Module 2: Social Network Structure, Measures & Visualization

Syllabus:

Lecture no	Content	Duration (Hr)	Self-Study (Hrs)
1	Basics of Social Network Structure - Nodes, Edges & Tie	1	1
2	Describing the Networks Measures - Degree Distribution	1	1
3	Density, Connectivity, Centralization, Tie Strength & Trust	1	1
4	Network Visualization - Graph Layout,	1	1
5	Visualizing Network features	1	1
6	Scale Issues. Social Media Network Analytics - Common Network Terms	1	1
7	Common Social Media Network Types	1	1
8	Types of Networks, Common Network Terminologies	1	1
9	Network Analytics Tools.	1	1

Theoretical Background:

- Social media network analytics is the process of collecting, analyzing, and interpreting data from social media platforms.
- This helps to gain insights into user behavior, engagement, and trends.
- Social media networks generate vast amounts of data daily, and analytics allows businesses, marketers, and researchers to harness this data for strategic for
    - Decision-making
    - Audience targeting
    - Measuring the effectiveness of their social media campaigns.

Key Definitions:

Social media network analytics is the process of collecting, analyzing, and interpreting data from social media platforms to gain insights into user behavior, engagement, and trends. Social media networks generate vast amounts of data daily, and analytics allows businesses, marketers, and researchers to harness this data for strategic decision-making, audience targeting, and measuring the effectiveness of their social media campaigns.

Lecture 1

Learning Objective: Learners shall be able to understand the basics of social network structure

Basics of Social Network Structure - Nodes, Edges & Tie

Social media network analytics is the process of collecting, analyzing, and interpreting data from social media platforms to gain insights into user behavior, engagement, and trends. Social media networks generate vast amounts of data daily, and analytics allows businesses, marketers, and researchers to harness this data for strategic decision-making, audience targeting, and measuring the effectiveness of their social media campaigns.

Key aspects of social media network analytics include:

Data Collection: Analytics tools gather data from various social media platforms, including Facebook, Twitter, Instagram, LinkedIn, YouTube, and more. The data may include metrics such as likes, shares, comments, follower counts, click-through rates, and impressions.

Monitoring and Listening: Social media analytics involves real-time monitoring and listening to conversations happening across different social channels. This helps organizations understand what people are saying about their brand, products, or services.

Performance Metrics: Social media analytics provides insights into the performance of individual posts, campaigns, or overall social media accounts. Marketers can analyze engagement metrics to evaluate the effectiveness of their content and campaigns.

Audience Insights: Analytics tools help identify the demographics and interests of social media audiences. This information assists in crafting targeted content and reaching the right audience.

Competitor Analysis: Social media analytics allows businesses to track their competitors' performance, content strategies, and engagement levels, helping them stay competitive in the market.

Sentiment Analysis: Sentiment analysis is used to understand the sentiment behind social media mentions, whether positive, negative, or neutral. This helps gauge public perception and sentiment toward a brand or topic.

Influencer Marketing: Analytics can help identify influential users with a large and engaged following. Brands can collaborate with these influencers to reach a wider audience and improve brand visibility.

Campaign Measurement: Analytics tools enable the measurement of the success of social media campaigns, helping marketers assess ROI and make data-driven decisions for future campaigns.

Social Listening for Crisis Management: Social media analytics can be used for social listening during times of crisis or brand reputation management. Organizations can monitor mentions and sentiment to respond quickly to emerging issues.

Overall, social media network analytics plays a crucial role in optimizing social media strategies, improving brand engagement, understanding customer preferences, and enhancing overall social media marketing efforts. By leveraging these insights, businesses can build stronger relationships with their audience and achieve their social media marketing objectives effectively.

Basically, there are two types of social networks:

Ego network Analysis

Complete network Analysis

1. Ego Network Analysis

Ego network Analysis is the one that finds the relationship among people. The analysis is done for a particular sample of people chosen from the whole population. This sampling is done randomly to analyze the relationship. The attributes involved in this ego network analysis are a person's size, diversity, etc.

