4. Social Network Analysis

EN6106 – Emerging Topics in Information Technology

Level III – Semester 6

Overview

4.1. Introduction to Social Networks

4.2. Goals of Analysis

4.3. Variables and Relations

4.4. Mathematical Foundations

4.5. Data Collection

4.6. Data Management

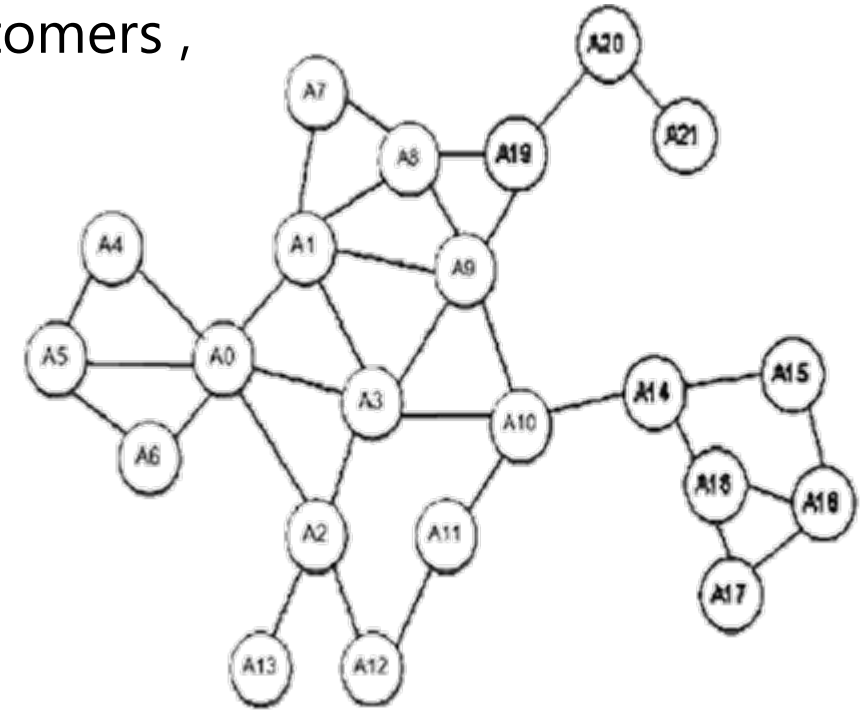
4.7. Multivariate Analysis

4.8. Visualizations

4.1 What is a Network ?

- Network Represent ,
  - Actors nodes vertices / points /Ties / edges / arcs / lines / links
- Connecting with pairs of actors
- Types of social relationships

  Family , Friendship , Customers ,



4.1 Introduction to Social Networks

- Social network analysis (SNA) is a field that uses mathematical and statistical tools to analyze the relationships and interactions among individuals, groups, or entities. In SNA, each individual or entity is represented as a node or vertex, and the connections or relationships between them are represented as edges or links.

4.1 Introduction to Social Networks

- SNA can be used to study a wide range of social phenomena, such as communication patterns, influence, collaboration, social support, and the spread of information or disease. By analyzing the patterns and structures of social networks, researchers can gain insights into how groups function, how ideas or behaviors spread, and how social change happens.

4.2 Goals of Analysis

- The goal of social network analysis is to gain insight into a community by examining the relationships that exist between its members, represented as a network.
- The analysis involves identifying the key individuals and groups within the network, as well as the connections and associations between them, which are also known as "components".
- Networks consist of nodes (points) that are linked by connections.
- In social network analysis, nodes represent people, and the links between them can include various social connections, such as friendships, family ties, or financial relationships.

4.2 What can social network analysis do?

- Social network analysis is a useful tool for understanding the scope and impact of gangs.
- It can provide insights into the reach and activity of gangs, as well as their influence on individuals and communities.
- One potential application of social network analysis is to identify individuals who may be at risk of becoming involved with gangs or being exploited by them.

4.2 What can social network analysis do?

Network analysis can be done "qualitatively" (i.e., by hand-drawn diagrams), but a more systematic approach can provide numerous benefits, including:

- Practicality: The approach is objective and replicable, and doesn't require extensive knowledge or training.
- Wider applications: The approach can help understand local gang issues and relationships with gang-associated individuals, and can be used for community impact statements and interventions like gang injunctions.
- Targeted responses: The approach can help differentiate between core and peripheral gang members, allowing for more effective work in tackling gangs and gang culture.
- Multiple uses: The analysis can be centrally collected and provided to local teams, who can then examine and manipulate the networks as needed to address particular issues.

