



# BUDDHA INSTITUTE OF TECHNOLOGY

## Gida Gorakhpur



### Department- Computer Science and Allied (ARTIFICIAL INTELLIGENCE MACHINE LEARNING)

Program & Semester- B.Tech 3<sup>rd</sup> Year (6<sup>th</sup> Semester)

Course and Code- Social Media Analytics and Data Analytics BCAM 061

#### Course Outcome

CO No.	Course Outcome	Bloom's Knowledge Level (KL)
CO 1	Understand basic concepts and need of social media analysis	
CO 2	Understand the fundamental of graphs and matrices in social media analysis	
CO 3	Understand networking fundamentals of social media analysis	
CO 4	Understand social networking and modelling concepts and methods.	
CO 5	Understand processing and visualizing social media data	

## UNIT-3

# NETWORK STRUCTURE & MODELLING

### 1. Define Network Structures and explain their types.

Network structure refers to the arrangement of nodes (users) and edges (relationships) in a social network. It helps in understanding how users are connected and how information flows across the network. A proper understanding of network structure is essential for analyzing communication patterns and identifying influential users in social media platforms.

#### Types of Network Structures:

- **Centralized Network:**  
One central node connects all other nodes.  
Fast communication but highly dependent on one node.
- **Decentralized Network:**  
Multiple hubs exist, distributing control.  
More reliable than centralized networks.
- **Distributed Network:**  
No central authority; all nodes are equal.  
Highly robust but complex.

Network structures are useful in identifying influencers, analyzing data flow, and improving marketing strategies. For example, centralized networks are useful for quick dissemination of information, while distributed networks ensure stability and fault tolerance. Thus, network structure plays a crucial role in social media analytics.

### 2. Explain Homophily in Social Networks.

Homophily is the tendency of individuals to connect with others who are similar to them in terms of interests, beliefs, or characteristics. It is a fundamental concept in social network analysis and explains how communities are formed in social media platforms.

#### Types of Homophily:

- **Status Homophily:** Based on age, gender, education.
- **Value Homophily:** Based on shared interests or opinions.

Homophily leads to the formation of clusters where similar individuals interact more frequently. This increases communication within groups but reduces interaction between different groups, sometimes creating echo chambers.

#### Applications:

- Recommendation systems
- Targeted advertising
- Community detection

For example, users on Instagram follow accounts with similar interests such as fitness or technology.

Understanding homophily helps businesses design better marketing strategies and improve user engagement. It also helps in predicting behavior and improving content recommendations.

### 3. Explain Clustering in Social Networks.

2024–25

Clustering refers to the grouping of nodes in a network such that nodes within the same group are more closely connected to each other than to nodes outside the group. It helps in identifying communities and understanding social interactions.

#### Key Concepts:

- **Cluster:** Group of closely connected nodes
- **Clustering Coefficient:** Measures strength of connections

Clustering naturally occurs in social media where people form groups based on common interests or relationships. High clustering indicates strong community bonding.

#### Importance:

- Detects communities
- Helps in marketing segmentation
- Improves recommendation systems

For example, students of a college form a cluster on social media platforms. Clustering helps analysts understand group behavior and communication patterns. It is widely used in social network analysis to study community structures.

### 4. Explain Structural Holes in Social Networks.

2024–25

Structural holes are gaps in a network where there is no direct connection between two groups. These gaps create opportunities for individuals to act as bridges between different communities.

#### Key Points:

- A bridging person is called a **broker**
- Broker connects disconnected groups
- Provides access to diverse information

Structural holes are important because they give strategic advantages to the broker. Such individuals can control the flow of information and influence multiple groups simultaneously.

#### Applications:

- Influencer identification
- Marketing strategies
- Information diffusion

For example, a user connected to both technical and non-technical communities can share knowledge across groups. Structural holes help in improving communication and spreading information efficiently in social networks.

## 5. Explain Snowball Sampling and Random Walks.

2024–25

Snowball sampling is a method where existing participants recruit new participants from their network. It is commonly used in social network studies where data is difficult to collect directly.

### Snowball Sampling:

- Starts with a small group
- Expands through connections
- Useful for hidden populations

Random walk is a technique where movement occurs randomly from one node to another in a network. It helps in unbiased sampling of large networks.

### Comparison:

- Snowball → Faster but biased
- Random Walk → Accurate but slower

### Applications:

- Data collection
- Network analysis
- Social research

Both methods are important for analyzing large-scale social networks and understanding user connectivity.

