Cover Times of Random Walks

R. Teal Witter

April 8, 2019

Abstract

abstract

Contents

1 Introduction and Background

"white screen" problem numerous applications many mathematical tools

1.1 Motivation

universal travel sequences graph connectivity protocol testing

1.2 Definitions

We consider an undirected graph G without self loops or multi-edges. Let $V(G)$ be the set of vertices and let $E(G)$ be the set of edges on G. Then n is the number of vertices $|V(G)|$.

We say $i \sim j$ for i and j in the vertex set $V(G)$ if (i,j) in $E(G)$. Let $c(i,j)$ be the non-negative weight of the edge between vertices i and j. If the graph is unweighted then $c(i,j) = 1$ if $i \sim j$.

Define the weight of vertex i

$$c(i) = \sum_{j:i\sim j} c(i,j).$$

We consider a random walk (X_t) on G. Call X_t the vertex that the walk is on at time t. The random walker begins at vertex X_0. At each step, the walker at vertex i moves to neighboring vertex j with probability $\frac{c(i,j)}{c(i)}$.

Let T_j be the number of steps until the first visit to j. Formally,

$$T_j = \min\{t : X_t = j\}.$$

Let the hitting time $\mathrm{t_{hit}}$ be the maximum expected time between two vertices on G. Formally,

$$\mathrm{t_{hit}} = \max_{i,j\in V(G)} \mathrm{E}_i[T_j].$$

Note that $\mathrm{E}_i[T_j] = \mathrm{E}[T_j | X_0 = i]$. Let the cover time be the time until all vertices have been visited by the random walk (X_t). We typically refer to the expected cover time $\mathrm{t_{cov}}$. Formally,

$$\mathrm{t_{cov}} = \mathrm{E}[\max_{j\in V(G)} T_j].$$

1.3 Outline

2 Cover Times of Structured Graphs

2.1 Small Example

In general, it is always possible to find the exact cover time of a graph using first step analysis. We demonstrate the strategy on a small graph suggested by [2].

Fig. 1 shows our small graph and the first step analysis for the random walk (X_t) started from the top left vertex. The black circle indicates the current vertex of the random walk and the empty circles indicate the visited vertices.

For a graph with n vertices, there are n possible choices for the current vertex and 2^{n-1} possible combinations of visited and unvisited vertices for each one. Of course, some states are inaccessible. But even on Fig. 1 where we exploit symmetry, there are almost a prohibitive number of cases. It is easy to see that the number of states grows exponentially with the number of vertices.

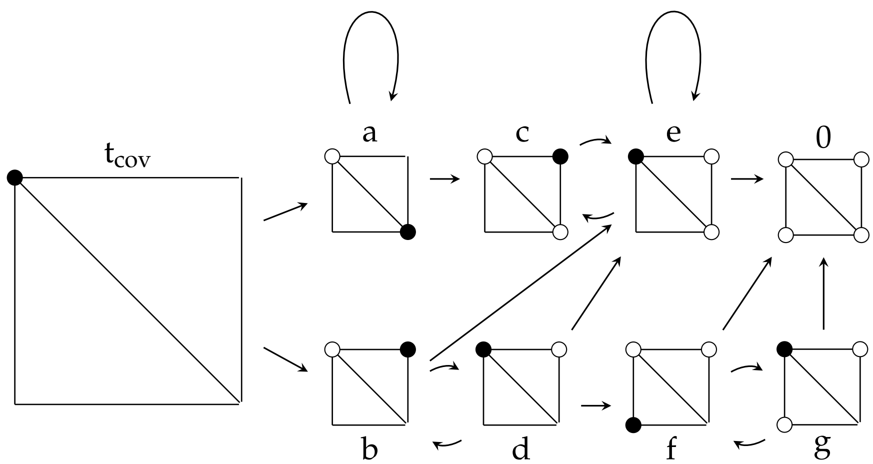

Figure 1: First step analysis applied to a small, asymmetric graph with $n = 4$.

If we let the label of each graph in Fig. 1 denote the expected cover time from that graph, we have the following system of equations:

$$\begin{aligned}
\mathrm{t_{cov}} &= 1 + \frac{1}{3}\mathrm{a} + \frac{2}{3}\mathrm{b} \\
\mathrm{a} &= 1 + \frac{1}{3}\mathrm{a} + \frac{2}{3}\mathrm{c} \\
\mathrm{b} &= 1 + \frac{1}{2}\mathrm{b} + \frac{1}{2}\mathrm{e} \\
\mathrm{c} &= 1 + 1\mathrm{e} \\
\mathrm{d} &= 1 + \frac{1}{3}\mathrm{a} + \frac{1}{3}\mathrm{e} + \frac{1}{3}\mathrm{f} \\
\mathrm{e} &= 1 + \frac{1}{3}\mathrm{c} + \frac{1}{3}\mathrm{e} + \frac{1}{3}\dot{0} \\
\mathrm{f} &= 1 + \frac{1}{2}\mathrm{g} + \frac{1}{2}\dot{0} \\
\mathrm{g} &= 1 + \frac{2}{3}\mathrm{f} + \frac{1}{3}\dot{0}
\end{aligned}$$

With a little linear algebra, we find that $\mathrm{t_{cov}} = \frac{43}{6}$.

