CME 305: Discrete Mathematics and Algorithms
Instructor: Professor Aaron Sidford (sidford@stanford.edu)
February 22, 2018

Lecture 13 - Random Walks[1]

In this lecture we continue discussion of random walks on connected undirected graphs initiated in the last lecture. We show how to compute and approximate various random walk related quantities and we draw connections between these quantities and concepts in electrical networks. Finally, we conclude this unit with a brief connection between these concepts and the question of computing a random spanning tree in an undirected gap.

1 Random Walk Recap

As with last lecture for the remainder of this lecture we let $G = (V, E, w)$ be a undirected simple graph with positive weights $w \in \mathbb{R}^E_{>0}$. Furthermore, we consider the standard random walk on this graph where we call a random sequence $v_1, ..., v_k$ of vertices a random walk on G if independently for all $i \in [k-1]$ we have

$$\Pr[v_{i+1} = v \,|\, v_i = u] = \begin{cases} \frac{w_{\{u,v\}}}{\deg(u)} & \text{if } \{u, v\} \in E \\ 0 & \text{otherwise} \end{cases}.$$

Furthermore, we let $\mathbf{W} \in \mathbb{R}^{V \times V}$ be defined as $\mathbf{W} = \mathbf{A}\mathbf{D}^{-1}$ so that for all $u, v \in V$ we have $\mathbf{W}_{vu} = \Pr[v_{i+1} = v \,|\, v_i = u]$. Consequently, $\mathbf{W}\vec{1}_u$ is the distribution of vertices resulting of one step of a random walk from u.

2 Hitting Time

We begin by discussing how to compute a natural relationship between vertices of a graph, the hitting time of a random walk. Formally, for all $s, t \in V$ we let H_{st} denote the hitting time from s to t, i.e. expected number of steps for a random walk started at s to first reach t. Here we provide a characterization of H_{st} in terms of the Laplacian of a graph and through this how how to compute hitting times efficiently given the ability to solve Laplacian systems efficiently, i.e. a Laplacian system solver.

The technique we use here to reason about the hitting times of a graph is fairly general and can be used to reason about various random walk related quantities. Let us fix $t \in V$ and define the vector $h \in \mathbb{R}^V$ to be such that for all $a \in V$ we have $h_a = H_{at}$ with $h_t \stackrel{\text{def}}{=} 0$, i.e. we say that the hitting time from t to t is 0. We wish to compute this vector h and to do it we will write down a natural recursive formula characterizing h.

Note that for every vertex $a \in V$ with $a \neq t$ it is the case that the hitting time from a to t is simply one plus the expected hitting time from the vertices resulting from one step of a random walk from a. In other words, the hitting time from a to t is given by

$$H_{at} = 1 + \sum_{v \in N(a)} \mathbf{W}_{va} H_{vt}.$$

Rearranging, we have that this is equivalent to saying

$$h^\top 1_a = 1 + h^\top \mathbf{W} 1_a$$

[1] These lecture notes are a work in progress and there may be typos, awkward language, omitted proof details, etc. Moreover, content from lectures might be missing. These notes are intended to converge to a polished superset of the material covered in class so if you would like anything clarified, please do not hesitate to post to Piazza and ask.

or

$$1_a^\top(\mathbf{I} - \mathbf{W}^\top)h = 1$$

for all $a \in V$. Similarly, for some unknown $c \in \mathbb{R}$ it is clearly the case that

$$h^\top 1_t = c + h^\top \mathbf{W} 1_t$$

or

$$1_t^\top(\mathbf{I} - \mathbf{W}^\top)h = c\,.$$

Simply writing all these equalities in matrix form we have that for some value $c' \in \mathbb{R}$

$$(\mathbf{I} - \mathbf{W}^\top)h = \vec{1} - c'\vec{1}_t\,.$$

However, since $\mathbf{W} = \mathbf{A}\mathbf{D}^{-1}$ and $\mathbf{D}$ is invertible, multiplying both sides of this by $\mathbf{D}$ yields that for some $c'' \in \mathbb{R}$

