Social and Technological Networks Edinburgh, 2015

Lecture 8. Spectral analysis of web graphs

Rik Sarkar *Class notes*

In this lecture, we covered:

- Adjacency matrix.
- Hub and authority rules as operations by the adjacency matrix.
- Convergence of normalized scores to eigen vectors.
- Scaled Pagerank as a matrix and convergence from Perron's theorem.
- Random walk interpretation

The adajacency matrix can be seen both as a representation of the graph and an "operator" that executes the hub and authority rules. The hub rule is equivalent to multiplying the authority vector by M, while the authority rule is equivalent to multiplying the hub vector by M^T.

In the operation, the alternate application of the two rules is critical. It is the combined operation equivalent to MM^T that converges on repetition to a relevant eigen vector.

The final value of hub score converges to

$$\frac{h^{\langle k \rangle}}{c_1^k} = q_1 z_1 + \left(\frac{c_2}{c_1}\right)^k q_2 z_2 + \dots.$$

We can think of $q_1 z_1$ as our target value, while everything else is "error" on top of the target, with $\left(\frac{c_2}{c_1}\right)$ as the largest coefficient. Through iterations, HITS is trying to bring these error terms down to zero.

The speed of this convergence depends on how small the ratio $\left(\frac{c_2}{c_1}\right)$ is (since it is the largest). A smaller ratio implies faster convergence, since then powers of all the error terms go to zero faster.

Spectral gap. What ratio of $\left(\frac{c_2}{c_1}\right)$ is small enough? Usually, anything like a constant is good. Since then a power $\left(\frac{c_2}{c_1}\right)^k$ becomes small in logarithmic number of steps. What is not good is if the ratio is arbitrarily close to 1. For example, if $\left(\frac{c_2}{c_1}\right) \approx \frac{n-1}{n}$, then a logarithmic number of steps do not suffice. It will require about $k = n$ iterations to get $\left(1 - \frac{1}{n}\right)^k$ down to a constant like $\frac{1}{e}$. Hence the spectral gap $|c_1| - |c_2|$ determines the speed of convergence.

Exercise 0.1. *Suppose* $\left(\frac{c_2}{c_1}\right) = m < 1$ *is a constant, show that the the number of iterations needed for* $\left(\frac{c_2}{c_1}\right)^k$ *to become smaller than a parameter* ε *is* $O(\log \varepsilon)$.

Exercise 0.2. *Suppose for a network, the eigen vector $c_1 = 1$, and spectral gap $c_1 - c_2 = \frac{1}{\ln n}$. In how many iterations will the coefficients (such as c_2/c_1) become smaller than ε?*

Exercise 0.3. *HITS ranks web pages as hubs and as authorities separately. It is not clear that this double ranking is so meaningful for web pages. In fact pagerank does not doe this at tall. But what can be a different application where this separate ranking can be useful?*

For the analysis of pagerank, we directly used Perron-Frobenius theorem to show that Pagerank converges to the eigen vector. For more discussion of the theorem see [2, 1]. It is also possible to show it using analysis of the type used for HITS.

Random walks. The random walks interpretation basically says that if we started a random walk in the network, then the pagerank value of a node is exactly the probability of the random walk being at that node after a large number of steps. So in this case the starting pagerank "value" at a node is the probability of the random walk starting at that node, and the values diffusing among the nodes is analogous to this probability spreading in the network until it reaches a steady state.

This also implies that pagerank in a sense looks for the probability of a user being at a certain page if they are clicking links randomly. Of course, this is not quite true, since all links on a page are not equivalent. Can you think of ways to assign different probabilities to different links? Can you think of reasons why we should not weight the link probabilities by their respective degrees or pageranks?

References

[1] Perron's theorem. https://en.wikipedia.org/wiki/Perron

[2] Proof: Perron-frobenius theorem. http://www.math.cornell.edu/˜web6720/Perron-Frobenius_Hannah Cairns.pdf.