For large graphs, the first step analysis as an approach to finding the expected cover time is obviously intractable. The rest of this section deals with particularly symmetric graphs for which it is possible find a solution for an arbitrary n. However, we must rely on expected cover time bounds for the vast majority of cases.

2.2 Complete Graph

We call the graph with an edge between each pair of vertices the complete graph. Note that a complete graph with n vertices has $\binom{n}{2}$ edges.

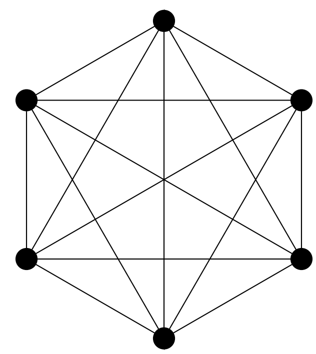

Figure 2: The complete graph with six vertices (and ten edges).

Our goal is to find the expected number of steps until we have visited all n vertices in the complete graph. Observe that the random walk can start at any vertex without loss of generality because the graph is symmetric. The strategy is to write the cover time $\mathrm{t_{cov}}$ in terms of the expected value of more simple random variables. Let X_i be the random variable that represents the number of steps to go from $i-1$ to i unique vertices (excluding the current vertex).

We write the cover time in terms of X_i and use linearity of expectation.

$$\mathrm{t_{cov}} = \mathrm{E}[X_1 + X_2 + X_3 + ... + X_{n+1}] = \sum_{i=1}^{n} \mathrm{E}[X_i]$$

The random variable X_i takes value k when $k-1$ steps lead us to already visited vertices and the k^{th} step is to a previously unvisited vertex. It follows that X_i is geometric with distribution

$$\mathrm{P}(X_i = k) = (1-p_i)^{k-1} p_i,$$

where p_i is the probability of moving to a previously unvisited vertex. Then $\mathrm{E}[X_i] = \frac{1}{p_i}$. We now find the probability p_i that we go to a previously unvisited vertex is $\frac{n-i}{n-1}$. This is because there are $n-i$ unvisited vertices and a total of $n-1$ adjacent vertices. It follows that $\mathrm{E}[X_i] = \frac{n-1}{n-i}$. Then

$$\begin{aligned}\mathrm{t_{cov}} &= \sum_{i=1}^{n} \frac{n-1}{n-i} \\ &= (n-1)\left(\frac{1}{n-1} + \frac{1}{n-2} + ... + 1\right).\end{aligned}$$

A natural interpretation of the cover time on a complete graph is the so-called coupon collecting problem. There are r different coupons that are randomly packaged in cereal boxes. An avid fan buys a box of cereal every day. We want to know how many days until the fan has collected each coupon. The strategy is to think of every coupon as a vertex on the complete graph. We then move from vertex to vertex with uniform probability. The difference between the coupon collecting problem and the complete graph is that we now have self-edges. Before, we had to leave our current vertex at each step. Now, we can stay in the same vertex (provided our collector found the same coupon two days in a row). We apply our approach to the complete graph and substitute the r possible coupons we can find for the $n-1$ vertices we could move to. Then

$$t_{\text{cov}} = r\left(\frac{1}{r} + \frac{1}{r-1} + \dots + 1\right).$$

Another more subtle application of the cover time on a complete graph is the cover time on an n-star.

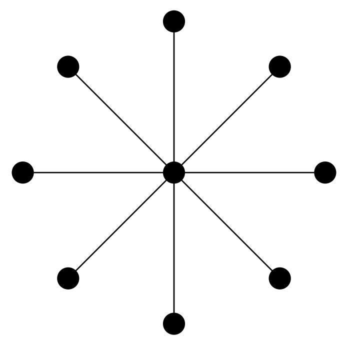

Figure 3: The nine-star.

