TTIC 31150/CMSC 31150
Mathematical Toolkit (Spring 2023)

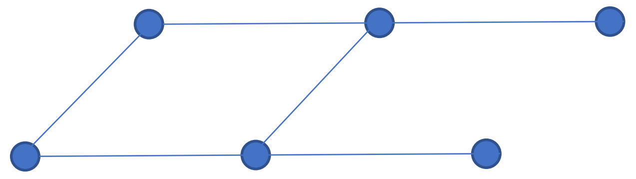
Avrim Blum and Ali Vakilian

Lecture 16: Random walks on graphs

Recap

- Randomized algorithms for routing to minimize congestion. Randomized complexity classes **RP** and **BPP**, connections to **P/poly**.
- Probability distributions over uncountably infinite spaces, σ-field (σ-algebra), measurability, random variables, CDFs and density functions.
- Gaussian random variables, properties. d-dimensional Gaussians.
- Dimensionality reduction and the Johnson-Lindenstrauss Lemma.
- Tail bounds for sum of independent squared-Gaussian RVs.
- Properties of the unit ball in high dimensions (volume, most points near the surface, most points near the equator, …)

Random Walks on Graphs



Imagine you are lost in a maze. How long will it take you to get out if you just walk around randomly?

General setup: Underlying undirected graph $G = (V, E)$, with n vertices and m edges. Starting from some initial vertex, at each step we move to a random neighbor of current node.

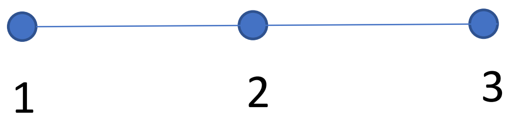
Quantities of interest:

- Hitting time H_{uv}: defined as $\mathbb{E}$[number of steps to reach v | start at u].
- Commute time C_{uv}: $\mathbb{E}$[number of steps to reach v and return to u| start at u].

  Let X_{uv}^{hit} = #steps to reach v starting from u. So, $H_{uv} = \mathbb{E}\left[X_{uv}^{hit}\right]$, $C_{uv} = \mathbb{E}\left[X_{uv}^{hit} + X_{vu}^{hit}\right] = H_{uv} + H_{vu}$.

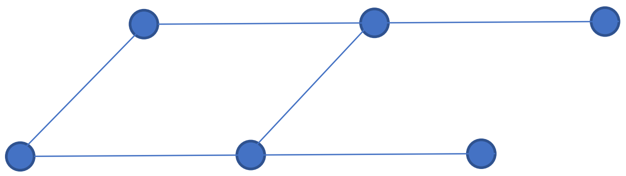
- Cover time from u, Cov_u: $\mathbb{E}$[number of steps to visit all of G| start at u].
- Cover time of G, Cov_G: $\max_u Cov_u$.

Example



H_{12}? H_{21}? H_{31}? Cov_G?

Cover-Time Theorem

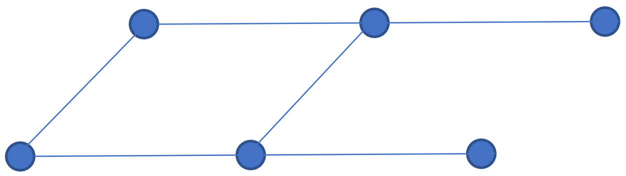


Theorem: If G is a connected graph with n vertices and m edges, then $Cov_G \leq 2m(n-1)$.

So, if you're lost in a maze and walk around randomly, you will visit all the nodes (and hence, the exit) in $O(mn)$ steps.

- On a line, this is tight: it really does take $\Theta(n^2)$ steps in expectation for a random walk to visit all the nodes.
- For some graphs, it is not tight. E.g., for a clique, the cover time is only $O(n \log n)$. Can you see why?
- An example of a graph where cover time is $\Omega(n^3)$ is "lollipop graph": clique of size $n/2$ connected to line of length $n/2$.

Cover-Time Theorem



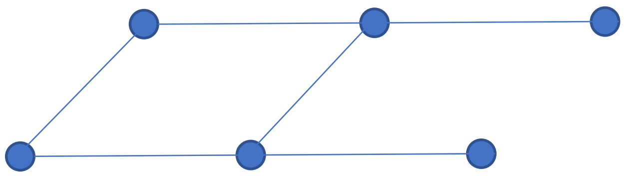
Theorem: If G is a connected graph with n vertices and m edges, then $Cov_G \leq 2m(n-1)$.

For convenience, lets consider current state as being on some edge $\{u, v\}$ headed in some direction (e.g., to v). The theorem will follow from the following key lemma:

Lemma: for any edge/direction (u, v), the expected number of steps between consecutive visits to (u, v) is $2m$.

