# INTRODUCTION TO COVER TIMES OF RANDOM WALKS

**Darren Yao**

ABSTRACT

In this paper, we explore the topic of cover times: the expected number of steps it takes a walk to visit all vertices of a graph. We start by introducing some basic definitions and results in graph theory and random walk cover times. We then calculate the cover times for three common types of graphs. For graphs where we are unable to explicitly compute the cover time, we can use a method called the spanning tree technique to find an upper bound. Lastly, we state some general bounds on cover times.

## 1 Introduction

When we're given a random walk starting from a specific vertex on a graph, there are some questions that we might want to ask: Does the random walk hit every vertex? If so, how long does it take? How does the starting vertex affect the time this takes? Are some vertices visited more often in the long term behavior of the random walk?

In this paper, the primary question we will focus on is the following:

*For a simple random walk $W_u$ on a simple, connected, undirected graph $G = (V, E)$, what is the expected number of steps required to visit all the vertices in G, maximized over all starting vertices u?*

## 2 Graph Theory Basics

As is usual for an expository paper, we must begin by confusing the reader with an extensive list of definitions.

First, we'll introduce graphs, because they are the topic of this paper.

**Definition 2.1.** A graph $G$ is a pair$(V, E)$, where $V$ is a set of vertices and $E$ is a set of edges.

In this paper, we will only consider graphs that meet the following conditions:

- **undirected:** edges can be traversed in both directions. This means that edges can be written as *unordered* pairs $(u, v)$, only once.
- **simple:** no edge goes from a vertex to itself, and every pair of vertices has at most one edge between them.
- **connected:** it is possible to reach any vertex from any other vertex by traversing some edges.

All graph vertices will be 0-indexed for notational convenience: this means that in a graph of $n$ vertices, we will number the vertices 0 through $n - 1$.

**Definition 2.2.** Each vertex $u$ has a **neighbor set** $N(u) \subseteq V$, consisting of all vertices that are connected to $u$ with an edge.

**Definition 2.3.** The **degree** of a vertex $u$ is the number of edges attached to $u$.

A well-known result in graph theory is the degree sum formula:

**Lemma 2.4** (Degree Sum Formula)**.** *In an undirected graph $G = (V, E)$,*

$$\sum_{u \in V} d(u) = 2|E|$$

*Proof.* The proof is by a simple double counting argument: the sum of the degrees of each vertex is equal to the number of vertex-edge pairs $(v, e)$ where $v$ is incident to $e$. Also note that each edge has two vertices incident to it. Therefore, the total number of $(v, e)$ pairs is also twice the number of edges. This completes the proof. ■

**Theorem 2.5.** *A random walk on an undirected graph $G = (V, E)$ that is finite, connected, and not bipartite converges to the stationary distribution $\pi$ where for any vertex $v \in V$, $\pi_v = \frac{d(v)}{2|E|}$.*

*Proof.* We know that

$$\sum_{v \in V} \pi_v \sum_{v \in V} \frac{d(V)}{2|E|} = 1$$

by the degree sum formula, therefore, $\pi$ is a proper probability distribution. If we let $P$ be the transition matrix of the random walk on $G$, then we can find the stationary distribution using $\pi = \pi P$, which we can write as

$$\pi_v = \sum_{u \in N(v)} \frac{d(u)}{2|E|} \frac{1}{d(u)} = \frac{d(v)}{2|E|}$$

The theorem follows from this. ■

## 3 Cover Times

Now, we will introduce some additional definitions relating specifically to random walks and cover times.

**Definition 3.1.** The **hitting time** $H[u, v]$ is the expected time it takes $W_u$ to visit vertex $v$.

**Definition 3.2.** The **first return time** $R[u] = H[u, u]$ is the expected time it takes $W_u$ to return to vertex $u$.

**Definition 3.3.** The **cover time from vertex** $u$ $COV_u[G]$ is the expected amount of time it takes the random walk $W_u$ to visit every vertex of $G$. The **cover time** $COV[G]$ is the maximum of $COV_u[G]$ over all starting vertices $u$.

We now introduce a result about stationary distributions of random walks.

**Theorem 3.4.** *In any finite, irreducible, and ergodic Markov chain, $\pi_j = \frac{1}{R[j]}$ for all $j$.*

With these results, we can proceed to calculating the cover times for three common types of graphs.

