Randomized Algorithms

Random Walks on Undirected Graphs

Joseph Chuang-Chieh Lin

Department of Computer Science & Engineering,
National Taiwan Ocean University

Spring 2026

Outline

1. The random-walk model
2. Stationary distribution
3. Hitting and commute times
4. Cover time
5. Application: s–t connectivity

Why random walks on graphs?

- A random walk is a Markov chain whose states are vertices of a graph.
- It is a simple model of local exploration: from the current vertex, choose a random neighbor.
- In this lecture note, we use graph structure to bound:
  - stationary distributions,
  - hitting and commute times,
  - cover times,
  - and a randomized low-space algorithm for s–t connectivity.

Guiding question

How long does local random exploration need to see a target vertex, or even all vertices?

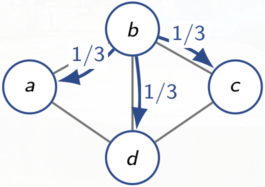
Random walk on an undirected graph

Let $G = (V, E)$ be a finite undirected graph.

Definition 1

At vertex u, the standard random walk moves to a uniformly chosen neighbor:

$$P_{u,v} = \begin{cases} 1/d(u), & (u,v) \in E, \\ 0, & \text{otherwise.} \end{cases}$$



The transition matrix is local

- The transition matrix P has one row per vertex.
- Row u has nonzero entries only at neighbors of u.
- The degree $d(u)$ controls the transition probabilities leaving u.

Example

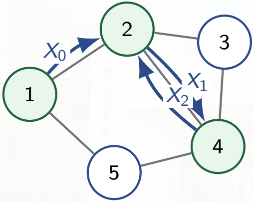
If $N(u) = \{v_1, v_2, v_3\}$, then

$$P_{u,v_1} = P_{u,v_2} = P_{u,v_3} = \frac{1}{3}.$$

Memoryless property

Given the current vertex, the previous trajectory carries no additional information for choosing the next vertex.

A sample walk



A walk is a random sequence

$$X_0, X_1, X_2, \ldots$$

with

$$\Pr[X_{t+1} = v \mid X_t = u] = P_{u,v}.$$

Note

The graph is fixed; the randomness is in the walk.

Structural assumptions

In the following discussions, we assume G is:

- finite,
- undirected,
- connected,
- non-bipartite.

Why?

Structural assumptions

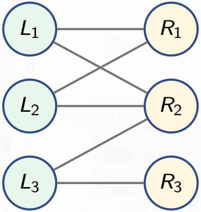
In the following discussions, we assume G is:

- finite,
- undirected,
- connected,
- non-bipartite.

Why?

- Connected $\Rightarrow$ every vertex can be reached from every other vertex.
- Non-bipartite $\Rightarrow$ the walk is aperiodic.
- Together, the *finite* Markov chain converges to a stationary distribution.

Bipartite graphs cause period two



bipartite: alternate sides

In a bipartite graph, every step switches side.

$$L \to R \to L \to R \to \cdots$$

So a return to the starting vertex can happen only at even times.

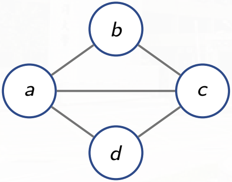
Consequence

The period is 2, so the chain is not aperiodic.

Non-bipartite means aperiodic

Lemma 1

A random walk on an undirected graph G is aperiodic if and only if G is not bipartite.



odd cycle present

- Every undirected edge gives a return of length 2: $v \to u \to v$.
- An odd cycle gives an odd return length.
- Having return lengths with gcd 1 means period 1 (i.e., *aperiodic*).

The stationary distribution

Theorem 1

For a finite, connected, non-bipartite undirected graph,

$$\pi_v = \frac{d(v)}{2|E|}$$

is the stationary distribution $\bar{\pi}$ of the simple random walk.

The stationary distribution

Theorem 1

For a finite, connected, non-bipartite undirected graph,

$$\pi_v = \frac{d(v)}{2|E|}$$

is the stationary distribution $\bar{\pi}$ of the simple random walk.

The walk spends more time at high-degree vertices, because they are incident to more directed edge-ends.

The stationary distribution

Theorem 1

For a finite, connected, non-bipartite undirected graph,

$$\pi_v = \frac{d(v)}{2|E|}$$

is the stationary distribution $\bar{\pi}$ of the simple random walk.

The walk spends more time at high-degree vertices, because they are incident to more directed edge-ends.

The “distribution” is proper:

The stationary distribution

Theorem 1

For a finite, connected, non-bipartite undirected graph,

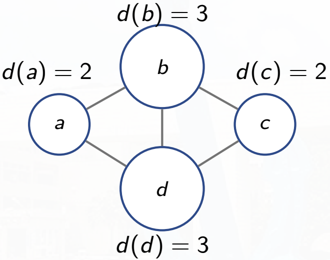
$$\pi_v = \frac{d(v)}{2|E|}$$

is the stationary distribution $\bar{\pi}$ of the simple random walk.