This analysis is done by traditional surveys. The surveys involve people being asked whom they interact with and their name of the relationship between them. It is not focused to find the relationship between everyone in the sample. It is an effort to find the density of the network in those samples. This hypothesis is tested using some statistical hypothesis testing techniques.

The following functions are served by Ego Networks:

Propagation of information efficiently.

Sensemaking from links, For example, Social links, relationships.

Access to resources, efficient connection path generation.

Community detection, identification of the formation of groups.

Analysis of the ties among individuals for social support.

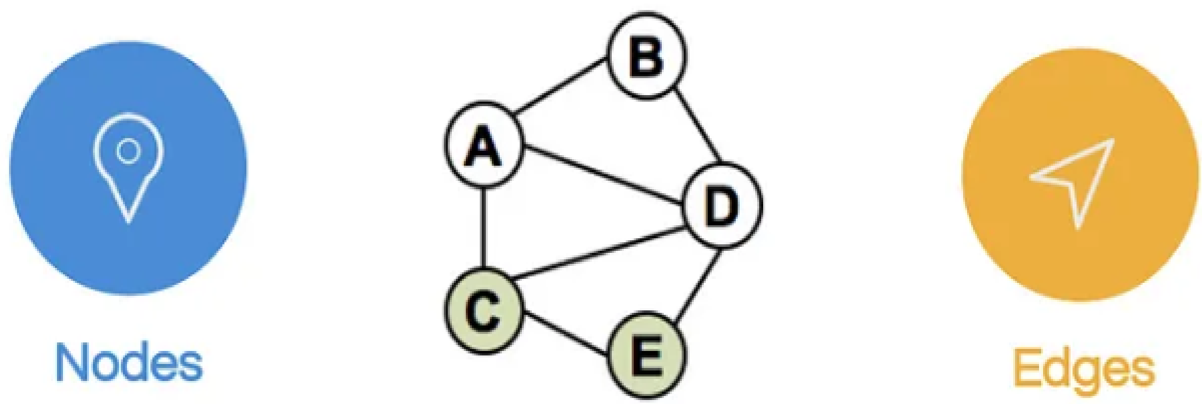
2. Complete Network Analysis

Complete network analysis is the analysis that is used in all network analyses. It analyses the relationship among the sample of people chosen from the large population. Subgroup analysis, centrality measure, and equivalence analysis are based on the complete network analysis. This analysis measure helps the organization or the company to make any decision with the help of their relationship. Testing the sample will show the relationship in the whole network since the sample is taken from a single set of domains.

Difference between Ego network analysis and Complete network analysis:

The difference between ego and complete network analysis is that the ego network focuses on collecting the relationship of people in the sample with the outside world whereas, in Complete network, it is focused on finding the relationship among the samples.

The majority of the network analysis will be done only for a particular domain or one organization. It is not focused on the relationships between the organizations. So many of the social network analysis measures use only Complete network analysis.



Nodes (A,B,C,D,E in the example) are usually representing entities in the network, and can hold self-properties (such as weight, size, position and any other attribute) and network-based properties (such as Degree- number of neighbors or Cluster- a connected component the node belongs to etc.)

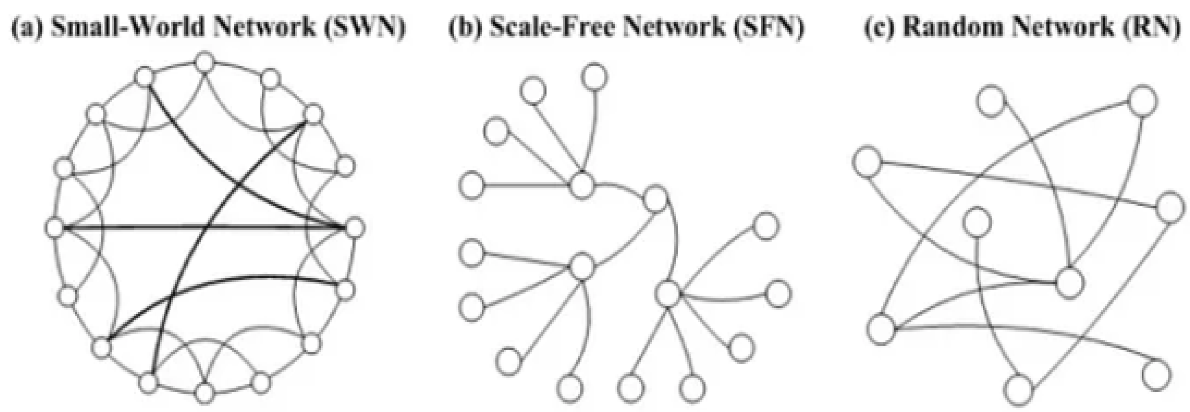
Edges represent the connections between the nodes, and might hold properties as well (such as weight representing the strength of the connection, direction in case of asymmetric relation or time if applicable).

