

Centrality and Network Analysis: A Perturbative Approach to Dynamical Importance

Nils Hultgren

Thesis Advisors:

Ivan Matic

Peter Mucha

Abstract:

The purpose of this paper is to investigate methods for analyzing networks from an algebraic perspective. The main focus will be on the dominant eigenpair of the adjacency matrix representing a graph, as many different centrality measures for networks are cast in terms of eigenvalue problems. We will see how it is affected by small perturbations to the graph and also propose methods for estimating these changes. This analysis will be justified from the ground up, only assuming basic elements of linear algebra.

1 Introduction

A popular folklore among mathematicians is that a relative measure of success is given by an individual's Erdős Number: the minimum number of connections through collaborators on published papers that it takes to reach Hungarian mathematician Paul Erdős. While this certainly doesn't completely describe one's worth in the mathematical community, it does beg the question of how valuable of a measure is it, and what are some better ways

to extract meaning from a network such as this (there has actually been a lot of research on the mathematics of the co-authorship network, see [10]). Answering questions such as these can provide insight into dozens of different fields, which is why network analysis is an increasingly exciting area of study. Relevant examples include social networks such as friendships between people on Facebook, or physical networks such as the system of railroads in the United States. Biological networks are everywhere, from the network of neurons in the brain to the Markov chain of hermit crabs leaving their shells to inhabit abandoned ones. After the completion of the human genome project, network analysis has become an enormous component of molecular biology (for more background of network analysis applications see [6]). Over the past twenty years, with advances in the speed and storage capability of computers, as well as the copious amounts of network data readily available, the field of network analysis has become increasingly pervasive and necessary. A frequently important concept, which we investigate in this paper, is the centrality of elements in a network. In addition, in most practical examples, networks are constantly evolving. As a result, we consider a perturbative approach to understanding how small changes to a network immediately affect its properties as a whole.

1.1 Background of Network Analysis

A network is any system consisting of different actors and connections between them. This can be represented by a mathematical graph. In some cases, it is more prudent to use a directed graph, for instance in network of games played between basketball teams a directed edge between two nodes would correspond to one team beating another. One of the fundamental tools for analyzing graphs is their adjacency matrices:

Definition: An *Adjacency Matrix* for network graph G is a matrix $A = [a_{ij}]$ such that

$$a_{ij} = \begin{cases} 1, & \text{if there is an edge connecting node } i \text{ to node } j \\ 0, & \text{otherwise} \end{cases}$$

In the case of an undirected graph, its adjacency matrix will be symmetric. In many instances involving very large networks, the adjacency matrix will be very sparse, and this property can be used to facilitate computational

analysis. This matrix can yield important information about the graph as a whole, as well as insight into the importance of each node. We are most interested in the set of eigenvalues of the adjacency matrix, which is called the *graph spectrum*. The largest magnitude eigenvalue is of particular interest and is called the *spectral radius*.

1.2 Centrality Measures

The relative importance of an actor in a network is often significant, but this concept is not inherently mathematical. What does it mean to be important, or central in a network? It may be clear that Grand Central station in New York is central to the network of the subway system but it remains to define this mathematically. An obvious place to start is the degree of the node, which is the total number of connections it has to other nodes. Each of the centrality measures we will discuss will be for node i of an n node graph $G = (V, E)$ where V is the set of vertices, and E is the set of edges. A comprehensive resource about centrality measures is [15].

Degree Centrality: $\text{Deg}(i)/(n - 1)$

In order to make better comparisons between graphs of different sizes the degree is standardized by dividing by $n - 1$, the maximum possible degree of any node. We also note here that the degree of node i can be calculated by the sum of row (or column) i in the adjacency matrix. Degree centrality is clearly a local measure, and does not necessarily indicate how important the node is to the graph overall. Another local centrality measure is the clustering coefficient.

$$\text{Clustering Coefficient: } \frac{\sum_{j < k \in N_i} a_{jk}}{\binom{\text{Deg}(i)}{2}}$$

Here N_i denotes the neighborhood of node i , or the set of all of the nodes adjacent to i . This measures how close a node and its neighbors are to being a *clique* or complete subgraph. The numerator counts the number of edges between nodes in this neighborhood, and the denominator is the total possible number of edges.

Another important quality is how close a node is to all of the other nodes in the graph.

Closeness Centrality: $\frac{n-1}{\sum_{j \in V} d(i, j)}$

Here $d(i, j)$ denotes the shortest path distance between nodes i and j , or geodesic distance. The smallest this sum could be is $n-1$ which would occur if i was adjacent to all the other nodes. So in this case we multiply by $n-1$ to standardize the measure. This would be undefined if i was not connected to some node j , but we are only concerned with connected graphs. Computing this value requires knowledge of the length of every shortest path between each node, which may not seem initially easy to find, but we can think of the adjacency matrix as a transition matrix for a markov chain. A_{ij}^k is the number of paths between i and j of length k (see [3]). Concerning geodesic paths between nodes, one may also wonder which nodes lie on many of these paths.

Betweenness Centrality: $\sum_{j < k} \Phi_i(j, k)$

$\Phi_i(j, k)$ is the proportion of shortest paths (there could be many of the same length) between nodes j and k that go through i . This measure is standardized by dividing by $\binom{n-1}{2}$, the total number of pairs of other nodes. If an internet server with a high betweenness score went down, it would potentially slow down the network because the geodesic distance between many of the servers would increase. For investigations into Closeness and Betweenness see [4] and [12].

One may also desire a measure that places importance on being connected to other high scoring nodes.

Eigenvector Centrality: $u(i) = \frac{\sum_{j \in N_i} u(j)}{\lambda}$

So here λ is a scalar, and each node's score is proportional to the sum of the scores of its neighbors. If u is represented as a vector where u_i is the eigenvector centrality of i , then we can rewrite this equation with the adjacency matrix as: $u\lambda = Au$. So u must be an eigenvalue of A , and in fact if we require all of its entries to be positive, then λ is the spectral radius of A . This fact is due to the Perron-Frobenius theorem which we prove in section 3. So eigenvector centrality is dependent on the spectral radius, and in fact, eigenvector centrality has been shown to be statistically correlated to other centrality measures especially closeness and degree (see [14]).

While each of these centrality measures describes a different quality, many of them are related to the graph spectrum. An important question then is how much does each individual node impact the spectral radius. This can be observed by measuring the change when the node is removed from the graph. Dynamical Importance is a value for observing this change, originally proposed in [11].

Definition: The *Dynamical Importance* of node i in a graph is the proportional change to the largest magnitude eigenvector of its adjacency matrix A when node i is removed from the graph: $I_i = \frac{\Delta\lambda}{\lambda}$

1.3 Symmetric Matrices

The remainder of this paper will build towards a perturbative analysis of Dynamical Importance, and for the present work we focus on undirected networks. Adjacency matrices for undirected networks are symmetric, so before attempting to analyze them we first need to quickly build a few elementary facts to be used in proofs involving symmetric matrices.

Let A be a symmetric matrix: $A = A^T$

Since A is symmetric, we know that any right eigenvector is the transpose of a left eigenvector with the same eigenvalue. If we assume $Ax = \lambda x$ then we have,

$$x^T A = x^T A^T = (Ax)^T = (\lambda x)^T = x^T \lambda$$

Now, we note that the eigenvectors of A are orthogonal because, for different eigenvectors x_i and x_j :

$$\begin{aligned} Ax_i &= \lambda_i x_i \\ x_j^T Ax_i &= x_j^T \lambda_i x_i \\ \lambda_j x_j^T x_i &= \lambda_i x_j^T x_i \end{aligned}$$

If $\lambda_i \neq \lambda_j$ then we know that $x_j^T x_i = 0$, and then we are done.

If $\lambda_i = \lambda_j$ then any linear combination of x_j and x_i will also be an eigenvector with the same eigenvalue: $A(x_j + cx_i) = Ax_j + cAx_i = \lambda_i x_j + c\lambda_i x_i = \lambda_i(x_j + cx_i)$

So if there are t eigenvectors for this eigenvalue, they will form a subspace whose dimension is the geometric multiplicity of the eigenvalue; and for a t -dimensional subspace, we can choose t orthogonal vectors. Thus, for an n -dimensional symmetric matrix, there are n orthogonal eigenvectors.

2 Power Method

The dominant eigenvector (the one corresponding to the largest eigenvalue) of the symmetric $n \times n$ adjacency matrix A is often of interest. It directly gives the eigenvector centrality information for all of the nodes in the network, and in addition, in our subsequent approximations of Dynamical Importance, we will need to use the components of this vector. Calculating the characteristic equation of a huge matrix would be extremely computationally costly, but there is a way around this issue. The power method is specified by an iteration which starts from any vector (except vectors in the codimension one space perpendicular to the dominant eigenvector) and approaches the dominant eigenvector of A :

$$u_{i+1} = \frac{Au_i}{\|Au_i\|} \quad (1)$$

Here we wish to prove that this converges to the dominant eigenpair. In general, this is true for any matrix, but we prove this for any symmetric matrix A . Also, we show the stronger fact that each iteration produces a vector closer to the dominant eigenvector (this is not necessarily true for a nonsymmetric matrix).

