CPS290: Algorithmic Foundations of Data Science **February 28, 2017**

Lecture 12 : Random Walks and Graph Centrality

Lecturer: Kamesh Munagala *Scribe: Kamesh Munagala*

Eigenvector Centrality

We now present a different type of connection between graphs and eigenvalues. In several applications, we wish to find *influential* vertices in a graph. For instance, suppose we model Twitter as a directed graph where there is an edge from user i to user j if i follows j. Then how can we measure importance of a user in the graph? One way of measuring it is look at the in-degree of a user, or the number of users following this user. However, such a measure does not give different amounts of importance to the different followers – if I have a follower who himself has a large number of followers, then I must be really important!

In sociology, importance of a user in a social network is termed *centraility*. As discussed above, *degree centrality* measures the importance of a user in terms of its degree in the graph. The concept of *eigenvector centrality* or *Bonacich centrality* generalizes this as follows: Let π_i denote the importance of user i in a directed social network, and let A denote the adjacency matrix, where $a_{ij} = 1$ if there is an edge from i to j. . A user's importance proportional to the sum of the importance of his or her neighbors.

$$\pi_i = \frac{1}{\lambda} \sum_{j=1}^{n} A_{ji} \pi_j$$

In matrix notation, this can be written as:

$$\lambda \vec{\pi}^T = \vec{\pi}^T A$$

In other words, π is the left eigenvector of A (or an eigenvector of A^T) corresponding to the eigenvalue λ. Of course, the constraint is that $\vec{\pi}$ have non-negative entries, and $\lambda \geq 0$. Is such a solution even possible? The answer lies in a surprising fact about positive matrices.

Theorem 1 (Perron-Frobenius Theorem)**.** *Let A be a square matrix with real, positive entries. Then, the largest eigenvalue λ_1 of this matrix is* real *and unique, meaning that for all other eigenvalues λ_i, we have $|\lambda_i| < \lambda_1$. Furthermore, there exists a unique (up to scaling) eigenvector $\vec{v^*}$ corresponding to λ_1 all of whose entries are real and positive, and this is the only eigenvector with all positive entries. Finally, the power iteration method $\vec{v_{t+1}} \leftarrow A\vec{v_t}$ starting at any initial vector $\vec{v_0}$ not orthogonal to $\vec{v^*}$ converges to $\vec{v^*}$ as $t \to \infty$.*

In fact, it is sufficient to assume A^k has all positive entries for some $k > 0$ for the above theorem to hold. For instance, in the case when A is the adjacency matrix of a graph, there are some entries that are zero. So it does not quite satisfy the Perron-Frobenius theorem. There are two ways to fix this. First, if we assume all entries a_{ij} are at least ϵ for a vanishingly small value ϵ, then the theorem will hold. In that case, there is a unique solution to the equation $\lambda \vec{\pi}^T = \vec{\pi}^T A$, and this can be computed by power iteration. Secondly, if the graph is connected but not bipartite, then A^k will have all positive entries for some $k > 0$, so that the statement of the theorem holds.

Random Walks

The above concept is intricately connected to the concept of *random walks* and *Markov chains*. Consider the following process on the graph: Suppose the location at the current step is vertex i, and let its out-degree be d_i. Then at the next step, the location is a vertex j chosen uniformly at random from the set of neighbors of i. Such a process is called a *random walk* or *Markov chain*, where the word *Markov* means that what happens at the next step depends on which vertex the process currently is at, and does not depend on how the process got to that vertex. Suppose we repeat this process for t steps and consider the vertex that the process stops at. This follows a distribution $\vec{\pi}_t$. Suppose we run the process for one more step, then we can calculate $\vec{\pi}_{t+1}$ as follows: The process is at vertex i with probability π_{it}, and conditioned on this, it transitions to vertex j with probability $\frac{1}{d_i}$. Therefore, the probability of being at vertex j at step $t+1$ is given by:

$$\pi_{jt+1} = \sum_{i|j\in N(i)} \frac{1}{d_i}\pi_{it} \qquad \forall j \in \{1, 2, \ldots, n\}, t$$

