Module 2

Social Network Structure, Measures & Visualization

Introduction

- 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 aspects of social media network analytics

1. Data Collection
2. Monitoring and Listening
3. Performance Metrics
4. Audience Insights
5. Competitor Analysis
6. Sentiment Analysis
7. Influencer Marketing
8. Campaign Measurement
9. Social Listening for Crisis Management

Types of Social Network Analysis

- Ego network Analysis
- Complete network Analysis

Common Networks Terms

- NETWORK
- SOCIAL NETWORKS
- SOCIAL NETWORK SITE
- SOCIAL NETWORKING
- SOCIAL NETWORK ANALYSIS

COMMON SOCIAL MEDIA NETWORK TYPES

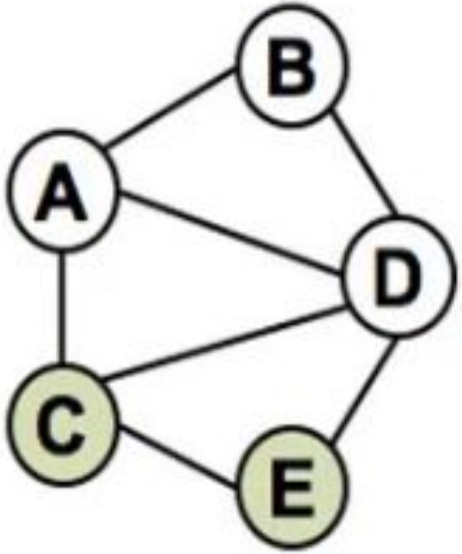
1. FRIENDSHIP NETWORKS
2. FOLLOW-FOLLOWING NETWORKS
3. FAN NETWORK
4. GROUP NETWORK
5. PROFESSIONAL NETWORKS
6. CONTENT NETWORKS
7. DATING NETWORKS
8. COAUTHORSHIP NETWORKS
9. COCOMMENTER NETWORKS
10. COLIKE
11. CO OCCURRENCE NETWORK
12. GEO COEXISTENCE NETWORK
13. HYPERLINK NETWORKS

Case Study

Investigating the Spread of Misinformation on Facebook: A Case Study on Vaccine-related Content

Network Theory

Nodes



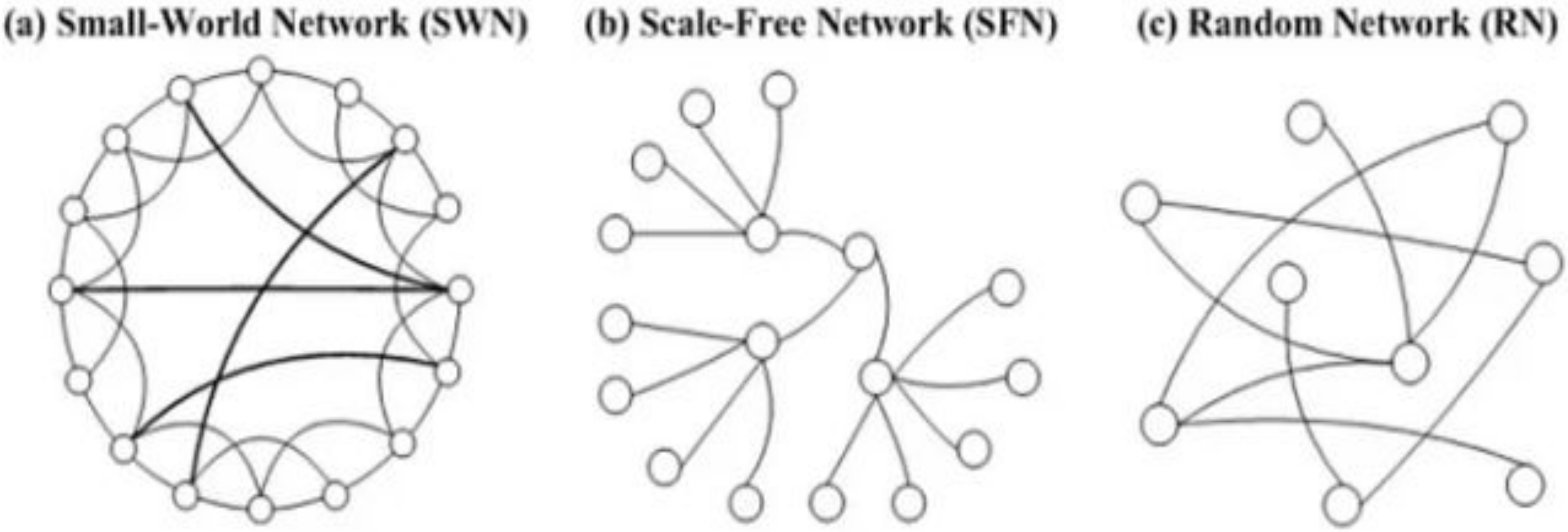
Edges

Network Theory

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).