Let the set of unique eigenvalues in increasing order of A be : $\{\lambda_1, \lambda_2 \dots \lambda_n\}$ corresponding to eigenvectors: $\{x_1, x_2 \dots x_n\}$ (which we can pick to be unit vectors).

Since the eigenvectors form an orthonormal basis for R^n we can express any vector:

$$u_1 = a_1x_1 + a_2x_2 \dots + a_nx_n$$

with scalars $a_1 \dots a_n$

So inserting this into our iteration equation yields:

$$u_2 = \frac{A(a_1x_1 + a_2x_2 \dots + a_nx_n)}{\|A(a_1x_1 + a_2x_2 \dots + a_nx_n)\|} = \frac{\lambda_1a_1x_1 + \lambda_2a_2x_2 \dots + \lambda_na_nx_n}{C_1}$$

where C_1 is a scalar which is the norm of the numerator and so $\|u_2\| = 1$. Iterating again yields:

$$u_3 = \frac{A(\lambda_1a_1x_1 + \lambda_2a_2x_2 \dots + \lambda_na_nx_n)}{C_1\|A(\lambda_1a_1x_1 + \lambda_2a_2x_2 \dots + \lambda_na_nx_n)/C_1\|} = \frac{\lambda_1^2a_1x_1 + \lambda_2^2a_2x_2 \dots + \lambda_n^2a_nx_n}{C_2}$$

And in general:

$$u_{i+1} = \frac{\lambda_1^i a_1 x_1 + \lambda_2^i a_2 x_2 \dots + \lambda_n^i a_n x_n}{C_i}$$

So as i goes to infinity, the λ_n term in the numerator dominates. More formally, since we are only concerned with the direction of the resulting vector (we know it will always be of length 1, due to the C_i scalar), we can divide out $\lambda_n^i a_n / C_i$ to get:

$$u'_{i+1} = \left(\frac{\lambda_1}{\lambda_n}\right)^i \frac{a_1}{a_n} x_1 + \left(\frac{\lambda_2}{\lambda_n}\right)^i \frac{a_2}{a_n} x_2 \dots + x_n$$

Here we are assuming that $a_n \neq 0$ since the starting vector is not orthogonal to x_n . So since λ_n is the largest eigenvalue we have:

$$\lim_{i \rightarrow \infty} u'_{i+1} = 0 \cdot \frac{a_1}{a_n} x_1 + 0 \cdot \frac{a_2}{a_n} x_2 \dots + x_n = x_n \quad (2)$$

So we have shown that the power method converges to the eigenvector corresponding to the largest eigenvalue, but in some of our later approximations, we only wish to iterate this method once, so we need to additionally prove that one iteration of the power method yields a vector which is closer (in this case we choose to use the Euclidean distance norm) to x_n than the original.

To make things easier, we can assume that u_1 is a unit vector. We have:

$$\begin{aligned} u_2 &= \frac{A(a_1 x_1 + a_2 x_2 \dots + a_n x_n)}{\|A(a_1 x_1 + a_2 x_2 \dots + a_n x_n)\|} = \frac{\lambda_1 a_1 x_1 + \lambda_2 a_2 x_2 \dots + \lambda_n a_n x_n}{\|\lambda_1 a_1 x_1 + \lambda_2 a_2 x_2 \dots + \lambda_n a_n x_n\|} \\ &= \frac{\lambda_1 a_1 x_1 + \lambda_2 a_2 x_2 \dots + \lambda_n a_n x_n}{\sqrt{(\lambda_1 a_1 x_1 + \lambda_2 a_2 x_2 \dots + \lambda_n a_n x_n)^T (\lambda_1 a_1 x_1 + \lambda_2 a_2 x_2 \dots + \lambda_n a_n x_n)}} \end{aligned}$$

And since $x_j^T x_i = 0$ and $x_i^T x_i = 1$ (since A is symmetric), we have:

$$u_2 = \frac{\lambda_1 a_1 x_1 + \lambda_2 a_2 x_2 \dots + \lambda_n a_n x_n}{\sqrt{\lambda_1^2 a_1^2 + \lambda_2^2 a_2^2 \dots + \lambda_n^2 a_n^2}}$$

We want to show that $\|u_2 - x_n\| < \|u_1 - x_n\|$.

$$\|u_2 - x_n\| = \|(\lambda_1 a_1 x_1 + \lambda_2 a_2 x_2 \dots + (\lambda_n a_n - C) x_n) / C\|$$

$$\|u_1 - x_n\| = \|a_1x_1 + a_2x_2 \dots + (a_n - 1)x_n\|$$

Where $C = \sqrt{\lambda_1^2 a_1^2 + \lambda_2^2 a_2^2 \dots + \lambda_n^2 a_n^2}$ And as before, $x_j^T x_i = 0$ and $x_i^T x_i = 1$ so we have :

$$\begin{aligned} \|u_2 - x_n\| &= |(\lambda_1^2 a_1^2 + \lambda_2^2 a_2^2 \dots + (\lambda_n a_n - C)^2)/C^2|^{1/2} \\ &= |(\lambda_1^2 a_1^2 + \lambda_2^2 a_2^2 \dots + \lambda_n^2 a_n^2 - 2C\lambda_n a_n + C^2)/C^2|^{1/2} \end{aligned}$$

Using our definition of C, this becomes:

$$\|u_2 - x_n\| = |(C^2 - \lambda_n^2 a_n^2 + C^2 - 2C\lambda_n a_n + \lambda_n^2 a_n^2)/C^2|^{1/2} = |2 - 2\lambda_n a_n/C|^{1/2} = |2(1 - \lambda_n a_n/C)|^{1/2}$$

Now turning our attention to $u_1 \dots$

$$\|u_1 - x_n\| = |a_1^2 + a_2^2 \dots + (a_n - 1)^2|^{1/2} = |a_1^2 + a_2^2 \dots + a_n^2 - 2a_n + 1|^{1/2}$$

Since u is a unit vector we know: $u^T u = a_1^2 + a_2^2 \dots + a_n^2 = 1$

$$\|u_1 - x_n\| = |1 - 2a_n + 1|^{1/2} = |2 - 2a_n|^{1/2}$$

Thus, we want to show that

$$|2(1 - \lambda_n a_n/C)| < |2 - 2a_n|$$

Or,

$$|1 - \frac{\lambda_n}{C} a_n| < |1 - a_n| \tag{3}$$

$$\begin{aligned} C &= \sqrt{\lambda_1^2 a_1^2 + \lambda_2^2 a_2^2 \dots + \lambda_n^2 a_n^2} < \sqrt{\lambda_n^2 a_1^2 + \lambda_n^2 a_2^2 \dots + \lambda_n^2 a_n^2} \\ &= \sqrt{\lambda_n^2 (a_1^2 + a_2^2 \dots + a_n^2)} = \lambda_n \end{aligned}$$

So we know:

$$\frac{\lambda_n}{C} > 1$$

Also, $|a_n| < 1$ so $1 - a_n \geq 0$ In addition, the left side of (3) is positive because:

$$C^2 = \lambda_1^2 a_1^2 + \lambda_2^2 a_2^2 \dots + \lambda_n^2 a_n^2 > \lambda_n^2 a_n^2$$

C is positive so $C > \lambda_n a_n$ and thus $\frac{\lambda_n}{C} a_n < 1$ so $1 - \frac{\lambda_n}{C} a_n > 0$. So equation (3) reduces to $1 - \frac{\lambda_n}{C} a_n < 1 - a_n$ and so $a_n < \lambda_n a_n / C$. This is obviously true for positive a_n . For negative a_n each iteration would be further from x_n , but this is exactly what we want because we divided out the factor $\lambda_n^i a_n / C_i$ to get to equation (2), so we know for negative a_n it will converge to $-x_n$. In addition, from equation (3) and our expression for C , we can see that the convergence rate will depend on the separation of λ_n from the second largest magnitude eigenvalue. In fact, the asymptotic rate of convergence is the rate that $|\lambda_{n-1} / \lambda_n|^k \rightarrow 0$ where λ_{n-1} is the second largest magnitude eigenvalue [7]. In general, this will converge quickly because it is often the case that there is a large separation between these two eigenvalues [11].

2.1 Search Engines

A typical application of the concept of centrality is the ranking of web pages by search engines. One of the most prominent (and lucrative) methods, Google PageRank, is a direct application of eigenvector centrality, computed via the Power Method.

2.1.1 Google PageRank

It has since become more complicated, but the original equation (given by [7]) for the ‘Google matrix’ is:

$$G = \alpha S + (1 - \alpha)(1/n)E$$

The nodes of the graph represented by G correspond to web pages and the links between them are edges. S_{ij} is the probability that a person randomly clicking on links will go from page i to page j . The ingenious part of PageRank which allowed it to be so successful is the addition of the second term. E is a matrix of all ones, and it is overlaid on top of S to represent the possibility that someone moves to another webpage that is not linked to from the one they are on. The probability that this happens is designated by $1 - \alpha$. In the context of the graph, this is adding lightly weighted edges between every node, thus making it a complete graph. While G is complete, the product Gx is still computed efficiently using the sparse matrix S . Also,

since G is strictly positive, its largest eigenvalue will be positive and simple, and have a strictly positive eigenvector (proved in section **3**), whereas this is not necessarily true for S . In addition, the Power Method applied to G will converge at the same rate that $\alpha^k \rightarrow 0$ (see [7]).