The walk spends more time at high-degree vertices, because they are incident to more directed edge-ends.

The "distribution" is proper: $\displaystyle\sum_{v \in V} \pi_v = \sum_{v \in V} \frac{d(v)}{2|E|} = 1.$

Stationary distribution: a concrete picture



$|E| = 5$, so $2|E| = 10$

$$\pi_a = \pi_c = \frac{2}{10}, \qquad \pi_b = \pi_d = \frac{3}{10}.$$

Stationarity guarantee

For every edge $(u, v) \in E$,

$$\pi_u P_{u,v} = \frac{d(u)}{2|E|} \cdot \frac{1}{d(u)} = \frac{1}{2|E|}.$$

Similarly, $\pi_v P_{v,u} = \frac{1}{2|E|}$. Thus,

$$\pi_u P_{u,v} = \pi_v P_{v,u}.$$

Stationarity guarantee

For every edge $(u, v) \in E$,

$$\pi_u P_{u,v} = \frac{d(u)}{2|E|} \cdot \frac{1}{d(u)} = \frac{1}{2|E|}.$$

Similarly, $\pi_v P_{v,u} = \frac{1}{2|E|}$. Thus,

$$\pi_u P_{u,v} = \pi_v P_{v,u}.$$

Let $N(v)$ denote the neighbors of v. The relation $\bar{\pi} = \bar{\pi} P$ is equivalent to

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

Hitting time, commute time, cover time

Definition 2

For vertices $u, v \in V$, let

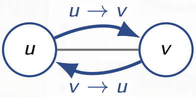
$$h_{u,v} = \mathbb{E}[\text{time to first reach } v \mid X_0 = u].$$

This is the *hitting time* from u to v.

Two related quantities

- **Commute time:**
$$h_{u,v} + h_{v,u}.$$
- **Cover time:** maximum expected time, over starting vertices, to visit every vertex.

Directed-edge viewpoint



Replace each undirected edge by two directed edges.

- There are $2|E|$ directed edges.
- The edge traversed at time t is a Markov-chain state.
- In stationarity, each directed edge has probability $1/(2|E|)$. [Exercise]

Adjacent commute time

Lemma 2

If $(u, v) \in E$, then
$$h_{u,v} + h_{v,u} \leq 2|E|.$$

Adjacent commute time

Lemma 2

If $(u, v) \in E$, then

$$h_{u,v} + h_{v,u} \leq 2|E|.$$

Core idea

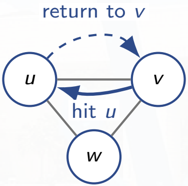
The expected return time to a directed edge $u \to v$ is

$$\frac{1}{1/(2|E|)} = 2|E|.$$

Returning to the directed edge $u \to v$ requires going from v to u and then back to v through that edge.

Why this bounds the whole commute

Suppose the walk has just traversed $u \to v$, so it is currently at v (state $u \to v$).



specific return via $u \to v$ is only one possible way

$$h_{v,u} + h_{u,v} \leq \mathbb{E}[\text{additional time until next traversal of } u \to v] = 2|E|.$$

A useful caution

Why the sum appears

The return time to a directed edge controls a complete round trip:

$$v \rightsquigarrow u \rightsquigarrow v.$$

The first traversal has already occurred; the bound concerns the additional time until the next one.

Cover time

Definition 3

The cover time C_G is

$$C_G = \max_{v \in V} \mathbb{E}[\text{time to visit all vertices} \mid X_0 = v].$$

Goal

Bound C_G using simpler quantities such as:

- number of edges $|E|$,
- number of vertices $|V|$,
- maximum hitting time $\max_{u \neq v} h_{u,v}$.

Spanning-tree tour bound

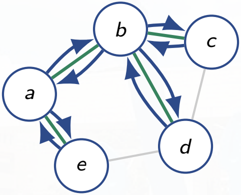
Lemma 3

For a connected undirected graph $G = (V, E)$,

$$C_G \leq 2|E|(|V| - 1).$$

- Choose any spanning tree T of G.
- A depth-first tour crosses each tree edge once in each direction.
- Each tree-edge commute costs at most $2|E|$ by Lemma 2.
- There are $|V| - 1$ tree edges.

