April 5, 2010
Franklin Kenter

PageRank, Spectral Graph Theory, and the Matrix Tree Theorem

Introduction

1 Introduction

In this lecture, we will go over the basics of the PageRank algorithm and how it relates to graph theory. Then, we will start our study in spectral graph theory by proving the Matrix Tree Theorem.

2 PageRank

There are two predominant ranking algorithms. PageRank and HITS. PageRank was developed by Brin and Page and is the foundation for what is now the Google search engine in 1997. The other, lesser known, is the HITS algorithm which focuses on "hubs" and "authorities" developed by Kleinburg [1] in 1999. Both of these algorithms are important, most notably, because they are able to capture the essence of a graph without doing any global pattern matching. Further, it should be noted that even though Kleinburg "did not receive any Google stock," mathematically, his algorithm is still very important today. Here, we will focus on the PageRank algorithm, for more information on the HITS algorithm, please see the references below.

As developed by Brin and Page, PageRank is a voting system whereby the weight of each vote is linearly propotional to the total value of the votes it receives. [3] This might seem paradoxial. However, this sets up a linear equation! More specifically, let $G = (V, E)$ be a directed graph. We seek the vector $\mathbf{f} \in \mathbb{R}^{|V|}$, indexed by the verticies in G, that satisfies the following:

$$\mathbf{f} = \alpha \mathbf{1}/n + (1 - \alpha)\mathbf{f}\mathbf{P}$$

$$\mathbf{f}\mathbf{1}^* = \mathbf{1}$$

Where α is a constant between 0 and 1, and $\mathbf{P}$ is the probability transition matrix of the directed graph.

In particular, $\mathbf{f}$ is an eignvalue of the matrix $\alpha\mathbf{1}^*\mathbf{1}/n+(1-\alpha)\mathbf{P}$ with eigenvalue 1. The PerronFrobenius Theorem guarentees that there exists an eigenvector with eigenvalue 1, and provided that $\mathbf{P}$ is strongly connected, all other eigenvalues have modulus strictly less than 1.

Before we continue, we should go over other interpretations of PageRank. Above we gave the vauge voter interpration. However there are two other interprations. First, as described by Brin and Page [3], PageRank also models a bored surfer that surfs the graph (i.e., internet) as follows: at each site, with probability α, the surfer jumps to a random node, each with equal probability, and with probability $1-\alpha$, the surfer clicks on a random link on the page we is currently navigating. This process creates a modified random walk on the original network with the modified probability transition matrix $\bar{\mathbf{P}} = \mathbf{f}(\alpha\mathbf{1}^*\mathbf{1}/n + (1-\alpha)\mathbf{P}$. The PageRank vector, $\mathbf{f}$, is the stationary distribution of this modified walk. Another, surprising interpretation is the impatient sufer. Suppose instead, the impatient surfer starts at a random node (or website), each node with equal probability. Then with probability α, the surfer stops surfing, and with probability $1-\alpha$, the surfer chooses a random incident outward arc (i.e. clicking a link on the page), each arc with equal probability, and consequently, moves to that new node (i.e., the new website). The impatient surfer repeats this process until he stops. The entries of the vector $\mathbf{f}$ represent the probability that the impatient surfer stops at the corresponding node.

To see the second interpretation above, let us look at:

$$\mathbf{f} = \alpha\mathbf{1}/n + (1-\alpha)\mathbf{f}\mathbf{P}$$

And solve for $\mathbf{f}$:

$$\mathbf{f} - (1-\alpha)\mathbf{f}\mathbf{P} = \alpha\mathbf{1}/n$$

$$\mathbf{f}(\mathbf{I} - (1-\alpha)\mathbf{P}) = \alpha\mathbf{1}/n$$

$$\mathbf{f} = \alpha\mathbf{1}/n\frac{\mathbf{I}}{(\mathbf{I} - (1-\alpha)\mathbf{P})}$$