## 4 Complete Graph

The **complete graph** on $n$ vertices, denoted by $K_n$, is the graph such that $E = \{(u, v) : u, v \in V, u \neq v\}$, or in other words, every pair of distinct vertices is connected by an edge. In the complete graph, $|E| = \binom{n}{2} = \frac{n(n-1)}{2}$.

**Theorem 4.1.** *[1]*
*For the complete graph on $n$ vertices $K_n = (V, E)$,*

*(a) $R[u] = n$ for any $u \in V$.*

*(b) $H[u, v] = n - 1$ for any $u, v \in V$ such that $u \neq v$*

*(c) $COV[K_n] = (n - 1)h(n - 1)$ where $h(k)$ is the $h^{th}$ harmonic number.*

*Proof.*

(a) $|E| = \frac{n(n-1)}{2}$ and $d(u) = n - 1$ for any $u \in V$. By Theorem 2.5, the stationary distribution is uniform, therefore by Theorem 3.4, $R[u] = n$.

(b) We examine the transition probabilities of the walk: $P_{u,u} = 0$, and $P_{u,v} = \frac{1}{n-1}$ for all $u, v$ such that $u \neq v$. Because of symmetry, we can treat this as a geometric random variable, where reaching vertex $v$ is the goal. The expected value of a geometric distribution with probability $\frac{1}{n-1}$, and therefore $H[u, v]$, is $n - 1$.

(c) We define $T(r)$ to be the expected time to visit $r$ distinct vertices. By symmetry of a complete graph, it doesn't matter what vertex we start from or which vertices we visit along the way, only the number of vertices that have been visited so far. Suppose it visits the $r'$th distinct vertex at time $t$ for some $r < n$. Then, there are $n - r$ unvisited vertices, and the time to go to an unvisited vertex is yet again the expected value of a geometric distribution, this time with probability $\frac{n-r}{n-1}$, the expectation of which is $\frac{n-1}{n-r}$. By linearity of expectation, we then have

$$COV[K_n] = \sum_{r=1}^{n-1} \frac{n-1}{n-r} = (n-1)h(n-1).$$

∎

# 5 Path Graph

The **path graph**, denoted by $P_n$ on $n$ vertices is the graph such that $E = \{(1,2),(2,3),\ldots,(n-1,n)\}$. It has $|E| = |V| - 1 = n - 1$.

**Theorem 5.1.** *[1]*

(a) $R[1] = R[n] = 2(n-1)$, *and* $R[i] = n - 1$ *for* $0 < i < n - 1$

(b) $H[i,j] = |i^2 - j^2|$. *In particular,* $H[1,n] = (n-1)^2$.

(c) $COV[P_n] = \begin{cases} \frac{5(n-1)^2}{4} & \text{if n odd} \\ \frac{5(n-1)^2}{4} - \frac{1}{4} & \text{if n even} \end{cases}$

*Proof.*

(a) $d(1) = d(n) = 1$, and $d(i) = 2$ for all $0 < i < n-1$, and $|E| = n-1$. Applying Theorems 2.5 and 3.4 completes the proof.

(b) We can express $R[n]$ in two different ways. By $(a)$, we know that $R[n] = 2(n-1)$, but we also know that $R[n] = 1 + H[n-1,n]$, because if a random walk is on vertex $n$ then it can only move to $n-1$. Now, for any $0 < r < n$, we have $H[r-1,r] = 2r-1$ because this is equivalent to $H[r-1,r]$ on $P_{r+1}$. Without loss of generality, let $i \leq j$. By linearity of expectation:

$$H[i,j] = \sum_{r=i+1}^{j} H[r-1,r] = \sum_{r=i+1}^{j} 2r - 1 = (j+i+1)(j-i) - (j-i) = j^2 - i^2.$$

We simply take the absolute value to account for the possibility that $i > j$.