Spanning tree tour picture



DFS tour of spanning tree: each tree edge twice

$$\sum_{(x,y)\in T} \left(h_{x,y} + h_{y,x}\right) \leq 2|E|(|V|-1).$$

Matthews' theorem

Lemma 4

Let G have n vertices. Then

$$C_G \leq H(n-1) \max_{u,v \in V : u \neq v} h_{u,v},$$

where

$$H(n-1) = \sum_{i=1}^{n-1} \frac{1}{i} \approx \ln n.$$

Matthews' theorem

Lemma 4

Let G have n vertices. Then

$$C_G \leq H(n-1) \max_{u,v \in V : u \neq v} h_{u,v},$$

where

$$H(n-1) = \sum_{i=1}^{n-1} \frac{1}{i} \approx \ln n.$$

Why and when is it useful?

If all pairwise hitting times are small, the cover time is like coupon collector with hitting-time cost.

Matthews' theorem: proof idea (1/3)

- Let $B = \max_{u \neq v} h_{u,v}$.
- Choose a uniformly random ordering of the vertices: $Z_1, Z_2, \ldots, Z_n$.
- Let T_j be the first time by which all of $Z_1, Z_2, \ldots, Z_j$ have been visited.
- For $j \geq 2$, the extra waiting interval satisfies

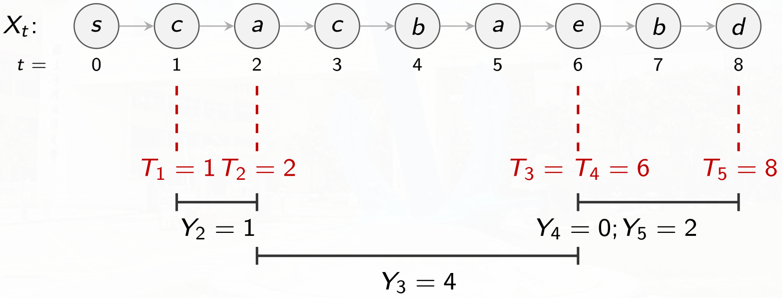
$$\mathbb{E}[T_j - T_{j-1}] \leq \frac{B}{j}. \quad \text{(Why?)}$$

  Summing gives a harmonic factor.
- Say the chain's history is given by $X_1, X_2, \ldots$. For $j \geq 2$, we consider

$$Y_j = \mathbb{E}[T_j - T_{j-1} \mid Z_1, Z_2, \ldots, Z_j; X_1, X_2, \ldots, X_{T_{j-1}}].$$

$$Z = (Z_1, Z_2, \ldots, Z_5) = (c, a, e, b, d), \qquad X = (X_0, X_1, \ldots, X_8) = (s, c, a, c, b, a, e, b, d).$$

Z_i order: $\boxed{Z_1 = c} \to \boxed{Z_2 = a} \to \boxed{Z_3 = e} \to \boxed{Z_4 = b} \to \boxed{Z_5 = d}$



- T_j: first time by which all of $Z_1, Z_2, \ldots, Z_j$ have been visited.
- Here b is visited at time 4, before e at time 6. Hence

$$T_4 = T_3 = 6, \qquad Y_4 = 0.$$

Matthews' theorem: proof idea (2/3)

The chain's history is given by $X_1, X_2, \ldots$. For $j \geq 2$, we consider

$$Y_j = \mathbb{E}[T_j - T_{j-1} \mid Z_1, Z_2, \ldots, Z_j; X_1, X_2, \ldots, X_{T_{j-1}}].$$

- **Note:** During the random walk, these vertices $Z_1, Z_2, \ldots$ are not necessarily visited in this same order.

Matthews' theorem: proof idea (2/3)

The chain's history is given by $X_1, X_2, \ldots$. For $j \geq 2$, we consider

$$Y_j = \mathbb{E}[T_j - T_{j-1} \mid Z_1, Z_2, \ldots, Z_j; X_1, X_2, \ldots, X_{T_{j-1}}].$$

- **Note:** During the random walk, these vertices $Z_1, Z_2, \ldots$ are not necessarily visited in this same order.

The expected number to cover the graph starting from vertex u:

$$\sum_{j=2} Y_j + \mathbb{E}[T_1].$$

Matthews' theorem: proof idea (2/3)

The chain's history is given by $X_1, X_2, \ldots$. For $j \geq 2$, we consider

$$Y_j = \mathbb{E}[T_j - T_{j-1} \mid Z_1, Z_2, \ldots, Z_j; X_1, X_2, \ldots, X_{T_{j-1}}].$$

- **Note:** During the random walk, these vertices $Z_1, Z_2, \ldots$ are not necessarily visited in this same order.

The expected number to cover the graph starting from vertex u:

$$\sum_{j=2} Y_j + \mathbb{E}[T_1].$$

- If $Z_1 = u$, which happens w.p. $\frac{1}{n}$, then $T_1 = 0$.
- Otherwise, $\mathbb{E}[T_1 \mid Z_1] = h_{u,Z_1} \leq B$.
- Hence, $\mathbb{E}[T_1] = (1 - \frac{1}{n})B$.

