

[illegible]

Internal Assessment Test 1 – September 2025

Sub:			Social Network Analysis				S u b C o d e:	B A D 7 1 4 D	Branch :	AINDS / CS (DS)	
Date:				Dur atio n:	90 mi nu tes	Ma x Ma rks :	50	S e m	VII		O B E
<u>Answer any FIVE Questions</u>									MAR KS	C O B T	
1		Compare Social Network Analysis with traditional statistical analysis. What advantages does SNA provide?							[10]	1	L5
2		Define Centrality. What are different types of Centrality. Explain with examples							[10]	1	L2
3		Compare the Random Network Model and the Ring Lattice Network Model. Which better represents real-world systems?							[10]	2	L2

4	a	Explain Jaccard Similarity & Cosine Similarity in detail.	[4]	1	L2
	b	Find out Jaccard Similarity for Set A= {0,3,7,8,9} & Set B= {0,1,2,5,6,9}.	[3]	1	L4
	c	Calculate Cosine Similarity Vector A= [8,6] and Vector B= [5,5]	[3]	1	L4
5	a	Explain the concept of aging in preferential attachment.	[5]	2	L1
	b	Give an example of a real-world scenario where this effect is observed.	[5]	2	L2
6		Explain Reciprocity in directed networks with an example. How is it measured?	[10]	2	L2

Answer Key

1. Compare Social Network Analysis (SNA) with Traditional Statistical Analysis

Aspect	Traditional Statistical Analysis	Social Network Analysis (SNA)
Focus	Individual attributes and aggregate trends	Relationships and interactions between entities
Data Type	Tabular, numeric	Relational (graphs: nodes & edges)
Unit of Analysis	Independent data points	Nodes and their connections
Assumption	Data points are independent	Data points (actors) are interdependent
Techniques	Regression, correlation, ANOVA	Centrality, community detection, shortest path, clustering coefficient
Output	Summary statistics, models	Network graphs, connectivity patterns

Compare Social Network Analysis (SNA) with Traditional Statistical Analysis

(CO1 – L5: Evaluate)

Comparison:

Aspect	Traditional Statistical Analysis	Social Network Analysis (SNA)
Focus	Individual attributes and aggregate trends	Relationships and interactions between entities
Data Type	Tabular, numeric	Relational (graphs: nodes & edges)
Unit of Analysis	Independent data points	Nodes and their connections
Assumption	Data points are independent	Data points (actors) are interdependent
Techniques	Regression, correlation, ANOVA	Centrality, community detection, shortest path, clustering coefficient
Output	Summary statistics, models	Network graphs, connectivity patterns

Advantages of SNA:

- Reveals **hidden structures** and **influence patterns**.
- Helps understand **information flow**, **community formation**, and **diffusion of innovation**.
- Captures **contextual and relational dependencies**.
- Widely applicable in **social media**, **epidemiology**, **citation networks**, etc.

Conclusion:

SNA provides deeper insights into relationships and network behaviors beyond what traditional statistics can uncover.

Definition:

Centrality measures the **importance or influence** of a node in a network.

Types:**1. Degree Centrality:**

- Number of direct connections a node has.
- Example: In a social network, a person with many friends has high degree centrality.
- Formula: $CD(v) = \deg(v)$

2. Closeness Centrality:

- Measures how close a node is to all others.
- Example: A node that can reach all others quickly.
- Formula: $CC(v) = 1 / \sum_u d(v, u)$

3. Betweenness Centrality:

- Indicates how often a node lies on the shortest path between other nodes.
- Example: A broker connecting two communities.
- Formula: $\sum_{s,t} \text{st} \text{st}(v)$

4. Eigenvector Centrality:

- Considers not just number of connections but **importance of connected nodes**.
- Example: Google's PageRank is based on this.

Conclusion:

Centrality quantifies influence in various ways—degree for connectivity, closeness for reachability, betweenness for brokerage, eigenvector for prestige.

3. Random Network Model vs Ring Lattice Network Model

Aspect	Random Network (Erdős–Rényi)	Ring Lattice Network (Watts–Strogatz base)
Structure	Nodes connected randomly	Nodes connected to fixed nearest neighbors
Clustering Coefficient	Low	High
Average Path Length	Short	Long
Realism	Models randomness	Models locality
Example	Randomly connected servers	People connected to their nearest friends

Which represents real-world systems better?

Out of Ring Lattice & Random Networks, Random Networks are close representative of real-world networks due to short Average path lengths.

4. (a) Jaccard Similarity & Cosine Similarity

(CO1 – L2/L4: Understand & Apply)

Jaccard Similarity:

- Measures overlap between two sets.
- Formula:

$$J(A,B) = \frac{|A \cap B|}{|A \cup B|}$$

Cosine Similarity:

- Measures the angle between two vectors in multi-dimensional space.
- Formula:

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} \Rightarrow \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$

b) Find Jaccard Similarity for

$A = \{0, 3, 7, 8, 9\}$, $B = \{0, 1, 2, 5, 6, 9\}$

- $A \cap B = \{0, 9\} \Rightarrow |A \cap B| = 2$
- $A \cup B = \{0, 1, 2, 3, 5, 6, 7, 8, 9\} \Rightarrow |A \cup B| = 9$

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{2}{9} = 0.22$$

c) Calculate Cosine Similarity for

$A = [8, 6]$, $B = [5, 5]$

$$A \cdot B = (8)(5) + (6)(5) = 40 + 30 = 70$$

$$\|A\| = \sqrt{(8^2 + 6^2)} = 10$$

$$\|B\| = \sqrt{(5^2 + 5^2)} = 7.07$$

$$70 / (10 \cdot 7.07)$$

Answer: Cosine Similarity = **0.99**

5 (a) Aging in Preferential Attachment

(CO2 – L1/L2: Recall & Understand)

Concept:

- In preferential attachment models (like Barabási–Albert), new nodes attach to high-degree nodes (“rich-get-richer”).
- Aging effect means older nodes gradually become less attractive for new links even if they have high degree.

Mathematically:

Attractiveness $A_i(t) \propto k_i(t) \times f(\text{age}_i)$, where $f(\text{age}_i)$ decreases with time.

b) Real-world Example:

- **Social media:** Older accounts (e.g., Orkut, MySpace) lose new followers to newer, trending profiles.
- **Citation networks:** Older papers get fewer new citations despite past influence.
- **YouTube:** Early viral videos lose attention to new trending content.

Conclusion:

Aging in preferential attachment models realistic social systems where popularity decays over time.

6. Reciprocity in Directed Networks

(CO2 – L2: Understand)

Definition:

Reciprocity measures how often **mutual relationships** occur in a directed network.

Formula:

$$r = L_{\leftrightarrow} / L$$

Where:

- $L_{\leftrightarrow}$ = Number of bidirectional (mutual) links
- L = Total number of directed links

Example:

- If $A \rightarrow B$ and $B \rightarrow A$, that's a reciprocal tie.
- If $A \rightarrow B$ only, it's non-reciprocal.

Interpretation:

- High reciprocity $\rightarrow$ strong mutual interactions (e.g., friendships).
- Low reciprocity $\rightarrow$ hierarchical or follower-based systems (e.g., Twitter).

Conclusion:

Reciprocity quantifies mutual connections, indicating **trust or bidirectional communication strength** in directed networks.