Real World Networks



Real World Networks

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.**

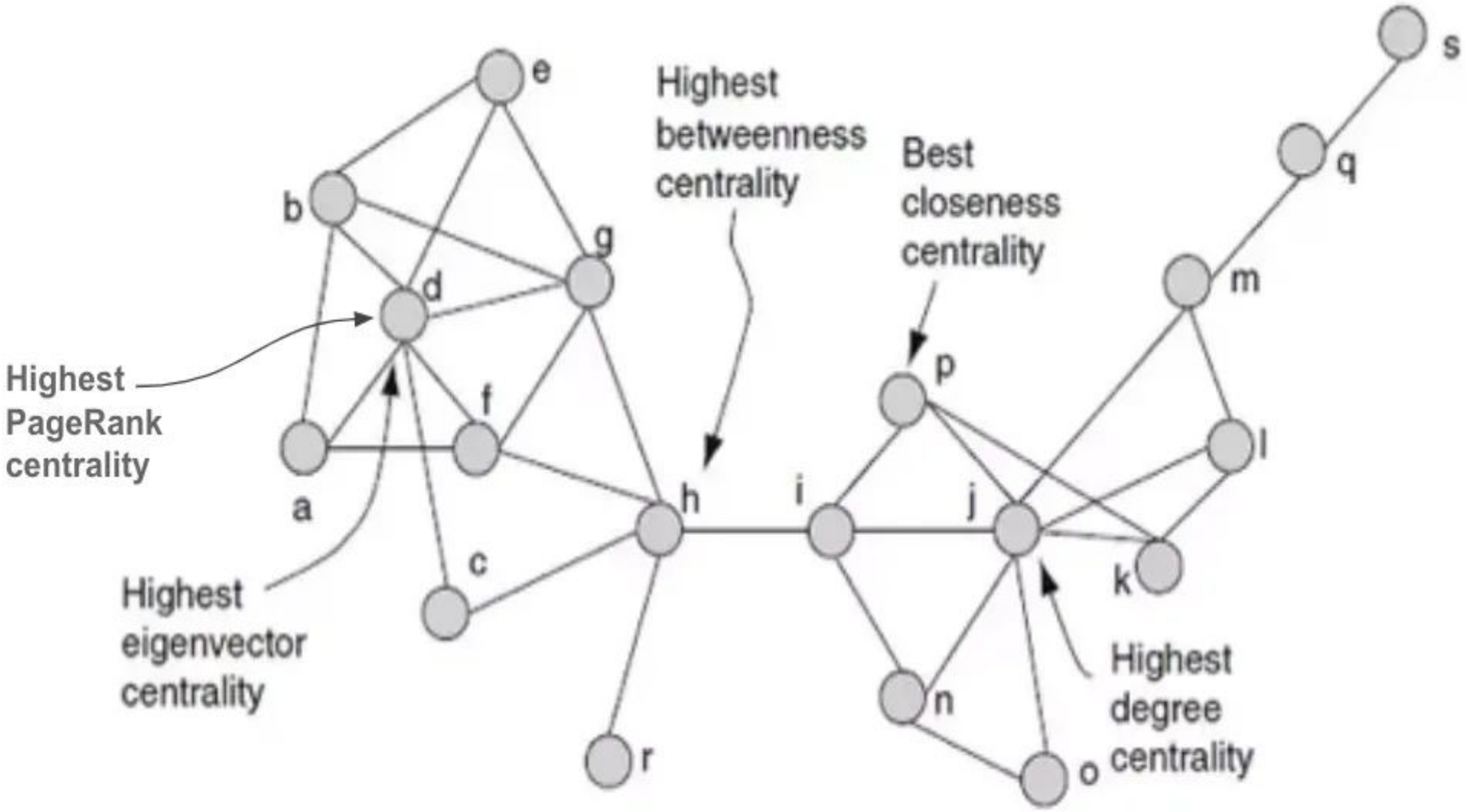
Random Network: A random network is like connecting people by flipping a coin. In a random network, nodes are linked together purely by chance, like throwing dice. There's no specific pattern or preference in creating connections. Random networks are used as a baseline t**o understand what features in real networks are due to chance and what might have some meaningful structure or organization**.