An n-star is a graph with one central vertex and $n-1$ adjacent vertices each with a single edge to the center. A random walk begins at the central vertex and moves with uniform probability to one of the $n-1$ adjacent vertices. Whereas in the complete graph and coupon collecting problems we restarted in one step, a random walk on the n-star probabilistically restarts in two steps: one step to each leaf and one step back. We apply our approach to the complete graph and conclude that

$$t_{\text{cov}} = 2(n-1)\left(\frac{1}{n-1} + \frac{1}{n-2} + \dots + 1\right).$$

2.3 Line Graph

A line graph is a set of vertices connected in a line. (Fig. 4 is an example with $n = N + 1$.)

Our goal is to determine the cover time of a line graph. Let (X_t) be the random walk started at vertex i. Notice that a random walk will have covered all vertices exactly when it has visited both endpoints. We can use this observation to break up the cover time into the time it takes to reach one of the endpoints and the time it takes to reach the other endpoint.

We will begin by considering the time (X_t) takes to visit either one of the endpoints from vertex i.

From $X_t = i$, we move with equal probability to $i+1$ and $i-1$ until we reach either 0 or N.

Figure 4: A line graph with a vertex at each integer from 0 to N.

Call e_i the expected number of steps until we reach either end of our random walk from state i. Our strategy is to write and solve a recurrence relation that uses what we know about the random walk. If $i = 0, N$, we are already at the endpoint and $e_i = 0$. Otherwise, we take a step. There's a half probability our step is to the right and a half probability it is to the left.

$$\begin{aligned} e_i &= 1 + \frac{1}{2}\left(e_{i+1}\right) + \frac{1}{2}\left(e_{i-1}\right) \\ 2e_i &= 2 + e_{i+1} - e_{i-1} \\ (e_{i+1} - e_i) &= (e_i - e_{i-1}) - 2 \end{aligned}$$

While the final line may look more complicated than what we started with, notice that the structure of the equality lends itself to telescoping.

To solve the telescoping equations, observe that the walk is symmetric: we could flip the walk over the vertical axis without changing e_i. This makes sense because our intuition tells us that the only identifying feature of i is its respective distances to the endpoints. The logical conclusion is that $e_i = e_{N-i}$. In particular, we have already seen that $e_0 = e_N = 0$ and it naturally follows that $e_1 = e_{N_1}$. We use these observations to find e_1:

$$\begin{aligned} (e_2 - e_1) &= (e_1 - e_0) - 2 \\ (e_3 - e_2) &= (e_2 - e_1) - 2 = (e_1 - e_0) - 4 \\ (e_4 - e_3) &= (e_3 - e_2) - 2 = (e_1 - e_0) - 6 \\ &\vdots \\ (e_N - e_{N-1}) &= (e_1 - e_0) - 2(N-1) \\ (0 - e_1) &= (e_1 - 0) - 2(N-1) \\ e_1 &= N - 1. \end{aligned}$$

(Note that the vertical dots denote inductive reasoning.) We now know e_1. However, our goal is to find the expected number of steps from an arbitrary point i. We use e_1 to find the general solution.

$$\begin{aligned} e_2 &= e_1 + (e_1 - e_0) - 2 = 2(N-2) \\ e_3 &= e_2 + (e_2 - e_1) - 2 = 2\left[2(N-2)\right] - (N-1) - 2 = 3(N-3) \\ &\vdots \\ e_i &= i(N-i) \end{aligned} \tag{1}$$

Theorem 1 follows from Eq. (1) and the observation that any line graph can be shifted so that the left endpoint is at 0 and the right endpoint is at N.

Theorem 1. *The expected number of steps from a state to either end of a line graph is the product of the distances from that state to each endpoint.*

We have the expected time until we reach one endpoint of our line graph. Now we want to find the expected time from one endpoint to the other $\mathrm{E}_0[T_N]$. By symmetry, $\mathrm{E}_0[T_N] = \mathrm{E}_N[T_0]$.

We write

$$\mathrm{E}_0[T_N] = \mathrm{E}_0[T_1] + \mathrm{E}_1[T_2] + ... + \mathrm{E}_i[T_{i+1}] + ... + \mathrm{E}_{N-1}[T_N].$$

where $\mathrm{E}_i[T_{i+1}]$ is the expected time from state i to $i+1$. Since there is only one edge incident to 0, we always move from 0 to 1. From state $i \neq 0$, we take one step in each direction with equal probability. In half the cases, we have arrived at $i+1$. In the other half of cases, we are in $i-1$. It will take us $\mathrm{E}_{i-1}[T_i]$ steps back to i and then another $\mathrm{E}_i[T_{i+1}]$ steps to $i+1$. Formally,

$$\begin{aligned}\mathrm{E}_i[T_{i+1}] &= 1 + \frac{1}{2}(\mathrm{E}_{i-1}[T_i] + \mathrm{E}_i[T_{i+1}]) \\ &= 2 + \mathrm{E}_{i-1}[T_i].\end{aligned}$$

