Notes 12: Random walk

1. Random walk

Random walk is a simple random process on vertices of a graph (which may be directed). The walk starts at an arbitrary vertex. If the graph has positive edge weights $w \in \mathbb{R}_+^E$ and the walk is currently at a vertex i, it will pick a random outgoing edge (i, j) with probability proportional to w_{ij} and move to j. This process is repeated indefinitely.

"Probability proportional to w_{ij}" means with probability $p_{ij} = w_{ij}/\sum_{j' \in V} w_{ij'}$. The matrix $P = (p_{ij})_{ij}$ is the transition matrix. Every row of this matrix sums to 1.

If $\pi_t(i)$ denotes the current probability of the walk to be at vertex i, then after one step, the probability of landing at vertex j will be $\sum_{i \in V} p_{ij}\pi_t(i)$. In matrix notation, we represent the current probability as a row vector $\pi_t^\top$, and the new probability after a step is $\pi_{t+1}^\top = \pi_t^\top P$. Given an initial distribution π_0 over vertices, the probability distribution at time t is $\pi_t^\top = \pi_0^\top P^t$.

If $A = (w_{ij})_{ij}$ is the adjacency matrix and D is the diagonal matrix of out-degrees, i.e. $D_{ii} = d(i) = \sum_{j' \in V} w_{ij'}$, then $P = D^{-1}A$.

Studying random walks has applications to random sampling. For example, you may want to sample a uniformly random deck of poker cards. Or given a graph, sample a uniformly random perfect matching (i.e. a collection M of edges so that every vertex is incident to exactly one edge in M).

2. Stationary distribution

A probability distribution π is stationary if $\pi^\top = \pi^\top P$.

Proposition 2.1. *An undirected graph always has a stationary distribution π whose probability at vertex i is proportional to its degree $d(i)$. That is, $\pi(i) = d(i)/\sum_{j \in V} d(j)$.*

Proof.

$$(\pi^\top P)(j) = \sum_{i \in V} p_{ij}\pi(i) = \sum_{i \in V} \frac{w_{ij}}{d(i)} \frac{d(i)}{d(V)} = \sum_{i \in V} \frac{w_{ij}}{d(V)} .$$

But $w_{ij} = w_{ji}$ for an undirected graph, so $\sum_{i \in V} w_{ij} = d(j)$, and the right-hand-side is $d(j)/d(V) = \pi(j)$. This means $\pi^\top P = \pi^\top$. □

Let us focus on the simpler setting of undirected graphs. Must random walk on an undirected graph converge to a unique stationary distribution?

(1) If an undirected graph is disconnected, it has different stationary distributions supported on different connected components. A random walk starting from a vertex will stay inside the connected component of that vertex.
(2) If an undirected graph is bipartite, then on even time step the random walk stays on the left side, while on odd times step it stays on the right side. The walk does not converge to any distribution.

Turns out these are the only obstructions to converging to the unique stationary distribution. We will see later that given any undirected, connected, non-bipartite graph, the random walk converges to the unique stationary distribution.

3. Spectrum

In this section we again focus on undirected graphs. The transition matrix $P = D^{-1}A$ is not symmetric, so *a prior* it may not have only real eigenvalues. Fortunately, it is similar to the normalized adjacency matrix $\mathcal{A} = D^{-1/2}AD^{-1/2}$ that is real symmetric, so they share the same spectrum (set of eigenvalues with multiplicities).

A stationary distribution π is a left-eigenvector of P with eigenvalue 1. For stationary distribution to be unique, a sufficient condition is that 1 is an eigenvalue of P with multiplicity 1. This is in turn equivalent to the following statements:

(1) 1 is an eigenvalue of $\mathcal{A}$ with multiplicity 1.
(2) 0 is an eigenvalue of normalized Laplacian $\mathcal{L} = I - \mathcal{A}$ with multiplicity 1.
(3) The graph is connected.

The above discussions imply that stationary distribution is unique if and only if the graph is connected.

We saw that even if the graph is connected, the walk may fail to converge to the unique stationary distribution if the graph is bipartite. How to see that from the spectrum of P?

Let $\lambda_1 \geqslant \ldots \geqslant \lambda_n$ be the eigenvalues P and $v_1, \ldots, v_n$ the corresponding left-eigenvectors. Since an undirected graph always has a stationary distribution π by Proposition 2.1, we know $\lambda_1 = 1$ and $v_1 = \pi$. It is the largest eigenvalue because all eigenvalues of $\mathcal{A}$ (and hence P) are between -1 and 1.

Any initial distribution $\pi_0 \in \mathbb{R}^V$ can be decomposed as a linear combination $\pi_0 = c_1 v_1 + \cdots + c_n v_n$ in the eigenbasis, the distribution $\pi_1 = \pi_0^\top P$ after one transition has the decomposition $\pi_0^\top P = c_1\lambda_1 v_1 + \cdots + c_n \lambda_n \mu_n$. Therefore after t steps, the walk distribution π_t becomes $c_1\lambda_1^t v_1 + \cdots + c_n\lambda_n^t v_n$.

