Eigenvector Centrality and Maximal Entropy Random Walks

Larry Fenn

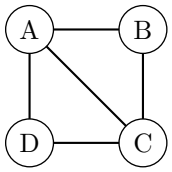
Contents

List of Figures

1 Introduction

Graph theory has established itself through its applicability in modeling physical, social, and information systems. One of the most common questions we might have of a graph is determining a measurement of how influential or "central" vertices in the graph are— for instance, the most connected person in a social network, or the most crucial intersection in a traffic grid. We begin by rigorously defining one of these centrality measures, eigenvector centrality, and then proceed to construct an approximation of eigenvector centrality. Next, we discuss random walks, and in particular maximal entropy random walks. Finally, we produce a local approximation for the maximum entropy random walk process.

2 Preliminaries



$$\begin{bmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

Figure 1: A graph and its adjacency matrix.

Throughout this document we will speak only of non-periodic, simple, finite, connected graphs. The results can be extended to more generalized cases, but they are beyond the scope of this document. For a graph G, and a particular labeling of G, we can construct the *adjacency matrix* $\boldsymbol{A}$ for G by setting $[\boldsymbol{A}]_{ij} = 1$ if there is an edge from vertex v_i to vertex v_j, and 0 otherwise. Since we are considering connected, non-periodic graphs only, every adjacency matrix must be irreducible (since to be reducible is equivalent to being able to distinguish separate connected components).

The adjacency matrix $\boldsymbol{A}$ can be thought of as representing a *linear operator*, which means that the following two operations are true for any two vectors $\vec{\boldsymbol{x}}$, $\vec{\boldsymbol{y}}$ and any scalar α:

- $\boldsymbol{A}\left(\vec{\boldsymbol{x}} + \vec{\boldsymbol{y}}\right) = \boldsymbol{A}\vec{\boldsymbol{x}} + \boldsymbol{A}\vec{\boldsymbol{y}}$
- $\boldsymbol{A}\left(\alpha\vec{\boldsymbol{x}}\right) = \alpha\boldsymbol{A}\vec{\boldsymbol{x}}$

For our purposes the vectors under consideration will just be ordered lists of numbers, each one tied to a certain vertex in the graph. In the same way that we constructed the adjacency matrix, the ith element of a vector will be the number associated to the vertex v_i. These two properties of a linear operator will come in handy many times in calculations involving $\boldsymbol{A}$.

For an $n \times n$ matrix $\boldsymbol{A}$, the scalar value λ and vector $\boldsymbol{\psi}$ that satisfy the equation $\boldsymbol{A\psi} = \lambda\boldsymbol{\psi}$ are the *eigenvalue* and *eigenvector* of $\boldsymbol{A}$. There are n such eigenvalue/eigenvector pairs, although they may not necessarily be distinct. If $\boldsymbol{A}$ is a diagonal matrix, observe that the eigenvalues of $\boldsymbol{A}$ are precisely the entries on the diagonal. Note that for any eigenvector and eigenvalue pair λ and $\boldsymbol{\psi}$, $\boldsymbol{A}^k\boldsymbol{\psi} = \boldsymbol{A}^{k-1}\boldsymbol{A\psi} = \boldsymbol{A}^{k-1}\lambda\boldsymbol{\psi} = \lambda\boldsymbol{A}^{k-1}\boldsymbol{\psi} = ... = \lambda^k\boldsymbol{\psi}$.

3 Spectral Graph Theory

The subject of spectral graph theory encompasses the study of the properties of a graph with relation to the eigenvalues and eigenvectors of matrices associated to that graph. For instance, while the adjacency matrix of a graph can be different depending on how that graph is labeled, it turns out that the spectrum (the collection of eigenvalues and eigenvectors) of its adjacency matrix stays the same.

3.1 Eigenvalues of the Adjacency Matrix

Since we are working only on undirected, simple, finite graphs, it must be the case that the adjacency matrix $\boldsymbol{A}$ for G is a finite, non-negative, real symmetric matrix. Hence, by the spectral theorem on finite symmetric matrices, there exists a real orthogonal matrix $\boldsymbol{Q}$ (orthogonal here meaning $\boldsymbol{Q}^T = \boldsymbol{Q}^{-1}$) such that $\boldsymbol{D} = \boldsymbol{Q}^T\boldsymbol{AQ} = \boldsymbol{Q}^{-1}\boldsymbol{AQ}$ is a diagonal matrix of real values. As a result, the eigenvalues of $\boldsymbol{A}$ are all real, and they are in fact the entries of $\boldsymbol{D}$. For future reference, we will label the eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$, where λ_n is the largest eigenvalue. Moreover, for a diagonal matrix we know that the eigenvectors form what is called an *eigenbasis*, which means that any vector can be rewritten as a linear combination of eigenvectors.