Centrality Measures

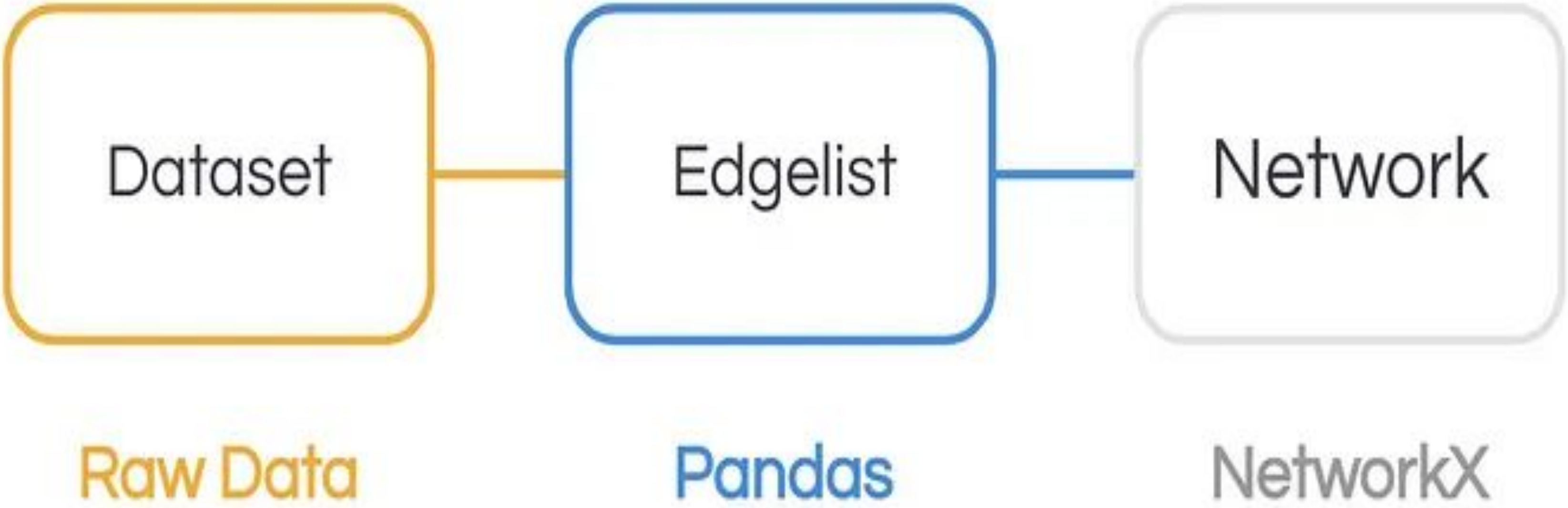
- **Degree** — the amount of neighbors of the node
- **EigenVector / PageRank** — iterative circles of neighbors
- **Closeness** — the level of closeness to all of the nodes
- **Betweenness** — the amount of short path going through the node

Centrality Measures

- **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.


Highest PageRank centrality
Highest eigenvector centrality
Highest betweenness centrality
Best closeness centrality
Highest degree centrality
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s

Building a Network



Degree Distribution

1. Definition: Degree distribution is a concept that helps us understand the **number of connections each node has in a social network.** In other words, it represents the pattern of degrees across all the nodes in the network.

2. Node Degree: **The degree of a node refers to the number of connections it has with other nodes. I**n the context of social networks, this could represent the number of friends, followers, or interactions an individual has within the network.

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.

Degree Distribution

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 Resilienc**e: 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 2 (connections with A and C)
- Node C: Degree 3 (connections with A, B, and D)
- Node D: Degree 2 (connections with C and E)
- Node E: Degree 2 (connections with D and F)
- Node F: Degree 1 (connection with E)

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.**

Density

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.

Density

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.

