Justification and Application of Eigenvector Centrality

Leo Spizzirri
With great debt to P. D. Straffin, Jr
Algebra in Geography: Eigenvectors of Networks[7]

03/06/2011

1 Introduction

Mathematics is only a hobby until it is figured out how to apply it to the world. This paper is a discussion of an application of linear algebra and graph theory to the discipline of geography. From this follows some obvious and some not-so obvious uses outside of theory. We will look at a method of indexing the *centrality* of a node in a graph and try to establish meaning to this value. We will focus especially on the Gould index (eigenvector centrality) of accessibility, demonstrating its calculation and then giving meaning and justification to its use an index. These ideas have their basis in geographical situations, but, with creativity can be applied to graphs describing any connections, such as social structures, capital flow or the spread of disease.

2 Terminology

- **Graph:** A collection of abstract objects, vertices and edges. Pairs of vertices can be connected by edges.
- **Centrality:** The relative importance of a node within a graph. There are various measures to determine this ranking, such as degree centrality and Gould's Index (eigenvector centrality).
- **Adjacency Matrix:** A matrix $\mathbf{A}$ related to a graph by $a_{ij} = 1$ if vertex i is connected to vertex j by an edge, and 0 if it is not.
- **Principle Eigenvalue:** The largest eigenvalue.
- **Nonnegative Matrix:** A matrix $\mathbf{A} = a_{ij}$ such that $a_{ij} \geq 0$ for all i, j.
- **Primitive:** A matrix $\mathbf{A}$ for which there exists a positive integer n such that the elements of $\mathbf{A}^n$ are strictly non-negative.
- **Connected Graph:** A graph is connected if every pair of vertices is conntected by some path.
- **Diameter of a Connected Graph:** This is the smallest integer n such that any vertex may reach any other vertex by a path of no more than n in size.

3 Geographic Motivaion

We are trying to get at the geographical importance of a node within a graph. In graph theory, this is referred to as centrality. The simplest (and, in many

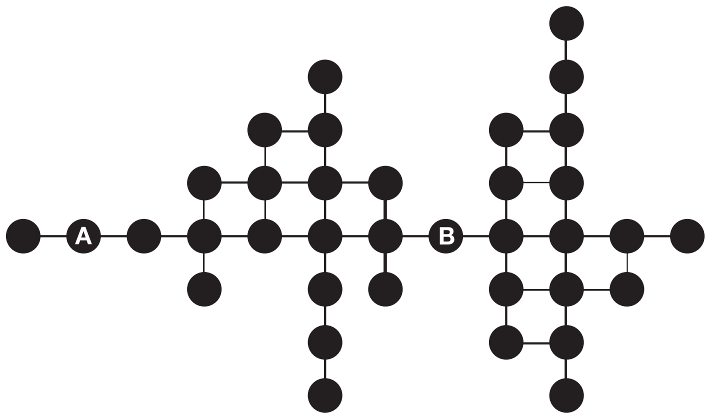

Figure 1: A Sample Graph

contexts, least useful) way to do this is the degree centrality. We define degree centrality as

$$C_D(v) = \frac{deg(v)}{n-1},$$

where $C_D(v)$ is the degree centrality of vertex v, n is the number of vertices in the graph and $deg(v)$ is the number of edges from v to another vertex. The issue with this is obvious. We are pretending that a single characteristic (number of connected edges) characterizes the entire node. For instance, in **Figure 1**, nodes A and B would both have a degree centrality of $0.0\overline{5}$. But our intuition seems to tell us they are not equally important to the graph. Clearly, in many contexts, say, flow of information, B would have much greater control over the entire graph than A. A more sophisticated measure is needed.

4 Gould Index of Accessibilty

One method more sophisticated than the degree of centrality was proposed by Peter Gould [4]. We examine this method, Gould's index of accessibility, also known as eigenvector centrality. First we look at the method, then we will attempt to justify it. We work with the graph in **Figure 2**.

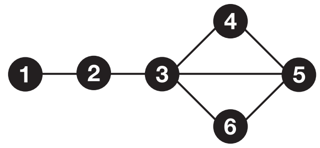

Figure 2: A Simple Graph

We begin with the adjacency matrix **A** of the graph, defined as

$$a_{ij} = \begin{cases} 1 & \text{if the vertices } i \text{ and } j \text{ are joined by an edge,} \\ 0 & \text{if the vertices } i \text{ and } j \text{ are not joined by an edge.} \end{cases}$$

It is usual to define the the entries a_{ii}, the diagonal, as 0. We have the adjacency matrix

