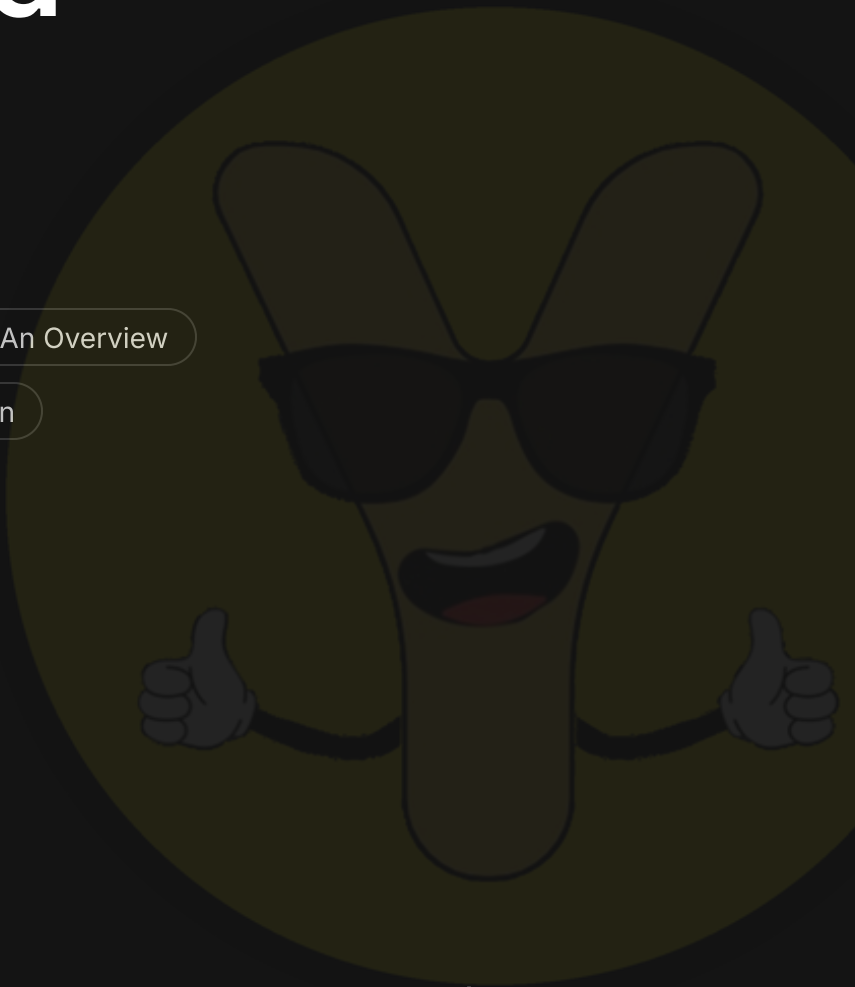
AKTU · 3RD YEAR · SEMESTER 6 · BCAM061

Social Media Analytics and Data Analysis

Unit 2 — Graphs and Matrices

- The adjacency matrix
- Paths
- Connectivity
- Distance and Breadth First Search
- Network Datasets: An Overview
- Nodes, ties, and influencers
- Making connections
- Link analysis
- Random graphs and network evolution
- Weighted Networks
- Hypergraphs