2.1.2 HITS

Though Google PageRank receives the most acclaim, there are many other approaches to Internet search algorithms, one unique example being the HITS algorithm developed by Jon Kleinberg in 1998. This algorithm is different from many other popular ones in that it assigns each node two different centrality scores: a hub score and an authority score. Hubs are nodes in a directed graph from which one can quickly travel to many other nodes. Authorities are nodes to which you can travel quickly from many other nodes. Although related, these scores are more complicated than just in-degree and out-degree.

Each node is given an initial authority score and hub score, in vectors x and y respectively, and the final scores are computed with this iterative algorithm:

$$\begin{aligned}x_i^{n+1} &= \sum_{j \in IN_i} y_j \\y_i^{n+1} &= \sum_{j \in OUT_i} x_j\end{aligned}$$

So the hub score for a node is the sum of the authority scores of the nodes it points to, and the authority score is the sum of the hub scores of the nodes pointing to it.

We can rewrite this iteration using the adjacency matrix A , and starting with just the authority vector y of all ones:

$$\begin{aligned}\text{While } x^n &\neq x^{n-1} \\x^n &= A^T y^{n-1}; \\y^n &= Ax^n; \\n &= n + 1; \\&\text{normalize } x \text{ and } y;\end{aligned}$$

There is a 1 in the adjacency matrix in the i th row and j th column if node i points to node j . So Ax produces y where each y_i is the sum of the

values of x that node i points to. Also, in A^T there is a 1 in the i th row and j th column if node j points to node i .

We can write out one iteration of the previous equations, so that now they are in this form:

$$\begin{aligned}x^n &= A^T A x^{n-1} \\ y^n &= A A^T y^{n-1}\end{aligned}$$

With the normalization of each vector after each iteration, we see that this is just the power method for each of the two matrices $A^T A$ and $A A^T$. So now we know that the HITS algorithm will actually always converge, since we showed earlier that the power method converges to the dominant eigenvalue, eigenvector pair. The the authority and hub scores are represented by the entries of the dominant eigenvectors of $A^T A$ and $A A^T$.

A convenient property of HITS is that while the adjacency matrix A for the directed graph is not symmetric, $A^T A$ and $A A^T$ both are symmetric. (obviously $(A^T A)^T = A^T A$). As we have seen so far in general symmetric matrices are easier to deal with, particularly because their eigenvectors are orthogonal.

3 Perron-Frobenius Theory

The purpose of this section will be to show that nonnegative irreducible matrices have a largest magnitude eigenvalue that is positive and simple, corresponding to a strictly positive eigenvector (meaning one that has all positive components) and this is the only nonnegative eigenvector. This is a particular result of the more general Perron-Frobenius theorem (see [8] and [1]).

3.1 A Proof for Positive Matrices

Before examining nonnegative matrices, we start with the simpler case of matrices with all strictly positive entries, denoted > 0 .

Lemma: Given $n \times n$ matrix $A > 0$, any nonnegative eigenvector of A must be strictly positive.

If x is a n dimensional vector with nonnegative entries then since A is strictly positive, as long as one entry of x is nonzero (it's not the zero vector), then Ax must be strictly positive. So if x has a 0 entry, then Ax cannot be a

multiple of x . Thus, for a nonnegative vector to be an eigenvector of A , it must be strictly positive.

Now we want to show that such a vector exists, and has a positive eigenvalue. Let S be the set of all n dimensional unit vectors strictly greater than 0. Consider the function $f : S \Rightarrow S : f(x) = \frac{Ax}{\|Ax\|}$

First, we must verify that the image of f is in S . Clearly $\frac{Ax}{\|Ax\|}$ is of length one. The i th element of Ax is $\sum_j A_{ij}x_j$, and since all of the elements of x and A are positive, this is strictly greater than 0. So f is a function from S to S .

Here, we could simply claim that since f is continuous from S to S , then by the Brouwer fixed point theorem, f must have a fixed point. However, we have not necessarily shown that f is continuous, nor do we wish to prove the general Brouwer fixed point theorem. Instead, since f is just the power method, we know that it converges to some x , and we can restrict ourselves to only symmetric A . Thus, from section 1, a single iteration can be no further from x , and so the only possible output of $f(x)$ is x .

$$\frac{Ax}{\|Ax\|} = x$$

So this x is a strictly positive eigenvector of A and it has eigenvalue $\|Ax\|$ which is strictly positive since $Ax \neq 0$.

Thus a positive eigenvalue exists for A , call it λ_n . Also, since this was generated using the power method, we know it is the spectral radius. Now we wish to show that this eigenvector is simple.

Assume λ_n has another eigenvector $v \neq 0$.

$$Av = \lambda_n v$$

We want to show that this v is a scalar multiple of x , or $v = xa$ for some $a \in \mathbb{R}$. Assume not, so x and v are linearly independent.

Lemma: $\exists s, t \in \mathbb{R}$ such that $sx + tv \geq 0$ with at least one 0 coordinate.

From section 1 we know any linear combination of x and v will also be an eigenvector for λ_n . Visually, this plane must intersect somewhere with the boundaries of the first quadrant.

So $sx + tv$ is a nonnegative eigenvector for λ_n with at least one 0 coordinate, however this is impossible because the eigenvectors for λ_n must be strictly positive. Thus, x is simple.

Now since we assumed A is symmetric it is easy to show that x is the only positive eigenvector. If $y > 0$ then $y^T x > 0$ but since A is symmetric all of its eigenvectors for different eigenvalues are orthogonal, and x is simple.

3.2 Extension to Irreducible Nonnegative Matrices

We now aim to make the same conclusions from **3.1** about nonnegative irreducible matrices.

Definition: A is a *reducible matrix* if there exists a permutation matrix P such that: $PAP^T = \begin{bmatrix} X & 0 \\ 0 & Z \end{bmatrix}$

Definition: A is *irreducible* if it is not reducible.

First we notice that a graph is connected if and only if its adjacency matrix is irreducible. If a graph is disconnected, we can separate it into at least two separate connected components (denoted C_i). If we create the adjacency matrix by putting all the nodes of each component in sequence, it will look

like this: $\begin{bmatrix} C_1 & 0 \\ 0 & C_2 \\ & & \ddots \end{bmatrix}$

Where for $i \neq j$ we know no nodes in C_i will have a connection to nodes in C_j so all of those spaces will be 0's. So if G is disconnected, we can just choose P to be a rearrangement of I which would rearrange the rows of A in that form. If the graph were connected, no matter what order the nodes are in, there would always be a 1 in a box between two components because all components are connected by at least one edge.

This leads us to another definition of reducible. If the graph is disconnected, there is at least one component that is not connected to the rest of the graph. Let S be the set of integers corresponding to these nodes in this component. $S \subset \{1, 2, \dots, n\}$ such that $(i \in S, j \notin S) \Rightarrow a_{ij} = 0$ This is just the statement that this component is not connected to the rest of the graph, and we know this is equivalent to A being reducible.

Now suppose we are given a nonnegative irreducible matrix A . We want to create a new matrix from A that is positive and has the same spectra so that we can use our proof from **3.1**.

Lemma: For nonnegative vector y , where $y_i = 0$ for some i , $(I + A)y$ has strictly less 0's than y .

Since $(I + A)$ and y are both nonnegative then it would be impossible for $(I + A)y$ to have more 0's than y since any positive component of y will also be positive in $(I + A)y$ due to the I .

Now we consider the case of $(I + A)y$ and y having equal number of 0's. This could only be possible if the same components of $(I + A)y$ were 0 as the ones that were 0 in y . We assume this is the case and find a contradiction:

$$y_j = 0 \Rightarrow y_j + (Ay)_j = 0 \text{ and so } (Ay)_j = 0$$

Let T be the set of the indices of the positive components of y : $T = \{t : y_t > 0\}$

$$\text{So for } j \in T, (Ay)_j = \sum_k a_{jk}y_k = 0 \text{ but for } k \notin T, y_k > 0$$

Thus, for this sum to be 0, every a_{jk} must be zero for $j \in T$ and $k \notin T$. This is exactly the definition we gave earlier for reducibility. Thus we have a contradiction.

Therefore, $(I + A)y$ has strictly less 0's than y . By the same logic, $(I + A)(I + A)y$ has strictly less than $(I + A)y$.

Note that y is not the zero vector so it has at most $n - 1$ zeros. Thus, $(I + A)^{n-1}y$ will have all positive components. All that we required of y was that it was nonnegative, so we could choose y to be $\begin{bmatrix} 1 & 0 & 0 & \dots \end{bmatrix}^T$ and then for $(I + A)^{n-1}y$ to be strictly positive, the first row of $(I + A)^{n-1}$ must be strictly positive. And so we can choose y accordingly and conclude that $(I + A)^{n-1} > 0$.