If two graphs are isomorphic then they must have the same eigenvalues. This statement can be turned around to provide us with a useful test to see if two graphs are not isomorphic: if two graphs do not have the same eigenvalues, then they must not be isomorphic. However, note that the converse is not necessarily true: two non-isomorphic graphs can have the same eigenvalues.[1]

(a) Eigenvalues: {-2,0,0,0,2}

(b) Eigenvalues: {-2,0,0,0,2}

Figure 2: Two non-isomorphic graphs with the same eigenvalues.

3.1.1 Perron-Frobenius Theorem

We know that eigenvalues of a graph must be real valued. In fact, there is a stronger result due to a special case of the **Perron-Frobenius Theorem**:

Theorem 3.1. *(Perron-Frobenius Theorem, non-negative irreducible case) If $\boldsymbol{A}$ is an irreducible, non-negative $n \times n$ matrix with spectral radius $\rho(\boldsymbol{A}) = r$, then the following statements hold:*

1. *r is a positive real number and it is an eigenvalue of $\boldsymbol{A}$*
2. *r is a simple eigenvalue (i.e. its algebraic multiplicity is 1)*
3. *$\boldsymbol{A}$ has a right eigenvector $\boldsymbol{\psi}_r$ with eigenvalue r whose components are all positive.*
4. *The **only** eigenvectors whose components are all positive are those associated with the eigenvalue r.*

Proof. See [2]. □

Remark Since r is positive and $r = \rho(\boldsymbol{A}) = \max_i\{|\lambda_i|\}$, it must be the case that what we called λ_n above is actually r, the spectral radius. Moreover, observe that since $\lambda_n = r$ is the maximum in absolute value of all the eigenvalues, it must be the case that $|\lambda_n| \geq |\lambda_i|$ for all the other eigenvalues.

3.1.2 Paths of Arbitrary Length

We can count the number of paths of length k between two vertices, v_i and v_j, using powers of the adjacency matrix.

Proposition 3.2. *The number of paths length k from v_i to v_j is given by $\left[\boldsymbol{A}^k\right]_{ij}$*

Proof. This is a consequence of the matrix multiplication algorithm: for example, the number of paths of length 2 between v_i and v_j for a graph on n vertices is given by the following:

$$\sum_{x=1}^{n} [\boldsymbol{A}]_{ix} [\boldsymbol{A}]_{xj}$$

This is exactly the matrix multiplication algorithm for multiplying row i of $\boldsymbol{A}$ with column j of $\boldsymbol{A}$, and this makes sense: we are adding up all the possible ways there are to go from v_i to v_j through a third vertex, v_x. If there is no connection either between v_i and v_x or v_x and v_j, then there is no possible route through v_x; but if there is a route from v_i to v_x to v_j, then our sum will effectively count it. This fact generalizes to $\boldsymbol{A}^k$; $[A^k]_{ij}$ is the number of distinct paths of length k from vertex v_i to vertex v_j. □

As a direct consequence of this fact, we can count the total number of possible paths of length k in G by adding up all the entries of $\boldsymbol{A}$: this is represented in matrix multiplication terms by $\mathbf{1}^T \boldsymbol{A}^k \mathbf{1}$.

In particular, the diagonal elements of $\boldsymbol{A}^k$ also refer to number of loops of length k that exist in the graph. Since the adjacency matrix is always diagonalizable, this implies that powers of $\boldsymbol{A}$ are also diagonalizable: $\boldsymbol{D}^2 = \boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q}\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q} = \boldsymbol{Q}^{-1}\boldsymbol{A}^2\boldsymbol{Q}$; and so on for any k. Since $\boldsymbol{D}$ is a diagonal matrix, this means that the eigenvalues of $\boldsymbol{A}^k$ are the eigenvalues of $\boldsymbol{A}$ raised to the kth power. This, combined with the fact that the trace of a matrix is the sum of its eigenvalues, leads us down the following deduction [1]:

