**Distribution Agreement**



Yiwen Guo April 7, 2017

Graph Laplacians For Directed Networks With Applications To Centrality Measures

by

Yiwen Guo

Michele Benzi
Adviser

Department of Mathematics and Computer Science

Michele Benzi
Adviser

Manuela Manetta
Committee Member

Jed Brody
Committee Member

2017

Graph Laplacians For Directed Networks With Applications To Centrality Measures

By

Yiwen Guo

Michele Benzi
Adviser

An abstract of
A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Mathematics and Computer Science

2017

Abstract

Graph Laplacians For Directed Networks With Applications To Centrality Measures
By Yiwen Guo

This thesis introduces a new approach to centrality measures by using the non-normalized graph Laplacians. It then compares and contrasts this approach with other existing techniques through small-scale and large-scale examples. Finally, it gives the conclusion and discusses some limitations.

Graph Laplacians For Directed Networks With Applications To Centrality Measures

By

Yiwen Guo

Michele Benzi
Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelors of Science with Honors

Department of Mathematics and Computer Science

2017

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

Have you ever used social media, such as Facebook, Twitter, and Instagram? If so, how many friends do you have? Or, in other words, how big is your social network? Social networks are connections and interpersonal relationships that a person has. More generally, a network contains some nodes representing individuals or entities and some edges between these nodes. In our daily lives, we have a variety of networks: transportation, food chain, web search, and so on. The following graphs illustrate some of these examples. In the first graph, every node represents a student from an elementary school and a student can choose up to two neighbors who will be sitting next to him or her. For instance, an edge that goes from HN2 to CE means that HN2 wishes CE to be his neighbor. We can also see that the bigger the node is, the more popular the student is in terms of being chosen to be other students' neighbor. In the second graph, every node stands for a species. If an edge points from a species to another one, it means that the latter eats the former.

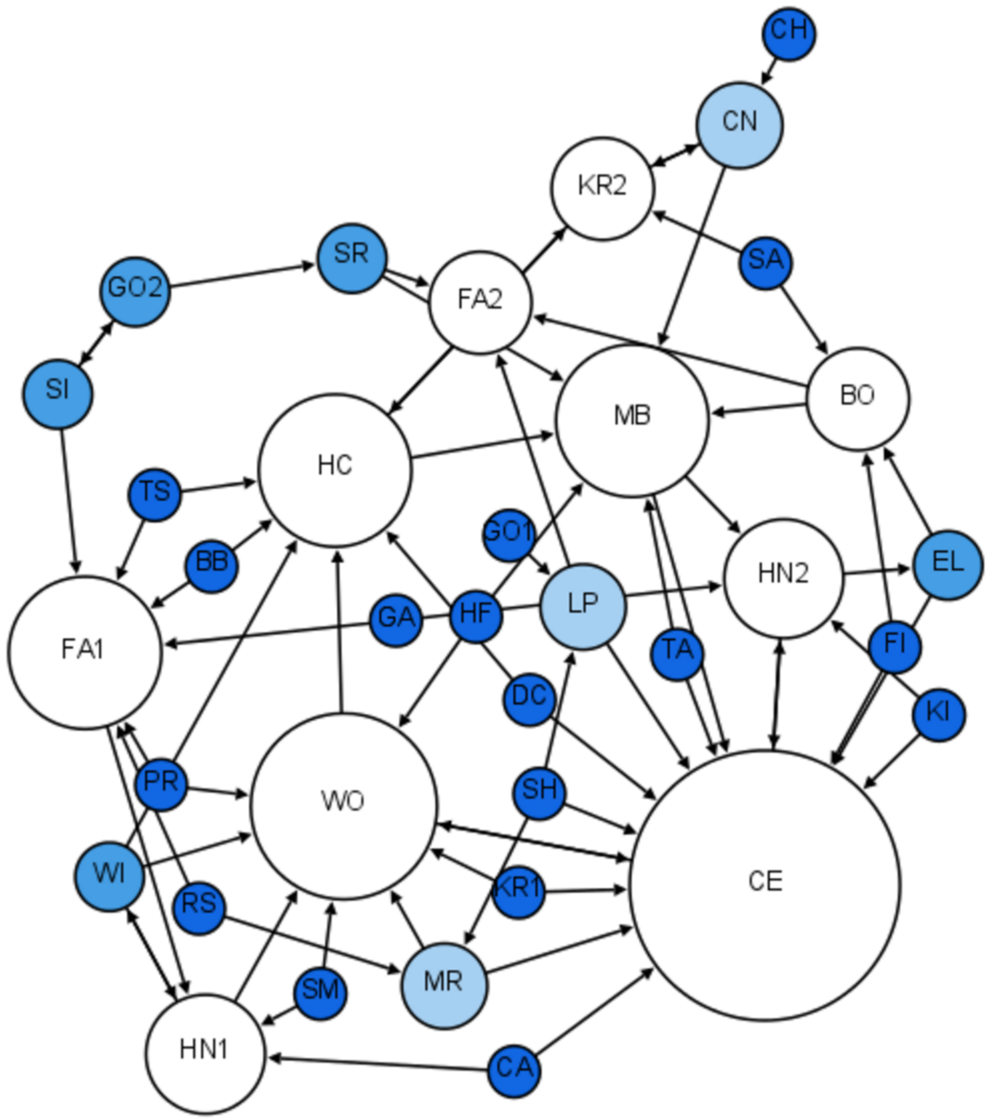


Figure 1.1: Jacob Moreno's sociogram of a group of elementary school students
This network can be found at https://commons.wikimedia.org/wiki/File:Moreno_Sociogram_1st_Grade.png

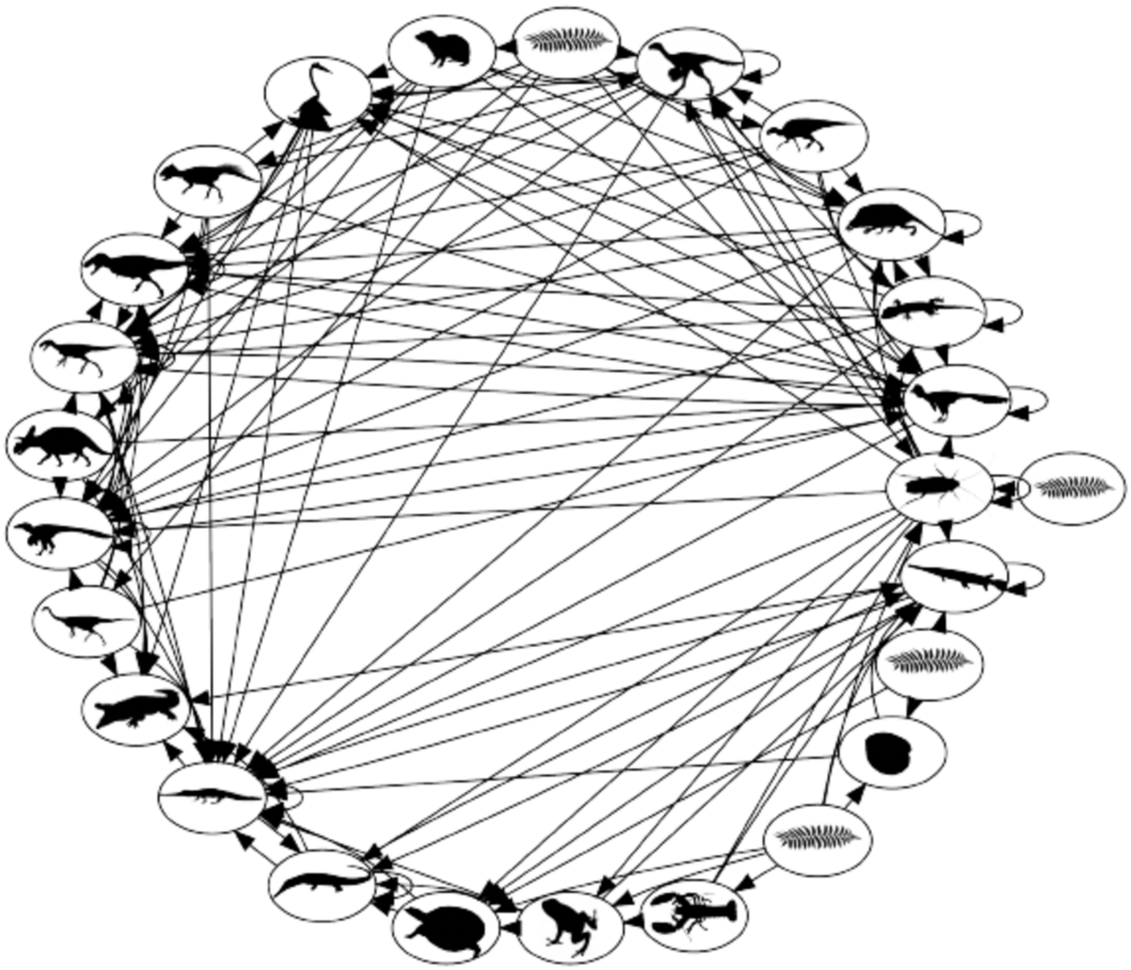

Figure 1.2: Food web network
This network can be found at https://www.researchgate.net/figure/235637860_fig1_Figure-5-Food-web-network-of-late-Cretaceous-terrestrial-communities-in-North-America

As shown above, networks are not only ubiquitous but also essential in our daily lives. Analyzing networks allows us to get a better understanding of the bigger world surrounding us. In this thesis, we are going to talk about centrality measures, or how to identify the most central or important nodes in a network.

