Mining & Learning on Graphs

Network Analysis

Yu Wang, Ph.D.
Assistant Professor
Computer and Information Science
University of Oregon
CS 410/510 - Fall 2024

Network Centrality

- Which nodes in the graph are "important"?

Summary

- Graph-theoretic measures
- Basic measures of centrality/prestige
  - Mostly from a social perspective
- Path-based
  - Closeness, Betweenness, Katz
- Eigenvector-based
  - Eigenvector, Katz, PageRank
- Others
  - Hubs and Authorities, Goodness and Fairness

Eigenvector Centrality

Phillip Bonacich, 1972.

- Importance of a node depends on the importance of its neighbors (note this is recursive)

$$c_i \leftarrow \sum_j \mathbf{A}_{ij} c_j$$

What is wrong with this type of recursive solution?

Eigenvector Centrality

- Importance of a node depends on the importance of its neighbors (note this is recursive)

$$c_i \leftarrow \sum_j \mathbf{A}_{ij} c_j$$

$$c_i = \alpha \sum_j \mathbf{A}_{ij} c_j$$

We can pick some value $\alpha < 1$

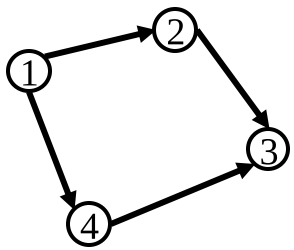
Network Centrality – Eigenvector Centrality

- Importance of a node depends on the importance of its neighbors (note this is recursive)

$$c_i \leftarrow \sum_j \mathbf{A}_{ij} c_j$$

$$c_i = \frac{1}{\lambda} \sum_j \mathbf{A}_{ij} c_j$$

$$\mathbf{A}\mathbf{c} = \lambda \boldsymbol{c}$$



$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0.1 \\ 0.2 \\ 0.5 \\ 0.3 \end{bmatrix} \longrightarrow \begin{bmatrix} 0.0 \\ 0.1 \\ 0.5 \\ 0.1 \end{bmatrix}$$

Eigen-vector Problem

Network Centrality – Eigenvector Centrality

Adjacency Matrix

$$\mathbf{A}c = \lambda c$$

We want our centrality value to be positive

Theorem 38.1 (Perron-Frobenius Theorem)

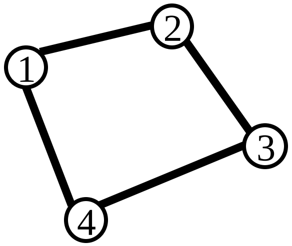
If a matrix $A \geq 0$ then,

Moreover if A is also irreducible then,

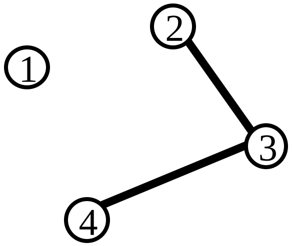
4. the eigenvector v associated with the eigenvalue $r(A)$ is strictly positive.
5. there exists no other positive eigenvector v (except scalar multiples of v) associated with $r(A)$.

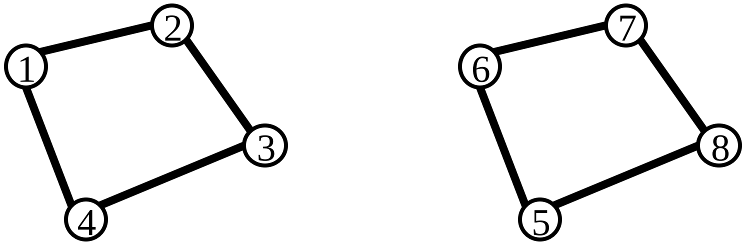
Network Centrality – Eigenvector Centrality

Connected



Disconnected





We can compute centrality in each connected component

Network Centrality – Eigenvector Centrality

Adjacency Matrix

$$Ac = \lambda c$$

We want our centrality matrix to be positive

If our adjacency matrix is not strongly connected, we can first partition our graph and only look at those strongly connected

