The cover time of random walks on random geometric graphs

Colin Cooper
Alan Frieze

$G = (V, E)$ is a connected graph. ($|V| = n, |E| = m$). For $u \in V$ let C_u be the expected time taken for a simple random walk $\mathcal{W}_u$ on G starting at u, to visit every vertex of G.

$G = (V, E)$ is a connected graph. ($|V| = n, |E| = m$). For $u \in V$ let C_u be the expected time taken for a simple random walk $\mathcal{W}_u$ on G starting at u, to visit every vertex of G.

The *cover time* C_G of G is defined as $C_G = \max_{u \in V} C_u$.

$G = (V, E)$ is a connected graph. ($|V| = n, |E| = m$). For $u \in V$ let C_u be the expected time taken for a simple random walk $\mathcal{W}_u$ on G starting at u, to visit every vertex of G.

The *cover time* C_G of G is defined as $C_G = \max_{u \in V} C_u$.

$C_G \leq 4m(n-1)$: Aleliunas,Karp,Lipton,Lovász,Rackoff (1979)

$G = (V, E)$ is a connected graph. ($|V| = n, |E| = m$). For $u \in V$ let C_u be the expected time taken for a simple random walk $\mathcal{W}_u$ on G starting at u, to visit every vertex of G.

The *cover time* C_G of G is defined as $C_G = \max_{u \in V} C_u$.

$C_G \leq 4m(n-1)$: Aleliunas,Karp,Lipton,Lovász,Rackoff (1979)

$(1 - o(1))n \ln n \leq C_G \leq (1 + o(1))\frac{4}{27}n^3$: Feige (1995)

Cooper and Frieze

- If $p = c\log n/n$ and $c > 1$ then w.h.p. $C_{G_{n,p}} \sim c\log\left(\frac{c}{c-1}\right) n\log n$.

Cooper and Frieze

- If $p = c \log n/n$ and $c > 1$ then w.h.p. $C_{G_{n,p}} \sim c \log\left(\frac{c}{c-1}\right) n \log n$.
- Let $c > 1$ and let x denote the solution in $(0, 1)$ of $x = 1 - e^{-cx}$. Let X_g be the giant component of $G_{n,p}$, $p = c/n$. Then w.h.p. $C_{X_g} \sim \frac{cx(2-x)}{4(cx-\log c)} n(\log n)^2$.

Cooper and Frieze

- If $p = c\log n/n$ and $c > 1$ then w.h.p. $C_{G_{n,p}} \sim c\log\left(\frac{c}{c-1}\right) n\log n$.
- Let $c > 1$ and let x denote the solution in $(0,1)$ of $x = 1 - e^{-cx}$. Let X_g be the giant component of $G_{n,p},\ p = c/n$. Then w.h.p. $C_{X_g} \sim \frac{cx(2-x)}{4(cx-\log c)} n(\log n)^2$.
- Let $G_{n,r}$ denote a random r-regular graph on vertex set $[n]$ with $r \geq 3$ then w.h.p. $C_{G_{n,r}} \sim \frac{r-1}{r-2} n\log n$.

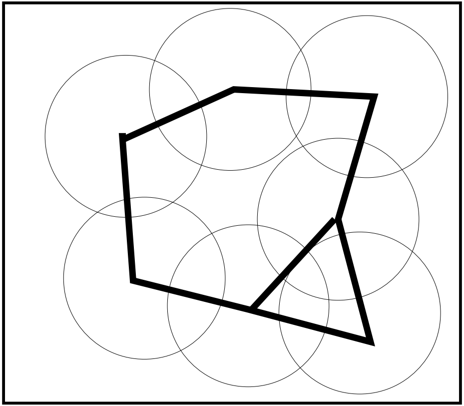
Cooper and Frieze

- If $p = c \log n/n$ and $c > 1$ then w.h.p. $C_{G_{n,p}} \sim c \log\left(\frac{c}{c-1}\right) n \log n$.
- Let $c > 1$ and let x denote the solution in $(0, 1)$ of $x = 1 - e^{-cx}$. Let X_g be the giant component of $G_{n,p}$, $p = c/n$. Then w.h.p. $C_{X_g} \sim \frac{cx(2-x)}{4(cx-\log c)} n(\log n)^2$.
- Let $G_{n,r}$ denote a random r-regular graph on vertex set $[n]$ with $r \geq 3$ then w.h.p. $C_{G_{n,r}} \sim \frac{r-1}{r-2} n \log n$.
- Let $G_m(n)$ denote a *preferential attachment graph* of average degree $2m$ then w.h.p. $C_{G_m} \sim \frac{2m}{m-1} n \log n$.