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}.$$

Using degree centrality we get the following ranking for the vertices n_i:

$$\begin{aligned} n_1 &= 0.2 \\ n_2 &= 0.4 \\ n_3 &= 0.8 \\ n_4 &= 0.4 \\ n_5 &= 0.6 \\ n_6 &= 0.4 \end{aligned}$$

The index that Gould proposes uses the (normalized) eigenvector from the principle eigenvalue. In this case we have eigenvalues $\lambda_0 = 2.70559$, $\lambda_1 = -1.851$, $\lambda_2 = -1.350$, $\lambda_3 = 1.056$, $\lambda_4 = -0.560$ and $\lambda_5 = 0$. λ_0 is the principle eigenvalue and we have eigenvector

$$\mathbf{v_0} = \begin{bmatrix} 0.092 \\ 0.249 \\ 0.581 \\ 0.405 \\ 0.514 \\ 0.405 \end{bmatrix}.$$

This is Gould's index of accessibility. The *ith* entry corresponds to the *ith* vertex and this is its accessibility rank. It seems to agree with intuition. The striking difference between this ranking and degree centrality is vertex 2 ranking below 4 and 6. Agreement with intuition is encouraging, but hardly convincing, so we proceed to justify granting this computation the title of index of accessibility.

5 Perron-Frobenius Theorem

To proceed we require several results from linear algebra. First, we define a new matrix

$$B = A + I = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}.$$

We have simply replaced the diagonal zeros with ones. This has the effect of giving eigenvectors of B that are exactly one greater than A, with identical eigenvectors. Clearly, B will be symmetric (as was A), and, as such, can be diagonalized by an orthogonal matrix. This, in turn, guarantees that the eigenvalues of B are real, so talking about their relative sizes is appropriate (we can choose the largest). Further, we will also be guaranteed real values populating the eigenvectors.

Finally, we will require *The Perron-Frobenius Theorem.* This needs two definitions. We call matrix $\mathbf{D} = d_{ij}$ **nonnegative** if $d_{ij} \geq 0$ for all i, j. A primitive matrix is a nonnegative square matrix $\mathbf{D}$ where there exists an integer $N > 0$ such that every value of $\mathbf{D}^N$ is strictly positive. We now state the theorem, a proof of which is found in the appendix.

Perron-Frobenius Theorem 1. *If* $\mathbf{M}$ *is an* $n \times n$ *nonnegative primitive matrix, then there is a largest eigenvalue* λ_0 *such that* *(i)* λ_0 *is positive*

(ii) λ_0 *has a unique (up to a constant) eigenvector* $\mathbf{v_1}$*, which may be taken to have all positive entries*

(iii) λ_0 *is non-degenerate*

(iv) $\lambda_0 > |\lambda|$ *for any eigenvalue* $\lambda \neq \lambda_0$

To see that we may apply the Perron-Frobenius Theorem, we examine $\mathbf{B}^k$. Taking $\mathbf{B}$ to the k^{th} power has the effect that the $ijth$ entry of $\mathbf{B}^k$ counts the number of ways of travelling from vertex i to vertex j by paths of length k, including stopovers. We illustrate how this works with the following graph.

This graph has the following adjacency matrix, with ascending powers:

$$\mathbf{M} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \quad \mathbf{M}^2 = \begin{bmatrix} 2 & 2 & 1 & 0 \\ 2 & 3 & 2 & 1 \\ 1 & 2 & 3 & 2 \\ 0 & 1 & 2 & 2 \end{bmatrix}, \quad \mathbf{M}^3 = \begin{bmatrix} 4 & 5 & 3 & 1 \\ 5 & 7 & 6 & 3 \\ 3 & 6 & 7 & 5 \\ 1 & 3 & 5 & 4 \end{bmatrix}$$

Let's look at the 3^{rd} entry of the first row of M^2. This was computed multiplying the 1^{st} row vector of M by the 3^{rd} column vector. The calculation picks out the connection $1 \rightarrow 2$ and $2 \rightarrow 3$, showing now a path length two connecting $1 \rightarrow 3$. Similarly, now that that path is established, computing $\mathbf{M}^3$ offers a path length 3 from $1 \rightarrow 4$. In this way $\mathbf{B}^k$ counts all paths of length k connecting any pair of vertices.

Now, the diameter of a connected graph is the minimum integer k such that any two vertices i, j, can be connected by a path no longer than k. Therefore, choosing $k \geq diameter(B)$, we guarantee that the entries of $\mathbf{B}^k$ are positive. Then $\mathbf{B}$ is primitive and Perron-Frobenius applies.