4.3 Variables and Relations

- Network – Graphs
- Nodes – Verticals , Actors
- Links – Edges , relations
- Clusters - Communities

4.3 Variables and Relations

Visualizing the relationships between the selected entity and all linked entities is made possible through the use of the Social Network graph.

- The Social Network graph shows entity-to-entity links and relationship clusters.
- Entity-to-entity links display all entities related to the main entity but without the linking attributes.
- Attribute information can be accessed through the Attribute Explorer tool.
- Relationship clusters group related entities and help identify patterns among clusters and relationships.

4.3 Graph

- A graph is a visual representation of a set of objects with links connecting pairs of objects.
- The objects are represented as points called vertices and the links between them are called edges.
- A graph consists of a pair of sets (V, E).
- V is the set of vertices and E is the set of edges connecting pairs of vertices.

4.3 Graph

- Vertex: Each node in a graph is represented as a vertex. It can be identified using an array index.
- Edge: A line or path connecting two vertices is called an edge. It can be represented using a two-dimensional array.
- Adjacency: Two vertices are adjacent if they are connected to each other through an edge.
- Path: A sequence of edges between two vertices is called a path.

4.3 Clusters

- Cluster analysis is important in detecting graph elements with "similar" properties, especially in large networks where identifying specific patterns or structures quickly is crucial.
- The process of grouping together elements/entities based on similarity measures is called cluster analysis.
- Similarity measures are usually calculated based on topological criteria, node location, or other characteristics of graph elements.

4.3 Clusters

- Nodes that are considered similar based on these measures are grouped into clusters.
- Each cluster contains elements that share common properties and characteristics.
- The collection of all clusters forms a clustering.

4.4 Mathematical Foundation

- Graph Theory
- Centrality Measures
- Matrix Algebra

Graph Theory

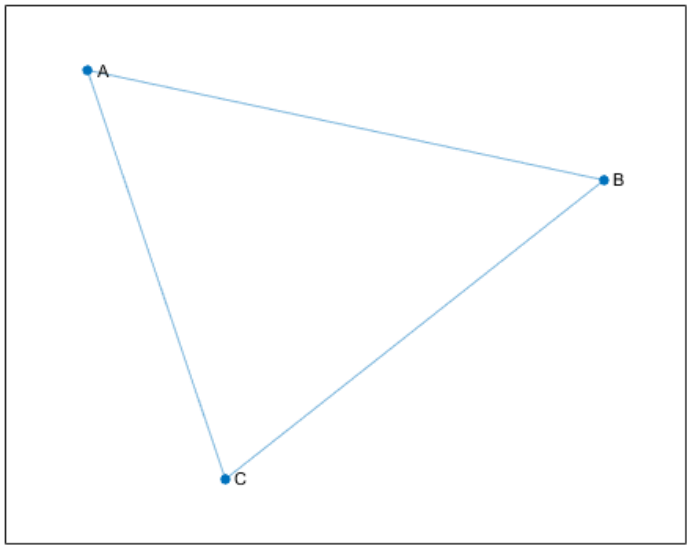
Definition of a graph, nodes, and edges.

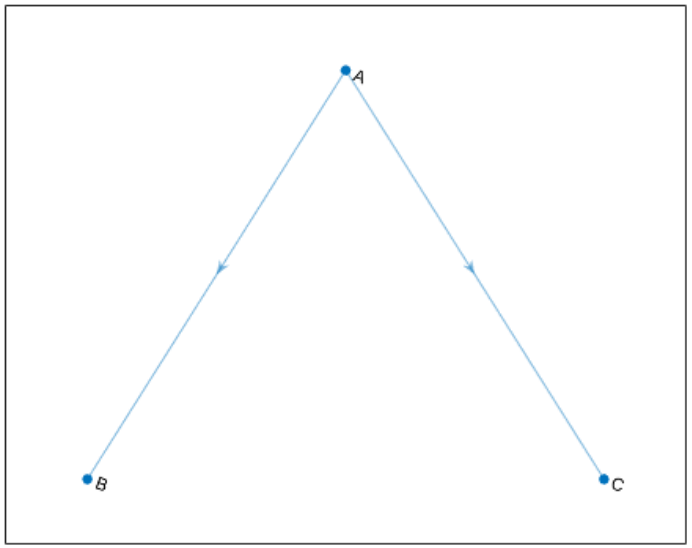
- Graph theory is the branch of mathematics that is most relevant to social network analysis. A graph is a set of nodes (also called vertices) connected by edges (also called links). In social network analysis, nodes represent individuals or groups, and edges represent relationships between them.

Types of graphs: directed and undirected.