Density

Calculating Density:

Density = (2 * Number of Edges) / (Number of Nodes * (Number of Nodes - 1))

The network's density can be calculated as follows:

1. Number of Nodes (N) = 6
2. Number of Edges (E) = 7 (A-B, A-C, B-C, C-D, C-E, D-E, E-F)
3. Density = (2 * 7) / (6 * (6 - 1)) = 14 / 30 ≈ 0.467

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

Degree

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)

```
A -- B -- C

|    |    |

D -- E    F

|    |    |

G -- H -- I
```

Density

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

```
A -- B -- C

|       |

D -- E    F

|       |

G -- H -- I
```

Connectivity

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.

Connectivity

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.

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

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.

Tie Strength

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.

In an environment where face-to-face interactions are absent, building and maintaining trust takes on a different dynamic.

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

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.

Network Visualization - Graph Layout

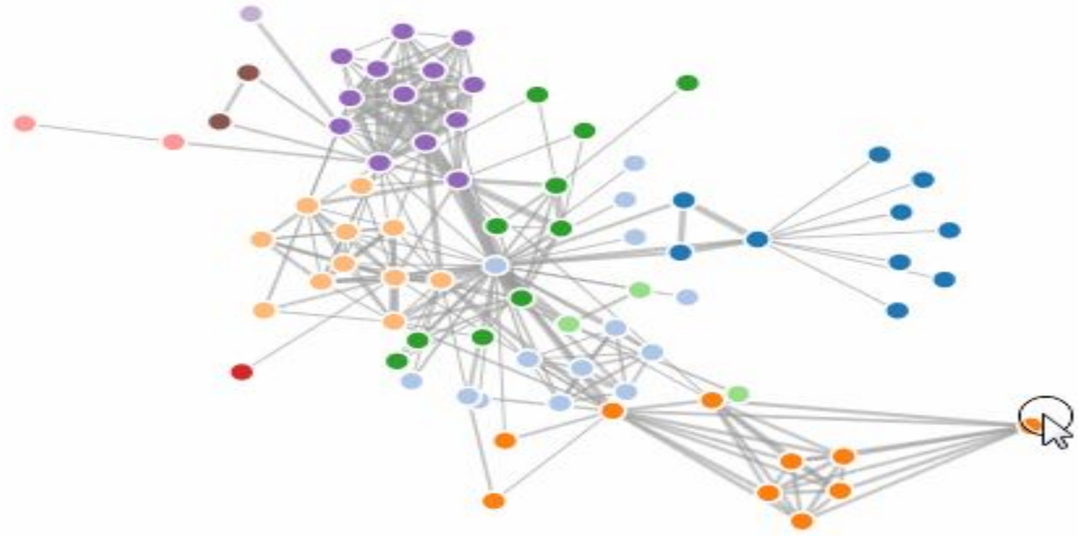
Force-Directed Layout: This is one of the most popular graph layout algorithms for social network visualization.

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.

Network Visualization - Graph Layout



Network Visualization - Graph Layout

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.

Network Visualization - Graph Layout

Network Visualization - Graph Layout

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.