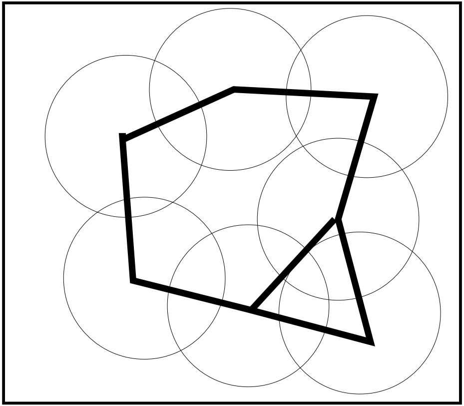
Cooper and Frieze

- If $p = c\log n/n$ and $c > 1$ then w.h.p. $C_{G_{n,p}} \sim c\log\left(\frac{c}{c-1}\right) n\log n$.
- Let $c > 1$ and let x denote the solution in $(0,1)$ of $x = 1 - e^{-cx}$. Let X_g be the giant component of $G_{n,p},\ p = c/n$. Then w.h.p. $C_{X_g} \sim \frac{cx(2-x)}{4(cx-\log c)} n(\log n)^2$.
- Let $G_{n,r}$ denote a random r-regular graph on vertex set $[n]$ with $r \geq 3$ then w.h.p. $C_{G_{n,r}} \sim \frac{r-1}{r-2} n\log n$.
- Let $G_m(n)$ denote a *preferential attachment graph* of average degree $2m$ then w.h.p. $C_{G_m} \sim \frac{2m}{m-1} n\log n$.
- Let $D_{n,p}$ denote a random *digraph* with independent edge probability p). If $p = c\log n/n$ and $c > 1$ then w.h.p. $C_{D_{n,p}} \sim c\log\left(\frac{c}{c-1}\right) n\log n$.

Random geometric graph $G = G(d, r, n)$ in d dimensions:
Sample n points V independently and uniformly at random from $[0, 1]^d$. For each point x draw a ball $D(x, r)$ of radius r about x.
$V(G) = V$ and $E(G) = \{\{v, w\} : \; w \neq v, \; w \in D(v, r)\}$



Random geometric graph $G = G(d, r, n)$ in d dimensions:
Sample n points V independently and uniformly at random from $[0, 1]^d$. For each point x draw a ball $D(x, r)$ of radius r about x.
$V(G) = V$ and $E(G) = \{\{v, w\} : \; w \neq v, \, w \in D(v, r)\}$



For simplicity we replace $[0, 1]^d$ by a torus.

Avin and Ercal $d = 2$

Theorem

$$C_G = \Theta(n \log n) \ w.h.p..$$

Avin and Ercal $d = 2$

Theorem

$$C_G = \Theta(n \log n) \; w.h.p..$$

Cooper and Frieze $d \geq 3$:

Theorem

Let $c > 1$ be constant, and let $r = \left(\frac{c \log n}{\Upsilon_d n}\right)^{1/d}$. Then w.h.p.

$$C_G \sim T_c = c \log\left(\frac{c}{c-1}\right) n \log n.$$

Υ_d is the volume of the unit ball in d dimensions.

First Visit Time Lemma.

Let $\pi_x = \frac{deg(x)}{2m}$ denote the steady state for a random walk on G.

First Visit Time Lemma.

Let $\pi_x = \frac{deg(x)}{2m}$ denote the steady state for a random walk on G.

Let the mixing time T be defined so that

$$\max_{u,x \in V} |P_u^{(t)}(x) - \pi_x| \le n^{-3}.$$

First Visit Time Lemma.

Let $\pi_x = \frac{deg(x)}{2m}$ denote the steady state for a random walk on G.

