Centrality in Networks

Introduction

Centrality measures are fundamental concepts in network analysis that help us identify the most important nodes in a network. While degree centrality is the most basic measure, there are several other centrality metrics that provide different perspectives on node importance.

Types of Centrality Measures

1. Degree Centrality

Degree centrality is the simplest centrality measure, calculated as the number of edges connected to a node.

Mathematical Definition: For node j, the degree d_j is calculated as:

$d_j = \sum_{i=1}^{n} A_{ij}$

where A_{ij} is the element of the adjacency matrix of the network.

Directed Networks: In directed networks, we distinguish between:

- **In-degree**: Number of edges pointing to a node $d_j^{(\text{in})} = \sum_{i=1}^{n} A_{ij}$
- **Out-degree**: Number of edges pointing from a node $d_i^{(out)} = \sum_{j=1}^{n} A_{ij}$

2. Closeness Centrality

Closeness centrality measures how close a node is to all other nodes in the network.

Mathematical Definition: For node j, closeness centrality c_j is calculated as:

$c_j = \frac{1}{\sum_{i=1}^{n} d_{ij}}$

where d_{ij} is the shortest path distance between node i and node j.

Interpretation:

- High closeness centrality means the node is close to all other nodes
- The node can reach all other nodes with short paths
- Useful for identifying nodes that can efficiently spread information

3. Betweenness Centrality

Betweenness centrality measures how often a node lies on the shortest paths between other nodes.

Mathematical Definition: For node j, betweenness centrality b_j is calculated as:

$$b_j = \sum_{s \neq t \neq j} \frac{\sigma_{st}(j)}{\sigma_{st}}$$

where:

- σ_{st} is the number of shortest paths between node s and node t
- $\sigma_{st}(j)$ is the number of shortest paths between node s and node t that pass through node j

Interpretation:

- High betweenness centrality means the node is a bridge between different parts of the network
- These nodes control the flow of information or resources
- Critical for network connectivity

4. Eigenvector Centrality

Eigenvector centrality considers not just the number of connections, but also the quality of those connections.

Mathematical Definition: For node j, eigenvector centrality x_j is calculated as:

$$x_j = \frac{1}{\lambda} \sum_{j=1}^{n} A_{ij} x_j$$

This can be written in matrix form as: $Av = \lambda v$

where:

- A_{ij} is the adjacency matrix element
- x_i is the eigenvector centrality of node i
- λ is the largest eigenvalue of the adjacency matrix

Interpretation:

- High eigenvector centrality means the node is connected to other important nodes
- Similar to degree centrality but with a feedback loop
- Useful for identifying influential nodes in social networks

5. PageRank Centrality

PageRank is a variant of eigenvector centrality that includes a damping factor to handle directed networks with sinks.

Mathematical Definition: For node i, PageRank centrality x_i is calculated as:

$$x_i = \frac{1-d}{N} + d \sum_{j=1}^{n} \frac{A_{ij}}{d_j^{out}} x_j$$

where:

- d is the damping factor (typically 0.85)
- N is the total number of nodes
- d_j^{out} is the out-degree of node j

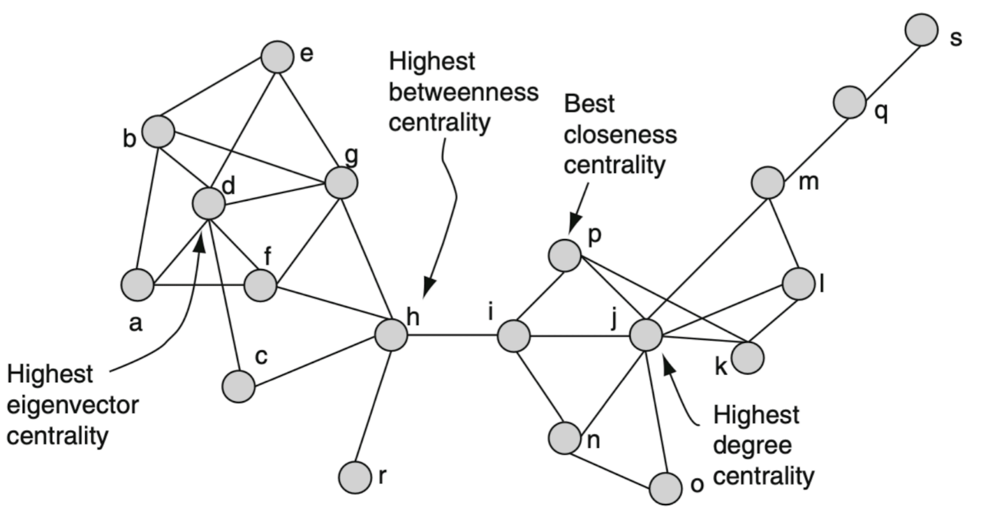
Interpretation:

- Combines random walk with teleportation
- Handles directed networks better than eigenvector centrality
- Originally developed for ranking web pages

Key Differences Between Centrality Measures

Centrality Type	What it measures	Best for identifying
Degree	Number of connections	Most connected nodes
Closeness	Average distance to others	Information spreaders
Betweenness	Bridge role in network	Network bottlenecks
Eigenvector	Connection to important nodes	Influential nodes
PageRank	Importance with damping	Web-like structures

D. Ortiz-Arroyo



Applications

1. **Social Networks**: Identifying influential users
2. **Transportation Networks**: Finding critical junctions
3. **Biological Networks**: Discovering key proteins or genes
4. **Information Networks**: Locating information hubs
5. **Economic Networks**: Finding key economic actors

Computational Considerations

- **Degree centrality**: $O(E)$ time complexity
- **Closeness centrality**: $O(V^2)$ for dense networks, $O(V+E)$ for sparse networks
- **Betweenness centrality**: $O(VE)$ for unweighted, $O(VE + V^2 \log V)$ for weighted
- **Eigenvector centrality**: $O(V^3)$ for exact computation, $O(V^2)$ for power iteration
- **PageRank**: $O(V^2)$ per iteration, typically converges in 10-20 iterations

References

- Eigenvector Centrality Video
- PageRank Video
- https://cambridgeuniversitypress.github.io/FirstCourseNetworkScience/