Network Visualization - Graph Layout

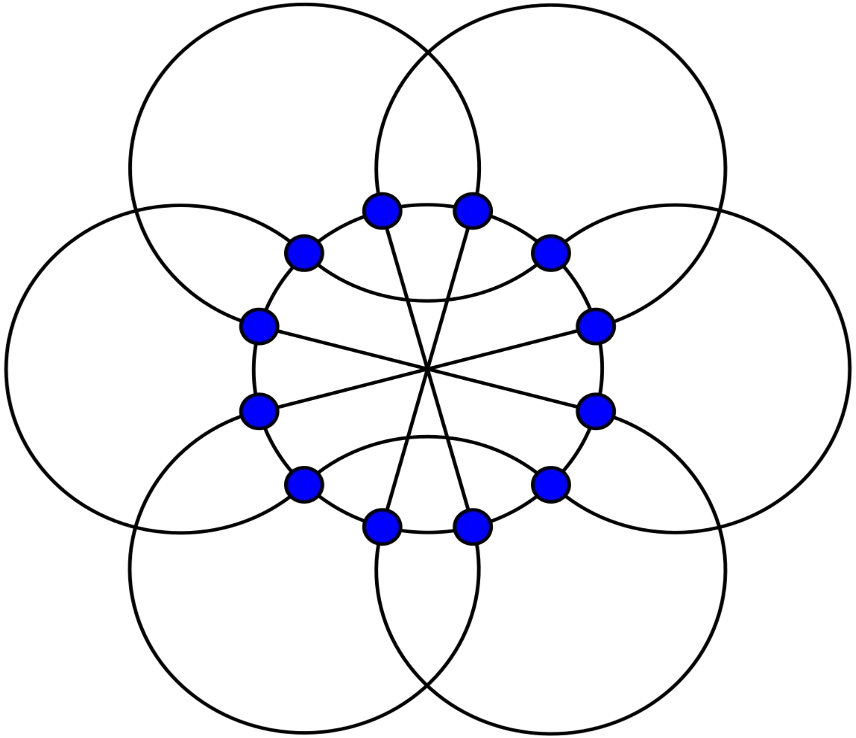
Friend D

|

Friend A-You-Friend B

|

Friend C



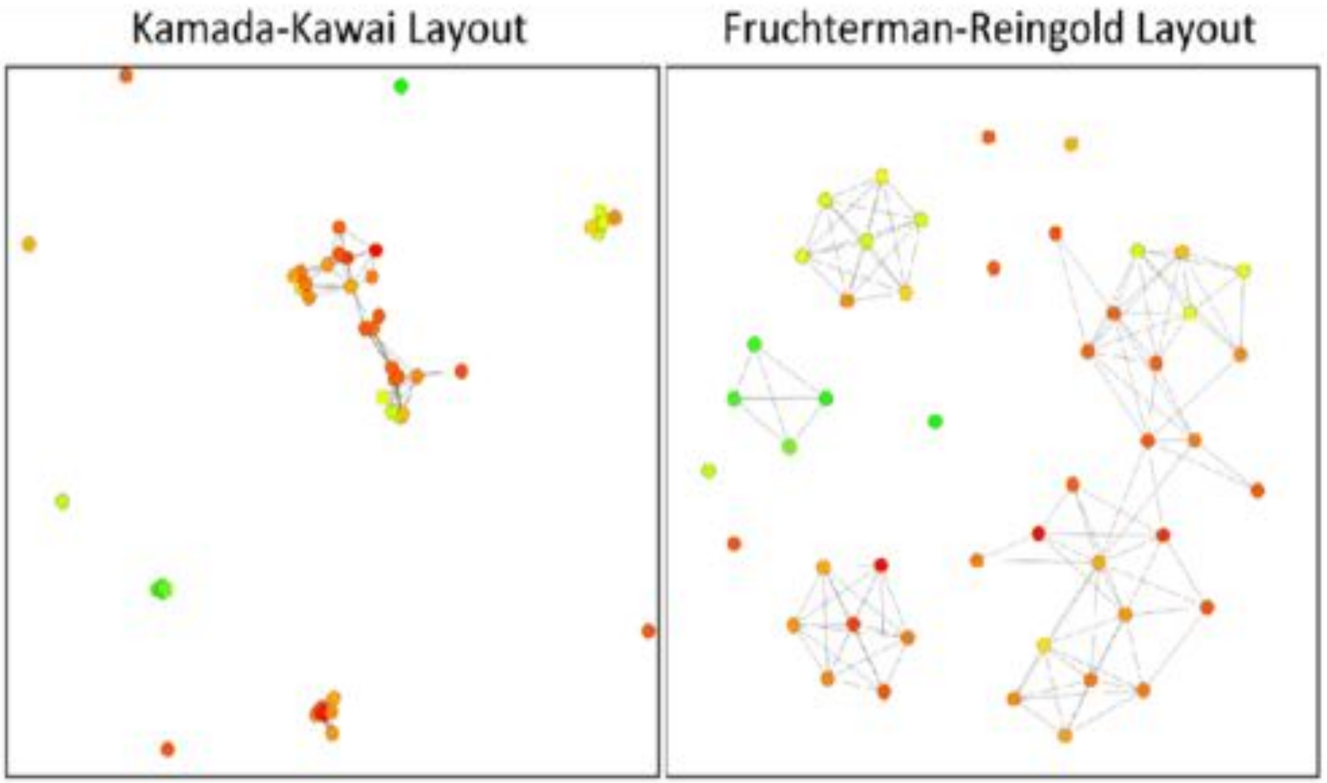
Network Visualization - Graph Layout

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.