$$(\mathbf{D} - \mathbf{A})h = d - c''\vec{1}_t$$

where $d \in \mathbb{R}^V$ with $d_i = \deg(i)$ for all $i \in V$. Now, we know that $\mathcal{L} = \mathbf{D} - \mathbf{A}$ and that $\ker(\mathcal{L}) = \text{span}(\vec{1})$. Consequently, $\text{im}(\mathcal{L})$ is the the orthogonal complement of $\text{span}(\vec{1})$, i.e. every vector orthogonal to the all ones vector. Consequently, there exists a vector x with

$$\mathcal{L}x = b \stackrel{\text{def}}{=} d - c''\vec{1}_t$$

if and only if

$$0 = b^\top \vec{1} = \|d\|_1 - c''\,.$$

Therefore, we know that $c'' = \|d\|_1$. Consequently, we know that

$$\mathcal{L}h = d - \|d\|_1\vec{1}_t\,.$$

However, we know that solutions to $\mathcal{L}h = b$ are unique up to adding the all ones vector and thus for any x with $\mathcal{L}x = d - \|d\|_1\vec{1}_t$ we have that $h = x + \alpha\vec{1}$ for some value of α. However, since we also know that $h(t) = 0$ we have that $x(t) + \alpha = 0$, i.e. $\alpha = -x(t)$. Putting this all together we have just shown the following.

Lemma 1. *If $h \in \mathbb{R}^V$ with $h_a = H_{at}$ for all $a \in V$ then there exists $x \in \mathbb{R}^V$ with $\mathcal{L}x = d - \|d\|_1\vec{1}_t$ and for all such x we have $h = x - x_t\vec{1}$.*

From this lemma we see that if we can solve linear systems in $\mathcal{L}$, i.e. we have a Laplacian system solver, then the time to solve linear systems in $\mathcal{L}$ is an upper bound on the time to compute h. Note that we showed this simply by writing constraints that h must obey and by the uniqueness of solutions to $\mathcal{L}$ we were able to show that h can be gleaned from solutions to the constraints we wrote. This is a fairly general technique.

Also note that we can write the same lemma a little more compactly using pseudoinverse notation. Formally we let $\mathcal{L}^\dagger$ denote the Moore-Penrose pseudoinverse of $\mathcal{L}$. For symmetric PSD matrices $\mathbf{M}$ the pseuoinverse $\mathbf{M}^\dagger$ is simply the matrix that is the inverse of $\mathbf{M}$ outside of the kernel of $\mathbf{M}$ and has the same kernel as $\mathbf{M}$. More formally, if $\mathbf{M} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$ for $\mathbf{U}$ with orthonormal columns and diagonal $\mathbf{\Lambda}$ then $\mathbf{M}^\dagger = \mathbf{U}\mathbf{\Lambda}^\dagger\mathbf{U}^\top$ where $\mathbf{\Lambda}^\dagger$ is the matrix where $\mathbf{\Lambda}$ wherever every non-zero entry is inverted and zero entries are kept the same.

Note that if there exists some x with $\mathcal{L}x = b$ then $\mathcal{L}(\mathcal{L}^\dagger b) = b$. Consequently, an immediate corollary of the above lemma is as follows.

Lemma 2. *For all $s, t \in V$ it is the case that $H_{st} = (\vec{1}_s - \vec{1}_t)^\top\mathcal{L}^\dagger(d - \|d\|_1\vec{1}_t)$.*

Proof. Let $x = \mathcal{L}^\dagger(d - \|d\|_1\vec{1}_t)$. Since $\vec{1} \perp d - \|d\|_1\vec{1}_t$ it is the case that $\mathcal{L}x = d - \|d\|_1\vec{1}_t$ and thus by the previous lemma

$$H_{st} = \vec{1}_s^\top h = \vec{1}_s^\top(x - x_t\vec{1}) = x_s - x_t = (\vec{1}_s - \vec{1}_t)^\top x\,.$$

□

2.1 Commute Times

We can use the analysis in the previous section to give an even simpler formula for the commute time between a and b. Recall that the commute time from a to b, defined by C_{ab} is given by $C_{ab} = H_{ab} + H_{ba}$, i.e. the expected number of steps for a random walk starting from a to get to b and back again. We show the following simple formula for C_{ab}.