Matthews' theorem: proof idea (3/3)

For the Y_i's:

- If Z_j is not the last vertex seen from $Z_1, Z_2, \ldots, Z_j$, then $Y_i = 0$ since $T_j = T_{j-1}$.
- If Z_j is the last vertex seen from $Z_1, Z_2, \ldots, Z_j$, then $Y_j \leq B$.
  - $\because$ Y_j: the hitting time h_{Z_k, Z_j} (say Z_k was visited last out of $Z_1, Z_2, \ldots, Z_{j-1}$ during the random walk).
  - Z_j was chosen according to a random permutation, independent of the random walk, so Z_j is last out of $Z_1, Z_2, \ldots, Z_{j-1}$ with probability $1/j$.

Matthews' theorem: proof idea (3/3)

For the Y_i's:

- If Z_j is not the last vertex seen from $Z_1, Z_2, \ldots, Z_j$, then $Y_i = 0$ since $T_j = T_{j-1}$.
- If Z_j is the last vertex seen from $Z_1, Z_2, \ldots, Z_j$, then $Y_j \leq B$.
  - $\because$ Y_j: the hitting time h_{Z_k,Z_j} (say Z_k was visited last out of $Z_1, Z_2, \ldots, Z_{j-1}$ during the random walk).
  - Z_j was chosen according to a random permutation, independent of the random walk, so Z_j is last out of $Z_1, Z_2, \ldots, Z_{j-1}$ with probability $1/j$.

$$\sum_{j=2}^{n} Y_j + \mathbb{E}[T_1] \leq \sum_{j=2}^{n} \frac{1}{j} B + \left(1 - \frac{1}{n}\right) B = \left(1 + \sum_{j=2}^{n} \frac{1}{j}\right) B - \frac{1}{n} B$$

$$= H(n-1)B.$$

Lemma 3 vs. Lemma 4

	Lemma 3	Lemma 4
Input needed	$\|E\|$ and $\|V\|$	maximum hitting time
Bound	$2\|E\|(\|V\|-1)$	$H(n-1)\max_{u\neq v} h_{u,v}$
Method	spanning-tree tour	random ordering / coupon collector
Best when	no hitting-time info	good hitting-time estimates exist

Note

Lemma 3: coarse but always available.

Lemma 4: can be sharper when the graph mixes or spreads quickly.

s–t connectivity with tiny memory

Given an undirected graph $G = (V, E)$ and two vertices s, t, decide whether s and t are connected.

- Breadth-first search or depth-first search solves this in linear time.
- But such deterministic graph search needs $\Theta(n)$ space to **remember visited vertices**.
    - Can be done in $O(n + m)$ time by BFS/DFS.
- The random-walk algorithm uses only the current vertex and a step counter.

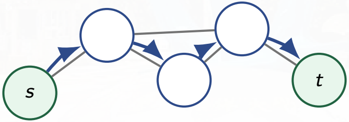
Goal

Trade time and error probability for very small space: $O(\log n)$ bits.

Algorithm 1: Testing connectivity by random-walks

s–t connectivity algorithm

1. Start a random walk from s.
2. If the walk reaches t within $2n^3$ steps, return “connected.”
3. Otherwise return “not connected.”



Correctness and one-sided error

- If there is no path from s to t, the walk can never reach t.
- Therefore returning "connected" is always correct.
- The only possible error is saying "not connected" even though a path exists.

Bounding the error

If s and t are connected, the expected time to reach t from s is at most the cover time of their component.

$$\mathbb{E}[T_{s\to t}] \le C_G \le 2nm < n^3.$$

Then Markov's inequality gives

$$\Pr[T_{s\to t} > 2n^3] \le \frac{1}{2}.$$

Space usage and amplification

- Store current vertex: $O(\log n)$ bits.
- Store step counter up to $2n^3$: $O(\log n)$ bits.
- Choose a random neighbor at each step.

Amplification

Repeat the random walk independently r times.

$$\Pr[\text{“false negative after } r \text{ runs”}] \leq 2^{-r}.$$

The space remains $O(\log n)$; the time grows by a factor of r.

Takeaways (1/2)

- Random walks on connected non-bipartite undirected graphs converge to
$$\pi_v = \frac{d(v)}{2|E|}.$$
- Directed-edge stationarity gives the adjacent commute-time bound
$$h_{u,v} + h_{v,u} \leq 2|E|.$$
- Cover time can be bounded by either spanning-tree tours or Matthews' theorem.

Takeaways (2/2)

- The random-walk s–t connectivity algorithm is simple, one-sided, and uses only $O(\log n)$ space.
- In fact, Reingold's theorem later showed that even the randomization can be removed. [Omer Reingold J. ACM 2008]

Discussions