We could also think of this from a graph theoretic perspective. Consider a connected graph G , with adjacency matrix A whose nodes represent people in a town each with a certain disease. After each iteration, if person i is connected to person j they will contract their diseases. If a disease starts with one person, since this is a connected graph, after $n - 1$ iterations, everyone will have that disease because $n - 1$ is the maximum path length between any two nodes. Consider the matrix D where $D_{ij} > 0$ if person i has disease j . If D starts as the identity, this corresponds to everyone having their own unique disease. Multiplying by $(I + A)$ corresponds to the exact iterative process described; we need the I because everyone retains their own disease at each iteration. Thus, $I(I + A)^{n-1}$ must result in D being strictly positive.

Now we have already shown that positive matrices have a simple positive eigenvalue of maximal magnitude with a corresponding positive eigenvector.

So this is true for $(I + A)^{n-1}$ but we now want to be able to deduce the same information about A .

Consider the positive dominant eigenvalue λ_n of $(I + A)^{n-1}$ with positive eigenvector v .

$$v\lambda_n = (I + A)^{n-1}v$$

Since all of the entries of $(I + A)$ are either 0 or greater than or equal to 1, then all of the entries of $(I + A)^{n-1}$ are greater than 1. Since v is strictly positive, we know that the i th element of $(I + A)^{n-1}v$ must be greater than the i th element of v . Consequently, we see that $\lambda_n > 1$.

Now, we also know that $I + A$ has the same eigenvectors as $(I + A)^{n-1}$ corresponding to eigenvalues that are the $n - 1$ root of the eigenvalues of $(I + A)^{n-1}$:

$$\begin{aligned} (I + A)x &= \lambda x \\ (I + A)^2x &= \lambda(I + A)x = \lambda^2x \\ (I + A)^ix &= \lambda^ix \end{aligned}$$

Thus $\lambda_n^{\frac{1}{n-1}}$ is an eigenvalue of $I + A$ with eigenvector v . And, since $\lambda_n > 1$, then $\lambda_n^{\frac{1}{n-1}} > 1$.

Since $I + A$ and $(I + A)^{n-1}$ have the same set of eigenvectors, we know that v is also a simple eigenvector for $I + A$.

So $I + A$ has a simple, strictly positive eigenvector v with eigenvalue $\lambda_n^{\frac{1}{n-1}} \geq 1$.

$I + A$ has the same eigenvectors as A , corresponding to the eigenvalues of A plus 1:

$$Ax = \lambda x \Rightarrow (I + A)x = x + \lambda x = (1 + \lambda)x$$

So $\lambda_n^{\frac{1}{n-1}} - 1$ is an eigenvalue of A with eigenvector v . $f(\lambda) = \lambda^{\frac{1}{n-1}} - 1$ is a monotone decreasing function so we can conclude that the eigenvalue $\lambda_n^{\frac{1}{n-1}} - 1$ for A is the maximum eigenvalue. Since A is nonnegative consisting of only 0's and 1's, and v is strictly positive, then the eigenvalue for v must be greater than 1. This is because A has at least one 1 in each row and in each column, so if $\lambda < 1$ then for each element of the eigenvector v_i , there must be some v_j in the eigenvector less than it. If we follow this chain for more than n steps, we must encounter a cycle, which is impossible. Thus $\lambda_n^{\frac{1}{n-1}} - 1 > 1$. Since $f(\lambda) = \lambda^{\frac{1}{n-1}} - 1$ must always be greater than -1 for

the eigenvalues of $(I + A)^{n-1}$, we see that this is also the greatest magnitude eigenvector. For another proof, using matrix games, see [1].

So finally, we can conclude that v is a strictly positive eigenvector of A corresponding to simple positive, greatest magnitude eigenvalue $\lambda_n^{\frac{1}{n-1}} - 1$.

4 First Order Approximations of Dynamical Importance

Armed with the results of Perron's Theorem we now proceed with an investigation of Dynamical Importance. Our goal is to give an approximation to the change to the dominant eigenvalue of the adjacency matrix when a node is removed. When removing a single node from a large network, the adjacency matrix of the graph will only change slightly. Consequently, it is reasonable to use a perturbative approach. We begin with a general approach adapted from [5].

If we remove a node, the adjacency matrix changes to: $A_{new} = A + \Delta A$. Here ΔA for removing node i is given by:

$$\Delta A = \begin{bmatrix} & & \vdots & & \\ & & & & \\ \dots & & -A_{ij} & \dots & \\ & & \vdots & & \end{bmatrix}$$

It is important to note here that $A + \Delta A$ will have row and column i all 0. We could simply remove this row and column entirely from the adjacency matrix and study the new smaller one. This would have the same set of eigenvalues with the multiplicity of 0 increased by 1, since e_i is an eigenvector of $A + \Delta A$ with 0 eigenvalue (e_i is a vector of all 0's with a 1 in the i th row). However, for the purpose of this paper, we choose to keep this row and column so that the remaining indices of the nodes do not change.

$Au = \lambda u$ for some eigenvector u associated with eigenvalue λ . Now we are interested in the largest eigenvalue of A_{new} . Since ΔA only changes a few entries of A which is a huge matrix, the change to the eigenvalue will be small and we can express the new eigenvalue and eigenvector thusly:

$\lambda_{new} = \lambda + \epsilon\lambda_1 + \epsilon^2\lambda_2 + \epsilon^3\lambda_3 + \dots$ where ϵ is small and we don't know λ_1 yet, and similarly: $u_{new} = u + \epsilon x_1 + \epsilon^2 x_2 + \epsilon^3 x_3 + \dots$

Plugging this into the eigenvalue equation yields:

$$A_{new}u_{new} = u_{new}\lambda_{new}$$

$$(A + \Delta A)(u + \epsilon x_1 + \epsilon^2 x_2 + \dots) = (\lambda + \epsilon \lambda_1 + \epsilon^2 \lambda_2 + \dots)(u + \epsilon x_1 + \epsilon^2 x_2 + \dots) \quad (4)$$

Now we can multiply this out, and since we are interested in a first order estimation, we will ignore the terms with powers of ϵ greater than 1, since these will be very small.

$$Au + \Delta Au + A\epsilon x_1 + \Delta A\epsilon x_1 = \lambda u + \epsilon \lambda_1 u + \lambda \epsilon x_1$$

Now we can cancel Au and λu . Also we can assume that $\Delta A\epsilon x_1$ is very small since ϵ is small and ΔA is extremely sparse (we will improve on this generalization later). However, we will still have x_1 terms which we do not know. Since A is symmetric, u^T is a left eigenvector for λ (from **1.3**) so we can multiply both sides by u^T yielding:

$$u^T \Delta Au + u^T A\epsilon x_1 = u^T \epsilon \lambda_1 u + u^T \lambda \epsilon x_1$$

We know $u^T A = u^T \lambda$ so we are left with:

$$u^T \Delta Au = u^T \epsilon \lambda_1 u$$

$$\frac{u^T \Delta Au}{\|u\|^2} = \epsilon \lambda_1 \quad (5)$$

This is a general linear algebra result for any small ΔA but for the specific case of removing a node, we can simplify this, starting with the fact that $\|u\|^2 = 1$

$$u^T \Delta Au = \begin{bmatrix} \dots & u^T & \dots \end{bmatrix} \begin{bmatrix} \vdots \\ \dots & -A_{ij} & \dots \\ \vdots \end{bmatrix} \begin{bmatrix} \vdots \\ u \\ \vdots \end{bmatrix}$$

Just concerning the first multiplication on the right, the only nonzero elements of the product will be in the i th row and also all rows the corresponding to indices of nodes adjacent to i . So we see that for removing node i ,

$$[\Delta Au]_j = \begin{cases} -u_i & \text{if } j \in N_i \\ -\lambda u_i & \text{if } j = i \end{cases}$$

$$\text{So } u^T \Delta Au = -\lambda u_i u_i + \sum_{j \in N_i} u_j (-u_i) = -u_i (\lambda u_i + \sum_{j \in N_i} u_j)$$

The sum of the elements of u corresponding to adjacent nodes to i is the same as the i th row of A times u , which is just λu_i . Thus, we have:

$$u^T \Delta Au = -u_i (\lambda u_i + \lambda u_i) = -2\lambda u_i^2 \quad (6)$$

4.1 Improved First Order Estimation

So we have simplified our previous estimate to no longer involve any matrix multiplication, but we can actually improve on it because that estimation was based only on the assumption that we only slightly change A . This approximation can be made more accurate by being more explicit about how exactly we are changing A . To get to equation (5) we used the fact that ΔA is sparse to conclude that $\Delta A \epsilon x_1$ was very small. By using the definition of ΔA we can improve on this and actually make this term precisely 0.

We know that $A + \Delta A$ has row $i = 0$ and column $i = 0$ since we are removing node i . The product $A_{new} u_{new}$ will therefore have a 0 as its i th element. In other words, we know $\lambda_{new} u_{new_i} = 0$

λ_{new} is positive (its the largest eigenvalue of A_{new}) so $u_{new_i} = 0$.