- There are different types of graphs, including directed graphs (where the edges have a direction) and undirected graphs (where the edges do not have a direction).
- Undirected Graphs

Directed Graphs





Degree of a node and its significance in network analysis.

- You should also be familiar with the concepts of degree (the number of edges incident to a node)

Paths and their relevance in analyzing connections between nodes.

- path (a sequence of nodes connected by edges).

Centrality Measures

- Another important concept in social network analysis is centrality. Centrality measures the importance of a node in a network. There are different types of centrality measures, including degree centrality, betweenness centrality, and eigenvector centrality.

Introduction to centrality and its importance in social network analysis.

- Centrality measures identify influential individuals, information brokers, and key connectors in social network analysis.
- They provide insights into information flow, power distribution, and control within the network.
- Different centrality measures (degree, betweenness, eigenvector, closeness) capture various aspects of node importance.
- Centrality analysis helps understand network structure, dynamics, and key actors.
- It finds applications in social sciences, organizational studies, marketing, and public health.
- Centrality analysis enhances our understanding of social interactions and network behavior.

Degree centrality: measuring node importance based on the number of edges.

It measures the number of connections or links a node has. Nodes with high degree centrality are highly connected and may have a greater influence or access to information.

Betweenness centrality: measuring node importance based on its position in connecting other nodes.

- It measures the extent to which a node lies on the shortest paths between other nodes. Nodes with high betweenness centrality act as bridges, controlling the flow of information between different parts of the network.

Eigenvector centrality: measuring node importance based on the importance of its neighboring nodes.

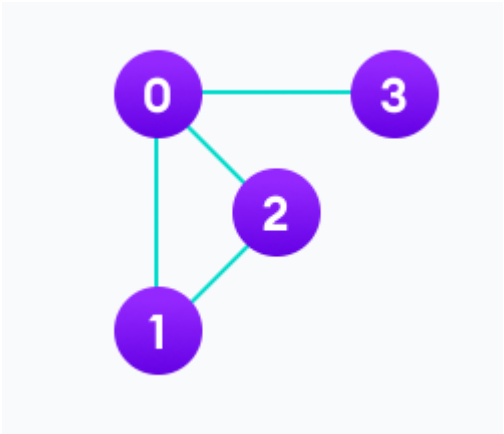
It considers both the number of connections a node has and the importance of those connections. Nodes with high eigenvector centrality are connected to other important nodes, indicating their overall influence or prestige within the network.

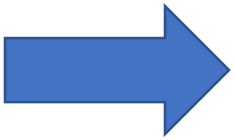
Closeness Centrality:

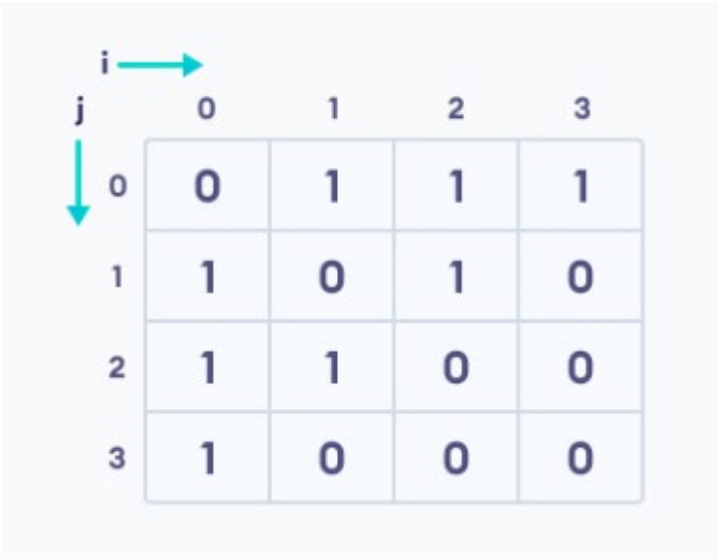
It measures how quickly a node can access or reach other nodes in the network. Nodes with high closeness centrality have shorter average path lengths to other nodes, enabling efficient information flow.

Matrix Algebra

Adjacency matrix: An adjacency matrix is a square matrix that represents a graph using boolean values (0's and 1's). Each element in the matrix indicates whether there is a direct path or edge between two vertices in the graph.