Theorem 38.1 (Perron-Frobenius Theorem)

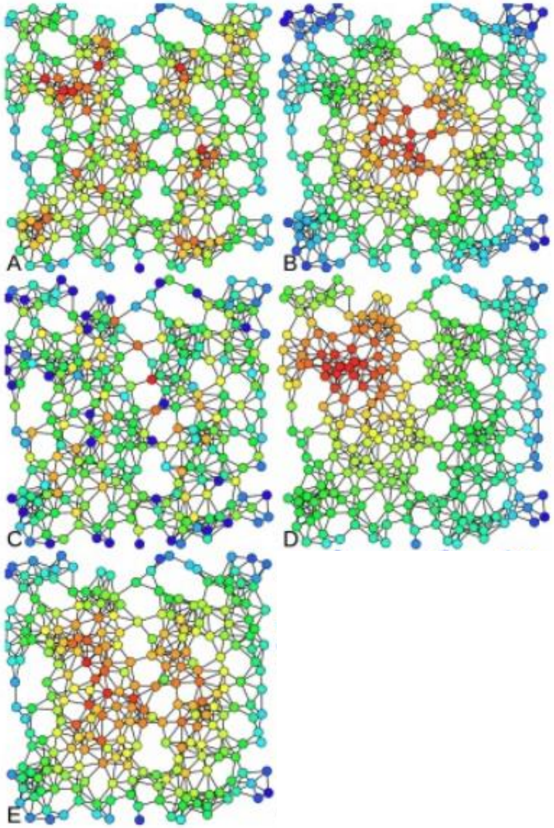
If a matrix $A \geq 0$ then,

Moreover if A is also irreducible then,

4. the eigenvector v associated with the eigenvalue $r(A)$ is strictly positive.
5. there exists no other positive eigenvector v (except scalar multiples of v) associated with $r(A)$.

Network Centrality – Eigenvector Centrality

Comparing with previous methods



A. Degree
B. Closeness
C. Betweenness
D. Eigenvector
E. Katz

PageRank

Google

pagerank

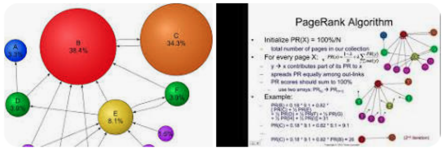
All Images Videos News Shopping Forums Web ⋮ More Tools

This search may be relevant to recent activity: pagerank ***centrality*** Your Search activity | Feedback

PageRank (PR) is an algorithm used by Google Search to rank web pages in their search engine results. It is named after both the term "web page" and co-founder Larry Page. PageRank is a way of measuring the importance of website pages.

Wikipedia
https://en.wikipedia.org › wiki › PageRank ⋮

PageRank - Wikipedia



About featured snippets • Feedback

People also ask ⋮

What is PageRank centrality?

Is PageRank still a ranking signal for Google?

How do you calculate PageRank?

What replaced PageRank?

Feedback

PageRank ⋮

PageRank is an algorithm used by Google Search to rank web pages in their search engine results. It is named after both the term "web page" and co-founder Larry Page. PageRank is a way of measuring the importance of website pages. Wikipedia

Feedback

Scholarly articles for **pagerank**

Inside **pagerank** - Bianchini - Cited by 747

A survey on **PageRank** computing - Berkhin - Cited by 620

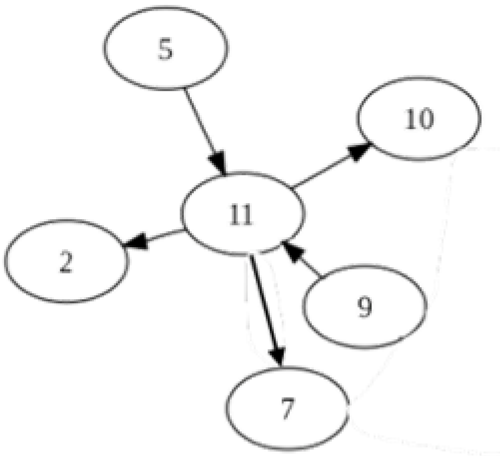
Efficient computation of **PageRank** - Haveliwala - Cited by 571