Ties on social media are connections between users. These could be friendships, follows, interactions, or group affiliations. Ties determine whose posts you see and influence the reach of your own posts. They shape your social experience online.

Small World Networks: A small world network is like having friends of friends.In a small world network, most people are not directly connected, but you can reach them through a few intermediaries.

Scale-Free Networks: A scale-free network is all about popularity and fame. In this network, a few individuals, like celebrities, have lots of connections (friends, followers, etc.), while most others have only a few.Scale-free networks are characterized by a few highly connected nodes, called "hubs," and many nodes with just a few connections.

Random Network: A random network is like connecting people by flipping a coin. In a random network, nodes are linked together purely by chance, like throwing dice. There's no specific pattern or preference in creating connections. Random networks are used as a baseline to understand what features in real networks are due to chance and what might have some meaningful structure or organization.



Learning from this lecture: Learners should be able to understand the basics of social network structure.

Lecture 2

Learning Objective: **Learners shall be able to understand the concept behind network measures.**

Describing the Networks Measures - Degree Distribution

Centrality Measures

Degree — the amount of neighbors of the node

EigenVector / PageRank — iterative circles of neighbors

Closeness — the level of closeness to all of the nodes

Betweenness — the amount of short path going through the node

- Degree Centrality: Imagine you have many friends, and each friend has a lot of other friends. Degree centrality measures how popular you are in your friend group. The more friends you have, the higher your degree centrality.
- Eigenvector Centrality / PageRank: Eigenvector centrality is like a popularity contest where not only the number of your friends matters but also how popular your friends are. If you are friends with popular people, your popularity will increase too. PageRank is like being ranked based on how popular your friends are and how many popular friends they have.
- Closeness Centrality: Closeness centrality is all about how close you are to all your friends. If you can quickly reach all your friends in just a few steps, you have high closeness centrality. It's like being at the center of your social circle.
- Betweenness Centrality: Betweenness centrality measures how important you are in connecting your friends with each other. If you act as a bridge, linking friends who are not directly connected, then you have high betweenness centrality. You play a crucial role in keeping everyone connected.

Degree Distribution

Types of Degree Distribution: Degree distribution can take various forms, and understanding these forms is crucial in social network analysis.

The two main types are:

  - Regular Degree Distribution: In this type, the degrees of nodes are relatively uniform across the network. Most nodes have similar degrees, creating a balanced structure.
  - Power-law Degree Distribution: Also known as the "scale-free" property, this type exhibits a few nodes with extremely high degrees (hubs) and many nodes with lower degrees. Power-law distributions are common in many real-world social networks.

- Importance of Degree Distribution: Degree distribution holds significant implications for understanding the network's characteristics:
- Identifying Influencers: Nodes with high degrees (hubs) are likely to be influential in the network, as they have a large number of connections.
- Network Resilience: Regular degree distributions can lead to robust networks, as there are no nodes with disproportionately high degrees that, if removed, could break the network.
- Scale-Free Networks: Power-law degree distributions are prevalent in many social networks, and understanding them is vital for analyzing network dynamics.