Let the mixing time T be defined so that

$$\max_{u,x \in V} |P_u^{(t)}(x) - \pi_x| \leq n^{-3}.$$

Let R_v be the expected number of visits by $\mathcal{W}_v$ within time T.

First Visit Time Lemma.

Let $\pi_x = \frac{deg(x)}{2m}$ denote the steady state for a random walk on G.

Let the mixing time T be defined so that

$$\max_{u,x \in V} |P_u^{(t)}(x) - \pi_x| \leq n^{-3}.$$

Let R_v be the expected number of visits by $\mathcal{W}_v$ within time T.

Fix $u, v \in V$. For $s \geq T$ let

$$\mathcal{A}_s(v) = \{\mathcal{W}_u \text{ does not visit } v \text{ in } [T, s]\}$$

First Visit Time Lemma.

Suppose that the connected graph $G = (V, E)$ has n vertices and m edges.

First Visit Time Lemma.

Suppose that the connected graph $G = (V, E)$ has n vertices and m edges.

Let $\pi_x = \frac{deg(x)}{2m}$ denote the steady state for a random walk on G.

First Visit Time Lemma.

Suppose that the connected graph $G = (V, E)$ has n vertices and m edges.

Let $\pi_x = \frac{deg(x)}{2m}$ denote the steady state for a random walk on G.

Let the mixing time T be defined so that

$$\max_{u,x \in V} |P_u^{(t)}(x) - \pi_x| \leq n^{-3}.$$

First Visit Time Lemma.

Suppose that the connected graph $G = (V, E)$ has n vertices and m edges.

Let $\pi_x = \frac{deg(x)}{2m}$ denote the steady state for a random walk on G.

Let the mixing time T be defined so that

$$\max_{u,x \in V} |P_u^{(t)}(x) - \pi_x| \leq n^{-3}.$$

Fix $u, v \in V$. For $s \geq T$ let

$$\mathcal{A}_s(v) = \{\mathcal{W}_u \textit{ does not visit } v \textit{ in } [T, s]\}$$

We try to get a good estimate of $\mathbf{Pr}(\mathcal{A}_s(v))$.

Write

$$H(s) = F(s)R(s)$$

where

$h_t = \mathbf{Pr}(\mathcal{W}_u(t) = v)$ and $H(s) = \sum_{t=T}^{\infty} h_t s^t$

$r_t = \mathbf{Pr}(\mathcal{W}_v(t) = v)$ and $R(s) = \sum_{t=0}^{\infty} r_t s^t$

f_t is the probability that the first visit of $\mathcal{W}_u$ to v in the period $[T, T+1, \ldots,]$ occurs at step t, and $F(s) = \sum_{t=T}^{\infty} f_t s^t$.

Write

$$H(s) = F(s)R(s)$$

where

$h_t = \mathbf{Pr}(\mathcal{W}_u(t) = v)$ and $H(s) = \sum_{t=T}^{\infty} h_t s^t$

$r_t = \mathbf{Pr}(\mathcal{W}_v(t) = v)$ and $R(s) = \sum_{t=0}^{\infty} r_t s^t$

f_t is the probability that the first visit of $\mathcal{W}_u$ to v in the period $[T, T+1, \ldots,]$ occurs at step t, and $F(s) = \sum_{t=T}^{\infty} f_t s^t$.

Note that

$$\mathbf{Pr}(\mathcal{A}_s(v)) = \sum_{t>s} f_t.$$

For

$$|z| \leq 1 + \lambda, \qquad \lambda = 1/KT$$

$$\begin{aligned} R(z) &= R_T(z) + \pi_v \frac{z^T}{1-z} + o(n^{-2}) \\ H(z) &= \pi_v \frac{z^T}{1-z} + o(n^{-2}) \end{aligned}$$

Now write

$$F(z) = \frac{H(z)}{R(z)} = \frac{B(z)}{A(z)}$$

where for $|z| \leq 1 + \lambda$

$$\begin{aligned} A(z) &= \pi_v z^T + (1 - z) R_T(z) + o(n^{-2}) \\ B(z) &= \pi_v z^T + o(n^{-2}) \end{aligned}$$

