Advanced Topics in Network Science

Lecture 06: Centrality

Sadamori Kojaku

What to Learn 📚

- What is centrality in networks? 🕸️
- How to operationalize centrality? 🔢
- How to find centrality in networks? 🔍
- Limitations of centrality ⚠️

Keywords: degree centrality, closeness centrality, betweenness centrality, eigenvector centrality, PageRank, Katz centrality, HITS, random walk

✍️ Pen and paper for centralities

What is centrality?

- A measure of how important/central a node is in a network
- Many definitions of centrality
- Let's consider some key ideas from history

Example from histories

The Golden Milestone (Milliarium Aureum) 🏛️

- Located in the Roman Forum
- Built by Emperor Augustus (1st emperor of Rome)
- Symbolized Rome as the center of the empire
- "All roads lead to Rome" is a reference to the Golden Milestone

Idea 1

A central node ~ A node that is connected to other nodes by a short distance 🤔

Closeness Centrality 🏃‍♀️

- Measure of how close a node is to all others.
- Centrality of a node i, denoted by c_i, is defined as

$$c_i = \frac{1}{\overline{d_i}}, \quad \overline{d_i} = \frac{1}{N-1}\sum_{j=1}^{N} d_{ij}$$

where

- d_{ij} is the shortest path length from node i to node j

Other Distance-based Centrality 🛤️

Harmonic Centrality

- Adjusts closeness for disconnected networks

$$c_i = \sum_{j \neq i} \frac{1}{d_{ij}}$$

Eccentricity Centrality

- Based on farthest distance from a node

$$c_i = \frac{1}{\max_j d_{ij}}$$

Betweenness Centrality 🌉

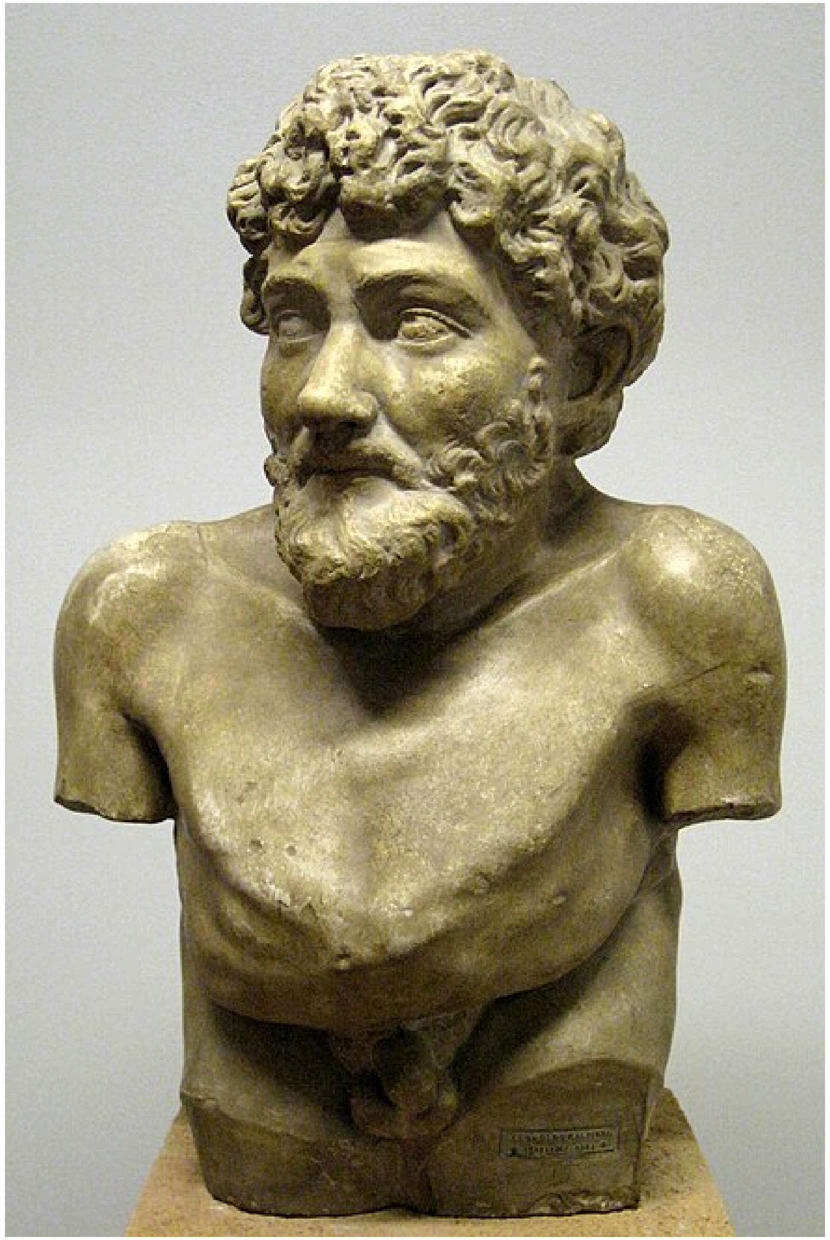
- Measures importance based on shortest paths