Lemma 3. *For all $a, b \in V$ it is the case that $C_{ab} = \|d\|_1 \cdot \left(\vec{1}_a - \vec{1}_b\right)^\top \mathcal{L}^\dagger(\vec{1}_a - \vec{1}_b)$.*

Proof. We have that

$$H_{ab} = (\vec{1}_a - \vec{1}_b)^\top \mathcal{L}^\dagger(d - \|d\|_1 \vec{1}_b) = \|d\|_1 \cdot \left(\vec{1}_a - \vec{1}_b\right)^\top \mathcal{L}^\dagger \left(\frac{1}{\|d\|_1} d - \vec{1}_b\right).$$

and

$$H_{ba} = (\vec{1}_b - \vec{1}_a)^\top \mathcal{L}^\dagger(d - \|d\|_1 \vec{1}_a) = \|d\|_1 \cdot \left(\vec{1}_a - \vec{1}_b\right)^\top \mathcal{L}^\dagger \left(\vec{1}_a - \frac{1}{\|d\|_1} d\right).$$

Summing these two quantities yields the result. □

Note that this lemma says that we can compute C_{ab} simply by looking at the quadratic form of $\mathcal{L}^\dagger$. Note that since $\mathcal{L}$ is PSD so is $\mathcal{L}^\dagger$and thus C_{ab} is simply given by the quadratic form of a PSD matrix. One nice consequence of this formula is that we can use things like Cheeger's inequality to easily reason about commute times.

For example, note that $\lambda_n(\mathcal{L}^\dagger) = \frac{1}{\lambda_2(\mathcal{L})}$ and therefore

$$(\vec{1}_a - \vec{1}_b)^\top \mathcal{L}^\dagger(\vec{1}_a - \vec{1}_b) = R_{\mathcal{L}^\dagger}(\vec{1}_a - \vec{1}_b) \cdot \|\vec{1}_a - \vec{1}_b\|_2^2 \leq \frac{2}{\lambda_2(\mathcal{L})}.$$

Consequently, if we say have a graph that is a path of length n and we wish to know the commute time between the endpoints of the path. Then it can be shown (see for example the homework) that λ_2 in this graph is at least $\Omega(\frac{1}{n^2})$ and therefore the commute time C satisfies

$$C \leq O(n) \cdot \frac{1}{\lambda_2(\mathcal{L})} = O(n^3)$$

i.e. the commute time between the endpoints is at most $O(n^3)$.

3 An Electrical Perspective

Another way to reason about random walks on undirected graphs and quantities like commute times in these graphs is by viewing the graph as an electric network. There is work that directly connects these notions, here we will show that these notions are equivalent by relating electric networks too Laplacian systems of equations and therefore random walks.

To build this equivalence we define the electrical network associated with an undirected graph as follows. For an undirected graph $G = (V, E)$ with positive weights on the edges $r \in \mathbb{R}^E_{>0}$. We interpret each edge $\{i, j\} \in E$ as a resistor and each vertex i as a junction where resistors may connect. We further interpret r_e as the resistance of the resistor connecting junction i to junction j when $e = \{i, j\}$.

Now the question we wish to ask about this electric network or network of resistors is if we force electric current to enter and leave the resistor network, where does it go? Formally, we assume we have a vector

$b \in \mathbb{R}^V$ specifying for every junction $i \in V$ how much electric current enters or leaves junction i. For example, we could imagine hooking up vertex i and j to endpoints of a battery so that $b = \vec{1}_i - \vec{1}_j$, i.e. one unit of current is flowing out of i and into the network and -1 unit of current is doing the same at j, i.e. one unit is coming in from the network at j. Now the question we wish to ask, is given this constraint of how much current enters and leaves every vertex, where does the electric current go? In other words, if for all $e = (i, j) \in E$ we let f_e denote the amount of current going from i to j what is f?