## 6. Explain the importance of Network Structures in Social Media Analytics.

Network structures represent the arrangement of nodes and edges in a social network. Nodes correspond to users or entities, while edges represent interactions such as friendships, follows, or communications. Understanding network structures is essential in social media analytics because it determines how information flows, how influence spreads, and how communities are formed within the network.

There are different types of network structures, each having its own characteristics and impact on communication.

- **Centralized Network:**

A single node acts as the hub connecting all other nodes. This structure allows fast dissemination of information but creates dependency on the central node.

- **Decentralized Network:**

Multiple hubs exist, distributing connections and responsibilities. This structure is more reliable and reduces the risk of failure compared to centralized systems.

- **Distributed Network:**

All nodes are equally connected with no central authority. It is highly robust and ensures fault tolerance but may involve complex communication paths.

The importance of network structures in social media analytics can be understood through several aspects.

- **Influence Identification:**

Network structures help identify key users or influencers who play a major role in spreading information.

- **Information Diffusion:**

The structure determines how quickly and widely information spreads across the network.

- **Community Detection:**

Clusters of users with similar interests can be identified based on structural patterns.

- **Targeted Marketing:**

Businesses use structural insights to focus on specific users or groups for effective promotion.

- **System Efficiency:**

Helps in improving network performance and identifying weak points.

For example, in a social media platform, influencers act as central nodes in a centralized structure, while online communities represent decentralized or distributed structures.

Thus, understanding network structures enables better analysis of user behavior, enhances decision-making, and supports efficient implementation of marketing and communication strategies in social media analytics.

## 7. Explain Equivalence in Network Structures with suitable example.

2024–25

Equivalence in network structures refers to identifying nodes that have similar roles or positions within a network. Instead of focusing only on direct connections, equivalence examines patterns of relationships to determine similarity among nodes. This concept is useful in simplifying complex networks and understanding user roles.

There are different types of equivalence that describe various forms of similarity among nodes.

- **Structural Equivalence:**

Nodes are structurally equivalent if they share identical connections with the same set of nodes. For example, two employees reporting to the same manager exhibit structural equivalence.

- **Regular Equivalence:**

Nodes are considered equivalent if they have similar relationships with similar types of nodes, even if they are not connected to the same individuals. For example, teachers in different schools perform similar roles.

- **Automorphic Equivalence:**

Nodes are interchangeable without altering the overall network structure.

Equivalence plays a significant role in social network analysis.

- **Role Identification:**

Helps in identifying users performing similar functions in a network.

- **Network Simplification:**  
Reduces complexity by grouping similar nodes together.
- **Recommendation Systems:**  
Suggests connections or content based on similar roles.
- **Behavior Analysis:**  
Helps in predicting actions of users with similar patterns.

For instance, in social media, two influencers with similar engagement levels and audience behavior can be considered equivalent even if they are not directly connected. This helps businesses identify multiple potential influencers for marketing campaigns.

Thus, equivalence provides a structured approach to understanding similarities in networks, making it easier to analyze large and complex social systems.

## 8. Explain Random Graphs and Network Evolution.

2024–25

Random graphs are networks in which edges between nodes are formed randomly based on a certain probability. They are used to model and analyze complex real-world systems such as social networks, communication systems, and biological networks. Network evolution refers to the process by which these networks grow and change over time.

In a random graph, each pair of nodes has a probability of being connected. This leads to the formation of unpredictable but analyzable structures.

- **Characteristics of Random Graphs:**
  - Edges are formed randomly
  - Degree distribution follows probability rules
  - Used for theoretical modeling

Different models are used to represent random graphs.

- **Erdős–Rényi Model:**  
Each pair of nodes has an equal probability of connection.
- **Barabási–Albert Model:**  
Based on preferential attachment, where nodes with more connections attract more links.

Network evolution explains how networks develop over time.

- **Growth:**  
New nodes continuously join the network.
- **Preferential Attachment:**  
Nodes with higher connectivity gain more links.
- **Formation of Hubs:**

Highly connected nodes emerge as central points in the network.

The study of network evolution helps in understanding real-world phenomena.

- Predicts how social networks expand
- Identifies influential users
- Helps in designing scalable systems

For example, in social media platforms, popular users gain more followers over time due to preferential attachment.

This results in the formation of hubs and scale-free networks.

Thus, random graphs and network evolution provide a strong theoretical foundation for analyzing and predicting the behavior of complex networks.