Now, we have guaranteed that for a connected graph we will find a well-defined principle vector v_0 whose entries are all positive. We consider a vector $\mathbf{x}$ that is not orthogonal to $\mathbf{v_0}$. We have

$$\mathbf{x} = \alpha_0 \mathbf{v_0} + \alpha_1 \mathbf{v_1} + \cdots + \alpha_n \mathbf{v_n}, \quad (\alpha_0 \neq 0).$$

Then

$$B^k\mathbf{x} = \lambda_0^k\alpha_0\mathbf{v_0} + \lambda_1^k\alpha_1\mathbf{v_1} + \cdots + \lambda_n^k\alpha_n\mathbf{v_n}, \quad (\alpha_0 \neq 0)$$

$$\frac{B^k\mathbf{x}}{\lambda_0^k} = \alpha_0\mathbf{v_0} + \frac{\lambda_1^k\alpha_1\mathbf{v_1}}{\lambda_0^k} + \cdots + \frac{\lambda_n^k\alpha_n\mathbf{v_n}}{\lambda_0^k}, \quad (\alpha_0 \neq 0).$$

Now, letting $k \to \infty$, $\frac{\mathbf{B}^k\mathbf{x}}{\lambda_0^k} \to \alpha_0\mathbf{v_0}$ since $\lambda_0 > \lambda_i, i \neq 0$. What we see here is that as we let k increase, the ratio of the components of $\mathbf{B}^k\mathbf{x}$ converge to $\mathbf{v}_0$. This fact will be essential to our justifications that follow.

6 Justification

Our first justification comes directly from the fact shown above, the convergence of $\mathbf{B}^k\mathbf{x}$ to the ratio of the components of $\mathbf{v_0}$). Even before Gould gave us his index, the row sums of $\mathbf{B}^k$ were an oft used measure of accessibility. As we saw above, entry j in row i gives all k length paths from i to j. So the row sum is the total number of k length paths to all other vertices from vertex i. If we take a column vector $\mathbf{x}$, whose entries are all 1, then the row sums of $\mathbf{B}^k$ are just $\mathbf{B}^k\mathbf{x}$. As we saw above, counting longer and longer paths will bring $\mathbf{B}^k\mathbf{x}$ to the same ratio as Gould's index.

As a second justification, let's imagine, cheerfully, the spread of a disease. Using the graph in **Figure 2**, we suppose there is a population of people at each vertex. One person at one vertex starts with the disease and at each time step spreads it to one person at each connected vertex. At each time step, each infected person follows this same pattern. If our initially infected person starts at vertex two, we represent this distribution $\mathbf{d}$ as

$$\mathbf{d} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Then, after k time steps, $\mathbf{B}^k\mathbf{d}$ represents the new distribution of infected people. We illustrate below.

$$\mathbf{Bd} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{B}^2\mathbf{d} = \begin{bmatrix} 2 \\ 3 \\ 2 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{B}^3\mathbf{d} = \begin{bmatrix} 5 \\ 7 \\ 8 \\ 4 \\ 5 \\ 4 \end{bmatrix}$$

When we normalize these vectors, an interesting pattern appears.
Let $\mathbf{d}_k = \frac{\mathbf{B}^k\mathbf{d}}{|\mathbf{B}^k\mathbf{d}|}$. We have

$$\mathbf{d}_1 = \begin{bmatrix} 0.577 \\ 0.577 \\ 0.577 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{d}_2 = \begin{bmatrix} 0.447 \\ 0.671 \\ 0.447 \\ 0.224 \\ 0.224 \\ 0.224 \end{bmatrix}, \quad \mathbf{d}_3 = \begin{bmatrix} 0.358 \\ 0.501 \\ 0.573 \\ 0.286 \\ 0.358 \\ 0.286 \end{bmatrix}, \quad \mathbf{d}_{30} = \begin{bmatrix} 0.092 \\ 0.249 \\ 0.581 \\ 0.405 \\ 0.514 \\ 0.405 \end{bmatrix}.$$

This will always happen. Consider $\mathbf{d}$, our initial distribution. It will always have at least one positive entry, so it cannot be orthogonal to $\mathbf{v_0}$ and, as shown above, the ratio of its entries will converge to that of $\mathbf{v_o}$, Gould's index. So, no matter where the disease starts (or how many start with it), the ratio of its distribution will settle into Gould's index, or eigenvector centrality ranking. The fact that such an intuitive process leads (relatively quickly) to this measure certainly adds weight to our case.

Our last justification uses degree centrality, discussed above. We define a vector $\mathbf{c}$ whose entries correspond to the degree centrality of the corresponding vertex (counting each as adjacent to itself). In this case, we have $\mathbf{c} = \begin{bmatrix} 2 & 3 & 5 & 3 & 4 & 3 \end{bmatrix}$.