Proposition 3.1. *If all non-trivial eigenvalues (λ_2 through λ_n) have magnitude strictly less than 1, then the random walk must converge to a unique stationary distribution.*

Proof.

$$\pi_t = c_1 \underbrace{\lambda_1^t v_1}_{=1^t\pi} + c_2\lambda_2^t v_2 + \cdots + c_n\lambda_n^t v_n \to c_1\pi$$

as the number of iterations t grows, since $\lambda_i^t \to 0$ for any $|\lambda_i| < 1$. Further $c_1 = 1$ (using the fact that π_t and π are distributions and both sum to 1). □

When is the smallest eigenvalue strictly bigger than 1?

Claim 3.2. *An undirected graph has a bipartite component if and only if -1 is an eigenvalue of its normalized adjacency matrix.*

We will not prove this claim. It can be proved using similar ideas as showing 1 is an eigenvalue of $\mathcal{A}$ with multiplicity 1 if and only if the graph is connected.

The proposition implies that random walk converges to a unique stationary distribution if and only if an undirected graph is connected and non-bipartite.

4. Mixing time

Continuing with the last section. Consider a connected, non-bipartite undirected graph. How quickly does the random walk converge to its stationary distribution? We need to bound $\|\pi_t - \pi\|_1$, which is twice the total variational distance (also known as statistical distance) between π_t and π.

Let $\beta = 1 - \max\{|\lambda_2|, |\lambda_n|\}$ be the spectral gap of P or $\mathcal{A}$ ("gap" to non-convergence to unique stationary distribution).

We can make use of the orthonormal eigenvectors $u_1, \ldots, u_n$ of $\mathcal{A}$ $(= D^{1/2}PD^{-1/2})$. They are related to the (not necessarily orthogonal) left-eigenvectors $v_1, \ldots, v_n$ of P via $v_i = D^{1/2}u_i$, because

$$v_i^\top P = (u_i^\top D^{1/2})(D^{-1/2}\mathcal{A}D^{1/2}) = u_i^\top \mathcal{A}D^{1/2} = \lambda_i u_i^\top D^{1/2} = \lambda_i v_i^\top \; .$$

4.1. Regular undirected graphs.

Suppose the undirected graph is regular: all vertices have the same degree d. In this case D is just dI, and we can assume P and $\mathcal{A}$ have the same eigenbasis, so we may take $v_i = u_i$.

The t-step distribution π_t was shown to be $\pi + c_2\lambda_2^t v_2 + \cdots + c_n\lambda_n^t v_n$. Then

$$\|\pi_t - \pi\|_2^2 = \left\| \sum_{2\leqslant i\leqslant n} c_i\lambda_i^t v_i \right\|_2^2 = \left\| \sum_{2\leqslant i\leqslant n} c_i\lambda_i^t u_i \right\|_2^2 = \sum_{2\leqslant i\leqslant n} c_i^2|\lambda_i|^{2t} \leqslant (1-\beta)^{2t} \sum_{i\leqslant 2\leqslant n} c_i^2 \; ,$$

where the last equality uses orthonormality of u_i's. And

$$\sum_{2\leqslant i\leqslant n} c_i^2 \leqslant \sum_{1\leqslant i\leqslant n} c_i^2 = \left\| \sum_{1\leqslant i\leqslant n} c_i u_i \right\|_2^2 = \|\pi_0\|_2^2 = \sum_{i\in V} \pi_0(i)^2 \leqslant \sum_{i\in V} \pi_0(i) = 1 \; ,$$

where the first equality uses orthonormality of u_i's, the last inequality uses $0 \leqslant \pi_0(i) \leqslant 1$ (because π_0 is a probability distribution). Therefore by Cauchy–Schwarz,

$$\|\pi_t - \pi\|_1 \leqslant \sqrt{n}\|\pi_t - \pi\|_2 \leqslant (1-\beta)^t\sqrt{n}\ ,$$

which is at most $1/4$ (say) when $t \geqslant 2(\ln n)/\beta$.

4.2. **General undirected graphs.** If the undirected graph is not regular, we can no longer assume $v_i = u_i$, but we can translate between to two eigenbases using $v_i = D^{1/2}u_i$. In this case

$$\|\pi_t - \pi\|_2 = \|D^{1/2}D^{-1/2}(\pi_t - \pi)\| \leqslant \|D^{1/2}\|_2\|D^{-1/2}(\pi_t - \pi)\|_2\ ,$$

where

$$\|D^{1/2}\|_2 = \sup_{x\in\mathbb{R}^V, \|x\|_2\leqslant 1} \|D^{1/2}x\|_2 = \max_{i\in V}\sqrt{d(i)}\ ,$$

and by previous calculations we have

$$\|D^{-1/2}(\pi_t - \pi)\|_2^2 = \left\|\sum_{2\leqslant i\leqslant n} c_i\lambda_i^t u_i\right\|_2^2 \leqslant (1-\beta)^{2t}\sum_{i\leqslant 2\leqslant n} c_i^2\ .$$