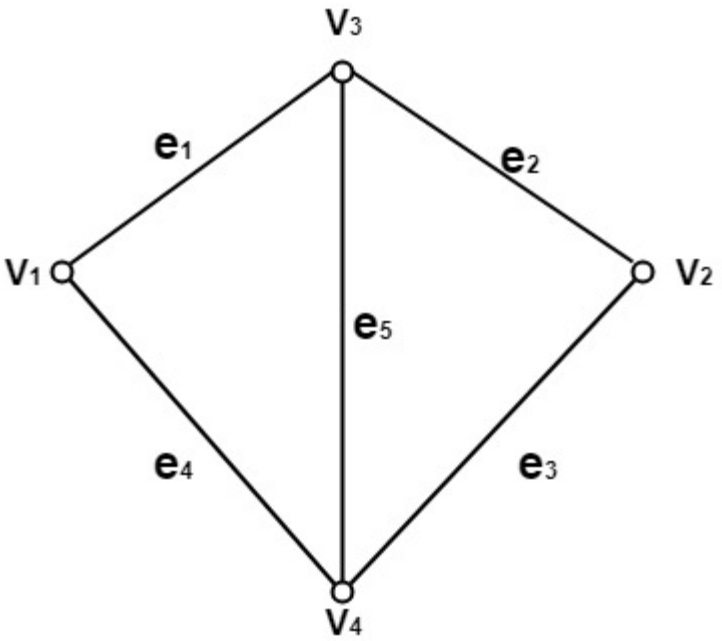
Incidence matrix: representation of connections between nodes and edges.

- If an Undirected Graph G consists of n vertices and m edges, then the incidence matrix is an n x m matrix C = [c_{ij}] and defined by

$$c_{ij} = \begin{bmatrix} 1, \text{if the vertex } V_i \text{incident by edge } e_j \\ 0, \text{otherwise} \end{bmatrix}$$

- There is a row for every vertex and a column for every edge in the incident matrix.
- The number of ones in an incidence matrix of the undirected graph (without loops) is equal to the sum of the degrees of all the vertices in a graph.

Incidence matrix [Example]



$M_I =$

	e_1	e_2	e_3	e_4	e_5
V_1	1	0	0	1	0
V_2	0	1	1	0	0
V_3	1	1	0	0	1
V_4	0	0	1	1	1

Laplacian matrix: measuring the connectivity and properties of a graph.

Modularity matrix: analyzing the community structure within a graph.

4.5 Data Collection

Data Collection Methods

- Overview of different methods for collecting data on social networks
- Surveys as a data collection method
- Observation as a data collection method
- Online data collection methods

- Surveys for Collecting Social Network Data:
  - Surveys are commonly used to collect data on social networks by asking respondents to identify their connections and describe their relationships.
  - Advantages: Allows data collection on a large number of individuals and relationships.
  - Limitations: Subject to biases such as social desirability and recall bias.

- Observation as a Data Collection Method:
  - Observation involves directly observing individuals or groups and recording their relationship dynamics.
  - Advantages: Provides rich and real-time data on relationship nature.
  - Limitations: Time-consuming and may not be feasible in all settings.

- Online Data Collection for Social Networks:
  - Online data collection involves gathering data from platforms like social media, online communities, and forums.
  - Advantages: Provides large amounts of diverse relationship data.
  - Limitations: Potential biases like selection bias and privacy concerns.

- Specialized Data Collection Methods:
  - Specialized methods, such as snowball sampling, name generators, and name interpreters, offer alternative approaches to gathering social network data.
  - Snowball Sampling: Starting with a few individuals and expanding the network through their connections.
  - Name Generators: Asking respondents to list their contacts.
  - Name Interpreters: Having respondents interpret the names of their contacts.

- Considerations for Choosing a Data Collection Method:
  - Factors to consider when selecting a method for collecting social network data.
  - Goals: Aligning the method with the research question.
  - Target Population: Considering the characteristics and accessibility of the population.
  - Resources: Assessing available resources for data collection.
  - Strengths and Limitations: Evaluating the advantages and disadvantages of each method.

- Conclusion:
  - Recap of social network analysis and the importance of data collection.
  - Summary of the three main methods: surveys, observation, and online data.
  - Emphasis on the need to consider method appropriateness based on research goals, population, and available resources.

4.6. Data Management

Data management in social network analysis is crucial for:

- Data Organization: Structuring data coherently to analyze relationships in the network effectively.
- Data Accuracy and Reliability: Standardizing data, removing errors and duplicates, ensuring trustworthy results.
- Data Accessibility and Retrieval: Easy access to specific information, efficient retrieval for analysis.
- Data Integration: Combining data from multiple sources, ensuring compatibility and consistency.
- Data Documentation and Reproducibility: Transparent documentation for sharing, collaboration, and replicability.
- Data Security and Privacy: Protecting sensitive information, adhering to ethical guidelines and regulations.
- Overall, effective data management ensures quality, organization, accessibility, and reliability of social network analysis, leading to accurate and meaningful insights.