A node is seen as important if it is a good authority or hub. An authority is a node gathering(receiving) information from other nodes. A hub is a node sending out(broadcasting) information to other nodes. A good authority is pointed at by good hubs and vice versa. Notice that this definition is iterative or circular. For more information, please refer to [1].

At this point, you may wonder why is it important to know the most central node in a given graph? For instance, imagine that we have a social network where every node represents an individual and any of the two nodes have an edge if they are acquaintances. If we were an advertisement company, we would wish to target the most influential people in this network so that the advertising information could spread out more quickly. Another example is virus transmission. Imagine again that we have a network with every node representing a patient and any of the two nodes will have an edge if virus transmission is possible between them. By identifying the most dangerous patient with the highest possibility of virus transmission in this network, we could try to isolate this particular patient to slow down the spread of diseases. Hence, we can see that centrality measures are essential in terms of flow of information.

The organization of the thesis is as follows. We begin by summarizing some basic concepts in graph theory, such as the definition of directed graphs, the adjacency matrix of a given graph, and its graph Laplacians. We then introduce a new approach to centrality measures, which is done by solving the linear systems:

$$L_{out}^T \mathbf{x}_{out} = \mathbf{0} \quad \text{and} \quad L_{in} \mathbf{x}_{in} = \mathbf{0}$$

where $L_{out}^T$ and $L_{in}$ are the graph Laplacians. We present and compare our solution with three other techniques: PageRank, HITS, and the Dominant Eigenvector Approach. Finally, we draw our conclusion and discuss some possible limitations that the new approach may have.

# Chapter 2

# Graph Theory

## 2.1 Basic Concepts

**Definition 1.** A graph $G = (V, E)$ is a set of nodes (vertices) $V$ with $|V| = n$ and a set of edges $E \subset \{(i, j)|i, j \in V\}$. (see [2]).

**Definition 2.** Given any two nodes $i, j \in V$, a graph is undirected if $(i, j) \in E$ implies $(j, i) \in E$. A graph is directed if $(i, j) \in E$ does not imply $(j, i) \in E$.

**Definition 3.** The in-degree $d^{in}$ of a node in $G$ is the number of directed edges ending in that node. The out-degree $d^{out}$ of a node is the number of directed edges starting at that node. Let $D_{out} = diag(d_1^{out}, d_2^{out}, ..., d_n^{out})$ be the out-degree diagonal matrix and $D_{in} = diag(d_1^{in}, d_2^{in}, ..., d_n^{in})$ be the in-degree diagonal matrix

.

**Definition 4.** The adjacency matrix $A$ of the graph is defined by $a_{ij} = 1$ if $(i, j)$ is an edge of $G$ and $a_{ij} = 0$ otherwise.

**Remark 1.** For an undirected graph, $A$ is symmetric. For a directed graph, $A$ is unsymmetric. Also, the row sums of $A$ give the out-degrees and the column sums of $A$ give the in-degrees.

**Definition 5.** A directed graph, also called digraph, is strongly connected if for any pair of nodes $i, j \in V$, there exists a directed path

$$i \to i_1 \to i_2 \to ... \to i_k \to j$$

(See Definition 7 below). A digraph is weakly connected if it is connected as an undirected graph (obtained by ignoring the orientation).

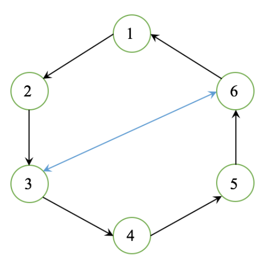


Figure 2.1: A strongly connected digraph


Figure 2.2: A weakly connected digraph

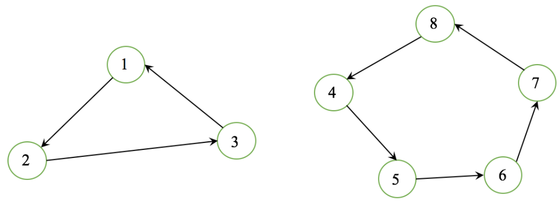


Figure 2.3: A digraph with eight nodes that is neither strongly nor weakly connected

**Definition 6.** A square matrix $A$ is said to be reducible if there exists a permutation matrix $\Pi$ such that

$$\Pi A \Pi^T = \begin{bmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{bmatrix}$$

where $A_{11}$ and $A_{22}$ are square. If there is no such $\Pi$, we say that $A$ is irreducible.

**Theorem 1.** A digraph is strongly connected if and only if the corresponding adjacency matrix $A$ is irreducible.

**Fact 1.** For any digraph $G = (V, E)$, there exists a permutation matrix $\Pi$ such that

$$\Pi A \Pi^T = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1p} \\ & A_{22} & \dots & A_{2p} \\ & & \ddots & \vdots \\ & & & A_{pp} \end{bmatrix}$$

where each $A_{ii}$ block is square and irreducible. The subgraphs of $G$ having adjacency matrices $A_{11}, \dots, A_{pp}$ are called the strongly connected components of $G$. The largest one (say $A_{11}$) is called the maximal strongly connected component.

**Definition 7.** A walk is a sequence of (directed) edges: $i \to i_1 \to i_2 \to ... \to i_k \to j$. The nodes $i_1, ..., i_k$ can be repeated. If none is repeated, we call this a (directed) path.

**Fact 2.** Let $A^k$ denote the $k$th power of the adjacency matrix $A$. Then $(A^k)_{ij}$ is the number of walks of length $k$ starting at node $i$ and ending at node $j$. In particular, $(A^k)_{ii}$ is the number of closed walks of length $k$ passing through node $i$. The distance $d(i, j)$ is the length of the shortest path between node $i$ and node $j$. If there is no such path starting at $i$ and ending at $j$, we set $d(i, j) = \infty$. Note that $d(i, j) \neq d(j, i)$ in general. Then a digraph is strongly connected if and only if $d(i, j)$ is finite for all $i, j \in V$. See [8].

## 2.2 Graph Laplacians

In this section, we first define Laplacians for undirected graphs and then describe two ways to extend this definition to the directed case. First of all, define $D$ as the diagonal matrix whose entries are the degrees of each node. Since graphs are undirected, the in-degrees of every node equal the out-degrees. We call $D$ the degree matrix of $G$. Then, the graph Laplacian is defined as $L = D - A$ where $A$ is the adjacency matrix. See [8].

Although we have a unique definition of Laplacians for undirected graphs, notice that the in-degrees and out-degrees in the directed case are generally not the same. Hence, we cannot simply use the same definition for directed graphs. In this section, we introduce two methods: (1) the symmetric Laplacian and (2) the nonsymmetric Laplacians. Refer to [9].

### 2.2.1 Symmetric Laplacian Via A Bipartite Graph Model

**Definition 8.** An undirected graph $G = (V, E)$ is bipartite if $V = V_1 \cup V_2$ with $V_1 \cap V_2 = \varnothing$, where $V_1, V_2$ are such that nodes in $V_1$ can be connected only with nodes in $V_2$ and nodes in $V_2$ can be connected only with nodes in $V_1$.

Any digraph of $n$ nodes can be represented uniquely by a bipartite graph on $2n$ nodes as follows: If the digraph is $G = (V, E)$, we construct the bipartite graph $\tilde{G} = (\tilde{V}, \tilde{E})$ where $\tilde{V} = V \cup V', V' = \{n + 1, n + 2, ..., 2n\}$, and $\tilde{E} = \{(i, j')|j' = n + j \text{ and } (i, j) \in E\}$. For instance,

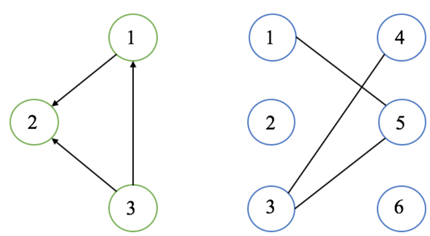


Figure 2.4: A digraph with its bipartite graph

The adjacency matrices for the digraph and for the bipartite graph are

$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}, \tilde{A} = \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}$$

Since $D = diag(1, 0, 2, 1, 2, 0)$,

$$\tilde{L} = D - \tilde{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -1 & -1 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 \\ -1 & 0 & -1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} D_{out} & -A \\ -A^T & D_{in} \end{bmatrix}$$

If we use the above definition of graph Laplacian, what is the meaning of the eigenvalues and eigenvectors of $\tilde{L}$ in terms of the original directed graph? Can we relate them to quantities related to $A$? Assume the digraph is strongly connected. Then $D_{out}$ and $D_{in}$ are invertible. Consider

$$
\begin{aligned}
\tilde{L} &= \begin{bmatrix} D_{out}^{-1/2} & 0 \\ 0 & D_{in}^{-1/2} \end{bmatrix} \begin{bmatrix} D_{out} & -A \\ -A^T & D_{in} \end{bmatrix} \begin{bmatrix} D_{out}^{-1/2} & 0 \\ 0 & D_{in}^{-1/2} \end{bmatrix} \\
&= \begin{bmatrix} I & -D_{out}^{-1/2} A D_{in}^{-1/2} \\ -D_{in}^{-1/2} A^T D_{out}^{-1/2} & I \end{bmatrix} \\
&= I_{2n} + \begin{bmatrix} 0 & -D_{out}^{-1/2} A D_{in}^{-1/2} \\ -D_{in}^{-1/2} A^T D_{out}^{-1/2} & \end{bmatrix}
\end{aligned}
\tag{2.1}
$$

The eigenvalues of this normalized and symmetric Laplacian $\tilde{L}$ are of the form

$$
\lambda_i = 1 \mp \sigma_i(D_{out}^{-1/2} A D_{in}^{-1/2}).
$$

Also let $\hat{L} = (D - A)D^{-1} = I_{2n} - AD^{-1}$ :

$$
\begin{aligned}
\begin{bmatrix} D_{out} & -A \\ -A^T & D_{in} \end{bmatrix} \begin{bmatrix} D_{out}^{-1} & 0 \\ 0 & D_{in}^{-1} \end{bmatrix} &= \begin{bmatrix} I & -AD_{in}^{-1} \\ -A^T D_{out}^{-1} & I \end{bmatrix} \\
&= I_{2n} - \begin{bmatrix} 0 & AD_{in}^{-1} \\ (D_{out}^{-1} A)^T & 0 \end{bmatrix}
\end{aligned}
\tag{2.2}
$$

Note that $\hat{L}$ is not symmetric and that there is no simple relationship between its eigenvalues and the singular values of $A$ or $AD_{in}^{-1}$ or $AD_{out}^{-1}$. For another definition of a symmetric Laplacian for directed graphs, see [6]. However, we will not make use of this definition in our thesis.

