Mathematical Toolkit **Spring 2023**

Lecture 16: May 15, 2023

Lecturer: Avrim Blum

1 Random Walks in Graphs: Setup

Imagine you are lost in a maze. How long will it take you to get out if you just move around randomly? In this class we will analyze this and other problems involving random walks on graphs.

General setup. We assume there is an underlying undirected graph $G = (V, E)$ with n vertices and m edges. We start at some initial vertex $s \in V$. Then, at each time step, we pick a random neighbor of our current node and move to it.

Quantities of interest. We will want to analyze the following quantities:

- The *hitting time* H_{uv}, defined as $\mathbb{E}[\text{number of steps to reach } v | \text{ start at } u]$.
- The *commute time* C_{uv}, defined as $\mathbb{E}[\text{number of steps to reach } v \text{ and then return to } u | \text{ start at } u]$.

  If we define X_{uv}^{hit} to be a random variable denoting the time for a walk starting at u to reach v, then $H_{uv} = \mathbb{E}[X_{uv}^{hit}]$ and $C_{uv} = \mathbb{E}[X_{uv}^{hit} + X_{vu}^{hit}] = H_{uv} + H_{vu}$.
- The *cover time from u*, Cov_u, which is the expected time to visit all nodes in the graph given that you start at node u.
- The *cover time Cov_G of G*, defined as $Cov_G = \max_u Cov_u$.

Theorem 1.1 *If G is a connected graph with n vertices and m edges then $Cov_G \leq 2m(n-1)$.*

We will prove Theorem 1.1 in two ways.

2 First proof of Theorem 1.1

For convenience, let's think of ourselves as at each time step being on some edge heading in some direction (as opposed to being at a node). That is, we are remembering both our current state and where we just came from. We will prove the theorem by analyzing the expected time between successive traversals of any given edge/direction. Notationally, we will say we are "on edge (u, v)" to mean we are on the $\{u, v\}$ edge in the direction towards v.

Lemma 2.1 *For any edge $\{u, v\}$ the expected number of steps between successive visits to (u, v) is $2m$.*

Note that Lemma 2.1 implies that if u and v are neighbors, then $C_{vu} \leq 2m$. That is because the expected time to go from v to u back to v is less than or equal to the expected time if we wait until we actually take the (u, v) edge, which is what the lemma is counting. (The fact that we initially just came from u is irrelevant). Before proving Lemma 2.1, let's first use it to prove Theorem 1.1.

Proof: [Theorem 1.1 using Lemma 2.1] Consider some spanning tree T of G and some fixed tour of the spanning tree. The tree has $n - 1$ edges and the tour traverses each edge twice. So,

$$\begin{aligned}
\mathbb{E}[\text{time to visit entire graph}] &\leq \mathbb{E}[\text{time to visit nodes in that order}] \\
&= \sum_{\{u,v\}\in T} H_{uv} + H_{vu} \\
&= \sum_{\{u,v\}\in T} C_{uv} \;\leq\; \sum_{\{u,v\}\in T} 2m \;=\; 2m(n-1).
\end{aligned}$$

■

Now, let's prove Lemma 2.1.

Proof: [Lemma 2.1] 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 is our probability distribution after one step?

Answer: it's the same. For any directed edge (v, w),

$$\begin{aligned}
\mathbb{P}[\text{on } (v, w) \text{ after 1 step}] &= \sum_{u:\{u,v\}\in E} \mathbb{P}[\text{on } (u, v) \text{ initially}] \cdot \frac{1}{deg(v)} \\
&= \frac{deg(v)}{2m} \cdot \frac{1}{deg(v)} \;=\; \frac{1}{2m}.
\end{aligned}$$

So, this is a fixed point, i.e., a stationary distribution of our random walk process.

Now, by linearity of expectation, this means that for any edge (u, v), in T steps the expected number of traversals of (u, v) is $\frac{T}{2m}$.

To prove the lemma, we want to invert this fact that in T steps the expected number of traversals is $\frac{T}{2m}$ to say that the expected time between two consecutive traversals of (u, v) is $2m$. Note that if our positions at different times t were *independent*, then this would follow immediately from the fact that expected value of a Geometric(p) random variable is $1/p$. However, they are not independent so we need to be careful. For example, if the graph was not connected and instead consisted of two pieces with $m/2$ edges each, then the expected time between consecutive traversals would be m, whereas the expected time to our first traversal if we start from the uniform distribution would be infinite.