Degree Distribution

```
A - B
 | / |
C - D
  |   |
 E - F
```

- Node A: Degree 2 (connections with B and C)
- Node B: Degree 3 (connections with A and C and D)
- Node C: Degree 4 (connections with A, B, D and E)
- Node D: Degree 3 (connections with B C and E)
- Node E: Degree 2 (connections with C and F)
- Node F: Degree 2 (connection with E and D)

Learning from this lecture: Learners should be able to understand the network measures like degree centrality.

Lecture 3

Learning Objective: **Learners shall be able to understand the network measures like density and connectivity.**

Density, Connectivity,

Density

Density is a crucial measure that helps us understand the level of interconnectedness and cohesion within a network.
By exploring network density, we can gain valuable insights into the structure and properties of social networks.

Interpretation

Density ranges from 0 to 1, where a density of 0 indicates a completely disconnected network with no edges, and a density of 1 represents a fully connected network where all possible edges are realized.
High density suggests a tightly connected network, where nodes have numerous connections with each other, facilitating efficient communication and information flow.
Low density indicates a more loosely connected network, where nodes have fewer connections, potentially resulting in information silos.
Network Properties:

Connectivity: High-density networks have strong connectivity, enabling rapid transmission of information or influence through the network.
Clustering: High-density networks often exhibit clustering, where nodes tend to form tightly-knit groups or communities, fostering local interactions and cohesion.

```
A - B
 | / |
C - D
 \ |
 E - F
```

Calculating Density:
Density = (2 * Number of Edges) / (Number of Nodes * (Number of Nodes - 1))
The network's density can be calculated as follows:
Number of Nodes (N) = 6
Number of Edges (E) = 7 (A-B, A-C, B-C, C-D, C-E, D-E, E-F)
Density = (2 * 7) / (6 * (6 - 1)) = 14 / 30 ≈ 0.467

```
A -- B -- C
|     |    |
 D -- E    F
|     |     |
 G -- H -- I
```

Degree Calculation:

- Node A: Degree 2 (connections with B and D)
- Node B: Degree 3 (connections with A, C, and E)
- Node C: Degree 2 (connections with B and F)
- Node D: Degree 3 (connections with A and E)
- Node E: Degree 3 (connections with B, D, and G)
- Node F: Degree 2 (connections with C and I)
- Node G: Degree 2 (connections with E and H)
- Node H: Degree 3 (connections with G, I, and I)
- Node I: Degree 2 (connections with F and H)

Density Calculation:

- Density = (2 * Number of Edges) / (Number of Nodes * (Number of Nodes - 1))
- Edges: A-B, A-D, B-C, B-E, C-F, D-E, D-G, E-G, E-H, F-I, H-I
- Number of Edges = 11
- Number of Nodes = 9
- Density = (2 * 11) / (9 * (9 - 1))
- Density ≈ 0.488

Connectivity in a social network refers to the extent to which nodes are linked to one another through direct or indirect connections, such as friendships, interactions, or collaborations.
A network with high connectivity has many paths between nodes, allowing information, influence, or resources to flow efficiently.
Types of Connectivity:

- Direct Connectivity: This refers to the direct relationships or connections between nodes. For example, in a friendship network, if person A is directly connected to person B, there is direct connectivity between them.
- Indirect Connectivity: Indirect connectivity involves connections between nodes through intermediate nodes. If person A is connected to person B through person C, there is indirect connectivity between A and B.

Learning from this lecture: Learners should be able to understand network measures like density and connectivity.

Lecture 4

Learning Objective: **Learners shall be able to understand the centralization, tie strength and trust.**

Centralization, Tie Strength & Trust

Centralization

Centralization is a network measure in social network analysis that quantifies the degree to which a network's resources, influence, or power is concentrated in a few central nodes.
It assesses the distribution of connections, control, or importance among the nodes in a network.
Centralization helps us understand the extent to which a network is hierarchical or decentralized, and it plays a crucial role in analyzing the dynamics of social networks.
Centralization in a social network refers to how much power or influence is concentrated in a few key individuals or nodes. Imagine you have a group of friends, and some friends are more popular or influential than others. If most decisions and information flow through those popular friends, the network is more centralized. On the other hand, if everyone in the group has an equal say and influence, the network is more decentralized.