Note that the lemma implies that if u and v are neighbors, then $C_{vu} \leq 2m$, because expected time $v \rightarrow u \rightarrow v$ is $\leq$ expected time starting from v to take (u, v) edge.

Cover-Time Theorem



Theorem: If G is a connected graph with n vertices and m edges, then $Cov_G \leq 2m(n-1)$.

For convenience, lets consider current state as being on some edge $\{u, v\}$ headed in some direction (e.g., to v). The theorem will follow from the following key lemma:

Lemma: for any edge/direction (u, v), the expected number of steps between consecutive visits to (u, v) is $2m$.

Proof of Theorem from Lemma:

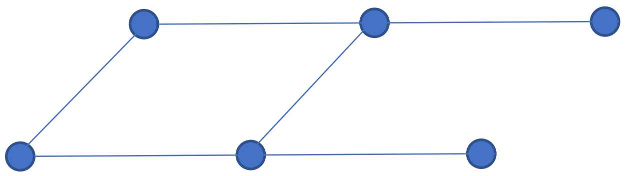
- Consider some spanning tree T of G and some fixed tour of T.

$\mathbb{E}$[time to visit G] $\leq$ $\mathbb{E}$[time to visit nodes in that order]

I.e., E[time from node 1 until visit node 2] + E[time from node 2 until visit node 3] + …

$$= \sum_{\{u,v\} \in T} H_{uv} + H_{vu} \quad = \sum_{\{u,v\} \in T} C_{uv} \quad = 2m(n-1)$$

Proof of key lemma



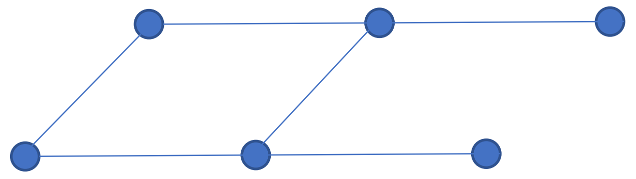
Lemma: for any edge/direction (u, v), the expected number of steps between consecutive visits to (u, v) is $2m$.

First:

- Suppose we started by picking an edge and direction uniformly at random (so our initial distribution has probability $\frac{1}{2m}$ on each directed edge). What does our distribution look like after 1 step?
- Answer: the same. (I.e., this is a stationary distribution)
- For any edge/dir (v, w), $\Pr[\text{on } (v, w) \text{ after 1 step}] = \sum_{u:\{u,v\}\in E} \Pr[\text{was on } (u, v)] \cdot \frac{1}{\deg(v)}$

$$= \frac{\deg(v)}{2m} \cdot \frac{1}{\deg(v)} = \frac{1}{2m}.$$

Proof of key lemma



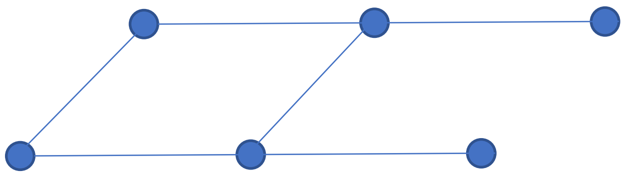
Lemma: for any edge/direction (u, v), the expected number of steps between consecutive visits to (u, v) is $2m$.

First:

- Suppose we started by picking an edge and direction uniformly at random (so our initial distribution has probability $\frac{1}{2m}$ on each directed edge). What does our distribution look like after 1 step?
- Answer: the same. (I.e., this is a stationary distribution)
- So, this means that for any directed edge (u, v), in T steps the expected number of traversals of that edge is $\frac{T}{2m}$, by linearity of expectation.

To prove the lemma, we want to invert this, to say that the expected gap between consecutive traversals is $2m$.

Proof of key lemma



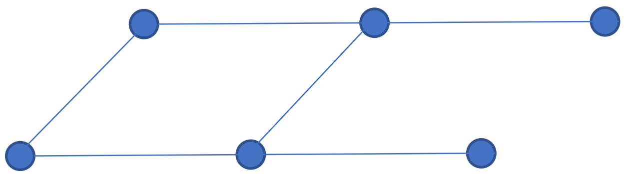
Lemma: for any edge/direction (u, v), the expected number of steps between consecutive visits to (u, v) is $2m$.

Note that if our positions at different times t were independent, then this would follow immediately from the fact that the expected value of a Geometric(p) R.V. is $1/p$.

However, these positions are not independent, so we need to be careful. E.g., if the graph consisted of two disconnected pieces with $m/2$ edges each, then the expected time between consecutive traversals would be m but the expected time to the first traversal would be infinite.

To prove the lemma, we want to invert this, to say that the expected gap between consecutive traversals is $2m$.

Proof of key lemma



Lemma: for any edge/direction (u, v), the expected number of steps between consecutive visits to (u, v) is $2m$.