So now we can redefine u_{new} with this in mind:

$u_{new} = u - u_i e_i + \epsilon y_1 + \epsilon^2 y_2 + \epsilon^3 y_3 + \dots$ where the i th element of y_1, y_2, \dots is 0. This new definition of u_{new} is a strategy originally proposed in [11]. So the new version of equation (4) is:

$$(A + \Delta A)(u - u_i e_i + \epsilon y_1 + \epsilon^2 y_2 + \dots) = (\lambda + \epsilon \lambda_1 + \epsilon^2 \lambda_2 + \dots)(u - u_i e_i + \epsilon y_1 + \epsilon^2 y_2 + \dots)$$

Throwing out terms with powers of ϵ greater than or equal to 2 now gives us:

$$Au + \Delta Au + A\epsilon y_1 + \Delta A\epsilon y_1 - (A + \Delta A)u_i e_i = \lambda u + \epsilon \lambda_1 u + \lambda \epsilon y_1 - \lambda u_i e_i - \epsilon \lambda_1 u_i e_i$$

And, since $A + \Delta A$ has column i all 0's, we know $(A + \Delta A)u_i e_i = 0$. Also, ΔA only has nonzero entries in the i th row and i th column, and $\Delta A_{ii} = 0$. Consequently, we know $\Delta A\epsilon y_1 = 0$ because $y_{1_i} = 0$. As before, we cancel Au and λu and left multiply by u^T . So now we have:

$$u^T \Delta Au + u^T A\epsilon y_1 = u^T \epsilon \lambda_1 u + u^T \lambda \epsilon y_1 - u^T \lambda u_i e_i - u^T \epsilon \lambda_1 u_i e_i$$

$$\begin{aligned}
u^T \Delta A u &= u^T \epsilon \lambda_1 u - u^T \lambda u_i e_i - u^T \epsilon \lambda_1 u_i e_i \\
\frac{u^T \Delta A u + u_i^2 \lambda}{\|u\|^2 - u_i^2} &= \epsilon \lambda_1
\end{aligned}$$

We can simplify this further by plugging in equation (6) to get:

$$= \epsilon \lambda_1 = \frac{-2\lambda u_i^2 + u_i^2 \lambda}{\|u\|^2 - u_i^2} = \frac{-\lambda u_i^2}{1 - u_i^2} \quad (7)$$

5 Application to Graph Laplacian

The Laplacian matrix L for a graph is another important tool for analyzing a network:

Definition: $L = D_A - A$ is the *Laplacian* of graph G where A is the Adjacency Matrix and D_A is the matrix with the sum of the rows of A (the degrees of the nodes) along the diagonal.

In this section we reproduce a result from [9], approximating the change to eigenvalues of the Laplacian matrix under perturbations to the graph. From our earlier result, we know that if we change L slightly, by adding or subtracting a small subgraph, we can approximate the change to any eigenvalue λ by using the associated eigenvector u :

$$\frac{u^T \Delta L u}{\|u\|^2} \simeq \Delta \lambda_L$$

We have a new adjacency matrix given by $A + \Delta A$, and the Laplacian matrix according to this is our new Laplacian; and we can form a ΔL where $L_{new} = L + \Delta L$.

$$\Delta L = D_{\Delta A} - \Delta A$$

So we have

$$\frac{u^T \Delta L u}{\|u\|^2} = \frac{u^T (D_{\Delta A} - \Delta A) u}{\|u\|^2}$$

Once again, we can choose u to be a unit vector.

$$\Delta \lambda_L \simeq u^T D_{\Delta A} u - u^T \Delta A u$$

$$\begin{aligned}
u^T D_{\Delta A} u &= \begin{bmatrix} \dots & u^T & \dots \end{bmatrix} \begin{bmatrix} \ddots & & \\ & \sum_j \Delta A_{ij} & \\ & & \ddots \end{bmatrix} \begin{bmatrix} u \\ \vdots \\ u \end{bmatrix} \\
&= \begin{bmatrix} \dots & u^T & \dots \end{bmatrix} \begin{bmatrix} \vdots \\ u_i \sum_j \Delta A_{ij} \\ \vdots \end{bmatrix} = \sum_i \left(u_i^2 \sum_j \Delta A_{ij} \right) \\
u^T \Delta A u &= \sum_i u_i \left(\sum_j \Delta A_{ij} u_j \right)
\end{aligned}$$

Thus

$$\begin{aligned}
\Delta \lambda_L &\simeq \sum_i \left(u_i^2 \sum_j \Delta A_{ij} \right) - \sum_i u_i \left(\sum_j \Delta A_{ij} u_j \right) \\
&\simeq \sum_i \left(u_i^2 \sum_j \Delta A_{ij} - u_i \sum_j \Delta A_{ij} u_j \right) \\
&\simeq \sum_i u_i \left(\sum_j \Delta A_{ij} u_i - \sum_j \Delta A_{ij} u_j \right) \\
&\simeq \sum_i u_i \left(\sum_j \Delta A_{ij} (u_i - u_j) \right) \\
&\simeq \sum_{ij} \Delta A_{ij} u_i (u_i - u_j) \tag{8}
\end{aligned}$$

Since ΔA is symmetric we know

$$\begin{aligned}
\sum_{ij} \Delta A_{ij} u_i (u_i - u_j) &= 2 \sum_{i < j} \Delta A_{ij} u_i (u_i - u_j) \\
&= \sum_{i < j} \Delta A_{ij} u_i^2 - 2 \sum_{i < j} \Delta A_{ij} u_j u_i + \sum_{i < j} \Delta A_{ij} u_i^2
\end{aligned}$$

Since clearly $\sum_{i < j} \Delta A_{ij} u_i^2 = \sum_{i < j} \Delta A_{ij} u_j^2$ we can simplify this to:

$$\Delta \lambda_L \simeq \sum_{i < j} \Delta A_{ij} (u_i - u_j)^2$$

This is precisely the formula given in [9]. Similar to section 4.1 the i th component of the new eigenvector for the Laplacian will also be 0 as it was for the Adjacency matrix, and we can use this idea to make our approximation more accurate. An approach using this concept is given in [16].

6 Change in the Eigenvector

Google's PageRank relies partly on the fact that a small change to the graph will not greatly affect the dominant eigenvector. Otherwise, they would have to update their matrix, whose size is on the order of billions, every single time someone added a new web page. In this section we use an approach from [13] to place a bound on how much the eigenvector can change based on the graph spectrum and the perturbation.

Definition: A matrix norm $\|\cdot\|$ is consistent if for all A, B $\|AB\| \leq \|A\|\|B\|$

Lemma: For any consistent matrix norm $\|\cdot\|$, $\rho(A) \leq \|A\|$ where $\rho(A)$ is the spectral radius of A .

Let $\|\cdot\|_z$ be consistent on $\mathbb{R}^{n \times n}$ and let $v(x)$ be a vector norm $\|\cdot\|$ on $\mathbb{R}^{n \times 1}$ defined by $v(x) = \|xy^T\|_z$ for some nonzero $y \in \mathbb{R}^{n \times 1}$. Let x be any eigenvector of A . We can normalize x according to v so that $v(x) = 1$. So $v(Ax) = \|Axy^T\|_z$ and $\|Axy^T\|_z \leq \|A\|_z \|xy^T\|_z$ since $\|\cdot\|_z$ is consistent. $\|A\|_z \|xy^T\|_z = \|A\|_z v(x) = \|A\|_z$. Thus $v(Ax) \leq \|A\|_z$, but $v(Ax) = v(\lambda x) = \lambda v(x) = \lambda$ so $\lambda \leq \|A\|_z$

Since x was an arbitrary eigenvector, we can conclude that for all eigenvalues of A this holds.

Let x and λ be the dominant eigenpair of A , and let $A_{new} = A + E$ where E is a small perturbation matrix. Consider the dominant eigenpair of A_{new} : $(A + E)\tilde{x} = \tilde{\lambda}\tilde{x}$.

Let Y be a matrix with columns according to vectors forming an orthonormal basis for $R(x)^\perp$ (This approach is from [13]). A property of the 2-norm for matrices is that if U is orthonormal then $\|AU\|_2 = \|A\|_2$.