We also have

$$\sum_{1\leqslant i\leqslant n} c_i^2 = \left\|\sum_{1\leqslant i\leqslant n} c_i u_i\right\|_2^2 = \left\|D^{-1/2}\sum_{1\leqslant i\leqslant n} c_i v_i\right\|_2^2 = \|D^{-1/2}\pi_0\|_2^2 \leqslant \|D^{-1/2}\|_2^2\|\pi_0\|_2^2 \leqslant \|D^{-1/2}\|_2^2\ ,$$

where the last inequality is by previous calculations, and

$$\|D^{-1/2}\|_2 = \sup_{x\in\mathbb{R}^V, \|x\|_2\leqslant 1} \|D^{-1/2}x\|_2 = \max_{i\in V}\sqrt{1/d(i)} = 1\Big/\left(\min_{i\in V}\sqrt{d(i)}\right)\ ,$$

Altogether,

$$\|\pi_t - \pi\|_2 \leqslant \|D^{1/2}\|_2(1-\beta)^t\sum_{2\leqslant i\leqslant n} c_i^2 \leqslant \|D^{1/2}\|_2\|D^{-1/2}\|_2(1-\beta)^t\ ,$$

So the general (irregular) undirected case has the extra factor

$$\|D^{1/2}\|_2\|D^{-1/2}\|_2 = \sqrt{\frac{\max_{i\in V} d(i)}{\min_{i\in V} d(i)}}\ .$$

5. Directed graphs

When will random walk on directed graphs converge to a unique stationary distribution?

We will assume all vertices has positive out-degree, for otherwise the random walk dies after reaching a vertex with zero out-degree, and there is no stationary distribution.

We argued that an undirected graph needs to be connected. For directed graphs, the counterpart is "strongly connected components": vertices i and j are in the same strongly connected component if there is a path from i to j, and vice versa.

Given a directed graph G, we can construct the strongly connected component directed graph H of G. The vertices of H are the strongly connected components of G, and every directed edge (i,j) in G induces the directed edge $(\sigma(i), \sigma(j))$ in H of the same weight, where $\sigma(i)$ denotes the strongly connected component of vertex i.

A vertex in H is a sink if it does not have any outgoing edge to a vertex different from itself. A random walk starting in a sink strongly connected component will stay in that component. If H has two sinks, then random walk on G cannot converge to a unique stationary distribution.

If H has a vertex that is not a sink, then random walk on G will "leak" probabilty mass to downstream strongly connected components. In this case, if G has a stationary distribution, all vertices in the source strongly connected component has zero probability.

For simplicity of discussion, we only want to consider convergence to the unique stationary distribution where all vertices have positive probability. For this to happen, H must have only a single vertex. In other words, G must itself be strongly connected.

Is strongly connectedness of G sufficient? We saw that if G is undirected, G also needs to be non-bipartite. For directed graph G, we need G to be non-periodic. Consider the directed k-cycle: Random walk on it has period k and does not converge. Let's say a directed graph is k-partite if its vertices can be partitioned into k disjoint subsets $V_0, \dots, V_{k-1}$, so that all edges go from V_i to $V_{(i+1) \bmod k}$. A directed is aperiodic if it is not k-partite for any integer $k > 1$.

Turns out once the directed graph with at least two vertices is strongly connected and aperiodic, then random walk must converge to a unique stationary distribution (with positive probability everywhere).

Theorem 5.1 (Fundamental Theorem of Markov Chains)**.** *Let G be a directed graph on at least two vertices. If G is strongly connected and aperiodic, then random walk converges to a unique stationary distribution π.*

We will not prove this theorem, but we will give a sketch below.

Does spectrum of P determine convergence to unique stationary distribution? For directed graphs, the transition matrix $P = D^{-1}A$ may have non-real eigenvalues. It is not hard to show that all its (complex) eigenvalues have magnitude at most 1. One can use (some version of) Perron–Frobenius theorem from linear algebra to show the following: Hypothesis of Theorem 5.1 implies that P has a unique eigenvalue of magnitude 1, and this eigenvalue is the real number 1 (so all other eigenvalues have magnitude strictly less than 1). By the same calculations in Proposition 3.1, random walk must converge to a unique stationary distribution.

6. Lazy random walk

One way to avoid bipartiteness or periodicity in random walk is to consider its lazy version: At every vertex, with probability 1/2, stay there, with probability 1/2, move normally.

If we denote by W the transition matrix of lazy random walk, then $W = I/2 + P/2$. If P has eigenvalues $\lambda_1, \dots, \lambda_n$, then W has eigenvalues $\mu_1, \dots, \mu_n$ where $\mu_i = (1 + \lambda_i)/2$.

In particular, if the graph is undirected (and hence all eigenvalues of P are real and between -1 and 1), then all eigenvalues of W are between 0 and 1. Equivalently, $0 \preccurlyeq W \preccurlyeq I$. Since W cannot have eigenvalue -1, this proves that lazy random walk will not suffer from bipartiteness of the graph.