Now, as a measure of centrality, this is too naive. However, a simple improvement follows the next line of reasoning. The connectedness of a vertex is not just how many other vertices it is connected too, but also how connected *those* vertices are. Thus, we add to each vertex its degree plus the degree of each vertex it is connected to. We can iterate this process, with our k^{th} iteration $\mathbf{c}_k$ being $\mathbf{c}_k = \mathbf{B}^{k-1}\mathbf{c}$. And, as before, the ratio of the entries of $\mathbf{c}_k$ go to Gould's index as k goes to infinity. In other words, $\lim_{k\to\infty} \frac{\mathbf{c}_k}{|\mathbf{c}_k|} = \mathbf{v}_0$. This idea, that connections have weight depending on who they are connected, coupled with an iterative computation, is central to some some modern uses of eigenvector centrality, as we will see in the next section.

So, we see, modeling different situations, we've arrived back at Gould's index each time. While a perfect description of what one might mean by 'connectedness' may not be possible, this certainly puts Gould's index in a font-runner position. We'll now look at some of its applications and examine in some detail how it can be adapted to the real world.

7 Usefulness

Eigenvector centrality (or similar modifications/extensions) crops up in vast array of fields. Sociologists were some of the first to use versions of eigenvector centrality (and other methods) to measure connection between players in social groups[3]. Eigenvector centrality has been used to study social networks of the lek-mating wire-tailed manakin and was helpful in predicting a males probability of social rise[1]. Work has been done to apply eigenvector centrality to the ranking of college football teams[5].

A more well known application comes from something very familiar to almost all of us: Google. Google's PageRank is the system by which the search engine ranks the pages in its search results (though the name PageRank actually comes from one of its key inventors, Larry Page). The graph describing the World Wide Web is incredibly, almost impossibly, complicated and sorting it out is no small feat. Here is the task in the words of Google founders, Sergey Brin and Larry Page:

> [W]e have taken on the audacious task of condensing every page on the World Wide Web into a single number, its PageRank. PageRank is a global ranking of all web pages, regardless of their content, based solely on their location in the Web's graph structure.
>
> Using PageRank, we are able to order search results so that more important and central Web pages are given preference. In experiments, this turns out to provide higher quality search results to users. Th intuition behind PageRank is that it uses information which is exernal to the Web pages themselves - their backlinks, which provide a kind of peer review. Furthermore, backlinks from "important" pages are more significant than backlinks from average pages. This is encompassed in the recursive definition of PageRank.[2]

Well, what is that recursive definition? We will describe a simplified version. To actually apply eigenvector centrality measure to the web, google has had to extend the approach to account for the variety of connections and structures of the internet. The basic definition of PageRank is as follows.

Let w be a web page and L_w be the number of links from w to other pages and C_w be the set of pages that link to w. Finally, we have a number k for normalization. Then PageRank $R(w)$ is:

$$R(w) = k \sum_{v \in C_w} \frac{R(v)}{L_v}$$

The definition is iterative. Each page can be given an arbitrary PageRank, then the iteration is run until it converges. We can say this another way. We let M be a square matrix such that $M_{v,w} = \frac{1}{L_v}$ if there is a link from v to w and 0 otherwise. We take R to be a vector of the current PageRanks of web pages. Then $R = kMR$. So, R is an eigenvector of M and, in now familiar fashion, iteration will bring R to converge to the dominant eigenvector of M. This is Google's version of the Gould index.

While they have made modifications to our basic model, such as making the impact of each connection inversely proportional to the number of connections (the factor $\frac{1}{L_v}$), the heart of the model follows the structure of eigenvector centrality. Obviously, there are greater sophistication necessary when dealing with something as vast as the internet, but Google has shown that large systems have an inherent 'intelligence' and PageRank (by way of Gould's index) helps us uncover it.

8 Summary and Conclusions

Gould's index has appeared in various forms, in many disciplines and by alternate routes. It's use as a measure has been shown to be invaluable in a wide variety of contexts, with the added benefit that it can be computed in a variety of ways. It also seems to fall in line with our intuitive understanding of connections.

The internet age seems ripe for application of eigenvector centrality and its associated ideas. With the seemingly endlessly increasing popularity of social media, being able to formally describe the ways in which people are connected is of great interest. One of the strengths of this kind of evaluation (in my opinion, though it may be debated) is its reliance on *collective* evaluation, as opposed to *expert* evaluation.[3] Collective evaluation is usually more fluid and fast. This is not strictly a feature of the web. I look forward to these methods being applied in new ways to things like public transportation planning, emergency infrastructure or social networking systems.