Ranking algorithm to determine the importance of the webpage

Ranking in Directed Graph

- Iterative Update Method
  - (similarity to Eigenvector centrality)

$$c_i \leftarrow \sum_{j \in N(i)} c_j = \sum_j \mathbf{A}_{ji} c_j$$



$$c_i^{t+1} = \sum_j \mathbf{A}_{ji} c_j^t$$
$$\mathbf{c}^{t+1} = \mathbf{A}^{\mathrm{T}} \mathbf{c}^{\mathrm{t}}$$
$$\mathbf{c}^0 = some\ initial\ vector$$

Issue:

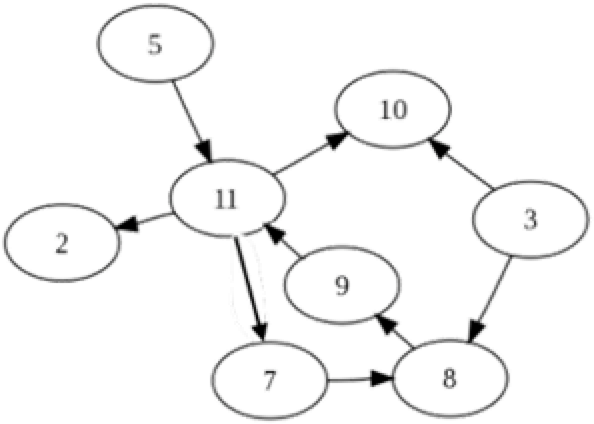
Keeps getting larger...

PageRank – Absorbing and Sourcing Node Issue

Ranking Directed Graph Problems

- Absorbing Nodes
- Source Nodes
- Cycles

$$\mathbf{c}^{t+1} = \mathbf{A}\mathbf{c}^{t}$$
$$\mathbf{c}^{0} = some\ initial\ vector$$

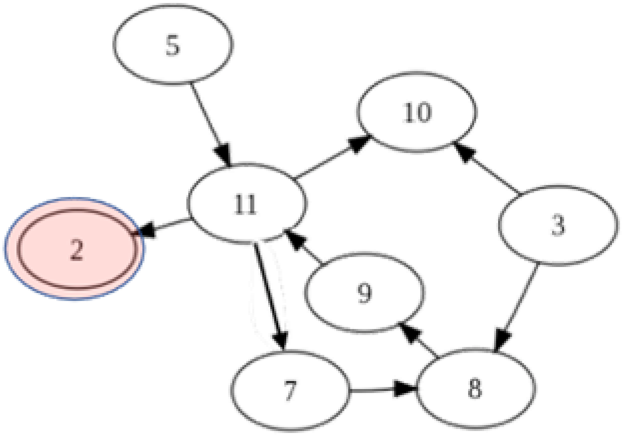


Is there any stable solution?