### 2.2.2 Nonsymmetric Laplacians

As mentioned earlier in this chapter, we cannot simply use $L = D - A$ because we do not know if $D$ is the out-degree matrix or the in-degree one. Just using as $D$ the degree matrix of the underlying undirected graph does not lead to a useful definition. Hence, we introduce two nonsymmetric Laplacians as follows.

**Definition 9.** $L_{in} = D_{in} - A$, $L_{out} = D_{out} - A$. Note that in general $\mathbf{1}^T L_{in} = \mathbf{0}$, $L_{in}\mathbf{1} \neq \mathbf{0}$; $L_{out}\mathbf{1} = \mathbf{0}$, $\mathbf{1}^T L_{out} \neq \mathbf{0}$.

Consider now the diffusion-type differential equations:

$$\begin{cases} \dot{\mathbf{x}} = -L_{out}\mathbf{x} \\ \mathbf{x}(0) = \mathbf{x}_0 \end{cases} \tag{2.3}$$

$$\begin{cases} \dot{\mathbf{x}} = -L_{in}\mathbf{x} \\ \mathbf{x}(0) = \mathbf{x}_0 \end{cases} \tag{2.4}$$

where $\dot{\mathbf{x}}$ denotes the derivative with respect to $t$ of $\mathbf{x}(t)$.

Assume that $G = (V, E)$ is strongly connected. In this case $A$ is an irreducible, nonnegative matrix. By the Perron-Frobenius theorem, the spectral radius $\rho(A) = \max\{|\lambda_i|; \lambda_i \in \sigma(A)\}$ is a simple eigenvalue of $A$ (i.e. $\rho(A) = \lambda_1 = \lambda_{max}(A)$ is real and $\geq 0$). Moreover, there exists $\mathbf{x} \in \mathbb{R}^n, \mathbf{x} = x_i$ with $x_i > 0$ for $i = 1, ..., n$ such that $A\mathbf{x} = \lambda_1\mathbf{x}$ $(= \rho(A)\mathbf{x})$. This $\mathbf{x}$ is unique up to normalization. As a consequence, the fact that $G$ is strongly connected implies that $0$ is a simple eigenvalue of $L_{in}, L_{out}$. In other words,

**Theorem 2.** dim ker$(L_{in})$ = dim ker$(L_{out}) = 1$.

Proof:

Since $G$ is strongly connected, the diagonal entries of $D_{in}, D_{out}$ are nonzero. Hence, $D_{in}, D_{out}$ are invertible. Therefore,

$$L_{out} = (I - AD_{out}^{-1})D_{out}, \quad L_{in} = D_{in}(I - D_{in}^{-1}A)$$

and dim ker$(L_{out})$ = dim ker$(I - AD_{out}^{-1})$, dim ker$(L_{in})$ = dim ker$(I - D_{in}^{-1}A)$.

Now, $A \geq 0, D_{in}^{-1} \geq 0, D_{out}^{-1} \geq 0$. So $AD_{out}^{-1}, D_{in}^{-1}A$ are both nonnegative and irreducible. Since $A$ is irreducible and multiplying an irreducible matrix by a nonsingular diagonal matrix does not affect the irreducibility, by the Perron-Frobenius theorem, both $AD_{out}^{-1}$ and $D_{in}^{-1}A$ have a simple dominant eigenvalue.

We claim that $1 \in \sigma(AD_{out}^{-1}), 1 \in \sigma(D_{in}^{-1}A)$, and all other eigenvalues of $AD_{out}^{-1}, D_{in}^{-1}A$ are less in magnitude than (or $\neq$) 1.

Recall that for any matrix $A \in \mathbb{C}^{n \times n}$ and any induced matrix norm, $\rho(A) \leq ||A||$ because

$$A\mathbf{x} = \lambda\mathbf{x} \Rightarrow \|A\mathbf{x}\| = \lambda\|\mathbf{x}\| \Rightarrow \lambda \leq \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|} \leq \|A\| = sup_{\mathbf{x}\neq\mathbf{0}}\frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}.$$

Now observe that since $\left\|AD_{out}^{-1}\right\|_\infty = 1$ and $AD_{out}\mathbf{1} = \mathbf{1}$, $\rho(AD_{out}^{-1}) = 1$. Hence, $\mathbf{1}$ is an eigenvector corresponding to $\lambda = 1$. Similarly for $D_{in}^{-1}A$ (just take the $\|\cdot\|_1$ norm, which is the $\|\cdot\|_\infty$ of the transpose). But if 1 is a simple eigenvalue of $AD_{out}^{-1}$, then 0 is simple eigenvalue of $I - AD_{out}^{-1}$. That is, dim ker$(I - AD_{out}^{-1}) = 1$. Likewise for dim ker$(I - D_{in}^{-1}A)$. Since multiplication by a nonsingular (diagonal) matrix does not change the rank, the Theorem is proved.

■

Since $L_{out}\mathbf{1} = (D_{out} - A)\mathbf{1} = D_{out}\mathbf{1} - A\mathbf{1} = D_{out}\mathbf{1} - D_{out}\mathbf{1} = 0$, we see that ker($L_{out}$) $=$ span$\{\mathbf{1}\} = \{\mathbf{x} = c\mathbf{1} | c \in \mathbb{R}\}$ . For $L_{in}$, we know that $\mathbf{1}^T L_{in} = \mathbf{0}^T$. Denote by $\mathbf{q}_1^{in}$ the (normalized) eigenvector of $L_{in}$ associated with the eigenvalue $\lambda = 0$. All we can say is that we can choose $\mathbf{q}_1^{in}$ to have positive entries (by the Perron-Frobenius theorem) if $G$ is strongly connected. Unless $G$ is undirected (in which case $L_{in} = L_{out} = L = D - A$), $\mathbf{q}_1^{in}$ is not a constant vector. Its entries are all positive and presumably they can tell us something about the graph.

Assume now for simplicity that $L_{in}, L_{out}$ are diagonalizable, then we can write:

$$L_{out} = X_{out}\Lambda_{out}X_{out}^{-1}, \quad L_{in} = X_{in}\Lambda_{in}X_{in}^{-1}$$

where $\Lambda_{out} = diag(0, \lambda_2(L_{out}), ..., \lambda_n(L_{out}))$, and $\Lambda_{in} = diag(0, \lambda_2(L_{in}), ..., \lambda_n(L_{in}))$. Note that

$$\lambda_i(L_{out}) \in \mathbb{C}, \quad \lambda_i(L_{out}) \neq 0, \quad \lambda_i(L_{in}) \in \mathbb{C}, \quad \lambda_i(L_{in}) \neq 0,$$

$$X_{out} = [1, \mathbf{q}_2^{out}, ..., \mathbf{q}_n^{out}] \in \mathbb{C}^{n\times n}, \quad X_{in} = [\mathbf{q}_1^{in}, \mathbf{q}_2^{in}, ..., \mathbf{q}_n^{in}] \in \mathbb{C}^n.$$

Also, denote by $\mathbf{y}_2^{out}, \mathbf{y}_2^{in}$ the left eigenvectors of $L_{out}, L_{in}$ corresponding to $\lambda = 0$. Then,

$$L_{out} = X_{out}\Lambda_{out}X_{out}^{-1} = \sum_{i=1}^{n} \lambda_i(L_{out})\mathbf{q}_i^{out}(\mathbf{y}_i^{out})^T \tag{2.5}$$

$$L_{in} = X_{in}\Lambda_{in}X_{in}^{-1} = \sum_{i=1}^{n} \lambda_i(L_{in})\mathbf{q}_i^{in}(\mathbf{y}_i^{in})^T \tag{2.6}$$

where the first term is a matrix of zeros in each of the sums.

From (2.5),

$$\begin{aligned} exp(-tL_{out}) &= \sum_{i=1}^{n} e^{-t\lambda_i(L_{out})}\mathbf{q}_i^{out}(\mathbf{y}_i^{out})^T \\ &= \mathbf{1}(\mathbf{y}_1^{out})^T + \sum_{i=2}^{n} e^{-t\lambda_i(L_{out})}\mathbf{q}_i^{out}(\mathbf{y}_i^{out})^T \end{aligned} \tag{2.7}$$

Since $L_{out}$ is an M-matrix of rank $n-1$, each $\lambda_i(L_{out})$ with $i \geq 2$ has positive real part ([3]). Therefore, as $t \to \infty$, $exp(-tL_{out}) \to \mathbf{1}(\mathbf{y}_1^{out})^T$. Thus, the solution to (2.5) is given by

$$\mathbf{x}(t) = exp(-tL_{out})\mathbf{x}_0 = [(\mathbf{y}_1^{out})^T\mathbf{x}_0]\mathbf{1} + o(t) \text{ as } t \to \infty.$$

So at steady state, the solution $\mathbf{x}(t)$ must reach "thermal equilibrium": that is, $\lim_{t\to\infty}\mathbf{x}(t) = c\mathbf{1}$ where $c = (\mathbf{y}_1^{out})^T\mathbf{x}_0$. The rate of approach to this equilibrium depends on $\gamma(L_{out}) = min\left\{|\lambda_i(L_{out})|, \lambda_i(L_{out}) \neq 0\right\}$.