To answer this question and characterize the electric current, lets list a few properties of the electric current $f \in \mathbb{R}^E$.

- **Kirchoff's current law**: $f \in \mathbb{R}^E$ must route the demands b, i.e. $\text{im}(f) = b$ or b gives the the total current imbalance at every vertex.
- **Ohm's law**: this classic law says the the electric current is actually given by potentials, or some real valued assignment of values to the vertices of the graph. Formally, this says that there must exist potentials, $v \in \mathbb{R}^V$ such that for all $(a, b) \in E$ it is the case that $f_{(a,b)} \cdot r_{(a,b)} = v_a - v_b$ in other words, the current times the resistance over an edge is equal to the potential difference over an edge.
- **Thompson's energy principal**: this is the idea that electric current follows the path of least resistance or the one that minimizes energy. Formally, it says that over all values $f \in \mathbb{R}^E$ the one that minimizes energy, i.e. $\sum_{e \in E} r_e \cdot f_e^2$ is the electric current.

As we will see certain pairs of these properties imply the third and suffice to characterize the electric current.

3.1 Krichoff's Law and Ohm's Law

Here we show that Kirchoff's current law and Ohm's law suffice to characterize the electric current in terms of Laplacian systems. To show this, let's look at a linear algebraic expression of these laws. Let $e_1, ..., e_m$ denote the edges of the graph / network and let us an impose an arbitrary ordering of the endpoints of each edge so that $e_i = (a_i, b_i)$ for all $i \in [m]$ where $a_i, b_i \in V$. We do this simply so that we have a canonical interpretation of $f(e_i)$ as the amount of current from a_i to b_i. Now recall that $\vec{\delta}_{e_i} = \vec{1}_{a_i} - \vec{1}_{b_i}$ and that $\mathbf{B} \in \mathbb{R}^{E \times V}$ is the matrix where row i is given by $\vec{\delta}_{e_i}^\top$.

Now let us look at $[\mathbf{B}^\top f]_a$ for arbitrary $f \in \mathbb{R}^E$ and $a \in V$ we have that

$$[\mathbf{B}^\top f]_a = \vec{1}_a^\top \left[\mathbf{B}^\top f\right] = \sum_{(a,b) \in E} f_{a,b} - \sum_{(b,a) \in E} f_{a,b} = \text{im}(f, a)\,.$$

Consequently, $\mathbf{B}^\top f = \text{im}(b)$ and the Kirchoff's current law that $\text{im}(f) = b$ we can write compactly as $\mathbf{B}^\top f = b$.

Next, let us consider Ohm's law. Note that for any $v \in \mathbb{R}^V$ we have that for $e_i \in E$ it is the case that $[\mathbf{B}v]_{e_e} = v_{a_i} - v_{b_i}$. Consequently, if we let $\mathbf{R} \in \mathbb{R}^{E \times E}$ be the diagonal matrix with $\mathbf{R}_{e,e} = r_e$ we have that the constraint that $f_{(a,b)} \cdot r_{(a,b)} = v_a - v_b$ for all $(a, b) \in E$ can be written more compactly as $\mathbf{R}f = \mathbf{B}v$. Consequently, we see that f obeys Ohm's law if and only if $f = \mathbf{R}^{-1}\mathbf{B}v$.