In a highly centralized network, a few individuals have a lot of control and can affect how information spreads and decisions are made. In a less centralized network, power is more evenly distributed among the group, making it more resilient and less dependent on a few key players.

Tie Strength

Tie strength is a concept used in social network analysis to describe the intensity or closeness of the relationship between two individuals in a social network.

In simple words, tie strength indicates how strong or weak the connection is between two people.
Example: Imagine you have two friends, Alice and Bob. The tie strength between Alice and Bob can be strong if they have a close, deep, and frequent relationship. They might share personal feelings, spend a lot of time together, and provide each other with strong emotional support.
On the other hand, tie strength can be weak if the connection between Alice and Bob is more casual or occasional. They might know each other and interact from time to time but do not share a deep emotional bond or spend much time together.
Characteristics of Tie Strength:

- Strong Tie: Strong ties are characterized by close relationships with high levels of emotional attachment, trust, and frequent interaction. People with strong ties are more likely to provide support, share important information, and help each other in times of need.
- Weak Tie: Weak ties are characterized by more distant or casual relationships with limited emotional involvement and infrequent interaction. While weak ties may not provide as much emotional support, they are valuable for accessing diverse information, ideas, and opportunities outside of one's immediate social circle.

Trust

In social media, users often share content, including news, articles, and opinions. The level of trust between users affects how others perceive the shared content.
Content shared by individuals with strong ties or high levels of trust is more likely to be considered credible and reliable by their network connections.
Trust is the cornerstone of any successful social interaction, and it becomes even more crucial in the virtual realm of social media networks.

Trust: Key Aspects

- Credibility of Sources
- Transparency and Authenticity
- Privacy and Data Security
- Consistency in Communication
- Engagement and Responsiveness
- User Reviews and Testimonials
- Mutual Interactions

Learning from this lecture: Learners should be able to understand centralization, tie strength and trust.

Lecture 5

Learning Objective: Learners shall be able to understand network visualization layout.

Network Visualization - Graph Layout

Network visualization involves representing connections and relationships between entities, often using a graph-based approach. Graph layout refers to how nodes (entities) and edges (connections) are arranged in a visual representation.

The goal of graph layout is to create a clear and understandable depiction of the network structure, making it easier to analyze and interpret the relationships.

Force-Directed Layout: This is one of the most popular graph layout algorithms for social network visualization.

It simulates physical forces (repulsion between nodes and attraction along edges) to position nodes in the graph.

Nodes that are connected by edges tend to be closer to each other, creating a visually intuitive representation of the network structure.

Force-directed layouts are aesthetically pleasing and can reveal clusters and communities within the network.

Hierarchical Layout: This layout arranges nodes in a tree-like structure, often used for visualizing hierarchical relationships within a network.

It's particularly useful when there is a clear parent-child relationship between nodes, such as in organizational structures.

Circular Layout: This layout arranges nodes in a circular pattern, with nodes connected by edges radiating outwards from the center.

It's effective for visualizing small to medium-sized networks with relatively simple connections.

Kamada-Kaway Layout: The Kamada-Kaway layout algorithm is a force-directed graph layout algorithm that aims to produce clear and aesthetically pleasing visualizations of graphs, including social networks.

It produces clear and visually appealing layouts.

Fruchterman-Reingold Layout: Similar to the force-directed layout, this algorithm simulates attractive and repulsive forces to position nodes.

It's particularly effective for showing clusters and communities within a network.

Spectral Layout: This algorithm uses eigenvectors of the graph's adjacency matrix to position nodes.

It can reveal underlying structural patterns in the network.

Grid Layout: This layout arranges nodes in a grid pattern, which can be useful for visualizing large networks in a more organized manner.

However, it might not capture complex relationships well.

Sugiyama Layout: Also known as layered or hierarchical layout, it's often used for directed acyclic graphs (DAGs) and organizes nodes into layers with edges flowing from top to bottom. It's useful for visualizing processes and dependencies.
Radial Layout: Similar to the circular layout, this arrangement places nodes on concentric circles, making it easier to show relationships between nodes.