As before, we can write this in matrix notation as follows: Let P denote the matrix where entry $p_{ij} = \frac{1}{d_i}$ if i has an edge to j, and 0 otherwise. This denotes the probability that the process at vertex i transitions to vertex j at the next step. In matrix notation, the above equation can be written as:

$$\vec{\pi_{t+1}}^T = \vec{\pi_t}^T P$$

where P is termed the *transition matrix*. Note that the sum of entries of each row in P is 1, since these represent the probabilities of transition out of a vertex. Such a matrix is termed *row stochastic*. It is easy to check that we can use *any* row stochastic matrix with non-negative entries to define the transition probabilities of the random walk on the underlying graph – the entry p_{ij} of the matrix would correspond to the probability of transitioning to j at the next step given the current vertex is i.

What we are interested in is the distribution over vertices of the process if it is run for a long time. Suppose there is a distribution $\vec{\pi^*}$ over vertices such that if the process is run for one more step starting from this distribution, then the distribution at the next step remains the same. This would mean

$$\vec{\pi^*}^T = \vec{\pi^*}^T P$$

This means P^T should have 1 as an eigenvalue, and the corresponding eigenvector is $\vec{\pi}$. It is not a priori clear this is true. But there is a simple way to check this. Instead of looking at eigenvectors of P^T, suppose we want to solve

$$P\vec{v} = \lambda\vec{v}$$

Since P is row-stochastic, summing the entries of a row yields 1. This means

$$P\vec{1} = \vec{1}$$

so that $\vec{1}$ is an eigenvector corresponding to the eigenvalue $\lambda = 1$. Note next that for any square matrix A, both A and A^T have the same eigenvalues. Therefore, 1 is an eigenvalue of P^T. Further, if we assume the entries of P are strictly positive, then by the Perron-Frobenius theorem, the only eigenvector with all positive entries corresponds to the largest eigenvalue. Therefore, $\lambda = 1$ is the largest eigenvalue of P and hence of P^T. Further, the corresponding eigenvector $\vec{\pi^*}$ of P^T has all positive entries, and is unique up to scaling. Therefore the relation $\vec{\pi^*}^T = \vec{\pi^*}^T P$ has a unique solution assuming all entries of P are strictly positive.

Finally, note that since $\lambda = 1$ is the largest eigenvalue, the power iteration method starting at any initial vector (not orthogonal to $\vec{\pi^*}$) converges to the largest eigenvector. But note that

$$\vec{\pi_{t+1}}^T = \vec{\pi_t}^T P = \vec{\pi_1}^T P^t$$

Therefore, the power iteration method is exactly how the probability distribution over vertices evolves if we run the random walk starting with some initial distribution $\vec{\pi_1}$ over the vertices. This means such a process converges to the distribution $\vec{\pi^*}$ over vertices in the limit as t goes to infinity. This distribution is therefore termed the *stationary distribution* of the Markov chain. Put differently, suppose you run the random walk for many steps t and look at the final vertex, its distribution will roughly follow $\vec{\pi^*}$.

Convergence Rate and Revisiting Cheeger's Inequality

There are two caveats to the discussion above. First, the Perron-Frobenius theorem needs all entries of the matrix to be positive, meaning there is a non-zero probability of transitioning from any vertex to any other vertex. A weaker assumption is that P^k has all positive entries for some $k > 0$ and that the underlying graph defined by non-zero p_{ij} is connected. Such a transition matrix is termed *ergodic*. The simplest way to ensure ergodicity holds for a connected transition matrix is to place self-loops at every vertex, so that the process stays at the current vertex with some probability each step. The stationary distribution exists even without such an assumption, but in general, there can be many such distributions. Further, it is easy to show that even if the stationary distribution is unique, this will only be the relative frequency of visiting vertices, and the distribution will never converge to this. For instance, think about running the random walk on a cycle of even length, starting at some vertex. The states at even time steps will always be different from the states at odd time steps, so there is no limiting distribution for such a walk.