$$c_i = \sum_{j<k} \frac{\sigma_{jk}(i)}{\sigma_{jk}}$$

- $\sigma_{jk}(i)$: number passing through node i
- σ_{jk}: number of shortest paths between j and k
- If $\sigma_{jk} = 1$ for all pairs (i.e., all pairs are connected by a single shortest path), c_i is the count of shortest paths through node i.

"A man is known by the company he keeps" 🤝

- Ancient Greek wisdom (Aesop)
- How do we define centrality based on the idea?



Idea 2

A node is important if it is connected to important nodes

Eigenvector Centrality 🌟

- **Key idea**: Important nodes are connected to other important nodes
- Let c_i be a *tentative* importance score of node i
- Let the importance score be proportional to the scores of nodes it is connected to

$$c_i = \lambda \sum_j A_{ij} c_j$$

or in matrix form

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

- **Question**: Can we solve this equation? If so, what is the solution 🤔?

Answer

- $\mathbf{c} = \lambda \mathbf{A} \mathbf{c}$ is an eigenvector equation
- The solution is an eigenvector of $\mathbf{A}$
- But here is a problem. The solution is not unique. Any eigenvector corresponding is a solution.
- Which one to choose?
  - We want the importance score to be all positive.
  - There is always one such eigenvector, the principal eigenvector (Peron-Frobenius theorem)

HITS Centrality 🎯

- Extension of eigenvector centrality to directed networks
- Introduces two kinds of importance score: **hub** and **authority** scores
- **Hub**: Connected from many authorities
    - Hub score: $x_i = \lambda_x \sum_j A_{ji} y_j$
    - $A_{ij} = 1$ or 0 if there is a directed edge from i to j or not
- **Authority**: Connected to many hubs
    - Authority score: $y_i = \lambda_y \sum_j A_{ij} x_j$
- Matrix form: $\mathbf{x} = \lambda_x \mathbf{A}^T \mathbf{y}, \quad \mathbf{y} = \lambda_y \mathbf{A} \mathbf{x}$
- **Question**: What are the solution to this equation?

Answer

- The eigenvectors of $\mathbf{A}^T\mathbf{A}$ and $\mathbf{A}\mathbf{A}^T$ are the solutions
- Take the principal eigenvector of $\mathbf{A}\mathbf{A}^T$ as the authority score and the principal eigenvector of $\mathbf{A}^T\mathbf{A}$ as the hub score.

Limitation of Eigenvector Centrality ⚠️

- Tends to concentrate importance on few well-connected nodes and may lead to most nodes having near-zero centrality scores
- Can underemphasize importance of less connected nodes

Katz Centrality: Addressing Limitations 🛠️

- Adds a base level of importance to all nodes

$$c_i = \beta + \lambda \sum_j A_{ij} c_j$$

- β: base score given to all nodes
- Provides more balanced centrality scores
- **Question**: Can we solve this equation? If so, what is the solution 🤔?

PageRank 🌐

- A node (web page) is important if it is visited by many random surfer who are browsing the web by clicking on links at random.
- The suerfer can teleport to any node at random with a probability of β.
- PageRank of a node i is the probability that a random surfer is at node i after many steps, i.e., the solution to the following equation:

$$c_i = \underbrace{\frac{\beta}{N}}_{\text{teleport}} + \underbrace{(1-\beta)\sum_j \frac{A_{ji}}{d_j^{\text{out}}} c_j}_{\text{click on a link}}$$

- d_j^{out}: out-degree of node j
- A_{ji}/d_j^{out}: Probability of moving from node j to node i
- c_i: the probability that a random surfer is at node i after many steps
- **Question**: Can we solve this equation? If so, what is the solution 🤔?

Degree-based Centrality 🔢

- Simplest form of centrality
- Count of edges connected to a node
- $c_i = d_i = \sum_j A_{ij}$

Hands-on Coding Exercise 🎓

Centrality Computation in Python 🐍

```
# Degree centrality
g.degree()

# Closeness centrality
g.closeness()

# Betweenness centrality
g.betweenness()

# Eigenvector centrality
g.eigenvector_centrality()

# PageRank
g.personalized_pagerank()
```

Key Takeaways 🗝️

- Centrality ≠ Universal Importance
- Context is crucial for interpretation
- Different measures highlight various network aspects
- Consider network structure and dynamics
- Use centrality as a tool, not absolute truth 🧰🤔