(c) We look at the case of $n+1$ to keep algebra simple. For $0 \leq r \leq n$, we let $a_r$ represent the expected amount of time a random walk starting from vertex $r$ requires to reach either end. Then, by basic transitions of the random walk, we have

$$a_r = \begin{cases} 0 & \text{if r = 0} \\ 1 + \frac{1}{2}a_{r-1} + \frac{1}{2}a_{r+1} & \text{if 0 < r < n-1} \\ 0 & \text{if r = n-1} \end{cases}$$

The solution to this system is $a_r = r(n-r)$: We have $a_0 = 0(n-0) = 0$ and $a_n = n(n-n) = 0$, as desired. Furthermore, $a_r - 1 - \frac{1}{2}a_{r-1} - \frac{1}{2}a_{r+1} = r(n-r) - 1 - \frac{1}{2}[(r-1)(n-(r-1)) + (r+1)(n-(r+1))] = 0$. Now we can calculate the cover time. Starting from any vertex, in order to cover the entire graph, the random walk must first reach one endpoint, then reach the other. From what we calculated previously,

$$COV_r[P_{n+1}] = a_r + H[0,n] = r(n-r) + n^2.$$

$COV[P_{n+1}]$ is the maximum over all $r$: when $n$ is even, this is attained at $r = \frac{n}{2}$, where $COV[P_{n+1}] = \frac{5n^2}{4}$. When $n$ is odd, this is attained at $r = \lfloor \frac{n}{2} \rfloor$ or $r = \lceil \frac{n}{2} \rceil$, in which case $COV[P_{n+1}] = \frac{5n^2}{4} - \frac{1}{4}$. However, because we calculated for $P_{n+1}$, we must replace the $n$'s with $n-1$'s to find the formula for $P_n$. This completes the proof.

∎

## 6 Cycle Graph

The **cycle graph** on $n$ vertices, denoted by $\mathbb{Z}_n$ is the graph such that $E = \{(1, 2), (2, 3), \ldots, (n-1, n), (n, 1)\}$. It is $P_n$ with an additional edge connecting vertex $n$ to vertex 1. In the cycle graph, $|E| = |V| = n$.

**Theorem 6.1.** *[1]*

(a) $R[u] = n$ *for any vertex* $u$.

(b) *For a pair of vertices* $u, v$ *such that the shorter distance between them is* $r \leq \frac{n}{2}$*, we have* $H[u, v] = r(n - r)$

(c) $COV[\mathbb{Z}_n] = \frac{n(n-1)}{2}$

*Proof.*

(a) By symmetry, and Theorem 2.5, we know that $\pi_u = \frac{1}{n}$ for all $n$. Applying Theorem 3.4 yields the desired result.

(b) We label the vertices of $\mathbb{Z}_n$ $(0, \ldots, n-1)$ in clockwise (arbitrarily) order. Then, we want to find $H[r, 0]$ for some $r$. On one path, the distance between $r$ and 0 is $r$, on the other path it is $n - r$. We can apply the same method we used in the proof of Theorem 5.1 to determine that $H[u, v] = r(n - r)$.

(c) The critical observation here is that the set of vertices the random walk has visited is always a single connected component. Let $T(r)$ be the expected time it takes a walk starting at some vertex to visit $r$ vertices of $\mathbb{Z}_n$. We can write

$$COV[\mathbb{Z}_n] = (T(n) - T(n-1)) + (T(n-1) - T(n-2) + \cdots + (T(2) - T(1)) + (T(1) - T(0)).$$

Suppose at some point, the random walk has visited $r < n$ different vertices. Then, without loss of generality, the visited vertices can be labeled $(1, 2, \ldots, r)$, and then $(r+1, r+2, \ldots, n-1, 0)$ for the unvisited vertices. The next time this walk reaches a new vertex, the new vertex is either $r + 1$ or 0. Therefore, $T(r+1) - T(r)$ is the same as the expected time for a walk on $P_n$ to reach an endpoint, starting on vertex $r$. This is just $r$. Summing all of the terms, we get

$$COV[\mathbb{Z}_n] = \sum_{r=0}^{n-1} r = \frac{n(n-1)}{2},$$

as desired.

■

## 7 Upper Bound with Spanning Tree Technique

Unfortunately, most other classes of graphs are not as nice, and we can't explicitly calculate their cover times. Instead, we find bounds. Here, we will discuss the spanning tree technique for finding an upper bound on the cover times.

First, we must introduce some necessary definitions:

**Definition 7.1.** A **tree** is an undirected graph that is connected and acyclic. All trees satisfy $|E| = |V| - 1$.

**Definition 7.2.** A **spanning tree** of a graph $G$ is a subgraph that includes all the vertices of $G$ but includes only the minimum possible number of edges to ensure connectivity. A graph can have multiple different spanning trees.