PageRank – Absorbing and Sourcing Node Issue

Absorbing Nodes

- Keeps receiving centrality from node 11 and does not pass it along

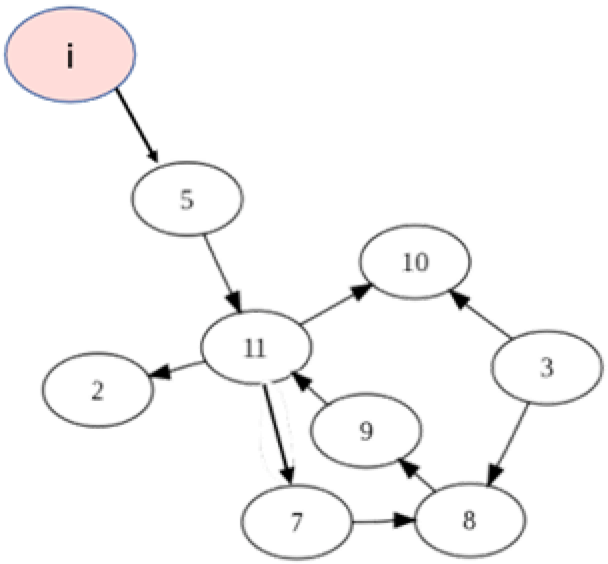


- Keeps increasing

PageRank – Absorbing and Sourcing Node Issue

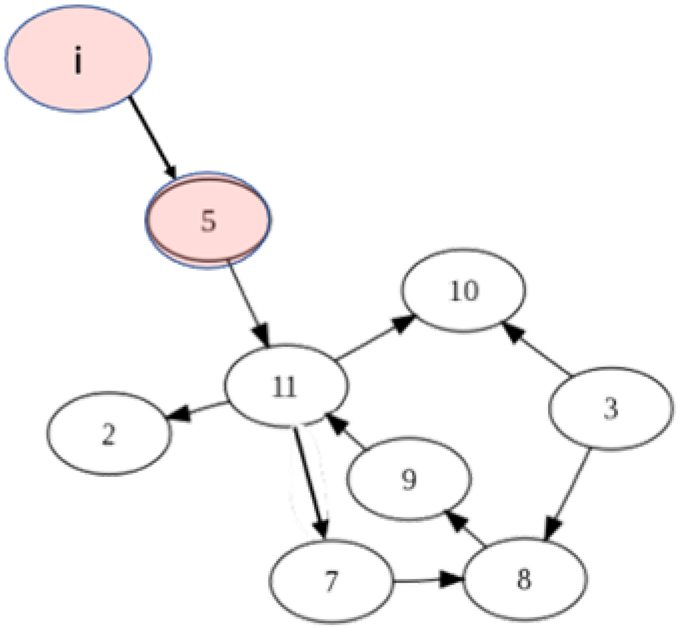
Source Nodes

- Node i initially passes its centrality to node 5
- Node i has no incoming links so centrality goes to 0



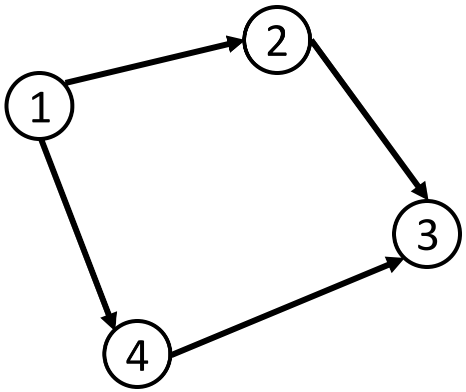
Source Nodes

- Node i initially passes its centrality to node 5
- Node i has no incoming links so centrality goes to 0
- Node 5 only receives centrality from node i, so it also then drops to 0 once node i is 0
- Cascading problem



PageRank – Absorbing and Sourcing Node Issue

How to remove absorbing or sourcing node?

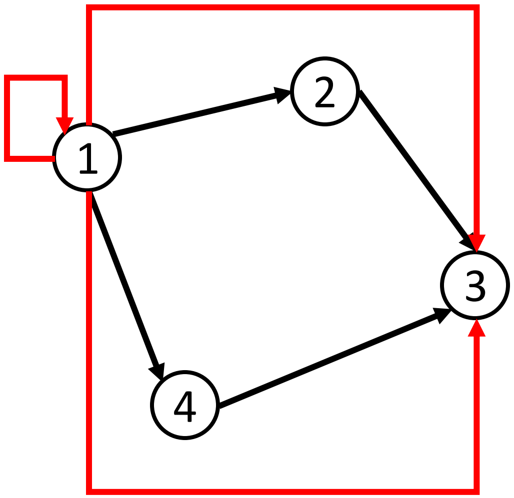


A

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

PageRank – Absorbing and Sourcing Node Issue

How to remove absorbing or sourcing node?