- $\sum_{i=1}^n \lambda_i = \mathrm{tr}(\boldsymbol{A}) = 0$, since G is simple (no self-loops)
- $\sum_{i=1}^n \lambda_i^2 = \mathrm{tr}(\boldsymbol{A}^2) = 2E(G)$, where $E(G)$ is the number of edges in G. This is because any loop of length 2 is precisely the loop that starts from a vertex, goes out along an edge, and comes back. The factor of 2 in the answer is due to the fact that any edge is shared by two vertices: so each edge is used twice in counting up all loops of length 2.
- $\sum_{i=1}^n \lambda_i^3 = \mathrm{tr}(\boldsymbol{A}^3) = 6T(G)$, where $T(G)$ is the number of triangles (loops of length three) in G. The 6 is because each triangle gets counted six times: for each of the three vertices in the triangle, there are two possible loops of length three to travel: one going "clockwise" and one going "counterclockwise", so to speak.
- In general, for $k > 2$, $\sum_{i=1}^n \lambda_i^k = \mathrm{tr}(\boldsymbol{A}^k) = 2 \cdot k \cdot C_k(G)$, where $C_k(G)$ is the number of loops of length k in the graph.

More generally, it should come as no surprise that the eigenvalues of $\boldsymbol{A}$ are also involved in computing the total number of paths of length k in $\boldsymbol{A}$.

Proposition 3.3. *The number of paths of length k is given by*

$$\sum_{i=1}^{n} a_i^2 \lambda_i^k \tag{1}$$

where $\{a_i\}$ are all constants and λ_i are the eigenvalues of $\boldsymbol{A}$.

Proof. [1] Recall that all the eigenvalues of $\boldsymbol{A}$ must be real, since $\boldsymbol{A}$ is symmetric. Let $\boldsymbol{\psi}_i$ denote the unit-length eigenvector associated with λ_i. Recall also that the eigenvectors of $\boldsymbol{A}$ form an eigenbasis; hence there is some linear combination such that

$$\mathbf{1} = \sum_{i=1}^{n} a_i \boldsymbol{\psi}_i$$

Inserting this into $\mathbf{1}^T \boldsymbol{A}^k \mathbf{1}$, we have the following expression for the total number of paths of length k (bear in mind that $\boldsymbol{\psi}_i^T \boldsymbol{\psi}_j = 0$ if $i \neq j$):

$$\left(\sum_{i=1}^{n} a_i \boldsymbol{\psi}_i^T\right) \boldsymbol{A}^k \left(\sum_{i=1}^{n} a_i \boldsymbol{\psi}_i\right) = \left(\sum_{i=1}^{n} a_i \boldsymbol{\psi}_i^T\right) \left(\sum_{i=1}^{n} a_i \lambda_i^k \boldsymbol{\psi}_i\right)$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j \lambda^k \boldsymbol{\psi}_i^T \boldsymbol{\psi}_j = \sum_{i=1}^{n} a_i^2 \lambda^k \left|\boldsymbol{\psi}_i\right| = \sum_{i=1}^{n} a_i^2 \lambda_i^k$$

□

Remark Since $|\lambda_i| \leq |\lambda_n|$ for all eigenvalues of $\boldsymbol{A}$, it must be the case that the expression $\sum_{i=1}^{n} a_i^2 \lambda_i^k$ is $O(\lambda_n^k)$; λ_n is the asymptotic growth rate for the number of paths of arbitrary length.

3.2 Eigenvector Centrality

The Perron-Frobenius theorem gave us both the eigenvalue $\rho(\boldsymbol{A})$ and an eigenvector with only positive components, $\boldsymbol{\psi}$.

Definition The *eigenvector centrality* of a vertex v_i is the ith entry in the eigenvector $\boldsymbol{\psi}$ associated with $\rho(\boldsymbol{A})$, denoted $[\boldsymbol{\psi}]_i$.

The eigenvector centrality is a kind of weighted average of influence, whereby a vertex connected to a "more influential" vertex would likewise be more influential; but a vertex connected to many "low influence" vertices would not carry as much influence.