Now, suppose that we define $w_e = \frac{1}{r_e}$ for all $e \in E$ and let $\mathbf{W} \in \mathbb{R}^{E \times E}$ be the diagonal matrix with $\mathbf{W}_{e,e} = w_e$ for all $e \in E$. With this we have that f obeys both Kirchoff's law and Ohm's law if and only if $\mathbf{B}^\top f = b$ and $f = \mathbf{W}\mathbf{B}v$ for some v. Combining these implies that $\mathbf{B}^\top \mathbf{W}\mathbf{B}v = b$. Consequently, in order for f to be the electric current there must be some $v \in \mathbb{R}^V$ with $\mathcal{L}v = b$ and moreover $f = \mathbf{W}\mathbf{B}v$ for one of these v vectors. However, the solutions to $\mathcal{L}v = b$ are unique up to adding $\alpha\vec{1}$ for some α and $\vec{1} \in \ker(\mathbf{B})$. Consequently, we have shown the following.

Lemma 4. *A vector $f \in \mathbb{R}^E$ is the unique electric current for b if and only if f satisfies Kirchoff's current law and Ohm's law which in turn happens if and only if for any vector v with $\mathcal{L}v = b$, which must exist, we have $f = \mathbf{R}^{-1}\mathbf{B}v$. In other words, $f = \mathbf{R}^{-1}\mathbf{B}\mathcal{L}^\dagger b$.*

3.2 Characterizing Commute Times

One nice consequence of the analysis in the previous section is that we can provide another characterization of commute times in the graph. Suppose $b = \vec{1}_s - \vec{1}_t$, i.e. the demands are to have one unit of current leave s and one unit enter at t. Let $f = \mathbf{R}^{-1}\mathbf{B}\mathcal{L}^\dagger b$ be the electric current routing b. What is the energy of f? We have that

$$\sum_{e \in E} r_e f_e^2 = f^\top \mathbf{R} f = b^\top \mathcal{L}^\dagger \mathbf{B}^\top \mathbf{R}^{-1} \mathbf{R} \mathbf{R}^{-1} \mathbf{B} \mathcal{L}^\dagger b = b^\top \mathcal{L}^\dagger \mathbf{B}^\top \mathbf{R}^{-1} \mathbf{B} \mathcal{L}^\dagger b$$
$$= b^\top \mathcal{L}^\dagger \mathcal{L} \mathcal{L}^\dagger b = b^\top \mathcal{L}^\dagger b = (\vec{1}_s - \vec{1}_t)^\top \mathcal{L}^\dagger (\vec{1}_s - \vec{1}_t)\,.$$

This quantity is known as the *effective resistance* between s and t and is a fundamental measure of the connection between s and t. Consequently, from our earlier analysis we see that the commute time between s and t is simply the effective resistance between them times the sum of the degrees of the vertices in the graph. Note that effective resistances have many applications and as we may discuss later in the class, sampling edges by effective resistance can be used to construct sparse approximate graphs that preserve many properties of graphs approximately, i.e. let us approximately sparsify the graph.

3.3 Thompson's Energy Principal

Here we show that Thompson's energy principal provides an alternative characterization of electric current. Whereas we have already shown that the electric current is the unique vector f routing b and obeying Ohm's law, here we show that f is also the unique minimum energy vector routing b.

Lemma 5. *If f minimizes $\sum_{e \in E} r_e f_e^2$ among all $f \in \mathbb{R}^E$ such that $\mathrm{im}(f) = b$ then f is the electric current routing b.*

Proof. Suppose $\mathrm{im}(f) = b$ and that f minimizes $e(f) = \sum_{e \in E} r_e f_e^2 = f^\top \mathbf{R} f$. Note that $\bigtriangledown e(f) = 2\mathbf{R}f$ and since f is minimal it must be the case that $\bigtriangledown e(f) \perp \ker(\mathbf{B}^\top)$, as if this was note the case then for the vector $g \in \ker(\mathbf{B}^\top)$ with g not perpendicular to $\bigtriangledown e(f)$ by moving infinitesimally along the direction g we would be able to decrease $e(f)$ while preserving $\mathbf{B}^\top f = b$. Now note that the space orthogonal to the kernel of $\mathbf{B}^\top$ is the image of $\mathbf{B}$ and therefore $\bigtriangledown e(f) \perp \ker(\mathbf{B}^\top)$ implies $\bigtriangledown e(f) \in \mathrm{im}(\mathbf{B})$, i.e. there exist $w \in \mathbb{R}^V$ with $2\mathbf{R}f = \mathbf{B}w$. By scaling this implies there exists $v \in \mathbb{R}^V$ with $\mathbf{R}f = \mathbf{B}v$ or $f = \mathbf{R}^{-1}\mathbf{B}v$. Consequently, $\mathbf{B}^\top f = b$ and $f = \mathbf{R}^{-1}\mathbf{B}v$ and thus f both routes the demands and is given by voltages and therefore is the electric current. □