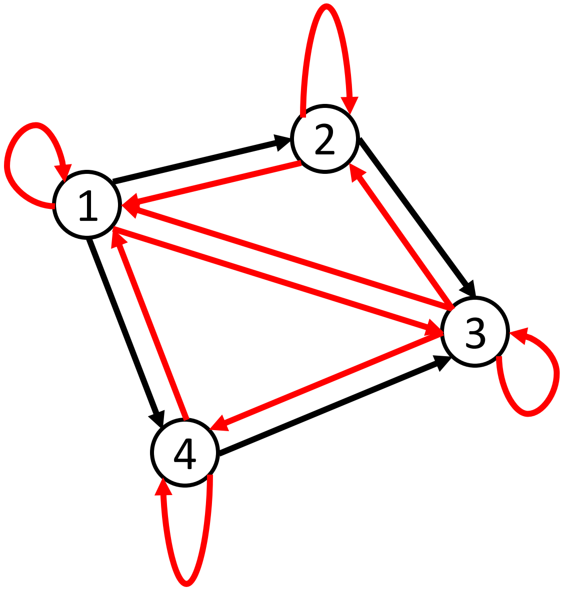


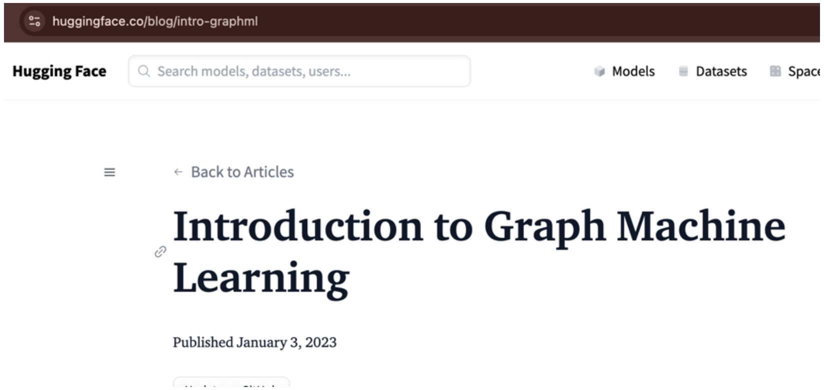
$$\mathbf{A} = \mathbf{A} + \mathbf{s}^{T}\mathbf{e}$$

$$\begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} + \underbrace{\begin{bmatrix} \mathbf{1} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} [\mathbf{1}\ \mathbf{1}\ \mathbf{1}\ \mathbf{1}]}_{\begin{bmatrix} \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}}$$

PageRank – Strongly Connected Component Issue

How to address strongly connected issue?