Now write

$$F(z) = \frac{H(z)}{R(z)} = \frac{B(z)}{A(z)}$$

where for $|z| \leq 1 + \lambda$

$$\begin{aligned} A(z) &= \pi_v z^T + (1-z)R_T(z) + o(n^{-2}) \\ B(z) &= \pi_v z^T + o(n^{-2}) \end{aligned}$$

$A(z)$ has a zero at

$$z_0 = 1 + \frac{\pi_v}{R_T(1) + o(1)}.$$

Now write

$$F(z) = \frac{H(z)}{R(z)} = \frac{B(z)}{A(z)}$$

$A(z)$ has a zero at

$$z_0 = 1 + \frac{\pi_v}{R_T(1) + o(1)}.$$

One can then show that

$$F(z) = \frac{B(z_0)/A'(z_0)}{z - z_0} + g(z)$$

where $g(z)$ is analytic inside $|z| \leq 1 + \lambda$.

One can then show that

$$F(z) = \frac{B(z_0)/A'(z_0)}{z - z_0} + g(z)$$

where $g(z)$ is analytic inside $|z| \leq 1 + \lambda$.

and so

$$\begin{aligned} f_t &= -\frac{B(z_0)/A'(z_0)}{z_0^{t+1}} + O((1+\lambda)^{-t}) \\ &\sim \frac{\pi_v/R_T}{(1+\pi_v/R_T)^{t+1}} + O((1+\lambda)^{-t}). \end{aligned}$$

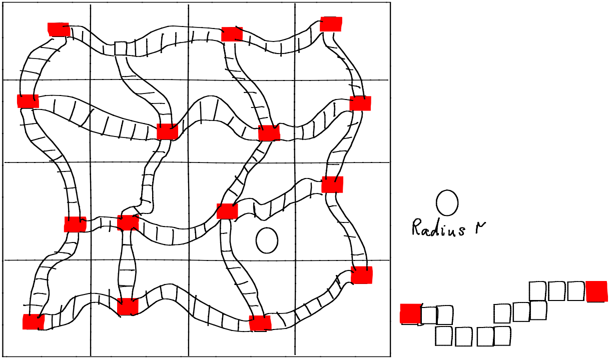
$$\mathbf{Pr}(\mathcal{A}_s(v)) = e^{-(1+o(1))\pi_v s/R_v}.$$

$$\mathbf{Pr}(\mathcal{A}_s(v)) = e^{-(1+o(1))\pi_v s/R_v}.$$

Most difficult task now is to show that $R_v = 1 + o(1)$ for all v.

Important Sub-Structure

Whp there is an embedded grid Γ made up of heavy sub-cubes, each contain $\Theta(\log n)$ vertices.
Edges of grid are sequences of heavy cubes, of same length.
Vertices of G are within $O(1)$ distance of Γ.



Given the grid Γ it is easy to use a Canonical Paths argument to show that

$$T = \tilde{O}(n^{2/d})$$

Given the grid Γ it is easy to use a Canonical Paths argument to show that

$$T = \tilde{O}(n^{2/d})$$

This estimate is not very good for $d = 2$. In this case it can be improved to $T = O(n/\log n)$.

Canonical Paths

We need two basic results on mixing times.

First let $\lambda_{\max}$ be the second largest eigenvalue of the transition matrix P. Then,

$$|P_u^{(t)}(x) - \pi_x| \leq \left(\frac{\pi_x}{\pi_u}\right)^{1/2} \lambda_{\max}^t.$$

Next, for each $x \neq y \in V$ let γ_{xy} be a *canonical* path from x to y in G. Then, we have that

$$\lambda_{\max} \leq 1 - \frac{1}{\rho},$$

where

$$\rho = \max_{e=\{x,y\}\in E(G)} \frac{1}{\pi(x)P(x,y)} \sum_{\gamma_{ab}\ni e} \pi(a)\pi(b)|\gamma_{ab}|,$$

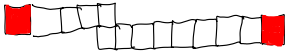
and $|\gamma_{ab}|$ is the length of the canonical path γ_{ab} from a to b.

Consider the grid Γ.It will help to fix a collection of points x_i in each red cube for $i = 1, 2, \ldots, M$.

