Betweenness and Clustering Coefficient

Name: Asodariya Harsh
Entry number: 2023csb1103

Question

What will be the betweenness and clustering coefficient of the given Impression Network?

To understand the question you have to understand the centrality of the graph. Centrality in network analysis refers to the measure of importance or significance of the nodes within a network based on their structural position and role to find out interaction and information flow; in simpler words, centrality is about figuring out which part of the network, like social connections or transportation routes, are most important or influential.

Now for the directed graph, there are mainly 2 ways to find out centrality:

1. Betweenness
2. Clustering Coefficient

Betweenness

Betweenness centrality measures the extent to which a node lies on the shortest paths between other nodes in the network. For each pair of nodes in the network

How it works: For each pair of nodes in the network, betweenness centrality counts the number of shortest paths that pass through a particular node. The more shortest paths a node lies on, the higher its betweenness centrality.

You can easily understand betweenness In a network of cities connected by highways, highways have high betweenness because they connect different clusters of cities. Imagine highways as the main roads linking big groups of cities. They're crucial for traveling between these groups. Meanwhile, roads within cities have low betweenness. That's because there are often many paths within a city leading to the same destination. Think of these roads like neighborhood streets—they're essential for getting around within the city, but they don't play as big a role in connecting different areas compared to highways.

Formula:

$$\text{Betweenness centrality} = \frac{\text{number of shortest paths from any node } s \text{ to any node } t \text{ having that node(v)}}{\text{total number of shortest paths from s to t}}$$

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

$C_B(v)$ represents the betweenness centrality of node v.

σ_{st} is the total number of shortest paths from node s to node t.

$\sigma_{st}(v)$ denotes the number of those shortest paths that pass through node v.

Significance:

1. Bridging Different Parts: Nodes with high betweenness centrality act as bridges between different parts of the network. They play a crucial role in connecting otherwise separate clusters or communities. Identifying these nodes helps understand how information, resources, or influence flows between different groups within the network.

2. Critical Infrastructure: In various networks like transportation, communication, or social networks, nodes with high betweenness centrality often represent critical infrastructure. Targeting these nodes could significantly disrupt the network's functionality. Understanding betweenness helps prioritize the protection or reinforcement of such critical nodes to enhance the network's resilience.

3. Information Flow:In social networks or information networks, nodes with high betweenness play a crucial role in information diffusion. They act as key influencers or gatekeepers, controlling the flow of information between different parts of the network. Understanding betweenness helps identify influential individuals or channels for targeted dissemination of information or ideas.

Clustering Coefficient

Clustering coefficient measures the degree to which nodes in a network tend to cluster together.

How it works: For a given node, the clustering coefficient is calculated based on the fraction of its neighbors that are also connected to each other. A high clustering coefficient indicates that the node's neighbors are well connected, forming a tightly-knit cluster or community.
It basically shows the importance of the node and how well its neighbours are connected with each other.

Formula:

$$C(v) = \frac{\text{number of triangles connected to node } v}{\text{number of all possible triangles connected to } v}$$

$$C(v) = \frac{2 \times n_C^2}{n_v \times (n_v - 1)}$$

$C(v)$ represents the clustering coefficient of node v.

n_C^2 represents the number of triangles in the neighborhood of node v.

n_v represents the number of neighbors of node v.

Significance:

1. Measure of Local Connectivity: The clustering coefficient quantifies the degree of local connectivity around a node in the network. A high clustering coefficient indicates that a node's neighbors tend to be interconnected, forming tightly knit clusters or communities. Understanding local connectivity is crucial for characterizing the network's structure and dynamics, as it reflects the presence of cohesive subgroups or functional modules within the network.

2. Identification of Network Hierarchy: In many real-world networks, such as social networks or biological networks, hierarchical structures emerge where nodes are organized into nested levels of connectivity. The clustering coefficient helps identify nodes that participate in such hierarchical organization, revealing the presence of densely interconnected clusters at different scales. This hierarchical organization provides insights into the network's functional specialization and evolutionary dynamics.

3. Assessment of Network Resilience: The clustering coefficient offers insights into the network's resilience against random failures or targeted attacks. Networks with high clustering coefficients tend to exhibit greater robustness, as tightly knit clusters provide redundancy and alternative pathways for information or resource flow. Analyzing the clustering coefficient helps identify critical nodes or clusters whose removal could disproportionately impact the network's connectivity and functionality.

References

1. https://youtu.be/qSqSzzFdhSE?si=kG0-GdYsSfedhCHZ

2. https://youtu.be/gLjvLytq9jY?si=F-QvIzynC5Y0SE5U