For the solution of (2.6), we find instead

$$\mathbf{x}(t) = exp(-tL_{in})\mathbf{x}_0 = [(\mathbf{y}_1^{in})^T\mathbf{x}_0]\mathbf{q}_1^{in} + o(t) \text{ as } t \to \infty.$$

So now the steady state is not a constant vector, but a multiple of the eigenvector $\mathbf{q}_1^{in}$, which is the eigenvetcor of $L_{in}$ associated with the zero eigenvalue. The rate of convergence is now governed by $\gamma(L_{in}) = min\left\{|\lambda_i(L_{in})|,\ \lambda_i(L_{in}) \neq 0\right\}$.

Now, given the solution $\mathbf{x}(t) = exp(-tL_{in})\mathbf{x}_0$, we would like to know if $\mathbf{x}(t)$ is monotonic. In other words, if $(\mathbf{x}_0)_i > (\mathbf{q}_1)_i$, where $L_{in}\mathbf{q}_1 = \mathbf{0}$ and $\mathbf{q}_1 > \mathbf{0}$, and is normalized, does $\mathbf{x}_i(t)$ decrease to $(\mathbf{q}_1)_i$ as $t \to \infty$? If $(\mathbf{x}_0)_i < (\mathbf{q}_1)_i$, does $\mathbf{x}_i(t)$ increase to $(\mathbf{q}_1)_i$ as $t \to \infty$?

In order to know the monoticity of the solution, we need to look at its derivative. Given $\mathbf{x}(t) = exp(-tL_{in})\mathbf{x}_0$,

$$\frac{d}{dt}\mathbf{x}(t) = -L_{in}e^{-tL_{in}}\mathbf{x}_0 \quad (= -L_{in}\mathbf{x}(t))$$

and we claim that

$$\frac{d}{dt}(-L_{in}e^{-tL_{in}})_{ij} < 0 \,\forall\, i = j \quad \text{ and } \quad \frac{d}{dt}(-L_{in}e^{-tL_{in}})_{ij} > 0 \,\forall\, i \neq j.$$

Argument: Let $t$ be small.

$$\begin{aligned} -L_{in}e^{-tL_{in}} &= -L_{in}(I - tL_{in} + \frac{t^2}{2!}L_{in}^2 - \frac{t^3}{3!}L_{in}^3 + ...) \\ &= -L_{in} + tL_{in}^2 - \frac{t^2}{2!}L_{in}^3 + \frac{t^3}{3!}L_{in}^4 - ... \\ &= -L_{in} + t(L_{in}^2 - \frac{t}{2!}L_{in}^3 + \frac{t^3}{3!}L_{in}^4 - ...) \\ &= -L_{in} + O(t) \end{aligned} \tag{2.8}$$

So if $t$ is small enough, the sign of the entries of $-L_{in}e^{-tL_{in}}$ is the same as the sign of the entries of $-L_{in}$. This implies that the diagonal entries are less than zero and the off-diagonal entries are bigger than zero.

Now let $t' > t, \Delta t = t' - t > 0$. Then,

$$
\begin{aligned}
-L_{in}e^{-t'L_{in}} &= -L_{in}e^{-(t'-t)L_{in}-tL_{in}} \\
&= (-L_{in}e^{-tL_{in}})e^{-\Delta tL_{in}} \\
&= (-L_{in}e^{-tL_{in}})(I - \Delta tL_{in} + \frac{(\Delta t)^2}{2!}L_{in}^2 - ...)
\end{aligned} \tag{2.9}
$$

For $\Delta t$ small enough, the signs of the entries of $-L_{in}e^{-t'L_{in}}$ will be equal to the signs of the entries of $-L_{in}e^{-tL_{in}}$. Hence, for all $t$,

$$
(-L_{in}e^{-tL_{in}})_{ij} \begin{cases} < 0 \text{ if } i = j \\ > 0 \text{ if } i \neq j \end{cases}
$$

Recall that

$$
exp(-tL_{in})\mathbf{x}_0 \to (1/n)(\mathbf{1}^T\mathbf{x}_0)\mathbf{q}_1 \quad as\ t \to \infty \tag{2.10}
$$

$$
exp(-tL_{in}) \to (1/\sqrt{n})\mathbf{q}_1\mathbf{1}^T \quad as\ t \to \infty \tag{2.11}
$$

From (2.11), as $t \to \infty$,

$$
(e^{-tL_{in}}\mathbf{x}_0)_i \begin{cases} \text{decreases to } (1/\sqrt{n})\mathbf{q}_1 \\ \text{increases to } (1/\sqrt{n})\mathbf{q}_1 \end{cases}
$$

depending on whether $(\mathbf{x}_0)_i > (\mathbf{q}_1)_i$ or $(\mathbf{x}_0)_i < (\mathbf{q}_1)_i$.

# Chapter 3

# Centrality Measures

In this chapter, we are mainly concerned with the question: how can we identify the most "important" nodes in a graph? Intuitively, you might think that this can be done by counting the number of in-degrees and out-degrees a node has and the one which has the most in and out-degrees is the most important. Although it is reasonable to think in this way, we will see later in this chapter that this simple approach is not sufficient to determine the importance of a node. In addition, how can we define "important"? These are the questions that we will address in this chapter. The techniques we touch upon in this chapter are the most popular and widely used ones, but there are also many other approaches to centrality measures such as [2] and [7]. We refer to [8] for additional discussion.

## 3.1 Existing Techniques For Centrality Measures

### 3.1.1 PageRank

**Definition 10.** Let $H = I - A^T D_{out}^{-1} = L_{out}^T D_{out}^{-1}$ and $K = I - AD_{in}^{-1} = L_{in} D_{in}^{-1}$, which are normalizations of $L_{out}, L_{in}$. Notice that $A^T D_{out}^{-1}$ and $AD_{in}^{-1}$ are column-stochastic.

Assume that $G$ is strongly connected. We know that $\lambda = 0$ is a simple eigenvalue of $H$, and the solution of

$$H\mathbf{x} = \mathbf{0} \quad \text{with} \quad x_i > 0 \quad \text{and } \mathbf{1}^T\mathbf{x} = \sum_{i=1}^{n} x_i = 1$$

is the stationary probability distribution of the Markov Chain, or random walk on $G$, described by the transition probability matrix $A^T D_{out}^{-1}$ [3]. To solve for **x**, we replace the last equation in $H\mathbf{x} = \mathbf{0}$ with the constraint $\sum_{i=1}^{n} x_i = 1$ so that we can get a unique solution. In this case, **x** is also known as the PageRank vector. Similarly, we do the same for $K\mathbf{x} = \mathbf{0}$ and the solution is called the reverse PageRank vector. One drawback of this method is that sometimes it does not distinguish between authority and hub ranking, as we will see in the next section.

### 3.1.2 HITS

In addition to the PageRank method mentioned, HITS (see [1]) is another method used to rank nodes. Let $A$ be the adjacency matrix of a digraph.

**Definition 11.** The singular value decomposition of $A$ is $A = U\Sigma V^T$ where $U$ and

$V$ are orthogonal matrices and $\Sigma$ is a diagonal matrix with nonnegative numbers on its diagonal, referred to as the singular values of $A$.

Then we can get:

$$A^T = (U\Sigma V^T)^T = V\Sigma^T U^T \text{ and } AA^T = U\Sigma^2 U^T, A^T A = V\Sigma^2 V^T.$$

Hence, the eigenvalues of $AA^T$ and $A^T A$ are the squared singular values of $A$. Let $U = [\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n] =$ left singular vectors of $A$ and $V = [\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n] =$ right singular vectors of $A$. Since $A = U\Sigma V^T$, $AV = U\Sigma$ and $A\mathbf{v}_i = \sigma_i \mathbf{u}_i$. Also, since $A^T = V\Sigma U^T$, $A^T U = V\Sigma$ and $A^T \mathbf{u}_i = \sigma_i \mathbf{v}_i$. Therefore,

$$AA^T \mathbf{u}_i = A(\sigma_i \mathbf{v}_i) = \sigma_i A\mathbf{v}_i = \sigma_i(\sigma_i \mathbf{u}_i) = \sigma_i^2 \mathbf{u}_i \text{ and}$$

$$A^T A\mathbf{v}_i = A^T(\sigma_i \mathbf{u}_i) = \sigma_i(\sigma_i \mathbf{v}_i) = \sigma_i^2 \mathbf{v}_i.$$

This shows that $\mathbf{u}_i$ is the ith eigenvector of $AA^T$ and $\mathbf{v}_i$ is the ith eigenvector of $A^T A$. If we denote $AA^T$ as the hub matrix and $A^T A$ as the authority matrix, then $\mathbf{u}_1$ gives the hub scores of nodes and $\mathbf{v}_1$ gives the authority scores of nodes. One problem of HITS is that in order to do the calculation, we need to make sure that the matrices $AA^T$ and $A^T A$ are irreducible. Note that the condition that $A$ is strongly connected is not enough for $AA^T$ and $A^T A$ to be irreducible.