We can normalize \tilde{x} in $\|\cdot\|_2$. Thus,

$$1 = \|\tilde{x}\|_2 = \left\| \begin{bmatrix} - & x^T & - \\ & [Y^T] & \end{bmatrix} \tilde{x} \right\|_2$$

This is a vector, so the 2-norm for $\mathbb{R}^{n \times 1}$ is just the usual Euclidean distance, the square root of the sum of the squares of the components.

$$1 = \left\| \begin{bmatrix} - & x^T & - \\ & [Y^T] & \end{bmatrix} \tilde{x} \right\|_2 = \sqrt{(x^T \tilde{x})^2 + \|Y^T \tilde{x}\|_2^2}$$

$$1 = (x^T \tilde{x})^2 + \|Y^T \tilde{x}\|_2^2$$

We know $\cos \angle(x, \tilde{x}) = \frac{x^T \tilde{x}}{\|x\| \|\tilde{x}\|}$ but in this case $\|x\| = \|\tilde{x}\| = 1$

So now we have

$$1 = \cos^2 \angle(x, \tilde{x}) + \|Y^T \tilde{x}\|_2^2$$

$$\sin^2 \angle(x, \tilde{x}) = \|Y^T \tilde{x}\|_2^2$$

Now we attempt to find a bound for this angle. Notice $(A + E)\tilde{x} = \tilde{\lambda}x$

$$\tilde{\lambda}\tilde{x} - A\tilde{x} = E\tilde{x}$$

We can multiply by Y^T to get

$$Y^T \tilde{\lambda}\tilde{x} - Y^T A\tilde{x} = Y^T E\tilde{x}$$

Y was defined as any orthonormal basis for $R(x)^\perp$, so if we are given that A is real and symmetric, then its eigenvectors are all orthogonal, so we could choose all of the eigenvectors of A besides x to form Y . So we have

$$\begin{bmatrix} - & x^T & - \\ & [Y^T] & \end{bmatrix} A \begin{bmatrix} | \\ x \\ | \\ [Y] \end{bmatrix} = \begin{bmatrix} \lambda & 0 & \dots \\ 0 & \lambda_2 & 0 \\ \vdots & 0 & \ddots \end{bmatrix}$$

i.e. A is diagonalizable. We can then use another matrix M to represent this.

$$\begin{bmatrix} - & x^T & - \\ & [Y^T] & \end{bmatrix} A \begin{bmatrix} | \\ x \\ | \\ [Y] \end{bmatrix} = \begin{bmatrix} \lambda & & 0 & \dots \\ 0 & \begin{bmatrix} M \end{bmatrix} & & \\ \vdots & & & \end{bmatrix}$$

Since Y is orthonormal $Y^T Y = I$ so since $Y^T A Y = M$, then $Y^T A Y Y^T = M Y^T$ so $Y^T A = M Y^T$ So returning to our earlier equation, we can substitute in M ,

$$Y^T \tilde{\lambda}\tilde{x} - M Y^T \tilde{x} = Y^T E\tilde{x}$$

$$\begin{aligned}
(\tilde{\lambda}I - M)Y^T \tilde{x} &= Y^T E \tilde{x} \\
Y^T \tilde{x} &= Y^T E \tilde{x} (\tilde{\lambda}I - M)^{-1}
\end{aligned}$$

Since the 2-norm is consistent, Y is orthonormal and \tilde{x} is of length 1 in the 2-norm, we have

$$\|Y^T \tilde{x}\|_2 = \|Y^T E \tilde{x} (\tilde{\lambda}I - M)^{-1}\|_2 \leq \|Y^T\|_2 \|E\|_2 \|\tilde{x}\|_2 \|(\tilde{\lambda}I - M)^{-1}\|_2 = \|E\|_2 \|(\tilde{\lambda}I - M)^{-1}\|_2$$

$$\sin \angle(x, \tilde{x}) \leq \|E\|_2 \|(\tilde{\lambda}I - M)^{-1}\|_2$$

$$\tilde{\lambda}I - M = \begin{bmatrix} \lambda_2 - \lambda & 0 & & \\ 0 & \lambda_3 - \lambda & & \\ & & \ddots & \\ & & & \ddots \end{bmatrix}$$

$$(\tilde{\lambda}I - M)^{-1} = \begin{bmatrix} \frac{1}{\lambda_2 - \lambda} & 0 & & \\ 0 & \frac{1}{\lambda_3 - \lambda} & & \\ & & \ddots & \\ & & & \ddots \end{bmatrix}$$

$\|A\|_2^2$ is the largest eigenvalue of AA^T so

$$\|(\tilde{\lambda}I - M)^{-1}\|_2^2 = \max \left\{ \frac{1}{(\lambda_i - \lambda)^2} \right\}$$

$$\|(\tilde{\lambda}I - M)^{-1}\|_2 = \frac{1}{\lambda - \lambda_{n-1}}$$

So finally,

$$\sin \angle(x, \tilde{x}) \leq \frac{\|E\|_2}{\lambda - \lambda_{n-1}} \quad (9)$$

Here λ_{n-1} is the second largest eigenvalue, and as stated earlier the separation between these values is often large [11].

Its important to realize here that the perturbation matrix E need not be symmetric for this to hold, but it was not necessary for the original matrix to be symmetric. We can show that this constant does not always hold when the original matrix is not symmetric with a simple counter example:

$$A = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & -1 & -1 \\ 0 & 0 & 1 & -1 & -1 & -1 \\ 0 & 0 & 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Which has the dominant eigenvector: $\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}^T$

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

A and $A+E$ still have the same eigenvalues, but the dominant eigenvector is now: $\begin{bmatrix} .9901 & -.099 & .099 & 0 & 0 & 0 & 0 \end{bmatrix}^T$

The sine of the angle between these two vectors is .1404, but our proof from earlier worked for consistent norms and $\|\cdot\|_\infty$ is consistent, yet the inequality fails because A was not symmetric:

$$\frac{\|E\|_\infty}{\lambda - \lambda_{n-1}} = \frac{.1}{2 - 1} = .1 < .1404$$

7 A More Accurate Approximation of Dynamical Importance

In order to find a more accurate approximation a natural approach would be to revisit equation (4) from section 4, but consider also consider terms with ϵ^2 terms.

$$Au + \Delta Au + A\epsilon x_1 + \Delta A\epsilon x_1 + A\epsilon^2 x_2 + \Delta A\epsilon^2 x_2 = \lambda u + \epsilon^2 \lambda_2 u + \epsilon^2 \lambda_1 x_1 + \epsilon^2 \lambda x_2 + \epsilon \lambda_1 u + \lambda \epsilon x_1$$

The problem that we will encounter here is that we will not be able to cancel out terms containing ϵx_1 . This is a possible approach but it would require knowing all of the eigenvectors of A (see [5]), which is not something that we want to assume. In 2010, Milanese et al. [9] proposes a way around this difficulty using the Power Method to approximate the change in u . In this paper we utilize this idea in a slightly different way. We abandon the series expansion approach of equation (4) for the simpler looking:

$$(A + \Delta A)(u + \Delta u) = (u + \Delta u)(\lambda + \Delta \lambda)$$

$$\Delta Au + A\Delta u + \Delta A\Delta u = \Delta\lambda u + \lambda\Delta u + \Delta\lambda\Delta u$$

We again multiply by u^T and solve for $\Delta\lambda$

$$u^T\Delta Au + u^T\Delta A\Delta u = u^T\Delta\lambda u + u^T\Delta\lambda\Delta u$$

$$\Delta\lambda = \frac{u^T\Delta Au + u^T\Delta A\Delta u}{1 + u^T\Delta u} \quad (10)$$

In the previous section we placed a bound on how much the dominant eigenvector can change based on the two largest eigenvalues and the perturbation matrix. So in general if the perturbation matrix is small, and the difference between the two largest eigenvalues is large the change to the eigenvector will be small, in which case it is reasonable to approximate it with one iteration of the power method (as proposed by [5]):

$$\begin{aligned} u + \Delta u &\simeq \frac{(A + \Delta A)u}{\|(A + \Delta A)u\|} = \frac{\lambda u + \Delta Au}{\|\lambda u + \Delta Au\|} \\ &= \frac{\lambda u + \Delta Au}{\sqrt{(\lambda u + \Delta Au)^T(\lambda u + \Delta Au)}} \end{aligned} \quad (11)$$

$$= \frac{\lambda u + \Delta Au}{\sqrt{\lambda^2\|u\|^2 + 2\lambda u^T\Delta Au + u^T\Delta A\Delta Au}} \quad (12)$$

Now since the u is a unit vector, and ΔA is very small (either very sparse, or has very small entries) we could claim the denominator will be very close to λ . We could approximate it as such, but we are more ambitious and retain these terms for now for a more accurate approximation. We represent the difference in the denominator from λ caused by these terms by some scalar x .

$$u + \Delta u \simeq \frac{\lambda u + \Delta Au}{\sqrt{\lambda^2\|u\|^2 + 2\lambda u^T\Delta Au + u^T\Delta A\Delta Au}} = \frac{\lambda u + \Delta Au}{\lambda + x}$$

$$\frac{\lambda u + \Delta Au}{\lambda + x} = \frac{\lambda u + \Delta Au + ux - ux}{\lambda + x} = \frac{u(\lambda + x) + \Delta Au - ux}{\lambda + x}$$

$$= u + \frac{\Delta Au - ux}{\lambda + x}$$

Thus $\Delta u \simeq \frac{\Delta Au - ux}{\lambda + x}$ and we have retained the terms from the denominator. It may have seemed necessary to forgo some accuracy by neglecting the small terms in the denominator in equation (12) in order to simplify the result, but we keep these x terms and we will see that they will actually all cancel in our final approximation!