Nonetheless, it turns out the intuition from Geometric random variables is indeed the right one. Let's consider our random walk process starting from the stationary uniform distribution, and let X_1 be a random variable denoting the time until we first reach edge (u, v). Let X_2 denote the time between our first traversal of (u, v) and our second traversal of (u, v), and so on for $X_3, X_4, \ldots$. Because the graph is connected, these R.V.'s are well-defined: we will indeed reach (u, v) with probability 1. In fact, these R.V.'s have bounded variance: as an extremely crude upper-bound, notice that wherever we are in the graph, it is possible to visit (u, v) within the next n steps and therefore our probability of doing so is at least some (possibly exponentially small) $\delta > 0$; this means that our R.V's are dominated by n times a Geometric(δ) R.V., which has finite variance. This means that as $T \to \infty$, the number of traversals observed $N \to \infty$ also, with probability 1.

Let's now apply Chebyshev's inequality to $X = \frac{X_1 + \ldots + X_N}{N}$. Let σ^2 be an upper-bound on $\mathsf{Var}[X_i]$. Since the X_i are independent, we have $\mathsf{Var}[X] \leq \frac{N\sigma^2}{N^2} = \frac{\sigma^2}{N}$. So, $\mathbb{P}[|X - \mathbb{E}[X]| \geq \varepsilon] \leq \frac{\sigma^2}{N\varepsilon^2}$. This tells us that for large N, with high probability the observed average gap length is close to its expectation. This means that the observed fraction of time-steps that are traversals, which is $1/$(observed average gap length) is also multiplicatively close to $1/\mathbb{E}[X]$. Since the *expected* fraction of traversals is $1/(2m)$, this means that as $N \to \infty$ we must have $\mathbb{E}[X] \to 2m$, since if a bounded R.V. is concentrated, it must be concentrated about its expectation. ■

So, this tells us that if we are lost in a maze, if we walk around randomly we will visit all the nodes (and hence, the exit, wherever it is) in $O(mn)$ steps. There are some graphs where this is tight: for instance, on a line it really does take $\Theta(n^2)$ steps in expectation for a random walk to go from the middle to one of the endpoints. There are other graphs where it is not: for instance, in a complete graph, the cover time is only $O(n \log n)$. Can you see why? An example of a graph that requires $\Omega(n^3)$ steps to cover is the "lollipop graph": a clique of size $n/2$ attached to the end of a line of length $n/2$.

3 Electrical networks and a second proof of Theorem 1.1

An electrical resistive network is a graph where on each edge we have a resistor of some resistance. If we connect up a battery of voltage V_{batt} to two nodes in this graph (assigning one a voltage of V_{batt} and the other a voltage of 0) then each node in the graph will have a voltage (also called its "potential") and each edge will have some current in some direction. Voltages and currents can be computed using the following two laws:

Kirchoff's law. Current is like water flow: for any node that is not connected to the battery, the total current in equals the total current out.

Ohm's law. $V = IR$. Here, V is the voltage difference across the resistor, R is the resistance of the resistor, and I is the current flow.

Intuitively, you can think of voltages as like "heights" with current flowing downhill, and the resistors like little water wheels that slow down the flow of the water (and use up energy).

The *effective resistance* R_{uv} between two nodes u and v is the resistance we would measure if we hooked up a battery between u and v and observed the amount of current that flows, i.e., V_{batt}/I. Some simple things we can see are that for two resistors in series, the resistances add:

```
 *-------R1--------*-------R2-------*
V0                V1               V2
```

To see this, say I is the current flowing left to right. Then $V_1 = V_0 - IR_1$ and $V_2 = V_1 - IR_2$ so $V_0 - V_2 = I(R_1 + R_2)$.

For two resistors R_1 and R_2 in parallel, like this:

```
     _____R1______
    /             \
V0*                *V1
    \_____R2______/
```

we get an effective resistance R satisfying $1/R = 1/R_1 + 1/R_2$. We can see this by noticing that the top path has current $I_1 = (V_0 - V_1)/R_1$ and the bottom path has current $I_2 = (V_0 - V_1)/R_2$, and the overall current $I = I_1 + I_2 = (V_0 - V_1)/R$.