Kamada-Kawai Layout
Fruchterman-Reingold Layout

Network Visualization - Graph Layout

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.

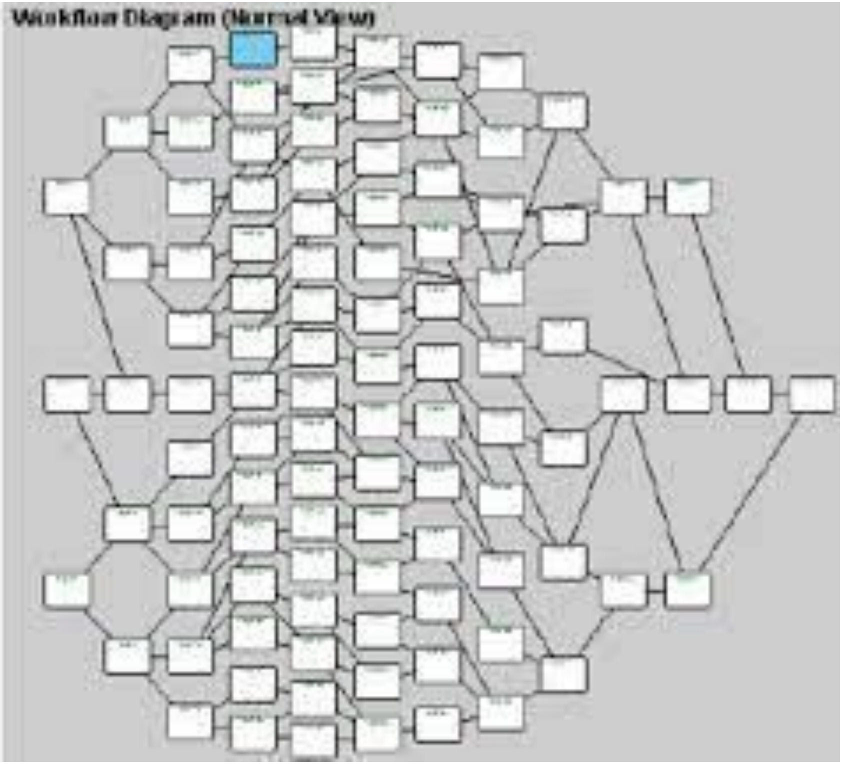
Network Visualization - Graph Layout

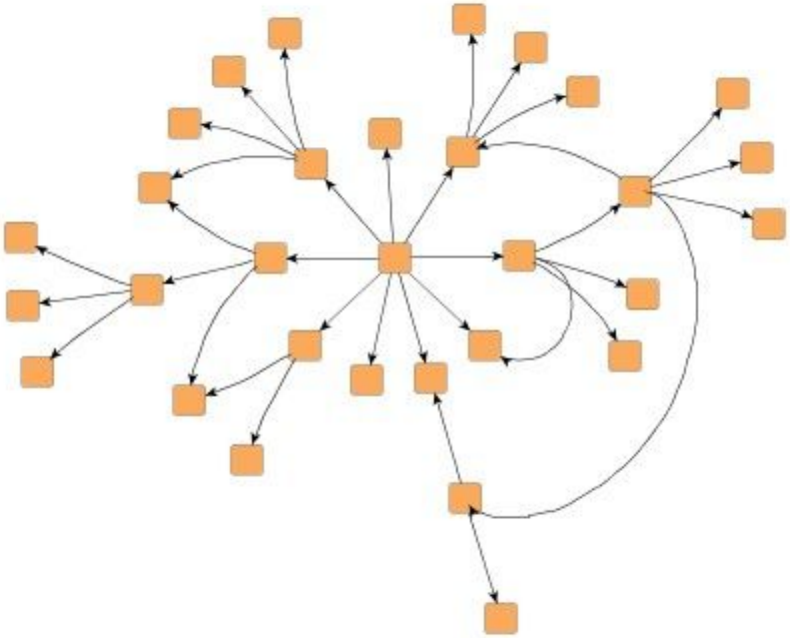
Sugiyama Layout: Also known as layered or hierarchical layout, it's often used for directed acyclic graphs (DAGs) and organizes nodes into layers with edges flowing from top to bottom.

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.

Network Visualization - Graph Layout





Visualizing Network features

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

Scale Issues

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.

Scale Issues

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.

Scale Issues

1. **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.
2. **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.
3. **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.