Now we can plug in Δu to equation (10):

$$\begin{aligned} \Delta \lambda &= \frac{u^T \Delta Au + u^T \Delta A \frac{\Delta Au - ux}{\lambda + x}}{1 + u^T \frac{\Delta Au - ux}{\lambda + x}} \\ \Delta \lambda &= \frac{u^T \Delta Au + \frac{1}{\lambda + x}(u^T \Delta A \Delta Au - xu^T \Delta Au)}{1 + \frac{1}{\lambda + x}(u^T \Delta Au - x)} \\ \Delta \lambda &= \frac{(\lambda + x)u^T \Delta Au + (u^T \Delta A \Delta Au - xu^T \Delta Au)}{\lambda + x + (u^T \Delta Au - x)} \end{aligned}$$

And so the x 's cancel and we are left with:

$$\Delta \lambda = \frac{\lambda u^T \Delta Au + u^T \Delta A \Delta Au}{\lambda + u^T \Delta Au} \quad (13)$$

So we have found a more accurate approximation for the change in λ for a small perturbation matrix ΔA . With our first order approximation we were also able to make it more accurate for the case of removing a single node by acknowledging that $u_{new_i} = 0$. Here we cannot use this technique to improve our estimate because we have actually already included this information because the i th element of $\lambda u + \Delta Au$ is already 0. This is because the i th row of ΔA is just the negation of the i th row of A , so the i th element of ΔAu will be $-\lambda u_i$.

As before we can simplify this nicely using our knowledge of ΔA . From equation (6), we know $u^T \Delta Au = -2\lambda u_i^2$.

So the only remaining thing that we need to simplify equation (13) is $u^T \Delta A \Delta Au$.

$$u^T \Delta A \Delta Au = \begin{bmatrix} \dots & u^T & \dots \end{bmatrix} \begin{bmatrix} \vdots & & \\ \dots & -A_{ij} & \dots \\ \vdots & & \end{bmatrix} \begin{bmatrix} \vdots & & \\ \dots & -A_{ij} & \dots \\ \vdots & & \end{bmatrix} \begin{bmatrix} \vdots \\ u \\ \vdots \end{bmatrix}$$

We already know that $\Delta Au_j = \begin{cases} -u_i & \text{if } j \in N_i \\ -\lambda u_i & \text{if } j = i \end{cases}$

So to utilize this we could rewrite $u^T \Delta A \Delta Au$ as $(\Delta Au)^T \Delta Au$

$$\begin{aligned} (\Delta Au)^T \Delta Au &= \sum_{j \in N_i} \{-u_i(-u_i)\} + (-\lambda u_i)^2 \\ &= u_i^2 \left(\sum_{j \in N_i} \{1\} + \lambda^2 u_i^2 \right) \end{aligned}$$

$\sum_{j \in N_i} 1$ is just counting the number of adjacent nodes to i . Thus:

$$u^T \Delta A \Delta Au = u_i^2 \times \text{Deg}(i) + \lambda^2 u_i^2 \quad (14)$$

So along with equation (5) we can plug this into our equation for $\Delta \lambda$ to get:

$$\begin{aligned} \Delta \lambda &= \frac{1}{1 + (-2\lambda u_i^2)/\lambda} \left(-2\lambda u_i^2 + \frac{u_i^2 \times \text{Deg}(i) + \lambda^2 u_i^2}{\lambda} \right) \\ \Delta \lambda &= \frac{\lambda u_i^2}{1 + (-2\lambda u_i^2)/\lambda} \left(-2 + \frac{\text{Deg}(i)}{\lambda^2} + 1 \right) \\ \Delta \lambda &= \frac{\lambda u_i^2}{1 - 2u_i^2} \left(\frac{\text{Deg}(i)}{\lambda^2} - 1 \right) \quad (15) \end{aligned}$$

So we have a second order approximation which only relies on the original eigen pair and the degree of the node. This is much more accurate than the previous estimations, but still remains computationally efficient. We compare this to our other two estimations on Erdős-Renyi random graphs (see [2]) generated in MATLAB.

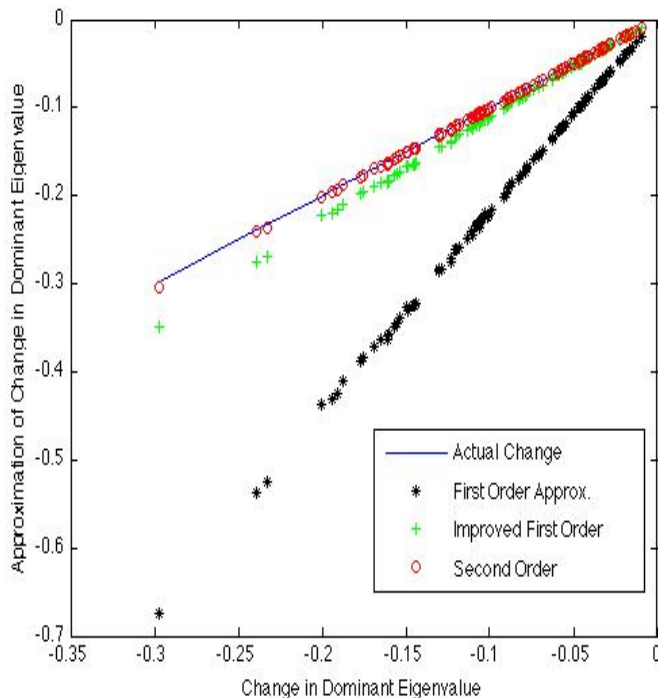


Figure 1. This is the result for an Erdos-Renyi 100 node random graph with probability of .1 for a connection between any two nodes. We removed each node and measured the change in the spectral radius (x-axis), and compared this to our three estimations of it (y-axis). The dotted blue line is the actual changes plotted against themselves. The first order estimation pictured is the one given by equation (6); the improved first order estimation is given by equation (7); and the second-order approximation is the one we proposed in equation (15). Its interesting to note that the error is smaller when the change to the spectral radius is small. This makes sense because our approximations were under the assumption that the perturbation would have a relatively small affect on the graph overall.

8 Analysis and Conclusions

We can compare our second order approximation to the one proposed in [9], which is

$$\Delta\lambda = u^T \Delta Au + u^T \Delta A \Delta Au / \lambda$$

or

$$\Delta\lambda = \lambda u_i^2 \left(\frac{\text{Deg}(i)}{\lambda^2} - 1 \right) \quad (16)$$

Their approximation followed a different approach but also used the power method to approximate the change in u . However, they chose to neglect the small terms in the denominator of (12) but if we include them as a scalar x as before, it does not cancel in their approximation, and it becomes:

$$\Delta\lambda = \frac{\lambda^2 u_i^2}{\lambda + x} \left(\frac{\text{Deg}(i)}{\lambda^2} - 1 \right)$$

Since the x 's cancelled in our approximation, we now examine exactly what this x is to investigate how our approximation could behave differently than the one from [9]. We revisit equation (12) but since we now have equations (6) and (14) we know exactly what those terms are:

$$\begin{aligned} \frac{\lambda u + \Delta Au}{\sqrt{\lambda^2 \|u\|^2 + 2\lambda u^T \Delta Au + u^T \Delta A \Delta Au}} &= \frac{\lambda u + \Delta Au}{\sqrt{\lambda^2 + 2\lambda(-2\lambda u_i^2) + u_i^2 \times \text{Deg}(i) + \lambda^2 u_i^2}} \\ &= \frac{\lambda u + \Delta Au}{\sqrt{\lambda^2 - 4\lambda^2 u_i^2 + u_i^2 \times \text{Deg}(i) + \lambda^2 u_i^2}} = \frac{\lambda u + \Delta Au}{\sqrt{\lambda^2 - u_i^2(\text{Deg}(i) - 3\lambda^2)}} \end{aligned}$$

It is frequently the case that the average degree of a graph scales with the largest eigenvalue. Consider for example the extreme case where every element of the eigenvector is the same. In this case, for u to be an eigenvector, every node must be exactly the same degree. So to approximate the error we replace $\text{Deg}(i)$ with λ , we see that the denominator is $\sqrt{\lambda^2 + u_i^2(3\lambda^2 - \lambda)}$

So as the average degree gets larger (and hence λ does also) we see that $(3\lambda^2 - \lambda)$ becomes large thus making x larger. Consequently, we expect the approximation in (16) to be sensitive to large average degrees in a manner in which ours from equation (15) is not. We similarly verify this in MATLAB.