DECODED BY

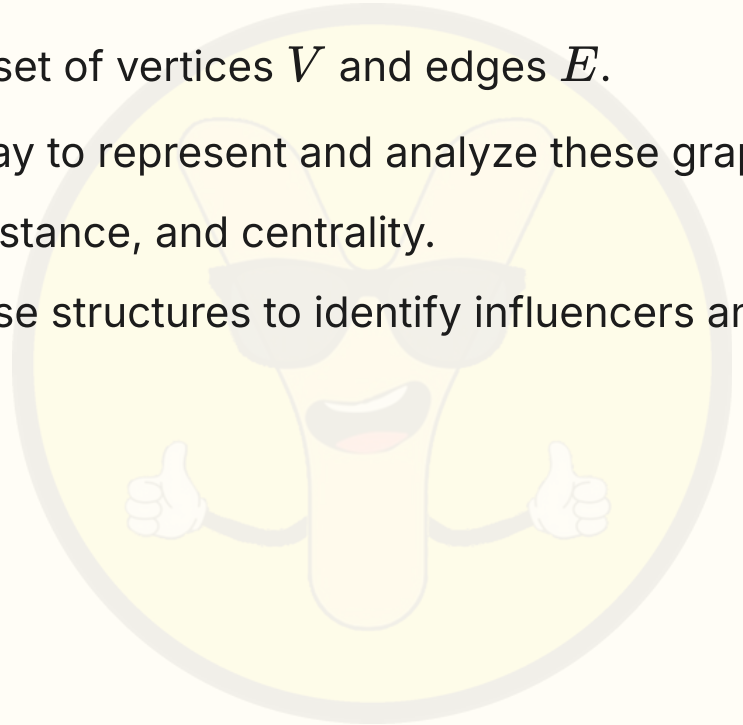
Chirag Jain

Founder, Ycotes · Gold Medalist in Artificial Intelligence, IIT Madras · taught 10,000+ students

Introduction to Graphs and Matrices

Social media networks are represented as graphs where users are nodes and interactions are edges.

- A graph $G = (V, E)$ consists of a set of vertices V and edges E.
- Matrices provide a mathematical way to represent and analyze these graph structures.
- Key metrics include connectivity, distance, and centrality.
- Social media analytics relies on these structures to identify influencers and communities.



The Adjacency Matrix

ADJACENCY MATRIX

A square matrix A of size $n \times n$ used to represent a finite graph.

- The element $A_{ij} = 1$ if there is an edge from vertex i to vertex j, and 0 otherwise.
- For undirected graphs, the matrix is symmetric ($A = A^T$).
- The diagonal elements A_{ii} are usually 0 unless self-loops exist.
- It is efficient for dense graphs but consumes $O(n^2)$ space.

Paths in Graphs

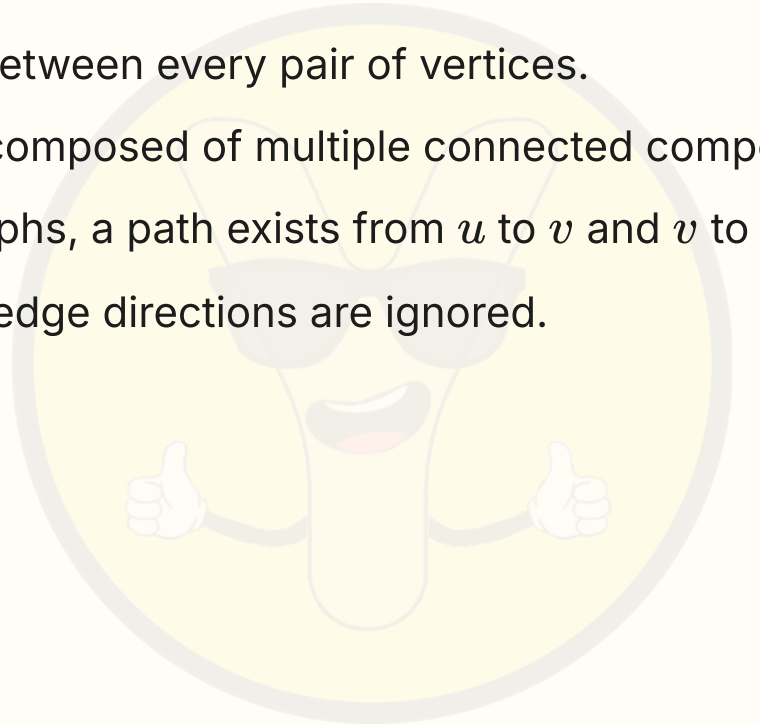
A path is a sequence of edges connecting a sequence of vertices.

- A path is a sequence of vertices $(v_0, v_1, ..., v_k)$ such that $(v_i, v_{i+1}) \in E$.
- Simple path: A path where all vertices are distinct.
- Cycle: A path that starts and ends at the same vertex.
- Path length: The number of edges in the path.
- In matrices, the number of paths of length k between i and j is given by $(A^k)_{ij}$.

Connectivity

Connectivity describes the reachability between nodes in a network.