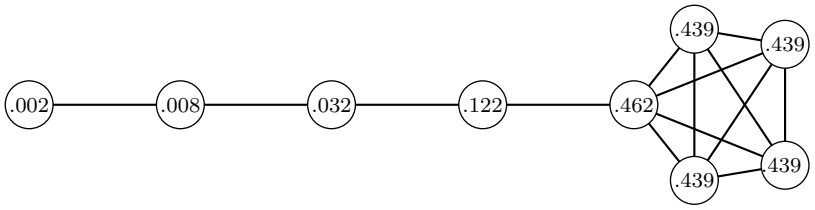


Figure 3: A graph with vertices labeled by their eigenvector centrality.

One way to derive this relationship is as follows: if we put on every vertex of the graph a score s_i then we desire the following property: that the score of an individual vertex is a (normalized) sum of the score of all of its neighbors. Call the list of all these scores the vector $\vec{\boldsymbol{s}}$. Putting aside the question of how to normalize the quantity, we have the following:

$$s_i = \frac{1}{S} \sum_{j=1}^{n} [\boldsymbol{A}]_{ij} \, s_j$$

Now, if we multiply through by S we have $Ss_i = \sum_{j=1}^{n} [\boldsymbol{A}]_{ij} s_j$; expressing the collection $\{s_i\}$ as the vector $\vec{\boldsymbol{s}}$, recall that the right hand side is exactly the matrix multiplication algorithm for the ith component of $\boldsymbol{A}\vec{\boldsymbol{s}}$. Therefore, $S\vec{\boldsymbol{s}} = \boldsymbol{A}\vec{\boldsymbol{s}}$, which has as solutions precisely the eigenvalues and eigenvectors of $\boldsymbol{A}$. Now, if we consider only a score assignment where all the vertices receive a *positive* score, we see that from the Perron-Frobenius theorem that there is in fact only one eigenvector $\boldsymbol{\psi}$ with all positive entries, and that it is associated precisely with the eigenvalue $\rho(\boldsymbol{A})$; hence eigenvector centrality is exactly what we are looking for in this scheme.

3.2.1 Applications

The primary utility of a centrality measure is to determine a measure for "influence" in a graph. In a social network, the centrality measure could represent clout; for websites, it has been used in the example of PageRank as a measure of relevancy. Additionally, it can be used alongside some applied mathematical model, such as predicting disease spread and devising a quarantine procedure: we could monitor the more influential nodes for sickness, and in the event they do become sick we will be well ahead of knowing that the disease is about to become a lot more widespread.

4 Power Iteration Algorithm

Computing the eigenvalues and eigenvectors of a matrix is, in general, difficult. In the naïve approach it requires computing the determinant of an $n \times n$ matrix, followed by solving an nth degree polynomial. There are special cases where determining the eigenvalues is made easier, such as the case of a diagonal matrix, but we cannot rely on our adjacency matrix $\boldsymbol{A}$ to have any special property. However, since we are interested primarily in the eigenvalue with the largest absolute value, we can take advantage of the *power iteration* algorithm, also known as "Von Mises" iteration. Given a matrix $\boldsymbol{A}$ and an initial "guess" $\vec{\boldsymbol{b}}_0$ for the eigenvector associated with the spectral radius, the algorithm is:

$$\vec{\boldsymbol{b}}_{k+1} = \frac{\boldsymbol{A}\vec{\boldsymbol{b}}_k}{\|\boldsymbol{A}\vec{\boldsymbol{b}}_k\|} \tag{2}$$

4.1 Algorithm Requirements

The algorithm will generate a sequence of vectors. It is possible for the algorithm to generate a sequence of vectors that doesn't converge, but has a subsequence of vectors that does converge.

Theorem 4.1. *If $\boldsymbol{A}$ has an eigenvalue λ_n (with eigenvector $\boldsymbol{\psi}_n$) strictly greater in magnitude than all of its other eigenvalues, and $\vec{\boldsymbol{b}}_0$ is a vector such that $\vec{\boldsymbol{b}}_0 \cdot \boldsymbol{\psi}_n \neq 0$, then the sequence of vectors $\{\vec{\boldsymbol{b}}_k\}$ defined by:*

$$\vec{\boldsymbol{b}}_{k+1} = \frac{\boldsymbol{A}\vec{\boldsymbol{b}}_k}{\|\boldsymbol{A}\vec{\boldsymbol{b}}_k\|} = \frac{\boldsymbol{A}^{k+1}\vec{\boldsymbol{b}}_0}{\|\boldsymbol{A}^{k+1}\vec{\boldsymbol{b}}_0\|}$$

has a convergent subsequence that converges to $\boldsymbol{\psi}_n$.

