Practical SocialNetwork Course - LAB 5.1 (Theory)

Khang Nguyen-Hoang

Ngày 30 tháng 12 năm 2024

1 Analysis of Basic Network Metrics

1.1 Average Degree

Definition: The average degree represents the average number of connections (edges) per node in a network. It provides insight into how interconnected the nodes are.

$$\text{Average Degree} = \frac{2m}{n}$$

where:

- m: Total number of edges in the network.
- n: Total number of nodes in the network.

Scope:

- Suitable for analyzing network density and general connectivity.
- Effective in comparing sparsity across networks.

What is a *good* value?

- A high average degree implies dense interconnections, often found in smaller, cohesive networks.
- A low average degree typically indicates sparse, large-scale networks like social networks.

1.2 Network Diameter

Definition: The diameter of a network is the maximum shortest path between any two nodes. It measures the longest distance information has to travel in the network.

$$\text{Diameter} = \max_{u,v \in V} d(u, v)$$

where $d(u, v)$ is the shortest path between nodes u and v. **Scope:**

- Useful in measuring network efficiency and resilience.
- Applicable in small-world and large-scale networks.

What is a *good* value?

- A small diameter indicates efficient communication (e.g., small-world networks).
- A large diameter implies decentralized and possibly fragmented networks.

1.3 Graph Density

Definition: Graph density measures the proportion of actual edges to the total possible edges in the network.

$$\text{Density} = \frac{2m}{n(n-1)}$$

For directed graphs:

$$\text{Density} = \frac{m}{n(n-1)}$$

Scope:

- Ideal for comparing networks of varying sizes.
- Often used to evaluate community structures and sparsity.

What is a *good* value?

- High density suggests strong interconnections (e.g., in closed groups).
- Low density is common in large-scale networks, such as social media graphs.

1.4 Connected Components

Definition: A connected component is a subgraph in which any two nodes are connected by a path, and no additional node is connected to the subgraph. **Scope:**

- Applicable in analyzing fragmentation and disconnected groups.
- Helps identify isolated communities.

What is a *good* value?

- A single connected component indicates high cohesiveness.
- Multiple components suggest fragmentation and decentralization.

1.5 Average Path Length

Definition: The average path length is the mean of the shortest paths between all pairs of nodes.

$$\text{APL} = \frac{\sum_{u \neq v} d(u, v)}{n(n-1)}$$

Scope:

- Relevant in understanding network efficiency.
- Commonly used in transportation and communication networks.

What is a *good* value?

- A low APL indicates high efficiency in information dissemination.
- A high APL reflects longer communication paths and potential inefficiencies.

1.6 Average Clustering Coefficient

Definition: Measures the likelihood that two neighbors of a node are also connected. For node v:

$$C(v) = \frac{2 \cdot \text{number of triangles through } v}{\deg(v)(\deg(v) - 1)}$$

The average clustering coefficient is:

$$C = \frac{1}{n} \sum_{v \in V} C(v)$$

Scope:

- Ideal for detecting tightly-knit groups in networks.
- Commonly used in social network analysis.

What is a *good* value?

- A high value suggests strong local clustering.
- A low value indicates sparse local connections.

2 Analysis of Centrality Metrics

2.1 Degree Centrality

Definition: The degree centrality of a node measures the number of edges connected to it.

$$C_D(v) = \frac{\deg(v)}{n - 1}$$

For directed graphs:

- **In-degree**: Number of incoming edges.
- **Out-degree**: Number of outgoing edges.

Scope:

- Highlights influential nodes based on direct connections.
- Often used in social media and citation networks.

What is a *good* value?

- High degree centrality identifies hubs and influential individuals.
- Low values are typical for peripheral nodes.

2.2 Betweenness Centrality

Definition: Measures the extent to which a node lies on the shortest paths between other nodes.

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where:

- σ_{st}: Total number of shortest paths between s and t.
- $\sigma_{st}(v)$: Number of shortest paths passing through v.

Scope:

- Essential for identifying bridging nodes or intermediaries.
- Applied in communication and transportation networks.

What is a *good* value?

- High betweenness centrality indicates a critical role in connectivity.
- Low values suggest limited importance in bridging.

2.3 Closeness Centrality

Definition: Measures how close a node is to all other nodes in the network.

$$C_C(v) = \frac{n-1}{\sum_{u \neq v} d(u,v)}$$

Scope:

- Ideal for analyzing reachability and influence.
- Common in network optimization tasks.

What is a *good* value?

- High values indicate central positioning in the network.
- Low values correspond to peripheral nodes.

2.4 Eigenvector Centrality

Definition: Measures the influence of a node based on the importance of its neighbors.

$$C_E(v) = \frac{1}{\lambda} \sum_{u \in \Gamma(v)} C_E(u)$$

where λ is the largest eigenvalue of the adjacency matrix. **Scope:**

- Suitable for identifying global influencers in networks.
- Commonly used in financial and citation networks.

What is a *good* value?

- High eigenvector centrality reflects global importance.
- Low values imply limited influence.

2.5 PageRank

Definition: Measures the importance of a node based on incoming connections, with iterative updates:

$$PR(v) = (1 - d) + d \sum_{u \in \Gamma(v)} \frac{PR(u)}{\deg(u)}$$

where d is the damping factor (typically 0.85). **Scope:**

- Widely applied in web search engines and ranking systems.
- Effective in directed networks.

What is a *good* value?

- High PageRank indicates significant importance (e.g., popular websites).
- Low values suggest less connectivity or influence.

2.6 HITS (Hub and Authority)

Definition: HITS distinguishes between two roles:

- **Hub**: Nodes pointing to many authoritative nodes.
- **Authority**: Nodes pointed to by many hubs.

Scope:

- Effective in directed graphs, particularly in web graphs.
- Identifies complementary roles in a network.

What is a *good* value?

- High hub scores identify central sources of information.
- High authority scores highlight credible destinations.

2.7 Eccentricity

Definition: Measures the longest shortest path from a node to any other node.

$$E(v) = \max_{u \in V} d(u, v)$$

Scope:

- Useful for identifying peripheral nodes.
- Common in analyzing diameter and efficiency.

What is a *good* value?

- Low eccentricity indicates central positioning.
- High values correspond to peripheral nodes or outliers.