### 3.1.3 The Dominant Eigenvector Approach

Last but not least, we introduce the dominant eigenvector approach for centrality measure. Given the adjacency matrix $A$, we compute $A\mathbf{x} = \lambda\mathbf{x}$ and find the eigenvector corresponding to the largest eigenvalue, which is called the dominant eigenvector for $A$. This is the hub centrality measure. Similarly, we compute $A^T\mathbf{y} = \lambda\mathbf{y}$ and find its corresponding dominant eigenvector, which represents the authority centrality measure.

## 3.2 A New Approach To Centrality Measure

Recall that in the previous chapter, we defined two nonsymmetric Laplacians $L_{out} = D_{out} - A, L_{in} = D_{in} - A$. Since $L_{out}\mathbf{1} = \mathbf{0}$, there exists $\mathbf{x}_{out} \in \mathbb{R}^n$ such that $(\mathbf{x}_{out})^T L_{out} = \mathbf{0}$ or $L_{out}^T\mathbf{x}_{out} = \mathbf{0}$. Similarly, since $\mathbf{1}^T L_{in} = \mathbf{0}^T$, there exists $\mathbf{x}_{in} \in \mathbb{R}^n$ such that $L_{in}\mathbf{x}_{in} = \mathbf{0}$. Moreover, by the Perron-Frobenius theorem, we can pick $\mathbf{x}_{out}, \mathbf{x}_{in}$ to be positive. So what do $\mathbf{x}_{out}, \mathbf{x}_{in}$ tell us about the graph? Before answering this question, we will first introduce some new concepts.

**Definition 12.** A node is a good authority (receiver) if it is pointed at by good hubs. Similarly, a node is a good hub (broadcaster) if it is pointed at by good authorities. Authorities and hubs can be not exclusive to each other. A node can be both.

We conjecture that the entries of $\mathbf{x}_{in}$ give the scores of authorities, which are receivers of information and that the entries of $\mathbf{x}_{out}$ give the scores of hubs, which are broadcasters of information. We will see some examples later in the next chapter.

Given $G = (V, E)$, undirected and connected, consider the following ODE:

$$\begin{cases} \dot{\mathbf{x}} = -L\mathbf{x} \\ \mathbf{x}(0) = \mathbf{x}_0 (\neq 0) \end{cases}$$

The solution of this ODE is as $t \to \infty$,

$$\mathbf{x}(t) = e^{-tL}\mathbf{x}_0 \to ((1/n)\mathbf{1}^T\mathbf{x}_0)\mathbf{1} \tag{3.1}$$

This constant vector is called the equilibrium or consensus. For instance, if $\mathbf{x}_0 = \mathbf{e}_1 = [1, 0, ..., 0]^T$, then $\mathbf{x}_\infty = [1/n, 1/n, ..., 1/n]^T$.

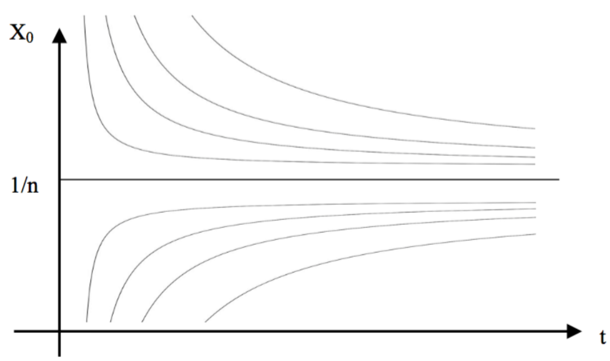


Figure 3.1: The steady-state of the solution for different initial values

If we rewrite $\dot{\mathbf{x}}$ using the definition of derivatives, we will get

$$\frac{\mathbf{x}(t+\Delta t) - \mathbf{x}(t)}{\Delta t} = -L\mathbf{x}(t) \Rightarrow \mathbf{x}_{t+1} = \mathbf{x}_t - \Delta t L\mathbf{x}_t = (I - \Delta t L)\mathbf{x}_t$$

If $\Delta t$ is small enough, this converges to the steady state: as $t \to \infty$, $\mathbf{x}_t \to \mathbf{x}_\infty$.

Now, given $G = (V, E)$, directed and strongly connected, consider again the following system of ODEs:

$$\begin{cases} \dot{\mathbf{x}} = -L_{out}\mathbf{x} \\ \mathbf{x}(0) = \mathbf{x}_0 \end{cases} \qquad (3.2) \qquad\qquad \begin{cases} \dot{\mathbf{x}} = -L_{in}\mathbf{x} \\ \mathbf{x}(0) = \mathbf{x}_0 \end{cases} \qquad (3.3)$$

The solution to $(3.2)$ converges to the same limit as $(3.1)$ and is called the concensus dynamics. However, for $(3.3)$,

$$\lim_{t\to\infty} \mathbf{x}(t) = (1/\sqrt{n})(\mathbf{q}_1^T\mathbf{x}_0)\mathbf{q}_1 \text{ where } L_{in}\mathbf{q}_1 = \mathbf{0}, \|\mathbf{q}_1\|_2 = 1.$$

Note that this limit is not constant in general. It is called the advection dynamics. Advection is the process where a distribution or mass is actively transported by a flow field. See [5]. Therefore, we can interpret the solution given by (3.3) as advection, where each node carries information and the latter moves from nodes to nodes. During this process, every node tries to gather information from all other nodes and thus (3.3) gives the authority scores.

For $L_{out}^T\mathbf{x} = \mathbf{0}$, $L_{out}^T = (D_{out} - A)^T = D_{out} - A^T$ where $A^T$ is the adjacency matrix for the reverse graph.

**Definition 13.** Given a directed graph $G$, its reverse graph is a directed graph on the same set of vertices with all of the edges reversed in direction. That is, if $G$ contains an edge $(i, j)$, then the reverse graph of $G$ contains an edge $(j, i)$.

Now, instead of gathering information from all other nodes, each node tries to send out information to all other nodes. Therefore, the solution of $L_{out}^T\mathbf{x} = \mathbf{0}$ gives the hub scores.

# Chapter 4

# Experiments And Comparison

## 4.1 Small-scale Graphs

In this section, we introduce three small-scale graphs and compute the score of each node by using four different methods. The results are shown as follows.

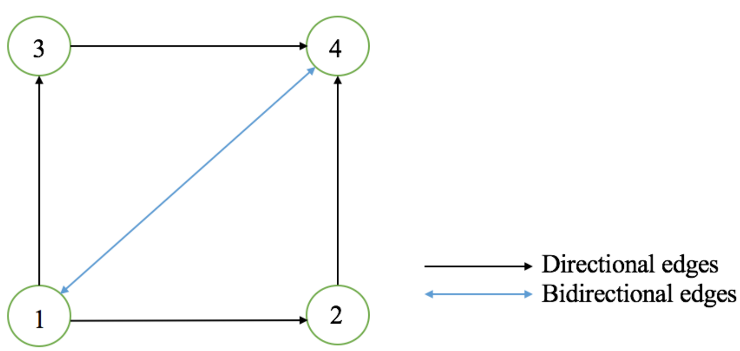


Figure 4.1: A four-node digraph

| Authority Ranking | | | | | | | |
|---|---|---|---|---|---|---|---|
| New Approach | | PageRank | | HITS | | Eigenvector | |
| Node | Score | Node | Score | Node | Score | Node | Score |
| 4 | 0.5000 | 1 | 0.3750 | - | - | 4 | 0.7442 |
| 1 | 0.1667 | 4 | 0.3750 | - | - | 1 | 0.4892 |
| 2 | 0.1667 | 2 | 0.1250 | - | - | 2 | 0.3215 |
| 3 | 0.1667 | 3 | 0.1250 | - | - | 3 | 0.3215 |

Table 4.1: Authority ranking for the graph in figure 4.1

| Hub Ranking | | | | | | | |
|---|---|---|---|---|---|---|---|
| New Approach | | Reverse PageRank | | HITS | | Eigenvector | |
| Node | Score | Node | Score | Node | Score | Node | Score |
| 1 | 0.5000 | 1 | 0.3750 | - | - | 1 | 0.7442 |
| 2 | 0.1667 | 4 | 0.3750 | - | - | 4 | 0.4892 |
| 3 | 0.1667 | 2 | 0.1250 | - | - | 2 | 0.3215 |
| 4 | 0.1667 | 3 | 0.1250 | - | - | 3 | 0.3215 |

Table 4.2: Hub ranking for the graph in figure 4.1

In this example, since $A^T A$ and $AA^T$ are not irreducible, we cannot use the ranking given by HITS. We can see that both our method and the dominant eigenvector approach rank node 4 as the most important in terms of authorities and node 1 as the most important in terms of hubs. This result seems reasonable since node 4 has three incoming edges and node 1 has three outcoming edges. It means that node 4 can be reached by any of the other three nodes and that node 1 can reach any other nodes. Hence, node 4 is regarded as the most important receiver of information and node 1 is the most important broadcaster of information. Note that PageRank and Reverse PageRank regard node 1 and 4 as equally important, which also makes sense. Recall that good authorities are those pointed to by good hubs and vice versa. However, this result implies one disadvantage of PageRank, which is its inability to distinguish between authorities and hubs in some cases.