Proof. For our purposes we will prove only the case involving irreducible non-negative symmetric matrices $\boldsymbol{A}$ (thus satisfying the hypothesis of the Perron-Frobenius theorem). As we have seen already, the eigenvectors of $\boldsymbol{A}$ form an eigenbasis $\{\boldsymbol{\psi}_i\}$; hence there is some linear combination such that:

$$\vec{\boldsymbol{b}}_0 = \sum_{i=1}^{n} a_i \boldsymbol{\psi}_i$$

With the hypothesis that $\vec{\boldsymbol{b}}_0 \cdot \boldsymbol{\psi}_n \neq 0$ we have that the coefficient $a_n \neq 0$; from the Perron-Frobenius theorem we also know that λ_n must be positive. Now, applying $\boldsymbol{A}^k$ to $\vec{\boldsymbol{b}}_0$, we have that:

$$\boldsymbol{A}^k\vec{\boldsymbol{b}}_0 = \boldsymbol{A}^k \sum_{i=1}^{n} a_i \boldsymbol{\psi}_i = \sum_{i=1}^{n} a_i \boldsymbol{A}^k \boldsymbol{\psi}_i = \sum_{i=1}^{n} a_i \lambda_i^k \boldsymbol{\psi}_i$$

Factoring out the term $a_n\lambda_n^k$:

$$\sum_{i=1}^{n} a_i \lambda_i^k \boldsymbol{\psi}_i = a_n \lambda_n^k \sum_{i=1}^{n} \frac{a_i \lambda_i^k}{a_n \lambda_n^k} \boldsymbol{\psi}_i = a_n \lambda_n^k \sum_{i=1}^{n} \frac{a_i}{a_n} \left(\frac{\lambda_i}{\lambda_n}\right)^k \boldsymbol{\psi}_i$$

Put another way, we have that:

$$\boldsymbol{A}^k\vec{\boldsymbol{b}}_0 = a_n\lambda_n^k \left(\boldsymbol{\psi}_n + \frac{a_{n-1}}{a_n}\left(\frac{\lambda_{n-1}}{\lambda_n}\right)^k \boldsymbol{\psi}_{n-1} + \cdots + \frac{a_1}{a_n}\left(\frac{\lambda_1}{\lambda_n}\right)^k \boldsymbol{\psi}_1 \right)$$

If we call λ_n the eigenvalue from the Perron-Frobenius theorem, then the quantity $\left|\frac{\lambda_i}{\lambda_n}\right| < 1$ for any $i \neq n$. Hence $\left(\frac{\lambda_i}{\lambda_n}\right)^k$ tends to zero as $k \to \infty$. Hence, $\boldsymbol{A}^k\vec{\boldsymbol{b}}_0 \to a_n\lambda_n^k\boldsymbol{\psi}_n$ as $k \to \infty$. Thus, the unit vector $\frac{\boldsymbol{A}^k\vec{\boldsymbol{b}}_0}{\|\boldsymbol{A}^k\vec{\boldsymbol{b}}_0\|}$ tends to $\boldsymbol{\psi}_n$ (times a possibly negative constant) as $k \to \infty$. □

Remark In particular, the rate of convergence is $\left|\frac{\lambda_{n-1}}{\lambda_n}\right|$.

Remark We can take advantage of the Perron-Frobenius theorem yet again and recognize that since $\boldsymbol{\psi}_n$ has all positive entries, then the initial "guess" vector of $\mathbf{1}$ satisfies the property that $\mathbf{1} \cdot \boldsymbol{\psi}_n \neq 0$.

5 Maximal Entropy Random Walks

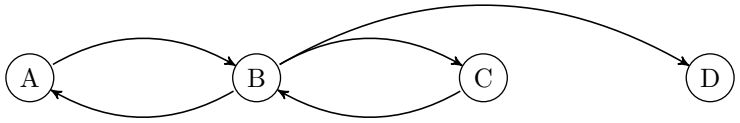
A *random walk* on a graph is a random traversal of the vertices of a graph. Without being bogged down in the formal details, we can think of a random walk as being determined by each of its steps. If we label the graph vertices $\{v_i\}$ then a single step of the random walk starting at position v_i is given by the following: for all the neighbors of v_i we call the probability that we step from v_i to v_j the *transition probability* P_{ij}. A random walk of length n starting from position v_i is given by all of the positions we hit after repeating this process n times. Different types of random walk processes can be defined based on how we describe the construction of the transition probabilities.