Next, the rate at which the distribution $\vec{\pi_t}$ converges to the stationary distribution is termed the *convergence rate*. Since the process is simply power iteration, it depends on the gap between the first eigenvalue (which is 1) and the second eigenvalue λ_2. Consider a d-regular undirected graph. For this graph, we have $\mathcal{L} = I - P$, where $\mathcal{L}$ is the normalized Laplacian and P is the transition matrix of the random walk. Therefore, the second smallest eigenvalue of $\mathcal{L}$, call this κ_2 is equal to $1 - \lambda_2$. Let ϕ denote the conductance of the underlying graph G. Then we know from Cheeger's inequality that

$$\kappa_2 = 1 - \lambda_2 \geq \frac{\phi^2}{2}$$

We have seen that the time taken for power iteration to converge is $O\left(\frac{\log n}{1-\lambda_2}\right)$, which is $O\left(\frac{\log n}{\phi^2}\right)$ by Cheeger's inequality. This means graphs with low conductance cuts need more time to achieve the stationary distribution. In a sense, the above result is intuitive. The conductance of a cut is the ratio of the edges crossing the cut to all the edges that originate in the cut. Take the cut S of conductance ϕ. Suppose the process is choosing a random edge within S the next step. Then, the probability that the process takes an edge of the cut is ϕ. Roughly speaking, this means it takes time $\frac{1}{\phi}$ in expectation to cross the cut. But unless it crosses the cut, the distribution over vertices cannot be close to the steady state distribution. This means the walk will take expected time $\Omega\left(\frac{1}{\phi}\right)$ to reach steady state. This would correspond to the "easy direction" of Cheeger's inequality.

However, this argument was only for one cut, and the graph has exponentially many cuts; if the random walk has to converge to steady state, it must have crossed all these exponentially many cuts at least once in expectation. It is therefore somewhat surprising and non-trivial that it is at most $O\left(\frac{\log n}{\phi^2}\right)$, and this corresponds to the "hard direction" of Cheeger's inequality.

PageRank

Note that one measure of eigenvector centrality can be obtained by considering the stationary distribution of the corresponding random walk. For an undirected graph, the stationary distribution is the same as *degree centrality*. It is easy to check that

$$\pi_i^* = \frac{d_i}{2|E|}$$

is the stationary distribution. To see this, note that

$$\sum_{j \in N(i)} \pi_j^* p_{ji} = \sum_{j \in N(i)} \frac{d_j}{2|E|} \frac{1}{d_j} = \frac{|N(i)|}{2|E|} = \frac{d_i}{2|E|} = \pi_i^*$$

Therefore, for an undirected graph, the eigenvector centrality measure obtained by the stationary distribution of the corresponding random walk is the same as the degree centrality.

For a directed graph, the correspondence is no longer true, and the random walk yields a more interesting stationary distribution. This is the idea behind the *PageRank* algorithm that is widely used to rank web pages, social networks, etc. Given the World Wide Web, we can make each page a vertex. There is an edge from i to j if page i has a link to page j. Eigenvector centrality means a webpage is deemed important if a lot of important pages have links to it. The PageRank algorithm models this as follows: There is a web surfer who is performing a random walk on the web graph. He starts at some page and follows links. We can now calculate the stationary distribution of this process. The catch is that the web graph is often disconnected, and further, the process can reach a page that does not have any links. The fix is the following:

- If a page i has no links, pretend the surfer restarts the process by going to a random page. This corresponds to setting $p_{ij} = 1/n$ for all j.
- At every step, the surfer does the random walk with probability $1 - \alpha$, and goes to a random page (that is, restarts the random walk) with probability α. This corresponds to setting

  $$p_{ij} = (1 - \alpha)\frac{a_{ij}}{d_i} + \frac{\alpha}{n}$$

  where $a_{ij} = 1$ if there is a link from i to j and 0 otherwise. The value of α is typically set to 0.15, and is termed the *teleportation probability*.

With these two changes, all entries of P are strictly positive so that we can apply the Perron-Frobenius theorem. Further, one can show that the second eigenvalue of P is at most $1 - \alpha$, implying power iteration will converge rapidly. This means the stationary distribution is unique, can be computed by performing the random walk, and the number of steps for which the walk must be performed is small. This method is termed the *PageRank algorithm* for computing centrality of web pages. Similar methods are used for ranking users in a social networks, and the power iteration method is implemented in practice on a large scale for this purpose.