Learning from this lecture: Learners should be able to understand **network visualization layout.**

Lecture 6

Learning Objective: **Learners shall be able to understand visualization of network features.**

Visualizing Network features
Reference: https://gwu-libraries.github.io/sfm-ui/posts/2017-09-08-sna

Learning from this lecture: Learners should be able to understand **visualization of network features.**

Lecture 7

Learning Objective: Learners shall be able to understand Scale Issues and Common Network Terms

Scale Issues. Social Media Network Analytics - Common Network Terms
The overwhelming amount of produced content and resulting network traffic gives rise to precarious scalability issues for social networks, such as handling a large number of users, infrastructure management, internal network traffic, content dissemination, and data storage.

1. **Node Overlapping and Clutter**: As the number of nodes in a network increases, it becomes difficult to position them without overlap. Overlapping nodes can make it hard to discern connections and identify individual nodes. Cluttered visualizations reduce the overall effectiveness of conveying information.
2. **Performance and Responsiveness**: Large networks can put a strain on visualization tools and systems, leading to slow performance and unresponsiveness. Interactivity, such as zooming and panning, might become sluggish, making it challenging for users to navigate and explore the network.
3. **Visual Noise and Distortion**: As the number of edges and nodes increases, the visual representation of connections can create a noisy and cluttered appearance. This can obscure important patterns and relationships, making it challenging to extract meaningful insights.
4. **Scalability of Layout Algorithms**: Layout algorithms that determine how nodes are

positioned and connected can struggle to handle large networks. Some algorithms might become inefficient, leading to long processing times or suboptimal visualizations.

5. **Limited Screen Real Estate**: Even with advanced visualization techniques, the limited size of screens and monitors can pose challenges for displaying large-scale networks effectively. Trying to fit a vast network into a confined space can lead to loss of detail and legibility.
6. **Edge Crossing and Routing**: When there are many edges in a network, they might cross over each other, making it difficult to follow connections. Edge routing becomes complex as the number of edges grows, affecting the overall readability of the visualization.
7. **Maintaining User Focus**: Presenting a large-scale network can overwhelm users and make it challenging for them to focus on specific parts of interest. Designing ways to guide user attention and provide context becomes crucial.

Strategies and techniques to address these challenges

- **Sampling**: Instead of visualizing the entire network, you can take a random or strategic sample of nodes and edges to create a more manageable visualization.
- **Hierarchical Visualization**: For very large networks, hierarchical layouts can be used to display networks at different levels of abstraction. This can help in visualizing clusters and subnetworks.
- **Aggregation**: Aggregating nodes into groups or communities can help reduce visual clutter while retaining the overall structure and relationships.
- **Dynamic Loading**: Loading only the visible portion of the network and dynamically loading additional nodes and edges as the user navigates can improve responsiveness.
- **High-Performance Visualization Tools**: Using specialized visualization libraries and tools designed for large-scale networks can significantly improve rendering speed and interactivity.

Common Networks Terms

NETWORK At a very basic level, a network is a group of nodes that are connected with links (Wasserman and Faust 1994). Nodes (also known as vertices) can represent anything, including individuals, organizations, countries, computers, websites, or any other entities. Links (also known as ties, edges, or arcs) represent the relationship among the nodes in a network. Networks can also exist among animals.

SOCIAL NETWORKS A social network is a group of nodes and links formed by social entities where nodes can represent social entities such as people and organizations. Links represent their relationships, such as friendship and trade relations. Social networks can exist both in the real and online worlds. A network among classmates is an example of real world social network. And a Twitter follow following network is an example of an online social media network. In a Twitter follow-following network, nodes are the Twitter users, and links among the nodes represents the follow-following relationship (i.e., who is following whom) among the users. The subject of this book is online social networks.

SOCIAL NETWORK SITE A social network site is a special-purpose software (or social media tool) designed to facilitate the creation and maintenance of social relations. Facebook, Google+, and LinkedIn are examples of social network sites. Different forms of social networks are discussed in a later section.

SOCIAL NETWORKING The act of forming, expanding, and maintaining social relations is called social networking. Using social network sites, users can, for example, form, expand, and maintain online social ties with family, friends, colleagues, and sometimes strangers.