Definition of Data Management in Social Networks.

- Data management in the context of social networks refers to the process of organizing, maintaining, and ensuring the quality of data collected on the relationships between individuals or groups within a network.

Purpose of Data Management:

- The purpose of data management in social networks is to ensure the integrity, usability, and accessibility of collected data.
- It involves practices such as data organization, accuracy enhancement, integration of multiple datasets, documentation, and data security. Effective data management enables researchers to analyze social networks efficiently and derive meaningful insights from the data.

- Methods of Data Management:
  - Manual Data Entry:
    - Entering data into a database or spreadsheet by hand.
    - Time-consuming and error-prone method suitable for small datasets or limited resources.
  - Spreadsheets:
    - using spreadsheets like Microsoft Excel or Google Sheets for data management.
    - Organizing data in rows and columns, basic calculations, and analysis capabilities. Limitations with large datasets and complex network analysis.
  - Specialized Software:
    - specialized software (e.g., Gephi, UCINET, Pajek) designed for social network analysis.
    - Details: Programs like Gephi, UCINET, and Pajek offer advanced features, such as importing data, visualizing network graphs, and performing network analysis techniques. Steeper learning curve and advanced technical skills may be required.

Best Practices for Effective Data Management:

- Standardizing Data:
  - Applying consistent formatting to data elements (e.g., name, age, gender).
  - Ensuring uniformity and compatibility of data for analysis purposes.
- Cleaning Data:
  - Removing duplicates, checking for errors, and filling in missing data.
  - Enhancing data accuracy and reliability by eliminating inconsistencies and addressing missing information.
- Storing Data:
  - Using secure and organized data storage systems.
  - Utilizing databases or cloud-based storage to ensure data accessibility, security, and efficient retrieval.
- Documenting Data:
  - Keeping track of data sources, collection methods, and relevant information.
  - Enhancing transparency, reproducibility, and research credibility by documenting the context and details of data collection.

4.7. Multivariate Analysis

Introduction to Multivariate Analysis:

- Statistical method used to analyze relationships between multiple variables in a network.
- Understanding variable relationships, identifying patterns and trends in data.

Types of Multivariate Analysis Techniques:

- Regression Analysis: - Examining relationships between two or more variables in a network. Purpose is understanding how changes in one variable relate to changes in another variable.
- Factor Analysis: - Identifying underlying factors or dimensions in a network. Purpose is reducing data complexity by grouping similar variables together.
- Cluster Analysis: - Grouping individuals or nodes in a network based on similarities in multiple variables. Purpose is identifying subgroups or communities within a network.
- Multidimensional Scaling: - Visualizing relationships between variables in a network. Purpose is identifying patterns or clusters of variables.

Considerations for Multivariate Analysis in Social Network Analysis:

- Variable Selection: Choosing relevant variables aligned with research question.
- Statistical Methods: Employing appropriate statistical techniques to test significance and control for confounding variables.

4.8 Visualizations

Introduction to Visualization in Social Network Analysis:

- Visualization as a fundamental aspect of social network analysis, enabling exploration and understanding of network structure and patterns.
- Importance: Identifying important nodes and actors, revealing clusters and communities within the network.

Types of Visualizations in Social Network Analysis:

Node-Link Diagrams:

- Common network visualization depicting nodes as circles/dots and relationships as lines/edges. Purpose is revealing overall network structure and highlighting centrality of individual nodes. - Example: Visualizing a social media network with user profiles as nodes and connections as edges.

Matrix Diagrams:

- Visual representation of node relationships in a matrix format. Purpose is uncovering patterns and clusters within the network. - Example: Displaying a co-authorship network matrix with authors as rows and columns, indicating collaboration strength.

Heat Maps:

- Matrix diagrams using color to represent the strength of relationships between nodes. Purpose is visualizing relationship strength in an intuitive and visually appealing manner.

  Example: Illustrating the intensity of interactions between individuals in an email communication network.

Force-Directed Layouts:

- Visualization technique positioning nodes based on their relationships using algorithms. Purpose is unveiling clusters, subgroups, and overall network structure.

  Example: Mapping a social network graph where individuals closer in relationships are located near each other.

Considerations for Effective Network Visualization:

- Choosing Appropriate Visualization Technique: Selecting the visualization method that best suits the data
- Clear Labeling and Design: Using descriptive labels, suitable color schemes, and legends to ensure ease of interpretation.
- Interactivity and Exploration: Incorporating interactive features to allow users to explore and manipulate the visualization.