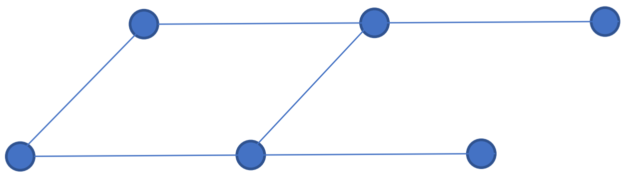
Note that if our positions at different times t were independent, then this would follow immediately from the fact that the expected value of a Geometric(p) R.V. is $1/p$.

However, these positions are not independent, so we need to be careful. E.g., if the graph consisted of two disconnected pieces with $m/2$ edges each, then the expected time between consecutive traversals would be m but the expected time to the first traversal would be infinite.

Still, the Geometric RV intuition turns out to be the right one.

To prove the lemma, we want to invert this, to say that the expected gap between consecutive traversals is $2m$.

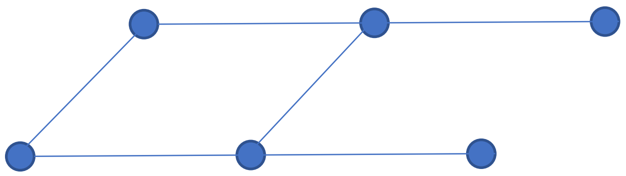
Proof of key lemma



Lemma: for any edge/direction (u, v), the expected number of steps between consecutive visits to (u, v) is $2m$.

- Consider our random walk process starting from the uniform distribution. Let X_1 be an RV denoting the time until we first reach (u, v). Then let X_2 denote the time from that point until our 2nd traversal of (u, v), etc.
- Because the graph is connected, we will indeed reach (u, v) with probability 1.
- In fact, these R.V.s have bounded variance:
  - Wherever you are, there is at least some (perhaps exponentially small) $\delta > 0$ probability that you reach (u, v) in the next n steps.
  - So, our process is dominated by n times a Geometric(δ) RV, which has finite variance.
- As $T \to \infty$, with probability 1 the number of traversals $N \to \infty$ too.

Proof of key lemma



Lemma: for any edge/direction (u, v), the expected number of steps between consecutive visits to (u, v) is $2m$.

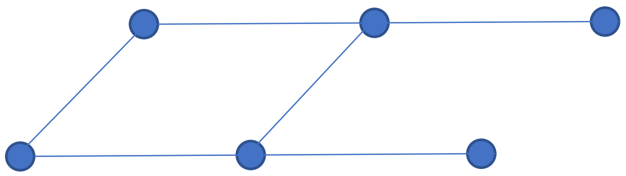
- Consider our random walk process starting from the uniform distribution. Let X_1 be an RV denoting the time until we first reach (u, v). Then let X_2 denote the time from that point until our 2nd traversal of (u, v), etc.

Now, let's apply Chebyshev to $X = \frac{X_1 + \cdots + X_N}{N}$. Let σ^2 be upper bound on $Var[X_i]$.

- Since X_i are independent, $Var[X] \leq \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N}$. So, $\mathbb{P}[|X - \mathbb{E}[X]| \geq \epsilon] \leq \frac{\sigma^2}{N\epsilon^2}$.
- So, for large N, whp the observed average gap length X is close to its expectation.

$$\ldots\ldots\ldots\ldots\; u - v \ldots\ldots\ldots\ldots\ldots u - v \ldots . u - v \ldots\ldots\ldots\ldots\ldots\ldots u - v \ldots\ldots\ldots\ldots u - v \ldots .$$

Proof of key lemma



Lemma: for any edge/direction (u, v), the expected number of steps between consecutive visits to (u, v) is $2m$.

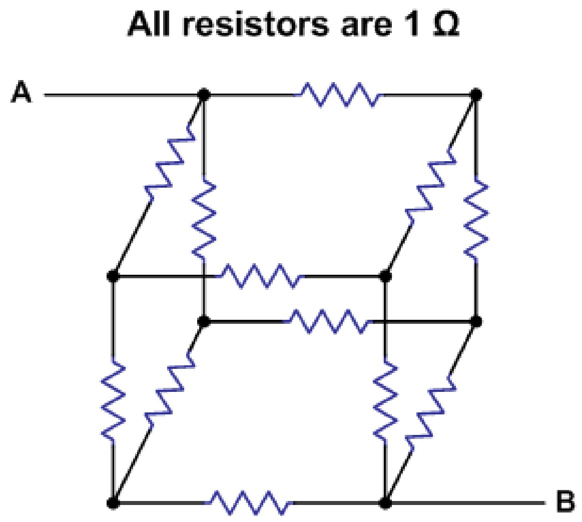
This means that the fraction of time steps that are traversals of (u, v), namely $1/X$, is also whp multiplicatively close to $1/\mathbb{E}[X]$.