SOCIAL NETWORK ANALYSIS Social network analysis is the science of studying and understanding social networks and social networking. It is a well established field with roots in a variety of disciplines including Graph Theory, Sociology, Information Science, and Communication Science.

Learning from this lecture: Learners should be able to understand Learners shall be able to understand Scale Issues and Common Network Terms

Lecture 8

Learning Objective: Learners shall be able to understand the types of network and common network terminologies.

Types of Networks, Common Network Terminologies

Common Social Media Network Types

1. FRIENDSHIP NETWORKS
2. FOLLOW-FOLLOWING NETWORKS
3. FAN NETWORK
4. GROUP NETWORK
5. PROFESSIONAL NETWORKS
6. CONTENT NETWORKS
7. DATING NETWORKS
8. COAUTHORSHIP NETWORKS
9. COCOMMENTER NETWORKS
10. COLIKE
11. CO OCCURRENCE NETWORK
12. GEO COEXISTENCE NETWORK
13. HYPERLINK NETWORKS

Learning from this lecture: Learners should be able to understand the types of network and common network terminologie**s.**

Lecture 9

Learning Objective: **Learners shall be able to understand Network Analytics Tools**

Network Analytics Tools

- NodeXL: NodeXL (an add-in for Microsoft Excel) is the free tool for social network analysis and visualization. It can help you construct and analyze Facebook networks (based on colikes and cocomments), Twitter networks (followers, followings, and tweets), and YouTube networks (user network and comments), among others.
- UCINET: UCINET is a social network analysis software application for windows operating system. It also includes Netdraw tool for network visualization. It can be downloaded and used for free for 90 days. https://sites.google.com/site/ucinetsoftware/home.
- Pajek: Pajek is a software application for analyzing and visualizing large networks (http://mrvar.fdv.uni-lj.si/pajek/). Pajek runs on Microsoft Windows operating systems and is free for noncommercial use.
- Netminer: Netminer (http://www.netminer.com/) is also a software application for large social network analysis and visualization. The application can used be for free for 28 days.
- Flocker: Flocker (http://flocker.outliers.es/) is a Twitter real-time retweets and mentions networks analytics tool.
- Reach: Reach is an online platform to map hashtag networks and identify the most influential accounts in the Twitter conversation: http://www.reachsocial.com/.
- Mentionmapp: This online tool is used to investigate Twitter mentions networks (http://mentionmapp.com)

Learning from this lecture: Learners should be able to understand the Network Analytics Tools

Objective Question:

What is the purpose of social media network analytics?

a) To create social media profiles

b) To generate random data

c) To analyze and interpret data from social media platforms

d) To design social media platforms

Which centrality measure focuses on how close a node is to all other nodes in a network?

a) Degree centrality

b) EigenVector centrality

c) Closeness centrality

d) Betweenness centrality

Which type of network visualization layout arranges nodes in a circular pattern?

a) Force-directed layout

b) Hierarchical layout

c) Circular layout

d) Spectral layout

What is the primary characteristic of a power-law degree distribution in a network?

a) All nodes have the same degree

b) A few nodes have high degrees, while most have low degrees

c) Nodes are arranged in a hierarchical structure

d) Nodes are connected in a circular pattern

Which term refers to the intensity or closeness of the relationship between two individuals in a social network?

a) Tie strength

b) Centrality measure

c) Node degree

d) Network density

Subjective Questions:

1. Explain the concept of tie strength in a social network. Provide an example to illustrate strong and weak ties.
2. Describe the differences between the regular degree distribution and the power-law degree distribution in a social network. How does each distribution affect network dynamics?
3. Discuss the challenges of visualizing large-scale social networks. Provide strategies and techniques that can be used to address these challenges.

4. Define centralization in the context of a social network. How does a highly centralized network differ from a decentralized network? Provide real-world examples to support your explanation.
5. Choose one common type of social media network (e.g., friendship network, follow-following network) and explain its characteristics and significance. How might organizations or individuals benefit from analyzing this type of network?