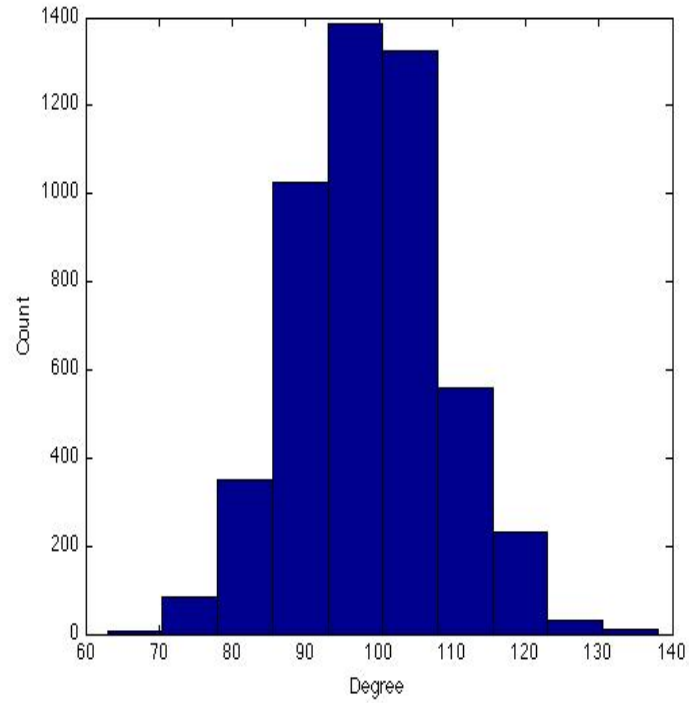


Figure 2. We generated a 5000 node Erdos-Renyi random graph with probability .02 of connections between nodes (and hence an average degree of 100). This is a histogram of the degrees of all of the nodes from the graph. We used this graph for the analysis depicted by Figures 3 and 4.

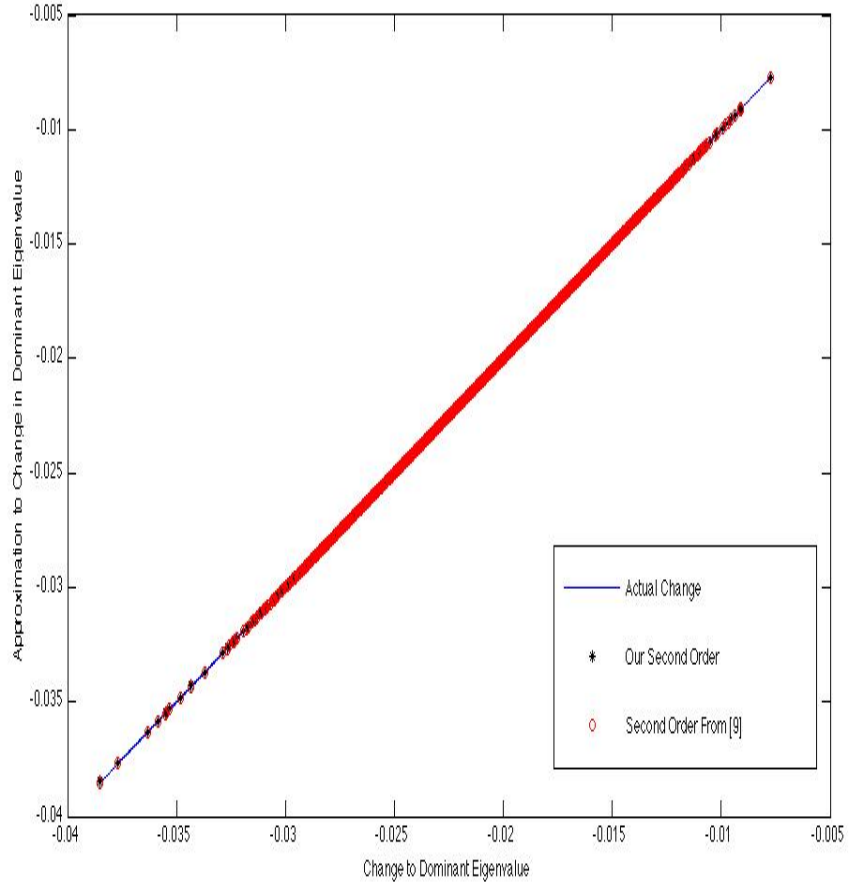


Figure 3. This is a plot of the actual change to the dominant eigenvalue when each node is removed versus the approximations to the change. Our approximation from (15) is the red circles, and the approximation given in (16) is the black stars. From this figure we see that both approximations are very accurate. For such a large network it is difficult to see a difference between them on this scale as we could in Figure 1, so we observe the percent error for each approximation in this example in Figure 4.

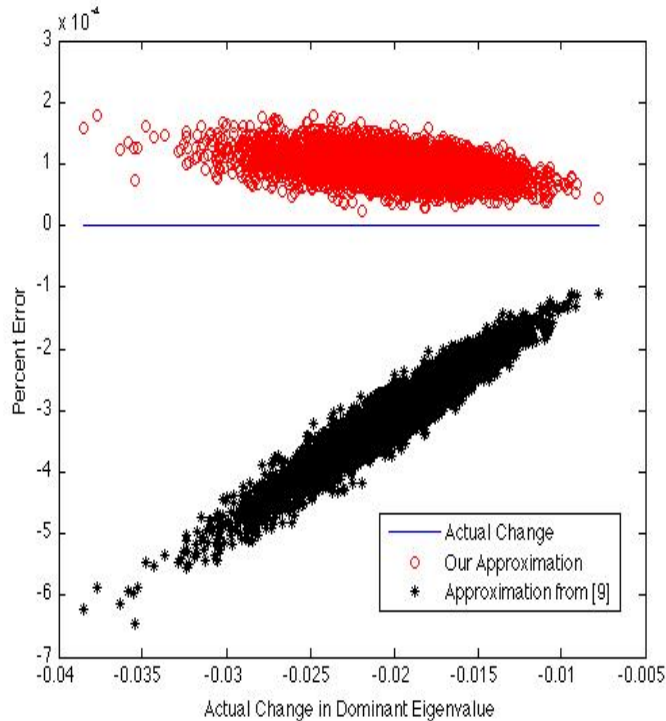


Figure 4. This is the same result as Figure 3, but this shows the percent difference plotted against the actual change. This is a more revealing depiction of the result, and we see that our approximation is more accurate in this case.

We have explored several approximations for dynamical importance, and their mathematical foundations. Our most accurate approximation, equation (15), is an original proposition which we justified in section 7 and represents an improvement over equation (16) (proposed in [9]) for graphs with large average degree. Our results could be applicable to large networks with high average degree, whose dynamics are related to their spectral radius. Our approximation could be used in importance ranking schemes for the nodes in such a network. In addition, these concepts could also be applied in a case where one endeavors to alter the spectral radius of a network to change its dynamical properties by successively removing nodes. A possible future study could be to analyze the performance of our approximation versus the one proposed in [9] for graphs with different distributions of degrees of nodes.

As more and more network data becomes accessible, computations such as these become more relevant for practically discerning meaning from huge amounts of information. In addition, an understanding of the mathematical foundations of network science in graph theory and linear algebra is necessary in order to strive to unravel the constant stream of open problems in network analysis provided by the continuous progression of technology.

Acknowledgments

I would like to acknowledge valued input and advice from Ivan Matic and I would also like to thank Peter Mucha for his guidance and mentorship.

References

- [1] R. Bapat and T. Raghavan. *Nonnegative Matrices and Applications*. (Cambridge University Press, New York, 1997).
- [2] B. Bollobas, *Modern Graph Theory*. (Springer, New York, 1998).
- [3] U. Brandes. *A Faster Algorithm for Betweenness Centrality*. Journal of Mathematical Sociology, 25, 163-177 (2001).
- [4] L. Freeman. *Centrality in Social Networks Conceptual Clarification*. Social Networks 1, 215-239 (1978).
- [5] L. Hinch. *Perturbation Methods*. (Cambridge University Press, New York, 1991).
- [6] E. Kolaczyk. *Statistical Analysis of Network Data : Methods and Models*. (Springer, New York, 2009).
- [7] A. Langville and C. Meyer. *Google's PageRank and Beyond: The Science of Search Engine Rankings*. (Princeton University Press, Princeton, 2006).
- [8] C. MacCluer. *The Many Proofs and Applications of Perrons Theorem*. SIAM Rev. 42, 487-498 (2000).
- [9] A. Milanese, J. Sun and T. Nishikawa. *Approximating Spectral Impact of Structural Perturbations in Large Networks*. Phys. Rev. E 81, 046112 (2010).

- [10] T. Odda. *On Properties of a Well-known Graph or What is Your Ramsey Number?* Annals of the New York Academy of Sciences 328, 166-172 (1979).
- [11] J. Restrepo, E. Ott and B. Hunt. *Characterizing the Dynamical Importance of Network Nodes and Links.* Phys. Rev. Lett. 97, 094102 (2006).
- [12] K. Stephenson and M. Zelen. *Rethinking Centrality: Methods and Examples.* Social Networks 11, 1 - 37 (1989).
- [13] G. Stewart. *Matrix Algorithms.* Vol. 2. (Society for Industrial and Applied Mathematics, Pennsylvania, 1998).
- [14] W. Valente, K. Coronges, C. Lakon and E. Costenbader. *How Correlated Are Network Centrality Measures?* Connect (Tor), Author Manuscript. (2008).
- [15] S. Wasserman and K. Faust. *Social Network Analysis: Methods and Applications* (Cambridge University Press, New York, 1994).
- [16] T. Watanabe and N. Masuda. *Enhancing the Spectral Gap of Networks by Node Removal.* Phys. Rev. E 82, 046102 (2010).