The first type of random walk is presented more as an example than anything else: for every vertex v_i of a graph, let the probability we travel from that vertex to any of its neighbors be $(\deg v_i)^{-1}$. This is sometimes called the "generic random walk": every step of the random walk effectively chooses equally between all of its possible next-steps. This random walk has been well-studied and has some interesting properties; for example, Pólya's Recurrence Theorem states that the generic random walk on the 1-dimensional and 2-dimensional infinite lattice is "recurrent" (i.e., it visits any point with probability one), but for higher dimensions it is not recurrent.

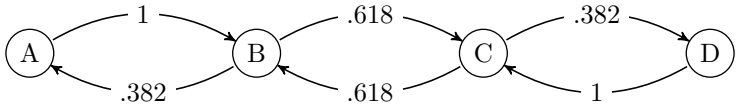
5.1 Construction

The random walk process we would like to discuss is called the *maximum entropy* random walk based primarily in analogy with information theory. For a fixed starting vertex v_i, for each of its neighbors v_j, we are seeking transition probabilities such that rather than choosing uniformly among neighbors (as is the case in the generic random walk), we are choosing uniformly among all possible *paths* from that vertex. This will typically not be the same as for the generic random walk; it will require some knowledge about the global structure of the graph. For example, consider the transition probabilities for a graph that just consists of a path of length four. Looking at one of the vertices of degree two; intuitively speaking, there are "more" paths going to vertices in one direction than in the other. This can be made rigorous by expressing the probability as the limit on path lengths; as we examine the paths of length $k = 1, 2, 3, ...$, we can see that the maximal entropy transition probabilities are precisely the limit as k goes to infinity.

(a) All paths of length 1 leaving vertex B.
1/2 go to A, 1/2 go to C.



(b) All paths of length 2 leaving vertex B.
2/3 of them go to B, 1/3 of them go to D.



(c) Maximal entropy probabilities.

Figure 4: Example: the path of length four.

For any vertex v_i and its neighbor v_j, we wish to determine the appropriate number P_{ij} such that all possible paths from v_i are equally likely. Recall that the entries in the matrix $\boldsymbol{A}^k$ indicate the number of paths from vertex v_i to vertex v_j of length k. Therefore, the number of paths of length k from v_i to any vertex is the sum across the ith row of $\boldsymbol{A}^k$; represented as $\sum_{x=1}^{n} \left[\boldsymbol{A}^k\right]_{ix}$ in summation terms and as $\left[\boldsymbol{A}^k \mathbf{1}\right]_i$ in matrix multiplication terms.

Thus, in order to determine P_{ij}, we must determine what the proportion is of all paths (in the limit, as path length goes to infinity) that leave vertex v_i and route through v_j compared to all paths that leave vertex v_i in total. This can be expressed in summation terms as [3]:

$$P_{ij} = \lim_{k\to\infty} \frac{[\boldsymbol{A}]_{ij} \sum_{x=1}^{n} \left[\boldsymbol{A}^{k-1}\right]_{jx}}{\sum_{j'=1}^{n} [\boldsymbol{A}]_{ij'} \sum_{x=1}^{n} \left[\boldsymbol{A}^{k-1}\right]_{j'x}} \tag{3}$$

The denominator is the expression for all paths from v_i: they have to route through some neighboring vertex $v_{j'}$.

Theorem 5.1.

$$P_{ij} = \frac{[\boldsymbol{A}]_{ij} [\boldsymbol{\psi}]_j}{\lambda [\boldsymbol{\psi}]_i} \tag{4}$$

Proof. Taking the limit, we first observe that we can rewrite the terms involving k in matrix multiplication terms:

$$\sum_{x=1}^{n} \left[\boldsymbol{A}^{k-1}\right]_{jx} = \left[\boldsymbol{A}^{k-1} \cdot \mathbf{1}\right]_{j}$$
$$\sum_{x=1}^{n} \left[\boldsymbol{A}^{k-1}\right]_{j'x} = \left[\boldsymbol{A}^{k-1} \cdot \mathbf{1}\right]_{j'}$$