It turns out that resistive networks and random walks have a lot in common. Here is one connection. Let's consider a resistive network G where every edge is a 1 Ohm resistor. Take a 1-volt battery and connect its poitive terminal to a set of vertices S and its negative terminal to a set of vertices T, where S and T are disjoint. So, all nodes in S will have voltage 1 and all nodes in T will have voltage 0. Let I_{uv} denote the current flowing on edge $\{u,v\}$ in the $u \to v$ direction (negative if it is flowing in the $v \to u$ direction). Then for any vertex $v \notin S \cup T$ we have $\sum_{u:\{u,v\}\in E} I_{uv} = 0$. This means that the voltage at v must be the average voltage of its neighbors. So, we have a linear system we can use to solve for voltages at every vertex in the graph. Now, consider instead a random walk where we start at some initial vertex and keep walking until we reach some node in $S \cup T$. Let p_v denote the probability that a random walk starting from v reaches a node in S before it reaches a node in T. So, $p_v = 1$ for $v \in S$, $p_v = 0$ for $v \in T$, and for any $v \notin S \cup T$ we have p_v is the average of its neighbors (since the first step of the walk is equally likely to go to any of its neighbors). Notice that these are the same equations. So, the voltage at v can be interpreted as the probability that a random walk starting from v would reach a node in S before it reaches a node in T.

As an aside, another way to think about this is the values p_v are the solution to the following optimization problem:

$$\text{Minimize} \sum_{\{u,v\}\in E} (p_u - p_v)^2 \text{ subject to } p_v = 1 \text{ for } v \in S \text{ and } p_v = 0 \text{ for } v \in T,$$

because for each $v \notin S \cup T$, the solution to this optimization will set p_v to the average of its neighbors.

Here is another connection: the commute time C_{uv} is directly connected to the effective resistance R_{uv}.

Theorem 3.1 *In a connected graph G with m edges, each of which is a unit resistor, for any two nodes u, v, we have $C_{uv} = 2mR_{uv}$.*

For example, on a line graph of n nodes and $n-1$ edges, the commute time between the two endpoints is exactly $2(n-1)^2$.

Note that if u and v are neighbors, then $R_{uv} \leq 1$ (because we have the resistor on this edge in parallel with the rest of the graph). By Theorem 3.1, this implies that $C_{uv} \leq 2m$. So, Theorem 3.1 gives another proof of Lemma 2.1.

To prove Theorem 3.1 we first prove the following lemma.

Lemma 3.2 *Fix some vertex v. For each node $x \neq v$, place a battery of voltage H_{xv} with positive terminal at x and negative terminal at v. Then $deg(x)$ current will flow out of each node $x \neq v$ and $2m - deg(v)$ current will flow into v.*

Proof: Let's define v to have voltage 0 so that each $x \neq v$ has voltage H_{xv}. Now, let us first think about the definition of H_{xv}. H_{xv} is the expected time for a random walk starting from x to reach v. Assuming $x \neq v$ (if $x = v$ then $H_{xv} = 0$), the very first step of that walk moves to a random neighbor w of x, after which the expected number of steps to go is just

H_{wv}. So, the expected length of the walk after the first step is just the average of H_{wv} over all neighbors w of x. That is, for $x \neq v$ we have:

$$\begin{aligned} H_{xv} &= 1 + \frac{1}{deg(x)} \sum_{w:\{x,w\}\in E} H_{wv}, \text{ or equivalently,} \\ deg(x) \cdot H_{xv} &= deg(x) + \sum_{w:\{x,w\}\in E} H_{wv}. \end{aligned} \tag{1}$$

Now, the current flowing on edge (x, w) is equal to $(V_x - V_w)/1$. So, the total current flowing out of node $x \neq v$ is equal to:

$$\begin{aligned} \sum_{w:\{x,w\}\in E} (V_x - V_w) &= \sum_{w:\{x,w\}\in E} (H_{xv} - H_{wv}) \\ &= deg(x) \cdot H_{xv} - \sum_{w:\{x,w\}\in E} H_{wv} \\ &= deg(x). \quad \text{(by equation (1))} \end{aligned}$$

■

Lemma 3.3 *Fix some vertex u. Suppose for each node $x \neq u$ we place a battery of voltage H_{xu} with negative terminal at x and positive terminal at u. Then $deg(x)$ current will flow into each node $x \neq u$ and $2m - deg(u)$ current will flow out of u.*

Proof: Same as for Lemma 3.2 (or, by symmetry). ■

Proof: [Theorem 3.1] Consider adding the voltages from Lemma 3.2 and Lemma 3.3. So, we have a voltage drop of $H_{uv} + H_{vu} = C_{uv}$ from u to v. Also, if we add the voltages, then currents add too by linearity. So, we get $2m$ units of current going out of node u into node v. Since no current flows into any other node, we can view this as a single battery between u and v. This means that the voltage drop equals $I \cdot R_{uv}$. So, we have $C_{uv} = 2m \cdot R_{uv}$ as desired. ■