Here, it looks silly to "divide" matrices. However, in this case we are dividing the identity by a linear polynomial in $\mathbf{P}$, so commutitivity is not an issue. However, we do this to emphasize the geometric series:

$$\frac{\mathbf{I}}{(\mathbf{I} - (1-\alpha)\mathbf{P})} = \sum_{n=1}^{\infty}(1-\alpha)^n\mathbf{P}^n$$

Hence, we get:

$$\mathbf{f} = \alpha \frac{\mathbf{1}}{n} \sum_{n=1}^{\infty} (1-\alpha)^n \mathbf{P}^n$$

Which corresponds exactly to the second interpretation above: α corresponds to the chance stopping. $\frac{\mathbf{1}}{n}$ corresponds to choosing a random initial node, and the sum corresponds to taking n steps before stopping.

One observation to make is that the vector $\frac{\mathbf{1}}{n}$ is, in fact, arbitrary. There is no reason we have to use $\frac{\mathbf{1}}{n}$. We could use a general probability distribution vector, $\mathbf{v}$. Hence would could define our "personalized PageRank" using a vector $\mathbf{v}$ by solving the following for $\mathbf{f}$:

$$\mathbf{f} = \alpha \mathbf{v} + (1-\alpha)\mathbf{f}\mathbf{P}$$

Or similarly,

$$\mathbf{f} = \alpha \mathbf{v} \sum_{n=1}^{\infty} (1-\alpha)^n \mathbf{P}^n$$

Now that we understand what PageRank is, let us emphasize ways to (and not to) compute it!

First, solving the initial equation, for a large directed graph, is expensive. PageRank is intended to be used upon very large networks- for example, the internet, with billions of nodes. It is not reasonable to solve a $1000000000 \times 1000000000$ matrix! Further, even when we solve for $\mathbf{f}$, we must invert a matrix which is not much different. There must be some other way, and indeed there is.

Brin and Page used the following recurrence:

$$\mathbf{f}_{t+1} = \alpha \frac{\mathbf{1}}{n} + (1-\alpha)\mathbf{f}_t \mathbf{P}$$

For some reasonable initia probability distributionl vector $\mathbf{f}_0$.

Hence, here, instead of solving a linear equation, or inverting a large matrix, we instead perform matrix multiplication several thousand times, after which $\mathbf{f}_t$ will be close to $\mathbf{f}$. Brin and Page have some results of their simulations here [3].

One thing to keep in mind is the secret role eigenvalues play in PageRank. For the standard random walk with probability transition matrix $\mathbf{P}$ (i.e., $\alpha = 0$), the random walk converages to a vector in the eigenspace corresponding to eigenvalue 1. Furthermore, the *rate of convergence* is determined by next largest eigenvalue in modulus. For example, if $alpha = 0$ the recurrance becomes:

$$\mathbf{f}_t = \mathbf{f}_0\mathbf{P}^t$$

Hence, if one were to decompose $\mathbf{f}_0\mathbf{P}^t$ using eigenspaces, we would see that the total error of $\mathbf{f}_0\mathbf{P}^t$ depends upon its next highest eigenvalue in modulus.

In that sense matrices of the form

$$\alpha\mathbf{1}\mathbf{1}^*/n\frac{\mathbf{I}}{(\mathbf{I}-(1-\alpha)\mathbf{P})}$$

Have received a lot of attention and are the study of a lot of research. This is the probability transition matrix of the modified random walk for the original form of PageRank, and its spectra, as mentioned above, plays a crutial role in the convergance of the modified random walk. In particular, α pushes all non-maximal eigenvalues away from 1, and hence, causes the process to converge faster. Though, one should be careful, for if α is too big, we loose much of the information captured in the PageRank.