Applying the same proof from the power iteration algorithm: represent $\mathbf{1}$ as a linear combination of eigenvectors:

$$\lim_{k\to\infty} \frac{\sum_{x=1}^{n} \left[\boldsymbol{A}^{k-1}\right]_{jx}}{\sum_{x=1}^{n} \left[\boldsymbol{A}^{k-1}\right]_{j'x}} = \lim_{k\to\infty} \frac{\left[\boldsymbol{A}^{k-1} \cdot \mathbf{1}\right]_{j}}{\left[\boldsymbol{A}^{k-1} \cdot \mathbf{1}\right]_{j'}} = \lim_{k\to\infty} \frac{a_n \lambda_n^{k-1} \left[\boldsymbol{\psi}\right]_{j} + O\left(\left(\frac{\lambda_{n-1}}{\lambda_n}\right)^{k}\right)}{a_n \lambda_n^{k-1} \left[\boldsymbol{\psi}\right]_{j'} + O\left(\left(\frac{\lambda_{n-1}}{\lambda_n}\right)^{k}\right)} = \frac{\left[\boldsymbol{\psi}\right]_{j}}{\left[\boldsymbol{\psi}\right]_{j'}}$$

Where $\boldsymbol{\psi}$ is the eigenvalue associated with $\rho(\boldsymbol{A})$ from the Perron-Frobenius theorem. So this allows us to express P_{ij} as:

$$P_{ij} = \frac{\left[\boldsymbol{A}\right]_{ij} \left[\boldsymbol{\psi}\right]_{j}}{\sum_{j'=1}^{n} \left[\boldsymbol{A}\right]_{ij'} \left[\boldsymbol{\psi}\right]_{j'}}$$

Notice now that the denominator is the familiar summation form of the operation $\left[\boldsymbol{A}\boldsymbol{\psi}\right]_{i}$; since $\boldsymbol{\psi}$ is the eigenvector we have that $\left[\boldsymbol{A}\boldsymbol{\psi}\right]_{i} = \lambda \left[\boldsymbol{\psi}\right]_{i}$; thus our final expression for P_{ij} follows. □

5.2 Local Approximation

As a practical matter, our expression for P_{ij} depends on entries in the adjacency matrix $\boldsymbol{A}$, and the eigenvalue/eigenvector pair associated with the spectral radius of $\boldsymbol{A}$. We have already established that approximation is possible for computing the eigenvector and eigenvalue; however, even in that situation it was required that we had access to all of $\boldsymbol{A}$. Now we turn to the question of computing P_{ij} even if we did not have the "global" information given to us in the form of the adjacency matrix $\boldsymbol{A}$.

This is a very practical consideration, as we can take as a direct example the concept of a social network graph: we may know all of our friends, but short of interviewing every single person on planet Earth we do not have a good grasp of the global adjacency matrix for the social network. One can imagine that there is indeed a global graph structure for all people currently living, but it is beyond intractable to try to determine and compute the entire global structure. Instead, we must appeal to "local" approximations of the terms containing the

spectral radius and its eigenvector ($[\boldsymbol{A}]_{ij}$ is assumed to be readily determined for some fixed i; we do not assume for all possible i, j that $[\boldsymbol{A}]_{ij}$ is known).

The way we will compute this local approximation is by taking advantage of the definition of the limit and of what $\boldsymbol{A}^k$ represents. Recall that P_{ij} is defined as the limit:

$$P_{ij} = \lim_{k\to\infty} \frac{[\boldsymbol{A}]_{ij} \sum_{x=1}^{n} \left[\boldsymbol{A}^{k-1}\right]_{jx}}{\sum_{j'=1}^{n} [\boldsymbol{A}]_{ij'} \sum_{x=1}^{n} \left[\boldsymbol{A}^{k-1}\right]_{j'x}}$$

Recall that $\left[\boldsymbol{A}^k\right]_{ij}$ represents the number of paths of length k from v_i to v_j. Hence, we can describe the *kth order approximation* of P_{ij}, P_{ij}^k, as follows [4]:

- $P_{ij}^0 = \dfrac{[\boldsymbol{A}]_{ij}}{\sum_{j'=1}^{n} [\boldsymbol{A}]_{ij'}} = \dfrac{[\boldsymbol{A}]_{ij}}{\deg v_i}$

  This approximation amounts to "uniformly select a neighbor".