## 9. Explain Citation Networks and their role in research.

2024–25

Citation networks are a type of network where nodes represent research papers and edges represent citations between them. These networks are widely used in academic research to understand how knowledge is developed and how ideas spread across different domains.

In a citation network, if one paper cites another, a directed edge is formed from the citing paper to the cited paper.

This creates a directed graph that reflects the flow of information.

- **Key Characteristics:**
  - Directed network
  - Nodes represent documents
  - Edges represent citations

Citation networks have several important properties.

- **Highly Connected Structure:**

Some papers receive a large number of citations, becoming central nodes.
- **Growth Over Time:**

New papers continuously add to the network.
- **Clustering:**

Papers in similar domains form clusters.

The role of citation networks in research is significant.

- **Identifying Influential Work:**

Highly cited papers indicate important contributions.
- **Tracking Knowledge Evolution:**

Helps in understanding how ideas develop over time.
- **Supporting Literature Review:**

Enables researchers to find relevant studies efficiently.

- **Measuring Impact:**

Used in ranking journals and researchers.

For example, a groundbreaking research paper becomes highly cited and forms a central node in the citation network.

Thus, citation networks provide valuable insights into the structure of academic knowledge, helping researchers analyze trends, identify key contributions, and understand the development of various fields.

## 10. Explain Ego-centered Networks and Dominance Hierarchies.

Ego-centered networks focus on a single node, known as the ego, and its direct connections, called alters. This type of network is useful for analyzing individual relationships and understanding how a particular user interacts within a larger system. Dominance hierarchies, on the other hand, represent the ranking of nodes based on influence, authority, or power within a network.

An ego-centered network consists of three main components.

- **Ego:**

The central node being analyzed

- **Alters:**

Nodes directly connected to the ego

- **Ties:**

Relationships between the ego and alters

This type of network helps in understanding personal influence and connectivity.

- Identifies close relationships
- Measures interaction strength
- Helps in targeted analysis

Dominance hierarchies describe how nodes are ranked in terms of importance or authority.

- Nodes at higher levels have more influence
- Common in organizational and social structures
- Reflects power distribution

These hierarchies are useful in analyzing leadership and influence patterns.

- Identifies key decision-makers
- Helps in understanding communication flow
- Supports organizational analysis

For example, in a corporate network, a manager may occupy a higher position in the dominance hierarchy, while employees form lower levels.

Thus, ego-centered networks and dominance hierarchies provide valuable insights into both individual relationships and overall power structures within social networks.

## 11. Explain Peer-to-Peer Networks and Recommender Networks.

Peer-to-peer networks and recommender networks are important concepts in modern social media systems. Peer-to-peer networks focus on decentralized communication, while recommender networks focus on personalized suggestions based on user behavior. Both play a crucial role in enhancing user experience and connectivity.

A peer-to-peer network is a distributed network where all nodes act as both clients and servers. There is no central authority, and nodes communicate directly with each other.

- **Characteristics of Peer-to-Peer Networks:**

- No central server
- Equal participation of nodes
- Direct communication between users

- **Advantages:**

- High reliability and fault tolerance
- Scalable system
- Efficient resource sharing

- **Limitations:**

- Security concerns
- Difficult to manage large networks

Recommender networks are systems designed to suggest items, content, or connections to users based on their preferences and behavior.

- **Types of Recommender Systems:**

- Content-based filtering
- Collaborative filtering
- Hybrid approaches

- **Working Principle:**

- Collect user data
- Analyze preferences
- Generate recommendations

- **Applications:**

- Product recommendations in e-commerce
- Friend suggestions in social media
- Content recommendations in streaming platforms

For example, platforms like Netflix and Amazon use recommender systems to enhance user engagement by suggesting relevant content.

Thus, peer-to-peer networks enable efficient communication without centralized control, while recommender networks improve personalization and user satisfaction in digital platforms.

**12. In a large social network, data collection is incomplete. Suggest and justify suitable sampling techniques for analysis.**

**2024–25**

In large-scale social networks, it is often impractical to collect complete data due to size, privacy restrictions, and continuous updates. Therefore, sampling techniques are required to obtain representative subsets of the network for analysis. Selecting appropriate sampling methods is crucial for ensuring accuracy and reliability of results.

Two commonly used techniques are snowball sampling and random walk. Each has its own advantages and limitations depending on the objective of analysis.

- **Snowball Sampling:**

This method begins with a small number of initial nodes and expands by including their neighbors.