We can use $\mathrm{E}_0[T_1] = 1$ to solve for $E_i[T_{i+1}]$:

$$\begin{aligned}E_i[T_{i+1}] &= E_{i-1}[T_i] + 2 = E_{i-2}[T_{i-1}] + 4 \\ &= E_0[T_1] + 2(i-1) = 2i - 1.\end{aligned}$$

Then

$$\begin{aligned}\mathrm{E}_0[T_N] &= 1 + 3 + ... + 2i - 1 + ... + 2(N-1) - 1 \\ &= N^2\end{aligned}$$

Then, by Theorem 1, for the random walk from state i on the line graph,

$$\mathrm{t_{cov}} = i(N-i) + N^2.$$

2.4 Simple Cycle

A simple cycle is a connected graph where each vertex has an edge to exactly two other vertices. (Fig. 5 is an example.)

Consider a random walk on a simple cycle as in Fig. 5. (Note that the starting vertex is arbitrary since the graph is symmetric.) Let (X_t) be the walk on the number line (between $-n$ and n) where $X_0 = 0$. The cover time $\mathrm{t_{cov}}$ of the simple cycle is the expected number of steps until we visit n distinct integers.

Let $\{a, a+1, \ldots, b-1, b\}$ denote the set of vertices we have visited. Since the walk is on a line, we can write $[a, b]$ for the range of the walk where a is the smallest and b is the largest vertex we have visited. Notice that if $a \leq -n$ or $b \geq n$, we have covered all n vertices on the simple cycle and conclude the random walk.

In a slight abuse of notation, let T_k be the number of steps until we reach k distinct vertices. Formally, $T_k = \min\{t : b - a + 1 = k\}$. Then by telescoping,

$$\mathrm{t_{cov}} = \mathrm{E}[T_n] = \mathrm{E}[(T_2 - T_1) + (T_3 - T_2) + \cdots + (T_k - T_{k-1}) + \cdots + (T_n - T_{n-1})]$$

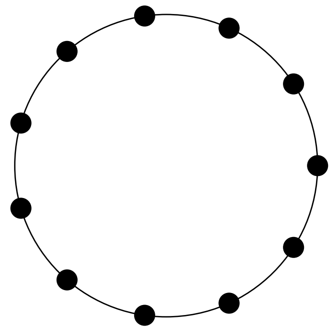

Figure 5: A cycle graph with $n = 11$ vertices.

At time T_k, we are either at integer a or b. Then $\mathrm{E}[T_{k+1} - T_k]$ is the expected time until we reach $a-1$ or $b+1$. Notice that we can shift our endpoints to 0 and $b+1-(a-1) = b-a+2 = k+1$. The expected number of steps to either $a-1$ or $b+1$ is $1(k+1-1) = k$ by Theorem 1.

By linearity of expectation,

$$\begin{aligned} \mathrm{t_{cov}} &= (\mathrm{E}[T_2] - \mathrm{E}[T_1]) + \cdots + (\mathrm{E}[T_{k+1}] - \mathrm{E}[T_k]) + \cdots + (\mathrm{E}[T_n] - \mathrm{E}[T_{n-1}]) \\ &= 1 + 2 + \cdots + k - 1 + \cdots + n - 1 = \frac{1}{2}n(n-1). \end{aligned}$$

3 Bounds

3.1 Matthews Method

We have seen that it is in general very hard to calculate the expected cover time of a random walk on a graph. The exceptions are symmetric or small graphs. In lieu of an exact solution, we want to bound the expected cover time.

The Matthews Method is an example of a remarkably good bound. In fact, it is tight for the complete graph: the expected time between any two vertices is $n-1$ so $\mathrm{t_{hit}} = n - 1$ and Theorem 2 gives us the main result of Section 2.2.

In addition, the proof uses strategies we employed on the simple cycle and follows an intuitive structure.

Theorem 2. *Let (X_t) be a random walk on a graph with n vertices. Then*

$$\mathrm{t_{cov}} \leq \mathrm{t_{hit}} \left(1 + \frac{1}{2} + \frac{1}{3} + ... + \frac{1}{n-1}\right).$$

Proof. Without loss of generality, label the vertices $\{1, 2, ..., n\}$. We may assume that the walk started at vertex n. Let σ be a uniform permutation of the unvisited $n-1$ vertices. The strategy is to look for the vertices in order σ.