Consider the grid Γ.It will help to fix a collection of points x_i in each red cube for $i = 1, 2, \ldots, M$.

We first define canonical paths between the x_i. We can in a natural way express $x_i = y(j_1, j_2, \ldots, j_d)$. The path from $y(j_1, j_2, \ldots, j_d)$ to $y(k_1, k_2, \ldots, k_d)$ goes

$$y(j_1, j_2, \ldots, j_d) \leftrightsquigarrow y(j_1 + 1, j_2, \ldots, j_d) \leftrightsquigarrow \cdots, y(k_1, j_2, \ldots, j_d)$$
$$\leftrightsquigarrow y(k_1, j_2 + 1, \ldots, j_d) \leftrightsquigarrow \cdots \leftrightsquigarrow, y(k_1, k_2, \ldots, k_d).$$

The $\leftrightsquigarrow$ represents a path in G that follows a 

choosing one vertex from each cube as necessary.

We obtain canonical paths for every pair of vertices by connecting each point x of V to its closest $x_i = \phi(x)$.

Each x_i is chosen by $O(\log n)$ points in this way.

Our path from x to y goes x to $\phi(x)$ to $\phi(y)$ to y. After this we find that each path has length $O(1/r)$ and each edge is in $\tilde{O}(1/r^{d+1})$ paths.

It follows that

$$\rho = \tilde{O} n \cdot 1 \cdot r^{-d-1} \cdot n^{-2} \cdot r^{-1} = \tilde{O}(1/(nr^{d+2})) = \tilde{O}(n^{2/d}).$$

Upper bound on cover time

Assuming that $R_v = 1 + o(1)$ for all v, we get an upper bound on C_G as follows:

$T_G(u)$ is the time taken to visit every vertex of G by $\mathcal{W}_u$.
U_t is the number of vertices of G which have not been visited by $\mathcal{W}_u$ at step t.

$$
\begin{aligned}
C_u = \mathbf{E}\,T_G(u) &= \sum_{t>0} \mathbf{Pr}(T_G(u) > t) \\
&\leq t + 1 + \sum_{s>t} \mathbf{E}\,U_s \\
&= t + 1 + \sum_{v \in V} \sum_{s>t} \mathbf{Pr}(\mathcal{A}_s(v))
\end{aligned}
$$

$$
\begin{aligned}
C_u = \mathbf{E}\, T_G(u) &= \sum_{t>0} \mathbf{Pr}(T_G(u) > t) \\
&\leq t + 1 + \sum_{s>t} \mathbf{E}\, U_s \\
&= t + 1 + \sum_{v\in V} \sum_{s>t} \mathbf{Pr}(\mathcal{A}_s(v)) \\
&\approx t + 1 + \sum_{v\in V} \sum_{s>t} e^{-s\pi_v} \\
&= t + 1 + \sum_{v\in V} \frac{e^{-t\pi_v}}{1 - e^{-\pi_v}} \\
&\approx t + 1 + m \sum_{v\in V} \frac{e^{-t\,\deg(v)/2m}}{\deg(v)}
\end{aligned}
$$

Suppose that $k = \alpha \ln n$. There are approximately

$$n\binom{n-1}{k}p^k(1-p)^{n-1-k} \sim n^{1-c+\alpha\ln(ce/\alpha)}$$

vertices of degree k.

Suppose that $k = \alpha \ln n$. There are approximately

$$n\binom{n-1}{k}p^k(1-p)^{n-1-k} \sim n^{1-c+\alpha\ln(ce/\alpha)}$$

vertices of degree k.

So, if $t = \tau n \ln n$ and $m \approx \frac{1}{2}cn\log n$,

$$C_u \leq t + 1 + \sum_{\alpha} n^{2-c+\alpha\ln(ce/\alpha)-\alpha\tau/c+o(1)}$$

Now

$$\max_{\alpha} 2 - c + \alpha\ln(ce/\alpha) - \alpha\tau/c = 2 - c + ce^{-\tau/c}$$

So,

$$C_u \leq \tau n \ln n + O(n^{2-c+ce^{-\tau/c}+o(1)})$$

and

$$C_u \leq (1+o(1))c\ln\left(\frac{c}{c-1}\right)n\ln n.$$

Lower bound on cover time

Lower bound on cover time

We can find a vertex u and a set of vertices S_0 such that at time $t^* \approx T_c$, the probability the set S_0 is covered by the walk $\mathcal{W}_u$ tends to zero.