- Starts with seed nodes
- Expands iteratively
- Suitable for hidden or hard-to-reach populations

**Advantages:**

- Easy to implement
- Useful for community detection
- Cost-effective

**Limitations:**

- Biased towards highly connected nodes
- May not represent entire network

- **Random Walk:**

This method involves moving randomly from one node to another in the network.

- Begins from a random node
- Moves to adjacent nodes randomly
- Continues for multiple iterations

**Advantages:**

- Produces less biased samples
- Covers wider network area

**Limitations:**

- Slower process
- Requires more computation

The choice between these techniques depends on the purpose of analysis. If the goal is to identify communities, snowball sampling is effective. If the aim is to estimate global network properties, random walk is more suitable.

Thus, appropriate sampling techniques help in analyzing large networks efficiently while maintaining a balance between accuracy and computational feasibility.

### 13. Analyze how equivalence concepts can be used to identify similar roles in a social network.

2024–25

Equivalence concepts are used in social network analysis to identify nodes that perform similar roles within a network. Instead of focusing only on direct connections, equivalence examines patterns of relationships, allowing analysts to group nodes based on similarity in behavior or position. This approach is useful in simplifying complex networks and understanding role-based interactions.

There are different types of equivalence that help in identifying similar roles.

- **Structural Equivalence:**  
Nodes are structurally equivalent if they share identical connections with the same set of nodes.
- **Regular Equivalence:**  
Nodes are considered equivalent if they have similar types of relationships with similar nodes.
- **Automorphic Equivalence:**  
Nodes can be interchanged without affecting the network structure.

These equivalence concepts have several practical applications in social network analysis.

- **Role Identification:**  
Helps in identifying users with similar responsibilities or influence.
- **Community Analysis:**  
Groups users with similar behavior patterns.
- **Recommendation Systems:**  
Suggests connections based on similarity of roles.
- **Network Simplification:**  
Reduces complexity by grouping similar nodes.

For example, in a social media platform, multiple influencers may have similar engagement patterns and audience types. Even if they are not directly connected, they can be considered equivalent based on their roles.

Thus, equivalence provides a structured method for identifying similarities in networks, enabling better analysis and decision-making in social media analytics.

### 14. A research community is evolving over time. Explain how network evolution and preferential attachment influence its structure.

2024–25

Research communities can be modeled as networks where nodes represent researchers or publications and edges represent collaborations or citations. Over time, these networks evolve as new nodes are added and new connections are formed. One of the key principles governing this evolution is preferential attachment.

Network evolution describes the process by which networks grow and change over time.

- **Growth:**  
New nodes continuously join the network.
- **Connection Formation:**  
New edges are created between nodes.
- **Structural Changes:**  
Network becomes more complex over time.

Preferential attachment is a mechanism where nodes with higher connectivity are more likely to receive new connections.

- “Rich get richer” phenomenon
- Nodes with more links attract more connections
- Leads to uneven distribution of links

The combined effect of network evolution and preferential attachment results in specific structural properties.

- **Formation of Hubs:**  
Highly connected nodes emerge as central points
- **Scale-Free Networks:**  
Degree distribution follows a power law
- **Influence Distribution:**  
Few nodes have high influence, many have low influence

For example, highly cited research papers continue to receive more citations, making them central nodes in citation networks.

Thus, network evolution and preferential attachment explain how networks grow, how influence is distributed, and why certain nodes become more important than others.

## 15. Evaluate how clustering and community detection help in improving marketing strategies on social media platforms.

2024–25

Clustering and community detection are important techniques in social network analysis that help identify groups of users with similar characteristics or interests. These techniques play a crucial role in improving marketing strategies by enabling targeted and personalized approaches.

Clustering refers to grouping nodes such that nodes within the same group are more densely connected compared to nodes outside the group. Community detection identifies these clusters within large networks.

- **Key Concepts:**
  - Cluster: Group of similar users
  - Community: Sub-network with strong internal connections
- **Methods:**

- Modularity-based approaches
- Graph partitioning
- Hierarchical clustering

These techniques are widely used in marketing strategies.

- **Targeted Advertising:**

Businesses can focus on specific user groups

- **Personalization:**

Content can be tailored based on user interests

- **Customer Segmentation:**

Helps divide users into meaningful categories

- **Improved Engagement:**

Relevant content increases user interaction

For example, a fitness brand can identify a cluster of users interested in health and target them with appropriate advertisements.

However, challenges exist such as dynamic network changes and overlapping communities. Despite these challenges, clustering and community detection significantly improve marketing efficiency.

Thus, these techniques provide valuable insights into user behavior and help businesses design effective, data-driven marketing strategies.