Let T_k be the first time that vertices $\sigma_1, \sigma_2, ..., \sigma_k$ have all been visited. For ease of notation, set $L_k = X_{T_k}$ to be the last state among $\sigma_1, \sigma_2, ..., \sigma_k$ to be visited (the first time they are all visited).

We begin by writing the expected cover time in a clever way (similar to the strategy we employed on the cycle graph):

$$\begin{aligned}
\mathrm{t}_{\mathrm{cov}} &= \mathrm{E}_n[T_{n-1}] \\
&= \mathrm{E}_n[T_1 + (T_2 - T_1) + ... + (T_k - T_{k-1}) + ... + (T_{n-1} - T_{n-2})].
\end{aligned}$$

Our goal is to bound the cover time in terms of the hitting time.

Consider the first term of the sum. Recall that the hitting time $\mathrm{t}_{\mathrm{hit}}$ is the maximum expected time from any vertex to any other. It follows that for any choice of σ_1, the expected time from n to σ_1 is less than or equal to $\mathrm{t}_{\mathrm{hit}}$. Formally for s an arbitrary state such that $1 \le s \le n-1$, $\mathrm{E}_n[T_1] = \mathrm{E}_n[T_s] \le \mathrm{t}_{\mathrm{hit}}$.

Now consider the k^{th} term of the sum for $1 < k \le n-1$. By conditional expectation,

$$\mathrm{E}_n[T_k - T_{k-1}] = \sum_{i=1}^{k} \mathrm{E}_n[T_k - T_{k-1} | L(k) = \sigma_i] P(L_k = \sigma_i).$$

Assume that the last vertex L_k to be visted among the first k vertices is not σ_k. Then the first time we visit the first $k-1$ vertices is also the first time we visit the first k vertices (since we must have already visited σ_k). Formally, $\mathrm{E}_n[T_k - T_{k-1} | L(k) \neq \sigma_k] = 0$. Then

$$\mathrm{E}_n[T_k - T_{k-1}] = 0 + \mathrm{E}_n[T_k - T_{k-1} | L(k) = \sigma_k] P(L_k = \sigma_k).$$

Since σ is a random permutation, the probability that we visit σ_k last is $\frac{1}{k}$. Formally, $P(L_k = \sigma_k) = \frac{1}{k}$.

Finally, consider the expected time between T_{k-1} and T_k given that last vertex we visit is σ_k. Certainly there exist r and s where $1 \le r \neq s \le n-1$ such that $L_{k-1} = r$ and $L_k = s$. Then $\mathrm{E}_n[T_k - T_{k-1} | L(k) = \sigma_k] = \mathrm{E}_r[T_s] \le \mathrm{t}_{\mathrm{hit}}$. So

$$\mathrm{E}_n[T_k - T_{k-1}] \le \mathrm{t}_{\mathrm{hit}} \frac{1}{k}.$$

Putting it all together,

$$\mathrm{t}_{\mathrm{cov}} \le \mathrm{t}_{\mathrm{hit}} \left(1 + \frac{1}{2} + ... + \frac{1}{k} + ... + \frac{1}{n-1}\right).$$

□

3.2 Spanning Tree Argument

[1] 3.2

Edge commute inequality

spanning trees

4 Distributional Aspects

4.1 Complete Graph

[5] Theorem 2.2.4 -> Example 2.2.5 -> Theorem 3.6.5 -> Example 3.6.6

5 Electrical Networks

5.1 Definitions

[1] 3.3

5.2 Commute Time

[1] Proposition 2.3 -> Corollary 2.8 -> Corollary 3.11
https://ocw.mit.edu/courses/mathematics/18-445-introduction-to-stochastic-processes-spring-2015/lecture-notes/MIT18_445S15_lecture8.pdf

5.3 Balanced Trees

[7] 11.3

6 Simulation

[1] [3] [4] [5] [6] [7]

References

[1] David Aldous and James Allen Fill. *Reversible Markov Chains and Random Walks on Graphs*. 2002.

[2] Gunnar Blom, Lars Holst, and Dennis Sandell. *Problems and Snapshots from the World of Probability*. Springer, 1994.

[3] Béla Bollobás. *Modern Graph Theory*. Springer, 1998.

[4] Peter G. Doyle and J. Laurie Snell. *Random Walks and Electric Networks*. Mathematical Assn of Amer, 1984.

[5] Rick Durrett. *Probability: Theory and Examples*. Cambridge University Press, 2011.

[6] Geoffrey Grimmett. *Probability on Graphs: Random Processes on Graphs and Lattices*. Cambridge University Press, 2010.

[7] David A. Levin and Yuval Peres. *Markov Chains and Mixing Times*. Cambridge University Press, 2017.