We know the expected fraction of time steps that are traversals is $\frac{1}{2m}$. And if a bounded RV is concentrated, it has to concentrate about its expectation. So, $\mathbb{E}[X] = 2m$.

- Since X_i are independent, $Var[X] \leq \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N}$. So, $\mathbb{P}[|X - \mathbb{E}[X]| \geq \epsilon] \leq \frac{\sigma^2}{N\epsilon^2}$.
- So, for large N, whp the observed average gap length X is close to its expectation.

$$\ldots\ldots\ldots\ldots u - v \ldots\ldots\ldots\ldots\ldots u - v \ldots. u - v \ldots\ldots\ldots\ldots\ldots\ldots u - v \ldots\ldots\ldots\ldots u - v \ldots.$$

Something completely different(?): electrical networks

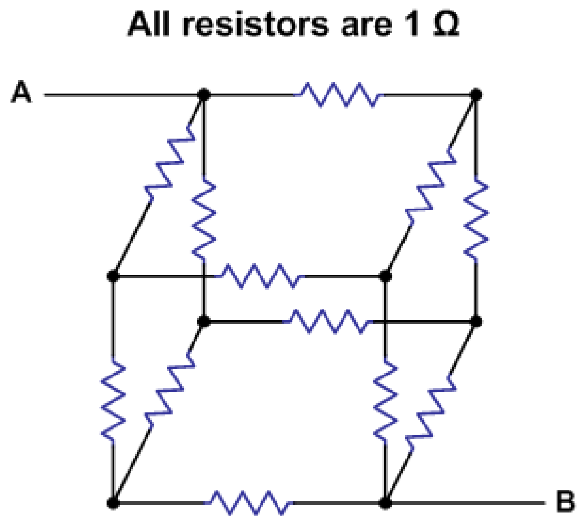


Consider a graph G where on each edge we have a resistor of some resistance.

- Say we connect a battery of some voltage V_{batt} between two nodes A and B (so $V_A - V_B = V_{batt}$, and let's for convenience say $V_B = 0$).
- Then each node in the graph will have a voltage (also called "potential") and each edge will have some current flowing in some direction.

Can think of voltage as like "height", and resistors like little water wheels or filters.

Something completely different(?): electrical networks



Voltages and currents can be computed using the following two rules.

- Kirchoff's law: current is like water flow: for any node not connected to the battery, flow in = flow out.
- Ohm's law: $V = IR$. Here, R is resistance, V is the voltage drop, and I is the current flow.

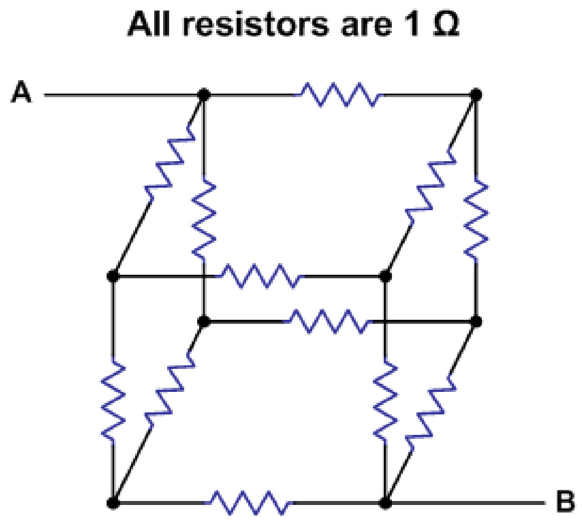
Effective resistance R_{uv} between u and v: connect up battery, measure current, $R_{uv} = \frac{V}{I}$.

Electrical networks and random walks

Consider a graph G, fix two distinguished nodes A,B.

Consider a random walk.

Let p_u be the probability a random walk starting from u reaches A before it reaches B.



Consider placing a 1-volt battery between A and B

Let V_u be the voltage at node u.

Then $p_u = V_u$.

- Solving for p_u: $p_A = 1$, $p_B = 0$, and for all $u \notin \{A, B\}$ we have $p_u = \frac{1}{\deg(u)} \sum_{v:\{u,v\}\in E} p_v$.
- Solving for V_u: $V_A = 1$, $V_B = 0$, and for all $u \notin \{A, B\}$ we have flow in = flow out, which means $\sum_{v:\{u,v\}\in E} \left(\frac{V_v - V_u}{1}\right) = 0$, so $V_u = \frac{1}{\deg(u)} \sum_{v:\{u,v\}\in E} V_v$.

Electrical networks and random walks

Next time: more connections (exact expression for commute time in terms of effective resistance), and rapid mixing.