One nice application of this lemma is that it allows us to easily show certain monotonicity properties of things like effective resistance. For example, suppose we add brand new edges to our electric circuit, does the effective resistance increase or decrease? Note that with the new edges we are only giving more freedom to the valid f with $\mathrm{im}(f) = b$ and thus the minimum energy f satisfying this has at most the energy it had before and therefore the energy of f cannot increase and neither can any effective resistance.

3.4 More

There are many more connections between electric networks and random walks and for further reading see the course webpage or email me.

4 Cover Time

We conclude by considering one more natural property of random walks on graphs that we would like to compute. Recall that the cover time of G, denoted $C(G)$, is the maximum over all starting vertices of the expected number of steps of a random walk needed before every vertex in the graph is visited at least once. Here we do not give an algorithm to compute $C(G)$ precisely but we do show that it can be computed approximately efficiently using techniques we have seen. Formally, we provide the following interesting result showing that $C(G)$ is well approximated by the maximum hitting time between any pair of vertices in the graph.

Lemma 6. *Let* $H_{\max} \stackrel{\text{def}}{=} \max_{a,b \in V} H_{ab}$ *then* $H_{\max} \leq C(G) \leq H_{\max} \cdot O(\log n)$.

Proof. Clearly, $H_{\max} \leq C(G)$ as visiting every vertex takes at least as much time as it takes to visit every pair. The more difficult and perhaps surprising direction is to show that $C(G) \leq H_{\max} \cdot O(\log n)$. To prove this, let $a \in V$ be arbitrary and suppose we still need to visit vertices $v_1, ..., v_k$. let h be a random variable denoting the number of steps until we visit at least half of the v_i. Furthermore, let $t(v_i)$ be a random variable for the number of steps of the random walk we take to visit v_i for the first time. Clearly

$$\sum_{i \in [k]} t(v_i) \geq \lfloor \frac{k}{2} \rfloor h$$

as we visit half the vertices after we have visited half the vertices. However by linearity of expectation this yields that

$$\mathbb{E}h \leq \frac{1}{\lfloor \frac{k}{2} \rfloor} \sum_{i \in [k]} \mathbb{E}t(v_i) = \frac{1}{\lfloor \frac{k}{2} \rfloor} \sum_{i \in [k]} H_{av_i} \leq \frac{k}{\lfloor \frac{k}{2} \rfloor} \cdot H_{\max} = H_{\max} \cdot O(1)\,.$$

Consequently, in expectation every $O(H_{\max})$ steps we visit half the remaining vertices and in $H_{\max} \cdot O(\log n)$ steps we visit all of them. □

Note that $H_{\max}$ is within a constant factor of the maximum expected commute time and therefore if we have a bound on the worst commute time between vertices of a graph (as we had for the path) we have a bound up to a $O(\log n)$ factor of the cover time of the graph.

One interesting application of this to computing a random spanning tree in a graph. Suppose we have an unweighted undirected graph and among the space of all spanning trees in a graph we wish to compute one uniformly at random. To do this consider the following algorithm, pick a random vertex in the graph with probability proportional to its degree and then run the standard random walk. Every time a vertex is visited for the first time (other than the first vertex) add the edge used to reach it to a set. Once all the vertices are visited output the associated set of edges, which in turn must be a tree. It can be shown that this tree is such a random spanning tree and thus we can compute a random spanning tree in an expected running time equal to the cover time of the graph.