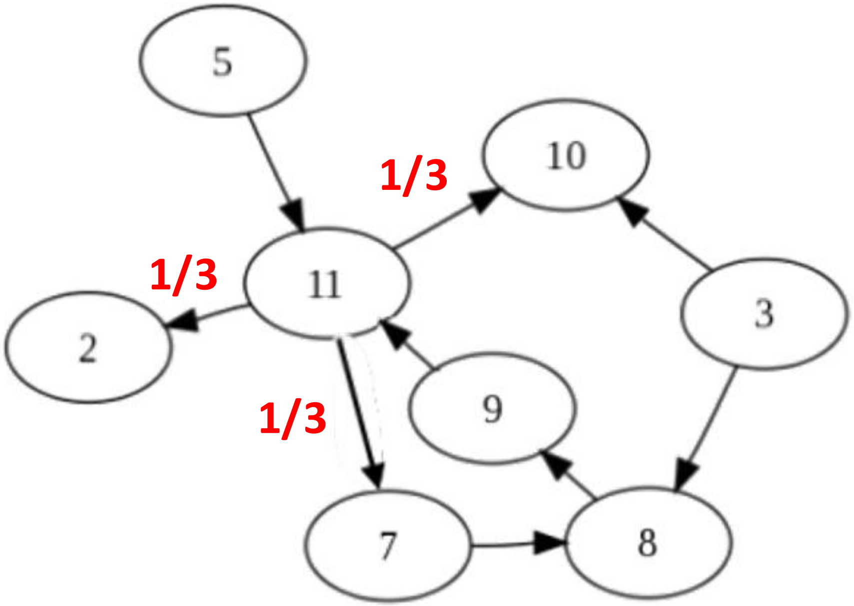


$$\mathbf{A} = \mathbf{A} + \mathbf{e}^{\mathrm{T}}\mathbf{e}$$

$$\begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} [1\ 1\ 1\ 1]$$

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

PageRank – Stochastic Issue



$$\mathbf{A} = \mathbf{D}^{-1}\mathbf{A}$$

$$\begin{bmatrix} 2^{-1} & 0 & 0 & 0 \\ 0 & 1^{-1} & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1^{-1} \end{bmatrix} \quad \begin{bmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

PageRank

Transition matrix:

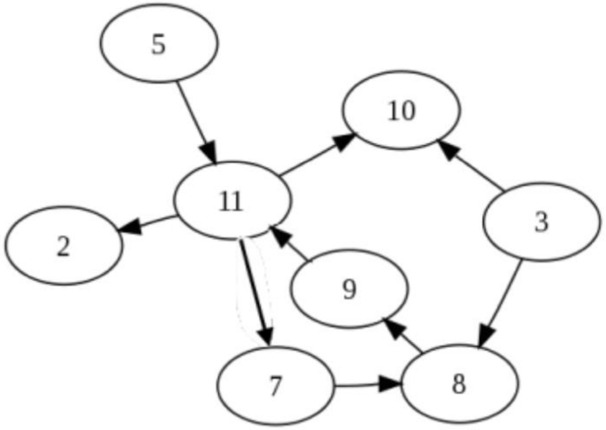
$$\mathbf{P} = \mathbf{D}^{-1}\mathbf{A}$$

Stochastic matrix:

$$\mathbf{P}' = \mathbf{P} + \frac{\mathbf{s}\mathbf{e}^T}{n}$$

PageRank matrix:

$$\mathbf{P}'' = \alpha\mathbf{P}' + (1-\alpha)\frac{\mathbf{e}\mathbf{e}^T}{n}$$



Perron-Frobenius Theorem

- Given a matrix that is
  - Stochastic (non-negative and rows sum to one)
  - Irreducible (i.e., strongly connected)

Then

We have a solution

$$\mathbf{c}^{t+1} = \mathbf{A}\mathbf{c}^{t}$$

$$\bar{\pi}\mathbf{P} = \bar{\pi}, \quad \text{where} \quad ||\bar{\pi}||_1 = 1$$

i.e, a stationary distribution of Markov chain

PageRank

$$\boldsymbol{c}^{t+1} = \mathbf{A}\mathbf{c}^{t}$$

$$\mathbf{A}'' = \alpha\mathbf{A}' + (1-\alpha)\frac{\mathbf{e}\mathbf{e}^{\mathrm{T}}}{n} = \alpha\mathbf{D}^{-\mathbf{1}}\mathbf{A}' + (1-\alpha)\frac{\mathbf{e}\mathbf{e}^{\mathrm{T}}}{n}$$

$$\boldsymbol{c}^{2} = \mathbf{A}''\mathbf{c}^{1}$$

$$\boldsymbol{c}^{3} = \mathbf{A}''\mathbf{c}^{2}$$

$$\vdots$$

$$\boldsymbol{c}^{20} = \mathbf{A}''\mathbf{c}^{19}$$

Converge →

$$\boldsymbol{c}^{n} = \boldsymbol{c}^{n-1}$$

Any Question?