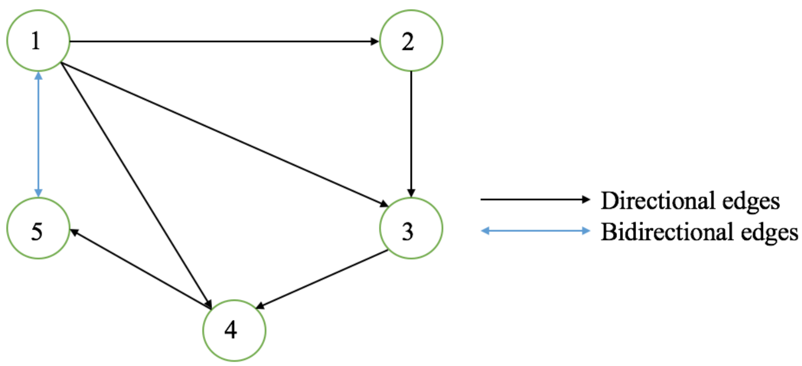


Figure 4.2: A five-node digraph

| Authority Ranking | | | | | | | |
|---|---|---|---|---|---|---|---|
| New Approach | | PageRank | | HITS | | Eigenvector | |
| Node | Score | Node | Score | Node | Score | Node | Score |
| 1 | 0.5000 | 1 | 0.2857 | - | - | 5 | 0.5896 |
| 5 | 0.2500 | 5 | 0.2857 | - | - | 4 | 0.5202 |
| 4 | 0.1250 | 4 | 0.2143 | - | - | 3 | 0.4138 |
| 2 | 0.0625 | 3 | 0.1429 | - | - | 1 | 0.3843 |
| 3 | 0.0625 | 2 | 0.0714 | - | - | 2 | 0.2505 |

Table 4.3: Authority ranking for the graph in figure 4.2

| Hub Ranking | | | | | | | |
|---|---|---|---|---|---|---|---|
| New Approach | | Reverse PageRank | | HITS | | Eigenvector | |
| Node | Score | Node | Score | Node | Score | Node | Score |
| 5 | 0.3636 | 5 | 0.3478 | - | - | 1 | 0.7637 |
| 4 | 0.2727 | 1 | 0.3478 | - | - | 5 | 0.4978 |
| 3 | 0.1818 | 4 | 0.1739 | - | - | 4 | 0.3245 |
| 1 | 0.0909 | 3 | 0.0870 | - | - | 3 | 0.2115 |
| 2 | 0.0909 | 2 | 0.0435 | - | - | 2 | 0.1379 |

Table 4.4: Hub ranking for the graph in figure 4.2

In this example, the authority ranking of the new approach and PageRank are consistent. However, there are some differences for hub ranking. From this example, we can see that the new approach shares some commonalities with PageRank, but it is also different from it. Again, since $A^T A$ and $AA^T$ are not irreducible, we cannot use HITS. Note that PageRank and Reverse PageRank view node 1 and 5 as equally important.

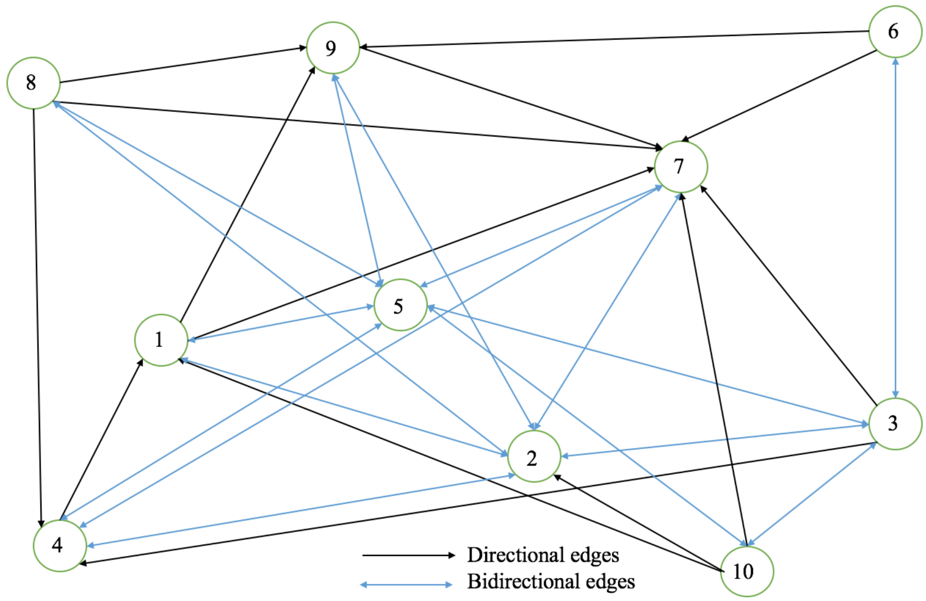


Figure 4.3: Knoke's data on information exchanges among organizations operating in the social welfare field from [10]

| Authority Ranking | | | | | | | |
|---|---|---|---|---|---|---|---|
| New Approach | | PageRank | | HITS | | Eigenvector | |
| Node | Score | Node | Score | Node | Score | Node | Score |
| 7 | 0.2500 | 7 | 0.1742 | 7 | 0.5581 | 7 | 0.5076 |
| 9 | 0.1321 | 5 | 0.1740 | 2 | 0.4204 | 5 | 0.4239 |
| 4 | 0.1321 | 2 | 0.1740 | 5 | 0.4204 | 2 | 0.4239 |
| 2 | 0.1249 | 4 | 0.1227 | 9 | 0.3145 | 4 | 0.3319 |
| 5 | 0.1070 | 9 | 0.0921 | 1 | 0.2696 | 9 | 0.2942 |
| 1 | 0.0992 | 1 | 0.0921 | 4 | 0.2498 | 1 | 0.2847 |
| 3 | 0.0567 | 3 | 0.0659 | 3 | 0.2493 | 3 | 0.2235 |
| 8 | 0.0464 | 8 | 0.0539 | 8 | 0.1366 | 8 | 0.1829 |
| 10 | 0.0327 | 10 | 0.0380 | 10 | 0.1296 | 10 | 0.1397 |
| 6 | 0.0189 | 6 | 0.0132 | 6 | 0.0588 | 6 | 0.0482 |

Table 4.5: Authority ranking for the graph in figure 4.3

| Hub Ranking | | | | | | | |
|---|---|---|---|---|---|---|---|
| New Approach | | Reverse PageRank | | HITS | | Eigenvector | |
| Node | Score | Node | Score | Node | Score | Node | Score |
| 6 | 0.2027 | 5 | 0.1905 | 8 | 0.3780 | 5 | 0.4478 |
| 10 | 0.1885 | 3 | 0.1698 | 10 | 0.3692 | 2 | 0.3698 |
| 3 | 0.1428 | 2 | 0.1345 | 5 | 0.3673 | 10 | 0.3618 |
| 8 | 0.1373 | 10 | 0.1120 | 2 | 0.3423 | 3 | 0.3436 |
| 5 | 0.0916 | 8 | 0.0816 | 1 | 0.3299 | 8 | 0.3391 |
| 2 | 0.0647 | 4 | 0.0696 | 4 | 0.3212 | 4 | 0.2876 |
| 4 | 0.0585 | 1 | 0.0642 | 3 | 0.3056 | 1 | 0.2770 |
| 1 | 0.0540 | 7 | 0.0638 | 9 | 0.2693 | 7 | 0.2385 |
| 9 | 0.0360 | 6 | 0.0602 | 6 | 0.2160 | 9 | 0.2279 |
| 7 | 0.0239 | 9 | 0.0535 | 7 | 0.2100 | 6 | 0.1748 |

Table 4.6: Hub ranking for the graph in figure 4.3

In this example [10], all four methods agree that node 7 is the most important receiver of information. However, we can see that they disagree about the most important broadcaster. Our method states that node 6 is the most important hub. This is counter-intuitive at first because node 6 has only three outcoming edges reaching out to node 6, 7, and 9. Recall that a good hub is one that is pointed to by a good authority. Hence, by this definition, it makes sense that node 6 is the most important node because it is pointed to by the most important authority, node 7. Reverse PageRank and the dominant eigenvector approach claim that node 5 is the most important hub, which also makes sense since node 5 can reach 7 nodes. In addition, HITS regards Node 8 as the most important hub since it can reach node 5, 2, and 7, which are important authorities. If we look at the top four nodes in authority ranking, our method has two overlaps with PageRank and the eigenvector approach, and three overlaps with HITS. Similarly, looking at the top four nodes in hub ranking, we can see that our method has two overlaps with PageRank and the eigenvector approach, and one overlap with HITS. These observations suggest that our method has some overlaps with the other three methods, but it also reveals some new information that the other three do not tell us.