- $P_{ij}^1 = \dfrac{[\boldsymbol{A}]_{ij} \sum_{x=1}^{n} [\boldsymbol{A}]_{jx}}{\sum_{j'=1}^{n} [\boldsymbol{A}]_{ij'} \sum_{x=1}^{n} [\boldsymbol{A}]_{j'x}} = \dfrac{[\boldsymbol{A}]_{ij} \deg v_j}{\sum_{j'=1}^{n} [\boldsymbol{A}]_{ij'} \deg v_{j'}}$

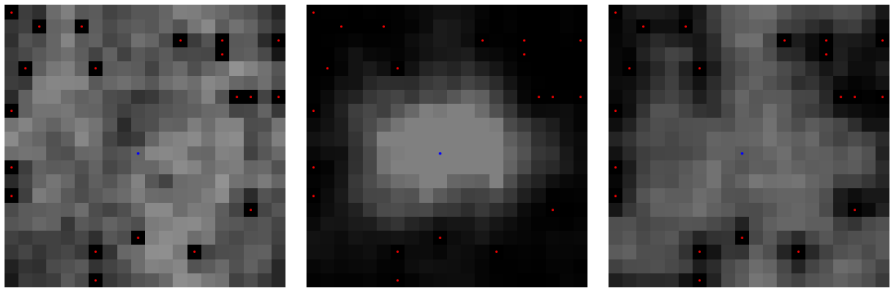
  This approximation amounts to "uniformly select a path of length two".

- $P_{ij}^2 = \dfrac{[\boldsymbol{A}]_{ij} \sum_{x=1}^{n} \left([\boldsymbol{A}]_{jx} \sum_{y=1}^{n} [\boldsymbol{A}]_{xy} \right)}{\sum_{j'=1}^{n} [\boldsymbol{A}]_{ij'} \sum_{x=1}^{n} \left([\boldsymbol{A}]_{j'x} \sum_{y=1}^{n} [\boldsymbol{A}]_{xy} \right)} = \dfrac{[\boldsymbol{A}]_{ij} \sum_{x=1}^{n} [\boldsymbol{A}]_{jx} \deg v_x}{\sum_{j'=1}^{n} [\boldsymbol{A}]_{ij'} \sum_{x=1}^{n} [\boldsymbol{A}]_{j'x} \deg v_x}$

  This approximation amounts to "uniformly select a path of length three".

And so on, although for practical purposes the second order approximation is already quite closely approximating the transition probabilities from the maximum entropy random walk [4].

5.3 Applications



(a) Generic. (b) Maximal entropy. (c) Approximation.

Figure 5: Frequency distribution for random walk encounters on a periodic lattice with selected nodes (in red) deleted. Highest eigenvector centrality node indicated in blue.

The primary feature of maximal entropy random walks that makes them useful is that their transition probabilities will create a process that has a steady state distribution that is clustered around the most "well-connected" area of the graph. Intuitively this is reasonable; as there are more paths going through a more well-connected part of the graph, there should be higher probabilities of entering there in a maximal entropy process. However, this is useful for a variety of reasons: for example, we could take this precisely to represent a quantitative measure of "well-connected", and use it to find cliques or communities in our social networks.

The usefulness of a local approximation to the maximal entropy random walk is that it enables us to devise procedures to "find" the areas of the graph that are well-connected, by approximating the maximal entropy random walk and following our approximations. Additionally, we can build estimates for the entries of the eigenvector $\boldsymbol{\psi}$ associated with the spectral radius of $\boldsymbol{A}$ without having to deal with the global matrix $\boldsymbol{A}$.

References

[1] Steven Kay Butler. *Eigenvalues and Structures of Graphs*. PhD thesis, University of California, San Diego, 2008.

[2] Hannah Cairns. A short proof of perron's theorem. `http://www.math.cornell.edu/~web6720/Perron-Frobenius_Hannah%20Cairns.pdf`, 2014.

[3] Jarosław Duda. *Extended Maximal Entropy Random Walk*. PhD thesis, Jagiellonian University, 2012.

[4] R Sinatra, J Gómez-Gardeñes, R Lambiotte, V Nicosia, and V Latora. Maximal-entropy random walks in complex networks with limited information. *Physical Review E*, 83(030103), 2011.