Social Network Analysis

Lecture 4: Centrality Measures (Including PageRank)

Dr. Hung-Nghiep Tran
nghiepth@uit.edu.vn

University of Information Technology, VNU-HCM, Vietnam
2025

Mục tiêu buổi học

- Understand the main types of centrality: *degree, closeness, betweenness, eigenvector, PageRank, HITS.*
- Distinguish between geometric, path-based, and spectral measures.
- Interpret what centralities capture in real-world networks, e.g., influence, brokerage, visibility.
- Explain PageRank as a random walk with teleportation and its motivation from Web search.
- Implement and compute centrality measures using networkx.

Câu hỏi kiểm tra trước

Quiz

1. Which of the following is a degree-based measure?
A. PageRank B. Closeness Centrality C. Degree Centrality D. Betweenness Centrality

2. Betweenness centrality measures…
A. Number of neighbors a node has B. How often a node lies on shortest paths
C. Eigenvalue of adjacency matrix D. Distance to all other nodes

3. What does PageRank add to eigenvector centrality?
A. Random teleportation B. Larger degree weight C. BFS search D. Power iteration

4. Closeness centrality is high when…
A. A node has many direct edges B. A node is close (short paths) to all others
C. A node is in a spider trap D. A node has high out-degree

5. Which centrality is most useful for finding “bridges” in a network?
A. Degree B. Betweenness C. Eigenvector D. HITS

Quiz

1. Which of the following is a degree-based measure?
A. PageRank B. Closeness Centrality **C. Degree Centrality** D. Betweenness Centrality

2. Betweenness centrality measures…
A. Number of neighbors a node has **B. How often a node lies on shortest paths**
C. Eigenvalue of adjacency matrix D. Distance to all other nodes

3. What does PageRank add to eigenvector centrality?
A. Random teleportation B. Larger degree weight C. BFS search D. Power iteration

4. Closeness centrality is high when…
A. A node has many direct edges **B. A node is close (short paths) to all others**
C. A node is in a spider trap D. A node has high out-degree

5. Which centrality is most useful for finding “bridges” in a network?
A. Degree **B. Betweenness** C. Eigenvector D. HITS

Đọc và diễn giải

Tài liệu đọc

- Reading:
  - Network Centrality
  - PageRank

Câu hỏi kiểm tra sau

Quiz

1. Why might eigenvector centrality assign low score to an important but isolated hub?
A. It doesn't count paths B. It depends on being connected to other central nodes
C. It ignores out-degree D. It uses harmonic mean

2. Which two problems does PageRank fix compared to plain eigenvector centrality?
A. Dead ends & spider traps B. High clustering & short paths
C. Duplicate edges & self-loops D. Long computation & memory

3. If teleportation probability = 0, PageRank reduces to…
A. Degree centrality B. Eigenvector centrality C. Betweenness centrality D. Closeness centrality

4. Which application best fits betweenness centrality?
A. Ranking sports players B. Identifying bridges in communication networks
C. Detecting influential websites D. Measuring assortativity

5. In HITS, "hubs" and "authorities" differ because…
A. Hubs are in-degree, authorities are out-degree B. Hubs point to good authorities; authorities are pointed to by hubs
C. Hubs use teleportation, authorities don't D. Hubs are based on closeness

Quiz

1. Why might eigenvector centrality assign low score to an important but isolated hub?
A. It doesn't count paths **B. It depends on being connected to other central nodes**
C. It ignores out-degree D. It uses harmonic mean

2. Which two problems does PageRank fix compared to plain eigenvector centrality?
A. Dead ends & spider traps B. High clustering & short paths
C. Duplicate edges & self-loops D. Long computation & memory

3. If teleportation probability = 0, PageRank reduces to…
A. Degree centrality **B. Eigenvector centrality** C. Betweenness centrality D. Closeness centrality

4. Which application best fits betweenness centrality?
A. Ranking sports players **B. Identifying bridges in communication networks**
C. Detecting influential websites D. Measuring assortativity

5. In HITS, "hubs" and "authorities" differ because…
A. Hubs are in-degree, authorities are out-degree **B. Hubs point to good authorities; authorities are pointed to by hubs**
C. Hubs use teleportation, authorities don't D. Hubs are based on closeness

Thảo luận

Chuẩn bị tuần trước

- Muddiest point: “Điểm nào mù mờ nhất sau khi đọc? (< 50 từ)”

Chuẩn bị tuần trước

- Trả lời ngắn câu hỏi định hướng:
  - Define degree centrality, betweenness centrality, and eigenvector centrality?
    - Degree: counts immediate neighbors.
    - Betweenness: measures brokerage by counting shortest paths through a node.
    - Eigenvector: influence based on being connected to other influential nodes, recursively.
  - Why might PageRank favor nodes with many incoming links from important neighbors?
    - Because votes are weighted: a link from an important page contributes more weight than from a random page.
  - Applications of betweenness centrality?
    - Identifying bridges in social networks, bottlenecks in transportation, or critical routers in communication.
  - Limitation of degree centrality?
    - Ignores global structure and influence; a node with many low-quality neighbors may seem “important.”
  - PageRank as random walk with teleportation?
    - A random surfer follows links with probability b, and teleports to any page with probability 1-b. This ensures convergence and avoids spider traps.

Hands-on lab

Setup

- Install python
  - with Anaconda
- Install IDE:
  - (Microsoft) vscode
  - Jupyter Notebook
- Install networkx
  - with pip
- Install gephi

Chuẩn bị cho tuần tới

Chuẩn bị trước tuần sau

- Reading:
  - (Stanford CS224W L15) Network Centrality
  - (Stanford CS224W L3) PageRank

- Câu hỏi định hướng (trả lời ngắn):
  - What does “community” mean in a network context?
  - Why are modularity-based methods popular for detecting communities?
  - Compare Louvain and spectral clustering: main intuition behind each?
  - What kind of “ground truth” communities might exist in real datasets (e.g., social, biological, citation)?
  - What challenges arise when evaluating community detection algorithms?

- Muddiest point: “Điểm nào mù mờ nhất sau khi đọc? (< 50 từ)”

Thank you for listening

Q & A