## 4.2 Larger Graphs

In this section, we compute the ranking scores for larger graphs. We first browse the sparse matrix collection created by Tim Davis and find those matrices that are binary and unsymmetric. We then find their largest strongly connected components and compute their rankings of nodes. Finally, we display the results in the following tables. Note that if $A^TA$ and $AA^T$ are not strongly connected, then HITS is not applicable.

| Description of Examples | | | |
|---|---|---|---|
| Example | ID | Number of Rows | Number of Columns |
| 1 | 1482 | 60 | 60 |
| 2 | 1456 | 31 | 31 |
| 3 | 2288 | 1300 | 1300 |
| 4 | 2286 | 70533 | 70533 |
| 5 | 2315 | 8490 | 8490 |

Table 4.7: Information on the test matrices

Example 1: This graph comes from a graph drawing contest in 2002.

| Authority Ranking | | | | | | | |
|---|---|---|---|---|---|---|---|
| New Approach | | PageRank | | HITS | | Eigenvector | |
| Node | Score | Node | Score | Node | Score | Node | Score |
| 57 | 0.1018 | 55 | 0.0903 | - | - | 51 | 0.3706 |
| 58 | 0.0837 | 51 | 0.0703 | - | - | 52 | 0.3367 |
| 51 | 0.0667 | 56 | 0.0691 | - | - | 50 | 0.3309 |
| 56 | 0.0657 | 49 | 0.0596 | - | - | 56 | 0.3230 |
| 55 | 0.0572 | 52 | 0.0590 | - | - | 49 | 0.3154 |
| 52 | 0.0561 | 57 | 0.0536 | - | - | 45 | 0.2929 |

Table 4.8: Authority ranking for example 1

| Hub Ranking | | | | | | | |
|---|---|---|---|---|---|---|---|
| New Approach | | Reverse PageRank | | HITS | | Eigenvector | |
| Node | Score | Node | Score | Node | Score | Node | Score |
| 9 | 0.1623 | 10 | 0.1404 | - | - | 53 | 0.3175 |
| 10 | 0.1224 | 11 | 0.1276 | - | - | 47 | 0.3125 |
| 11 | 0.1112 | 12 | 0.1036 | - | - | 54 | 0.2806 |
| 13 | 0.0969 | 9 | 0.0931 | - | - | 49 | 0.2418 |
| 36 | 0.0946 | 4 | 0.0668 | - | - | 48 | 0.2409 |
| 12 | 0.0602 | 31 | 0.0631 | - | - | 41 | 0.2317 |

Table 4.9: Hub ranking for example 1

| Number of Intersection Between Two Sets | | | |
|---|---|---|---|
| | New Approach and PageRank | New Approach and HITS | New Approach and eigenvector |
| Authority | 5 | NA | 3 |
| Hub | 4 | NA | 0 |

Table 4.10: Number of elements in common for example 1

Example 2: This graph was constructed by expanding a 200-page response set to a search engine query "California" by Jon Kleinberg.

| Authority Ranking | | | | | | | |
|---|---|---|---|---|---|---|---|
| New Approach | | PageRank | | HITS | | Eigenvector | |
| Node | Score | Node | Score | Node | Score | Node | Score |
| 6 | 0.1458 | 4 | 0.1542 | - | - | 4 | 0.3542 |
| 27 | 0.0833 | 6 | 0.0899 | - | - | 14 | 0.3465 |
| 18 | 0.0625 | 1 | 0.0771 | - | - | 5 | 0.3465 |
| 1 | 0.0417 | 3 | 0.0771 | - | - | 6 | 0.3218 |
| 10 | 0.0417 | 27 | 0.0514 | - | - | 1 | 0.2616 |
| 20 | 0.0417 | 18 | 0.0385 | - | - | 27 | 0.2105 |
| 19 | 0.0417 | 14 | 0.0385 | - | - | 3 | 0.1953 |
| 31 | 0.0417 | 5 | 0.0385 | - | - | 18 | 0.1775 |
| 11 | 0.0417 | 11 | 0.0257 | - | - | 20 | 0.1578 |
| 4 | 0.0417 | 12 | 0.0257 | - | - | 28 | 0.1578 |

Table 4.11: Authority ranking for example 2

| Hub Ranking | | | | | | | |
|---|---|---|---|---|---|---|---|
| New Approach | | Reverse PageRank | | HITS | | Eigenvector | |
| Node | Score | Node | Score | Node | Score | Node | Score |
| 21 | 0.1171 | 4 | 0.2513 | - | - | 3 | 0.3707 |
| 20 | 0.1081 | 3 | 0.0838 | - | - | 24 | 0.3394 |
| 4 | 0.1081 | 21 | 0.0681 | - | - | 14 | 0.3024 |
| 19 | 0.1081 | 19 | 0.0628 | - | - | 4 | 0.2753 |
| 28 | 0.0450 | 20 | 0.0628 | - | - | 9 | 0.2591 |
| 6 | 0.0360 | 6 | 0.0628 | - | - | 22 | 0.2348 |
| 3 | 0.0360 | 14 | 0.0366 | - | - | 7 | 0.2348 |
| 24 | 0.0360 | 1 | 0.0314 | - | - | 25 | 0.2348 |
| 13 | 0.0360 | 28 | 0.0262 | - | - | 23 | 0.2348 |
| 15 | 0.0360 | 5 | 0.0262 | - | - | 5 | 0.2248 |

Table 4.12: Hub ranking for example 2

| Number of Intersection Between Two Sets | | | |
|---|---|---|---|
| | New Approach and PageRank | New Approach and HITS | New Approach and eigenvector |
| Authority | 6 | NA | 6 |
| Hub | 7 | NA | 2 |

Table 4.13: Number of elements in common for example 2

Example 3: Wikipedia vote network - Wikipedia is a free encyclopedia written collaboratively by volunteers around the world. A small part of Wikipedia contributors are administrators, who are users with access to additional technical features that aid in maintenance. In order for a user to become an administrator a Request for adminship (RfA) is issued and the Wikipedia community via a public discussion or a vote decides who to promote to adminship. The network contains all the users and discussion from the inception of Wikipedia till January 2008. Nodes in the network represent Wikipedia users and a directed edge from node $i$ to node $j$ represents that user $i$ voted on user $j$.

| Authority Ranking | | | | | | | |
|---|---|---|---|---|---|---|---|
| New Approach | | PageRank | | HITS | | Eigenvector | |
| Node | Score | Node | Score | Node | Score | Node | Score |
| 4 | 0.0676 | 2 | 0.0302 | 604 | 0.2732 | 604 | 0.2276 |
| 2 | 0.0528 | 3 | 0.0296 | 393 | 0.2071 | 393 | 0.2193 |
| 3 | 0.0358 | 8 | 0.0223 | 299 | 0.2067 | 299 | 0.1924 |
| 516 | 0.0246 | 9 | 0.0184 | 291 | 0.1612 | 291 | 0.1657 |
| 6 | 0.0203 | 4 | 0.0172 | 344 | 0.1535 | 81 | 0.1424 |
| 78 | 0.0192 | 21 | 0.0125 | 588 | 0.1425 | 199 | 0.1361 |
| 17 | 0.0152 | 17 | 0.0125 | 757 | 0.1235 | 588 | 0.1298 |
| 44 | 0.0145 | 78 | 0.0110 | 676 | 0.1184 | 344 | 0.1257 |
| 30 | 0.0138 | 6 | 0.0103 | 770 | 0.1182 | 407 | 0.1179 |
| 7 | 0.0137 | 26 | 0.0101 | 691 | 0.1172 | 537 | 0.1135 |

Table 4.14: Authority ranking for example 3

| Hub Ranking | | | | | | | |
|---|---|---|---|---|---|---|---|
| New Approach | | Reverse PageRank | | HITS | | Eigenvector | |
| Node | Score | Node | Score | Node | Score | Node | Score |
| 1294 | 0.0562 | 1186 | 0.0150 | 578 | 0.1065 | 578 | 0.1357 |
| 1186 | 0.0484 | 1218 | 0.0096 | 858 | 0.1050 | 858 | 0.1261 |
| 719 | 0.0149 | 1294 | 0.0087 | 393 | 0.1011 | 5 | 0.1136 |
| 1205 | 0.0103 | 578 | 0.0071 | 757 | 0.0992 | 880 | 0.1108 |
| 961 | 0.0099 | 5 | 0.0068 | 719 | 0.0983 | 393 | 0.1100 |
| 871 | 0.0095 | 858 | 0.0067 | 562 | 0.0978 | 562 | 0.1097 |
| 1113 | 0.0092 | 901 | 0.0050 | 333 | 0.973 | 719 | 0.1078 |
| 1057 | 0.0090 | 880 | 0.0048 | 880 | 0.0939 | 178 | 0.1038 |
| 811 | 0.0089 | 719 | 0.0046 | 183 | 0.0926 | 604 | 0.1019 |
| 1157 | 0.0087 | 562 | 0.0044 | 604 | 0.0907 | 757 | 0.1005 |

Table 4.15: Hub ranking for example 3

| Number of Intersection Between Two Sets | | | |
|---|---|---|---|
| | New Approach and PageRank | New Approach and HITS | New Approach and eigenvector |
| Authority | 6 | 0 | 0 |
| Hub | 2 | 1 | 1 |

Table 4.16: Number of elements in common for example 3

Even though we have only two nodes in common with PageRank in this example, note that they are among the top three ranked nodes. This implies that our method and PageRank do agree on the most important top nodes. Also notice that HITS and the eigenvector approach agree on top four and two nodes respectively for authority and hub ranking.

