and does not affect the relation between probabilities for other paths. In particular, the conditional probability that the walk chooses a certain path for the say n next steps is always larger than the corresponding unconditional probability. Therefore the proof of Proposition 3.4 goes through for this conditional case noting that we have to reduce the expressions in (1) and (2) by 2. This, however, will clearly not upset things. (Remember that the paths of length $r + 1$ between u' and it's second neighbours are disjoint so that no more than two of them can pass w or w'.) Thus the above conditional expectation is also whp at most $(1 + o(1))N$. The cases $f(N) \geq O(N^{1/2})$ are carried out similarly. Thus, whp

$$\mathbf{E}^* R_i R_j \leq \frac{1}{i}\frac{1}{j}(1 + o(1))N^2$$

so that, since $\mathbf{E}^* R_k \geq \frac{1}{k}(1 - o(1))N$ whp by Lemma 1.5,

$$\mathbf{Cov}^*(R_i, R_j) \leq \frac{1}{i}\frac{1}{j}o(1)N^2.$$

whp. Therefore

$$2\sum\sum_{i<j} \mathbf{Cov}^*(R_i, R_j) \leq o(1)(N \log N)^2$$

whp so that

$$\mathbf{Var}^*(C) \leq o((N \log N)^2)$$

whp as desired. We have proved the following theorem.

THEOREM 3.5 *Let G_N be a random graph with distribution $\mathcal{G}(N, f(N)/N)$ and assume that $f(N)$ is of higher order than $(\log N)^3$ and let C be the cover time for a simple random walk on G_N. Then*

$$(1 - \epsilon)N \log N \leq C \leq (1 + \epsilon)N \log N$$

whp for any fixed $\epsilon > 0$.

Combining Theorem 2.2, Theorem 3.3 and Theorem 3.5 yields Theorem 1.1.

Remarks. For the cases where $f(N)$ is between $\log N$ and $(\log N)^3$ it is an open question whether or not there is actually a threshold for the cover time. By being a little more careful with the details of the proof of Proposition 3.2 or by using [3, Lemma X.7] it is possible to show that the number of paths of length $r+3$ between u and v is $(1-o(1))f(N)^{r+3}/N$ for these cases as well. However, the graph is not completely regular, not even asymptotically in the sense that all paths are not asymptotically equally probable for the random walk. This is what makes these cases different. It is therefore not clear that the result of Proposition 3.4 holds for these cases.

For the case where $f(N)$ is of order $\log N$ it would be interesting to come up with an exact expression for the expected cover time and to find out if this is also a threshold.

I have in this paper chosen to work with the $\mathcal{G}(N,p)$ model where the presence or absence of different edges are independent. Another standard random graph model is the $\mathcal{G}(N,M)$ model where exactly M uniformly chosen edges will be present in the graph. Letting $M = Nf(N)/2$, the degrees of the individual vertices will be hypergeometric random variables with expectation $f(N)$ whereas they are binomial in the $\mathcal{G}(N,p)$ case. However, as N gets large this difference will be of no importance and all the arguments will go through virtually unchanged. Therefore the results of this paper are valid also for the $\mathcal{G}(N,M)$ model.

Acknowledgment. I am grateful to Olle Häggström who came up with the idea to work with this problem and who also pointed out some references for me. I am also grateful to the referee for valuable comments.

References

1. D. Aldous, An Introduction to Covering Problems for Random Walks on Graphs, *J. Theor. Probab.* **2** (1989), 87-89.

2. A. Barbour, L. Holst and S. Janson, "Poisson Approximation," Oxford University Press, New York, 1992.

3. B. Bollobás, "Random Graphs," Academic Press, London, 1985.

4. A. Broder and A. Karlin, Bounds on the Cover Time, *J. Theor. Probab.* **2** (1989), 101-120.

5. D. Coppersmith, P. Tetali and P. Winkler, Collisions Among Random Walks on a Graph, *SIAM J. Discrete Math.* **6** (1993), 363-374.

6. P. Doyle and J. Snell, "Random Walks and Electrical Networks," The Mathematical Association of America, Washington DC, 1984.

7. U. Feige, A Tight Upper Bound on the Cover Time for Random Walks on Graphs, *Random Struct. Alg.* **6** (1995), 51-54.

8. U. Feige, A Tight Lower Bound on the Cover Time for Random Walks on Graphs, *Random Struct. Alg.* **6** (1995), 433-438.

9. G. GRIMMETT, H. KESTEN AND Y. ZHANG, Random Walk on the Infinite Cluster of the Percolation Model, *Pr. Th. Rel. Fields* **96** (1993), 33-44.

10. J. KAHN, N. LINIAL, N. NISAN AND M. SAKS, On the Cover Time of Random Walks on Graphs, *J. Theor. Probab.* **2** (1989), 121-128.

11. R. LYONS, R. PEMANTLE AND Y. PERES, Ergodic Theory on Galton-Watson Trees: Speed of Random Walk and Dimension of Harmonic Measure, *Erg. Th. Dynamical Syst.* **15** (1995), 593-620.

12. R. LYONS, R. PEMANTLE AND Y. PERES, Biased Random Walks on Galton-Watson Trees, *Pr. Th. Rel. Fields* **106** (1996), 249-264.

13. P. MATTHEWS, Covering Problems for Brownian Motion on Spheres, *Ann. Probab.* **16** (1988), 189-199.

14. P. MATTHEWS, Some Sample Path Properties of a Random Walk on the Cube, *J. Theor. Probab.* **2** (1989) 129-146.

15. P. TETALI, Random Walks and the Effective Resistance of Networks, *J. Theor. Probab.* **4** (1991), 101-109.

16. D. ZUCKERMAN, Covering Times of Random Walks on Bounded Degree Trees and Other Graphs, *J. Theor. Probab.* **2** (1989), 147-158.