Trees have many useful properties that we will take advantage of. One particular result of interest is the following:

**Proposition 7.3.** *In a tree* $T = (V, E)$*, there exists a walk* $\sigma = (v_0, v_1, \ldots, v_{2|V|-2})$ *such that each edge of* $T$ *is traversed once in each direction.*

*Proof.* The depth-first search started from any arbitrary vertex suffices. Refer to [4] for a description of the depth-first search algorithm. ■

Given this walk, we can upper bound the cover time as follows:

$$COV[G] \leq \sum_{i=0}^{2|V|-3} H[v_i, v_{i+1}].$$

Now, we can present our main result.

**Theorem 7.4.** *[8] For a graph $G$, such that $n = |V|$ and $m = |E|$, we have*

$$COV[G] < 4mn$$

*Proof.* By Theorems 3.4 and 2.5, we have

$$R[v] = \frac{2m}{d(v)}$$

for any $v \in V$. We can also write

$$R[v] = \sum_{u \in N(v)} \frac{1}{d(v)}(1 + H[u, v])$$

If we set the two equations equal to each other, we have

$$\frac{2m}{d(v)} = \sum_{u \in N(v)} \frac{1}{d(v)}(1 + H[u, v]).$$

Simplifying, we have the bound $H[u, v] < 2m$.

Now, since the walk $\sigma$ is $2n - 2$ steps long, we have

$$COV[G] \leq \sum_{i=0}^{2n-3} H[v_i, v_{i+1}] \leq 2m(2n - 2) < 4mn.$$

■

## 8 General Results

In this section, we will present several results regarding cover times on graphs in the general case. The proofs are rather long and arduous, so we will omit them.

First, we have the upper and lower bounds from Matthews' technique; this method uses electrical network theory and yields a bound within a constant factor of the precise value.

**Theorem 8.1** (Matthews' Upper Bound [7]). *For a graph $G = (V, E)$,*

$$COV[G] \leq H^*[G]h(n)$$

*where $H^*[G] = max_{u,v \in V} H[u, v]$ and $h(n)$ is the $n'$th harmonic number.*

**Theorem 8.2** (Matthews' Lower Bound [7]). *For a graph $G = (V, E)$,*

$$COV[G] \geq max_{A \in V} H_*[A]h(|A| - 1)$$

*where $H_*[A] = min_{u,v \in A, u \neq V} H[u, v]$.*

We also have some asymptotic general bounds on cover times. [1]

**Theorem 8.3** (Asymptotic Upper Bound [5]). *For any connected graph $G$ on $n$ vertices,*

$$COV[G] < CyCOG[G] \leq (1 + o(1))\frac{4}{27}n^3$$

*where $CyCOG[G]$ is the cyclic cover time: the expected time it takes to visit all the graph vertices in a specific cyclic order, minimized over all cyclic orders.*

**Theorem 8.4** (Asymptotic Lower Bound [6]). *For any connected graph $G$ on $n$ vertices, and any starting vertex $u \in V$:*

$$COV[G] \geq COV_u[G] \geq (1 + o(1))n \log n.$$

[1] Logarithms are base $e$ unless otherwise stated.

## References


[1] M. Abdullah: The Cover Time of Random Walks on Graphs. King's College London. `https://arxiv.org/pdf/1202.5569.pdf`

[2] A. Levin, Y. Peres, E. L. Wilmer: Markov Chains and Mixing Times. `https://pages.uoregon.edu/dlevin/MARKOV/markovmixing.pdf`

[3] R. Rubinfeld: Randomness and Computation. Massachusetts Institute of Technology. `https://people.csail.mit.edu/ronitt/COURSE/S14/Handouts/Lec_6_scribe.pdf`

[4] J. Kleinberg, E. Tardos, Algorithm Design, Addison Wesley (2005).

[5] U. Feige, A tight upper bound for the cover time of random walks on graphs. Random Structures and Algorithms, 6 51-54 (1995).

[6] U. Feige, A tight lower bound for the cover time of random walks on graphs. Random Structures and Algorithms, 6 433-438 (1995).

[7] P. Matthews, Covering problems for Brownian motion on spheres. Ann. Prob., 16:189-199 (1988).

[8] M. Mitzenmacher, E. Upfal Probability and Computing: Randomized Algorithms and Probabilistic Analysis, Cambridge University Press (2005).