Example 4: Slashdot social network - Slashdot is a technology-related news website know for its specific user community. The website features user-submitted and editor-evaluated current primarily technology oriented news. In 2002 Slashdot introduced the Slashdot Zoo feature which allows users to tag each other as friends or foes. The network contains friend/foe links between the users of Slashdot.

| Authority Ranking | | | | | | | |
|---|---|---|---|---|---|---|---|
| New Approach | | PageRank | | HITS | | Eigenvector | |
| Node | Score | Node | Score | Node | Score | Node | Score |
| 70355 | 4.3438e-06 | 70328 | 0.2126 | 47 | 0.1136 | 47 | 0.1173 |
| 70354 | 3.8454e-06 | 70355 | 0.1942 | 2479 | 0.1070 | 2479 | 0.1165 |
| 70328 | 3.3611e-06 | 70354 | 0.1768 | 395 | 0.1059 | 4669 | 0.1152 |
| 69948 | 1.9369e-06 | 69948 | 0.1072 | 4669 | 0.1042 | 395 | 0.1141 |
| 65219 | 7.2635e-07 | 65219 | 0.0550 | 193 | 0.1025 | 193 | 0.1046 |
| 1 | 5.8329e-13 | 31818 | 0.0444 | 338 | 0.0945 | 1710 | 0.0965 |
| 395 | 1.4025e-15 | 48538 | 0.0027 | 1710 | 0.0943 | 338 | 0.0962 |
| 401 | 9.6160e-16 | 8092 | 0.0027 | 4978 | 0.0881 | 4978 | 0.0888 |
| 4669 | 8.3182e-16 | 3132 | 0.0023 | 2525 | 0.0820 | 2525 | 0.0845 |
| 2479 | 7.6943e-16 | 35100 | 0.0022 | 321 | 0.0817 | 321 | 0.0841 |

Table 4.17: Authority ranking for example 4

| Hub Ranking | | | | | | | |
|---|---|---|---|---|---|---|---|
| New Approach | | Reverse PageRank | | HITS | | Eigenvector | |
| Node | Score | Node | Score | Node | Score | Node | Score |
| 69948 | 4.9881e-06 | 70328 | 0.2055 | 47 | 0.1218 | 47 | 0.1221 |
| 70355 | 4.4344e-06 | 70355 | 0.1878 | 395 | 0.1217 | 395 | 0.1146 |
| 26105 | 2.1227e-07 | 70354 | 0.1710 | 4669 | 0.1210 | 2479 | 0.1132 |
| 50067 | 2.1227e-07 | 69948 | 0.1037 | 2479 | 0.1204 | 4669 | 0.1100 |
| 48538 | 2.1227e-07 | 65219 | 0.0533 | 193 | 0.1016 | 193 | 0.1063 |
| 8092 | 2.1227e-07 | 31818 | 0.0468 | 338 | 0.0955 | 338 | 0.1002 |
| 3132 | 2.1227e-07 | 8092 | 0.0036 | 1710 | 0.936 | 1710 | 0.0985 |
| 35100 | 2.1227e-07 | 48530 | 0.0036 | 4978 | 0.0874 | 4978 | 0.0915 |
| 9045 | 2.1227e-07 | 56540 | 0.0027 | 2473 | 0.0833 | 321 | 0.0869 |
| 8341 | 2.1227e-07 | 48539 | 0.0027 | 2525 | 0.0832 | 2525 | 0.0865 |

Table 4.18: Hub ranking for example 4

| Number of Intersection Between Two Sets | | | |
|---|---|---|---|
| | New Approach and PageRank | New Approach and HITS | New Approach and eigenvector |
| Authority | 5 | 1 | 1 |
| Hub | 3 | 0 | 0 |

Table 4.19: Number of elements in common for example 4

Example 5: Gnutella peer-to-peer network - A sequence of snapshots of the Gnutella peer-to-peer file sharing network from August 2002. Nodes represent hosts in the Gnutella network topology and edges represent connections between the Gnutella hosts.

| Authority Ranking | | | | | | | |
|---|---|---|---|---|---|---|---|
| New Approach | | PageRank | | HITS | | Eigenvector | |
| Node | Score | Node | Score | Node | Score | Node | Score |
| 7165 | 0.0047 | 5720 | 0.0028 | - | - | 703 | 0.0491 |
| 7295 | 0.0045 | 7165 | 0.0027 | - | - | 112 | 0.0480 |
| 8438 | 0.0040 | 2371 | 0.0026 | - | - | 7840 | 0.0456 |
| 3091 | 0.0040 | 7295 | 0.0026 | - | - | 7390 | 0.0387 |
| 4447 | 0.0033 | 6901 | 0.0024 | - | - | 6926 | 0.0379 |
| 5793 | 0.0030 | 8438 | 0.0023 | - | - | 5239 | 0.0367 |
| 7171 | 0.0028 | 3091 | 0.0023 | - | - | 872 | 0.0365 |
| 7538 | 0.0027 | 6779 | 0.0019 | - | - | 7135 | 0.0361 |
| 6348 | 0.0025 | 4447 | 0.0019 | - | - | 3885 | 0.0360 |
| 7220 | 0.0025 | 3530 | 0.0018 | - | - | 6956 | 0.0353 |

Table 4.20: Authority ranking for example 5

| Hub Ranking | | | | | | | |
|---|---|---|---|---|---|---|---|
| New Approach | | Reverse PageRank | | HITS | | Eigenvector | |
| Node | Score | Node | Score | Node | Score | Node | Score |
| 1443 | 0.0040 | 518 | 0.0026 | - | - | 641 | 0.0798 |
| 629 | 0.0030 | 1443 | 0.0016 | - | - | 1383 | 0.0716 |
| 415 | 0.0030 | 238 | 0.0016 | - | - | 830 | 0.0695 |
| 1344 | 0.0027 | 2412 | 0.0014 | - | - | 390 | 0.0687 |
| 183 | 0.0026 | 390 | 0.0013 | - | - | 562 | 0.0682 |
| 610 | 0.0025 | 1264 | 0.0012 | - | - | 629 | 0.0668 |
| 180 | 0.0024 | 915 | 0.0012 | - | - | 1208 | 0.0664 |
| 6229 | 0.0023 | 629 | 0.0012 | - | - | 55 | 0.0658 |
| 371 | 0.0023 | 415 | 0.0012 | - | - | 1253 | 0.0639 |
| 1704 | 0.0020 | 1599 | 0.0012 | - | - | 1145 | 0.0620 |

Table 4.21: Hub ranking for example 5

| Number of Intersection Between Two Sets | | | |
|---|---|---|---|
| | New Approach and PageRank | New Approach and HITS | New Approach and eigenvector |
| Authority | 5 | NA | 0 |
| Hub | 3 | NA | 0 |

Table 4.22: Number of elements in common for example 5

# Chapter 5

# Conclusion And Discussion

According to the results in the previous chapter, we can see that the new approach is similar to PageRank and Reverse PageRank, but is very different from HITS and the dominant eigenvector approach. However, we see that the latter two methods yield similar results, which imply that the four methods discussed in this thesis can be divided into two groups. Since the new approach and PageRank(Reverse PageRank) are both based on graph Laplacians, even though the former depends on non-normalized ones and the latter on normalized ones, they rank nodes in similar fashions. Since HITS and the dominant eigenvector approach are both based on the adjacency matrix $A$, it is reasonable that their results agree with each other.

Even though the new approach shares some commonalities with PageRank, it is also different from it and thus sheds light on some new information that PageRank cannot provide. One interpretation of the new approach is in terns of advection, where each node carries some information and moves from nodes to nodes until it reaches an equilibrium or steady state. At this equilibrium, the score computed is a measure of how well the information is advected to one node from all the others (for authorities) and from one node to all the others (for hubs). In other words, the score computed represents how influential a node is in terms of spreading out information, or gathering information.

However, this method does have some limitations [5]. We say that a graph is balanced if its in-degrees are equal to its out-degrees. Recall from the second chapter where we talked about advection and consensus dynamics, the limit for

$$\dot{\mathbf{x}} = -L_{in}\mathbf{x},\ \mathbf{x}(0) = \mathbf{x}_0$$

goes to

$$\lim_{t\to\infty}\mathbf{x}(t) = \frac{1}{\sqrt{n}}(\mathbf{q}_1^T\mathbf{x}_0)\mathbf{q}_1 \text{ where } L_{in}\mathbf{q}_1 = \mathbf{0}, \|\mathbf{q}_1\|_2 = 1.$$

However, if the graph is balanced,

$$\mathbf{x}(t) \to ((1/n)\mathbf{1}^T\mathbf{x}_0)\mathbf{1}(= \text{constant}).$$

Therefore, in this case, $L_{out} = L_{in}$, which implies that the advection dynamics is equivalent to the consensus dynamics. Hence, $\mathbf{x}(t)$ tends to a constant vector as $t \to \infty$. Hence, if a graph is a balanced one, then no ranking is possible.

# Bibliography


[1] M. Benzi, E. Estrada, and C. Klymko, *Ranking hubs and authorities using matrix functions*, Linear Algebra Appl. 438 (2013) 2447-2474.

[2] M. Benzi, and C. Klymko, *Total communicability as a centrality measure*, Journal of Complex Networks (2013) 1, 124-149. doi: 10.1093/comnet/cnt007.

[3] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Acedemic Press, New York, 1979.

[4] D. Boley, G. Ranjan, and Z. Zhang, *Commute Times for a Directed Graph using an Asymmetric Laplacian*, Linear Algebra Appl. 435 (2011) 224-242.

[5] A. Chapman, and M. Mesbahi, *Advection on Graphs*, 50th IEEE Conference on Decision and Control and European Control Conference, 2011, pp. 1461-1466.

[6] F. R. K. Chung, *Laplacians and the Cheeger inequality for directed graphs*, Annals of Combinatorics, 9 (2005), pp. 1-19.

[7] F. Chung, *The heat kernel as the pagerank of a graph*, PNAS Vol. 104, No. 50: 19735-19740.

[8] E. Estrada, and P. A. Knight, *A First Course in Network Theory*, Oxford University Press, 2015.

[9] Y. Li, and Z. Zhang, *Digraph Laplacian and the Degree of Asymmetry*, Internet Mathematics Vol. 8, No. 4: 381-401.

[10] Z. Li, G. Wen, Z. Duan, and W. Ren, *Designing Fully Distributed Consensus Protocols for Linear Multi-agent Systems with Directed Graphs*, pp. 1-16.

[11] C. W. Wu, *Synchronization in networks of nonlinear dynamical systems coupled via a directed graph*, Nonlinearity 18 (2005) 1057-1064, doi: 10.1088/0951-7715/18/3/007.