9 Appendix

9.1 Proof of Perron-Frobenius Theorem

Perron-Frobenius Theorem 2. *If* $\mathbf{M}$ *is an* $n \times n$ *nonnegative primitive matrix, then there is a largest eigenvalue* λ_0 *such that*

(i) λ_0 *is positive*

(ii) λ_0 *has a unique (up to a constant) eigenvector* $\mathbf{v_1}$*, which may be taken to have all positive entries*

(iii) λ_0 *is non-degenerate*

(iv) $\lambda_0 > |\lambda|$ *for any eigenvalue* $\lambda \neq \lambda_0$

Proof. (i) Let λ_i be the eigenvalues of $\mathbf{M}$. We have $\sum \lambda = \mathbf{Tr}\ \mathbf{M} > 0$. Since the eigenvalues of $\mathbf{M}$ are real, then we have $\lambda_0 > 0$.
(ii) Let v_j be any real, normalized eigenvector of λ_0, then

$$\lambda_0 \mathbf{v} = \sum_j m_{ij} v_j, \quad (i = 1, 2...n),$$

and let $u_j = |v_j|$. Then

$$0 < \lambda_0 = \sum_{ij} a_{ij} v_i v_j = |\sum_{ij} a_{ij} v_i v_j| \leq \sum_{ij} a_{ij} u_i u_j$$

The variational principle gives us that the right hand side is less than or equal to λ_0, and equality if and only if $\mathbf{u}$ is an eigenvector of λ_0. If $u_i = 0$ for any i, then since $m_{ij} > 0$, then all $u_j = 0$, which cannot be true (if the matrix is primitive, not positive, we look at $\mathbf{M}^k \mathbf{u} = \lambda^k \mathbf{u}$ and apply the same argument).

(iii)If *lambda*$_0$ is degenerate, since $\mathbf{M}$ is real symmetric we can find two real orthonormal eigenvectors $\mathbf{u}$**,** $\mathbf{v}$ belonging to λ. Suppose that $\mathbf{u_i} < 0$ for some i. Adding equations

$$\lambda_0 \mathbf{u} = \sum_j m_{ij} u_j, \quad (i = 1, 2...n), \quad \lambda_0 |\mathbf{u}| = \sum_j m_{ij} |u_j|, \quad (i = 1, 2...n),$$

we get 0=$\lambda_0(u_i + |u_i|) = \sum_j m_{ij}(u_j + |u_j|)$, so it follows $u_j + |u_j| = 0$ for all j. Thus we have either $u_j = |u_j| > 0$ for every j, or $u_j = -|u_j| < 0$ for every j. The same applies to $\mathbf{v}$. Thus $\sum_j v_j u_j = \pm \sum_j |v_j u_j| \neq 0$, so $\mathbf{u}$ and $\mathbf{v}$ cannot be orthogonal and λ_0 is non-degenerate.

(iv) Let $\mathbf{w}$ be a normalized eigenvector of $\lambda \neq \lambda_0$, $\lambda \neq \lambda_0$. Now the non-degeneracy of λ_0, together with the variational property gives

$$\lambda_0 > \sum_{ij} m_{ij}|w_i||w_j| \geq |\sum_{ij} m_{ij} w_i w_i| = |\lambda|.$$

This proof is largely taken from F. Ninio's here [6].

□

10 References

References

[1] J.G. Blake, B. A. Loiselle, D. B. McDonald, P. G. Parker and T.B. Ryder, Social Networks in the Lekmating Wire-tailed Manakin (Pipra Filicauda), *Proceedings of The Royal Society B*, (April, 2008)

[2] S. Brin and L. Page, The PageRank Citation Ranking: Bringing Order to the Web, *Stanford InfoLab*, (Jan., 1998)

[3] M. Franceschet, PageRank: Standing on the Shoulders of Giants, *http://arxiv.org/pdf/1002.2858v3*, (Dec., 2010)

[4] P. R. Gould, On the Geographical Interpretation of Eigenvalues, *Transactions of the Institute of British Geographers*, No. 42 (Dec., 1967), pp. 53-86

[5] J.P. Keener, The Perron-Frobenius Theorem and the Ranking of Football Teams, *SIAM Review*, Vol. 35, No. 1 (Mar., 1993), pp. 80-93

[6] F. Ninio, A Simple Proof of the Perron-Frobenius Theorem for Positive Symmetric Matrices, *IOPscience*, (April, 1976)

[7] P. D. Straffin, Jr, Algebra in Geography: Eigenvectors of Networks, *Mathematics Magazine*, Vol. 53, No. 5(Nov., 1980), 269-276