Hence $T_G(u) > t_0$ w.h.p. which implies that $C_G \geq (1 - o(1))t^*$.

We construct S_0 as follows. Let $k_1 = (c - 1)\log n$ and let $S_1 = \{v : d(v) = k_1\}$.

Let $A = \left\{(u, v) : \; u \notin S_1, v \in S_1, \eta(u, v) \geq 1/(\log n)^2\right\}$, where $\eta(u, v)$ is the probability that $\mathcal{W}_u$ visits v during the mixing time. It can be shown that w.h.p. $|A| = \tilde{O}(T|S_1|)$.

By simple counting, we see that there exists $u \notin S_1$ such that $|\{v \in S_1 : (u,v) \in A\}| = \tilde{O}(T|S_1|/n) = o(|S_1|)$. We choose such a u and let $S_2 = \{v \in S_1 : (u,v) \notin A\}$.

We then take an independent subset S_0 of S_2. Because the maximum degree of G is $O(\log n)$ we can choose $|S_0| = \Omega(|S_2|/\log n)$.

We then let Z_0 denote the number of vertices in S_0 that are not visited in time $[1, t^*]$.

We prove that

$$\mathbf{E}(Z_0) \to \infty \text{ and } \mathbf{E}(Z_0^2) = (1 + o(1))\mathbf{E}(Z_0)^2$$

and use the Chebyshev inequality.

Expected number of returns within time T

$p_{\mathrm{esc}}(v, B)$, is the probability that $\mathcal{W}_v$ does not return to v before reaching set B.

$$p_{\mathrm{esc}}(v, B) = \frac{1}{d(v) \mathrm{R}_{\mathrm{EFF}}(v, B)},$$

where $\mathrm{R}_{\mathrm{EFF}}(a, B)$ is the *effective resistance* between v and B.

Thus $R_v(B)$, the expected number of returns to v before reaching B is given by

$$R_v(B) = \frac{1}{p_{\mathrm{esc}}(v, B)} = d(v) \mathrm{R}_{\mathrm{EFF}}(v, B).$$

Let

$$\eta(w, v) = \mathbf{Pr}(\exists\, 1 \le t \le T : \mathcal{W}_w(t) = v).$$

$$F = F(v) = \left\{ w : \eta(w, v) \le 1/\log^2 n \right\}$$

p_v is the probability of a return to v by $\mathcal{W}_v$ within time T.

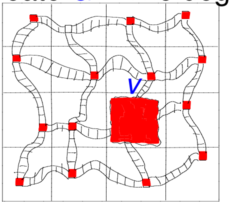
$$\begin{aligned} p_v &\le 1 - p_{\text{esc}}(v, F(v)) + 1/(\log n)^2. \\ R_v &\le \frac{1}{1 - p_v}. \end{aligned}$$

Enough to show that $p_{\text{esc}}(v, F(v)) = 1 - o(1)$.

It can be shown that $|F(v)| = n - o(n)$.

$$p_{\text{esc}}(v, F) = \frac{1}{d(v)\text{R}_{\text{EFF}}(v, B)}.$$

Raleigh's Theorem states that deleting edges increases effective resistance.
We remove edges to create G^*. The degree of v is unchanged and G^* looks like:



Because $d \geq 3$, walk has $\Omega(1)$ chance of getting far from v.
Because F is large, there is a $1 - o(1)$ chance that such a walk enters F. During $O(1)$ returns (in expectation)to cube containing v before entering F, walk has $o(1)$ chance of reaching v.

Open Questions

1. Determine the cover time for $d=2$. Known to be $\Theta(n \log n)$, but what is the constant.

2. How concentrated are the various quantities around their means.

3. Remove some of the "annoying" technical conditions for using main lemma:

$$T_{\pi_w} = O(1) \quad \text{and} \quad |R_T(z)| = \Omega(1), \ |z| \le 1 + \frac{1}{KT}$$

Thank You