# Distance Measures in Social Networks

**Path**: A path is a sequence of nodes connected by edges.

**Path Length**: Path length is the number of steps or hops in a path from one node to another.

**Shortest Path**: The shortest path between two nodes is the path with the fewest hops.

**Breadth-First Search (BFS)**: BFS is an efficient algorithm to find the distance from a node to all other nodes in a network.

**Average Shortest Path Length**: It's the average distance between any pair of nodes in the network.

**Diameter**: The diameter is the maximum possible distance between any two nodes in the network.

**Eccentricity**: Eccentricity of a node is the largest distance between that node and all other nodes in the network.

**Radius**: The radius is the minimum eccentricity in the network.

**Periphery**: The periphery consists of nodes with eccentricity equal to the diameter; they are typically far from other nodes.

**Center**: The center consists of nodes with eccentricity equal to the radius; they are typically central and close to other nodes.

**Karate Club Network**: An example network used to illustrate these concepts, where nodes represent members of a karate club.

**NetworkX**: A Python library for working with networks and graphs.

**Node Centrality**: Identifying central nodes in a network based on their connectivity and distances.

**Sensitivity of Center Definition**: The center of a network can change based on the distances between nodes, making it sensitive to small changes in the graph.

These concepts help in understanding and analyzing the structure of social networks, including identifying central nodes, peripheral nodes, and the overall network characteristics.

# Network Evolution and Dynamics:

1. **Network Evolution and Dynamics**:

Networks are dynamic structures that change over time.

Different models of network evolution describe how networks change from start to finish.

1. **Link Prediction Problem**:

The problem of predicting which edges will form in a given network's future.

Applications include friend recommendation algorithms, predicting new connections on social networks.

1. **Common Neighbors**:

Simple measure: Count of nodes that are neighbors of both nodes being considered.

Number of common neighbors indicates potential connection likelihood.

1. **Jaccard Coefficient**:

Normalized version of common neighbors.

Considers the ratio of common neighbors to the total number of neighbors.

Reflects potential connection strength more accurately.

1. **Resource Allocation Index**:

Measures the allocation of a resource (e.g., information) from one node to another via common neighbors.

Penalizes nodes with more neighbors, reducing the resource share.

Evaluates how effectively information is transmitted through common neighbors.

1. **Adamic-Adar Index**:

Similar to the Resource Allocation Index, but divides by the logarithm of the degree.

Incorporates logarithmic scaling to decrease the impact of high-degree nodes.

1. **Preferential Attachment Score**:

Based on preferential attachment model.

The product of degrees of two nodes being considered.

Reflects the likelihood of nodes with higher degrees forming connections.

1. **Community Structure**:

Network nodes are grouped into communities.

Nodes within the same community are more likely to connect.

1. **Common Neighbors with Community Bonus**:

Similar to common neighbors, but rewards shared neighbors in the same community.

Reflects increased likelihood of connection within the same community.

1. **Resource Allocation Soundarajan-Hopcroft Score**:

Incorporates community information into resource allocation index.

Only considers common neighbors in the same community as nodes being evaluated.

1. **Machine Learning Application**:

Measures can serve as features for machine learning-based link prediction.

Different measures may yield different results; a classifier can use them for predictions.

## Summary:

Link prediction aims to predict future connections in a network.

Various measures, including common neighbors, Jaccard coefficient, resource allocation, Adamic-Adar index, preferential attachment score, and community-based measures, are used.

The choice of measure may depend on the available data and the specific problem being addressed.