Now that we see have eigenvalues play a crutial role in graph theory, we can turn our attention to another important result in graph theory: the Matrix Tree Theorem:

3 The Matrix Tree Theorem

First, let us define some special matrices:

Definition 1. *The diagonal degree matrix,* $\mathbf{D}$ *of a graph, is a* $|V(G)| \times |V(G)|$ *indexed by the verticies such that* $\mathbf{D}_{uv} = d_v$ *whenever* $u = v$*, and =0 otherwise.*

Definition 2. *The adjacency matrix,* $\mathbf{A}$ *of a graph, is a* $|V(G)| \times |V(G)|$ *indexed by the verticies such that* $\mathbf{A}_{uv} = 1$ *whenever* $\{u, v\} \in E(G)$*, and 0 otherwise.*

Definition 3. *Given an orientation of the edges in G (i.e., for each edges, designate one end to the head and the other to be the tail). The oriented incidence matrix,* $\mathbf{B}$ *of a graph, is a* $|V(G)| \times |E(G)|$ *indexed by the verticies and edges such that* $\mathbf{A}_{ue} = 1$ *if v is the head of e, =-1 if v is the tail of e, and =0 otherwise.*

Now, let us define the combinatorial Laplacian:

Definition 4. *The combinatorial Laplacian,* $\mathbf{L}$*, of a graph, is the* $|V(G)| \times |V(G)|$ *indexed by the verticies such that* $\mathbf{L} := \mathbf{D} - \mathbf{A}$*. Where* $\mathbf{D}$ *and* $\mathbf{A}$ *have the same indices as prescribed by* $\mathbf{L}$*.*

Now, this seemlying innocent matrix, is in fact, very special, it produces Kirchhoff's Matrix Tree Theorem:

Theorem 1. *Kirchhoff 1847 Let $\mathbf{L}_v$ be the matrix $\mathbf{L}$ with the row and column corresponding to some vertex v deleted.*

Then, regardless of the choice for v, the number of spanning trees of a graph G is $Det\mathbf{L}_v$.

Proof. Designate an orientation of the edges in G Let $\mathbf{B}$ be the oriented incidence matrix of the graph, and let $\mathbf{B}_v$ be the same matrix with the row corresponding to v deleted.

The following facts are left as an exercise:

Fact 1 $\mathbf{L} = \mathbf{B}\mathbf{B}^*$

Fact 2 $\mathbf{L}_v = \mathbf{B}_v\mathbf{B}_v^*$

So $Det\mathbf{L}_v = Det\mathbf{B}_v\mathbf{B}_v^*$

Now, for $X \subset E$, let $\mathbf{B}_{v,X}$ denote the $n-1 \times n-1$ matrix with only the columns in X. Then,

$$\mathbf{L}_v = \mathbf{B}_v\mathbf{B}_v^* = \sum_{X \subset E, |X|=n-1} (Det\mathbf{B}_{v,X})^2$$

Fact 3 (Also left to the reader.) $|Det\mathbf{B}_{v,X}| = 1$ if X forms a spanning tree, =0 otherwise.

Using fact 3, the sum

$$\sum_{X \subset E, |X|=n-1} (Det\mathbf{B}_{v,X})^2$$

counts all spanning trees. Hence, the claim is proved. □

One key remark: $Det\mathbf{A} = \prod_i \lambda_i$ where $\lambda_1 \leq \ldots \lambda_n$ are the eigenvalues of $\mathbf{A}$. Hence, the Matrix Tree Theorem is, in fact, a result in spectral graph theory.

References

[1] Kleinberg, Jon (1999). ”Authoritative sources in a hyperlinked environment”. Journal of the ACM 46 (5): 604632. *http://www.cs.cornell.edu/home/kleinber/auth.pdf*

[2] Kirchhoff, G. ”ber die Auflsung der Gleichungen, auf welche man bei der untersuchung der linearen verteilung galvanischer Strme gefhrt wird.” Ann. Phys. Chem. 72, 497-508, 1847.

[3] Brin, S.; Page L. (1997). "The Anatomy of a Large-Scale Hypertextual Web Search Engine" *http://infolab.stanford.edu/ backrub/google.html*