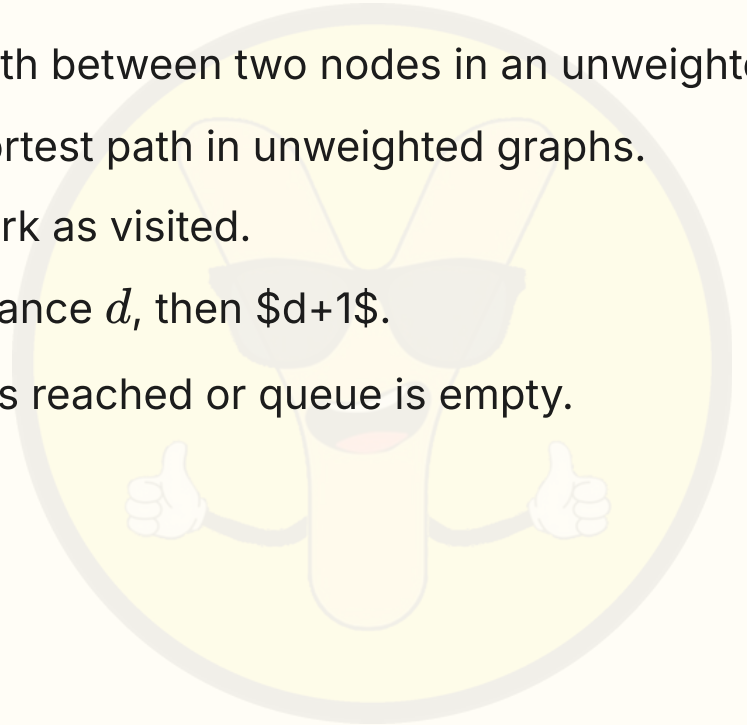
- Connected graph: There is a path between every pair of vertices.
- Disconnected graph: The graph is composed of multiple connected components.
- Strong connectivity: In directed graphs, a path exists from u to v and v to u.
- Weak connectivity: A path exists if edge directions are ignored.



Distance and Breadth First Search (BFS)

Distance is the shortest path length between two nodes.

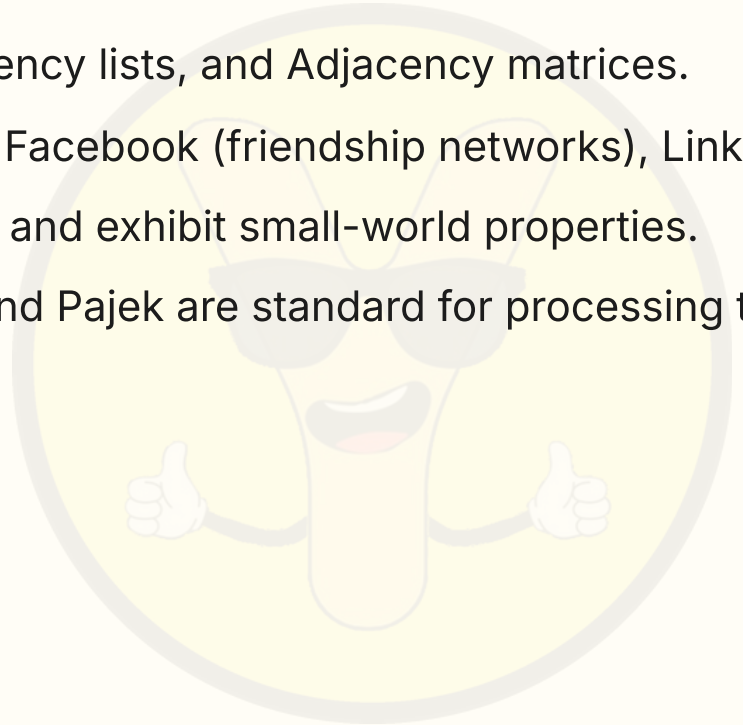
- Geodesic distance: The shortest path between two nodes in an unweighted graph.
- BFS Algorithm: Used to find the shortest path in unweighted graphs.
- Step 1: Start at source node and mark as visited.
- Step 2: Explore all neighbors at distance d, then $d+1$.
- Step 3: Stop when the target node is reached or queue is empty.



Network Datasets: An Overview

Real-world social media data often comes in specific formats.

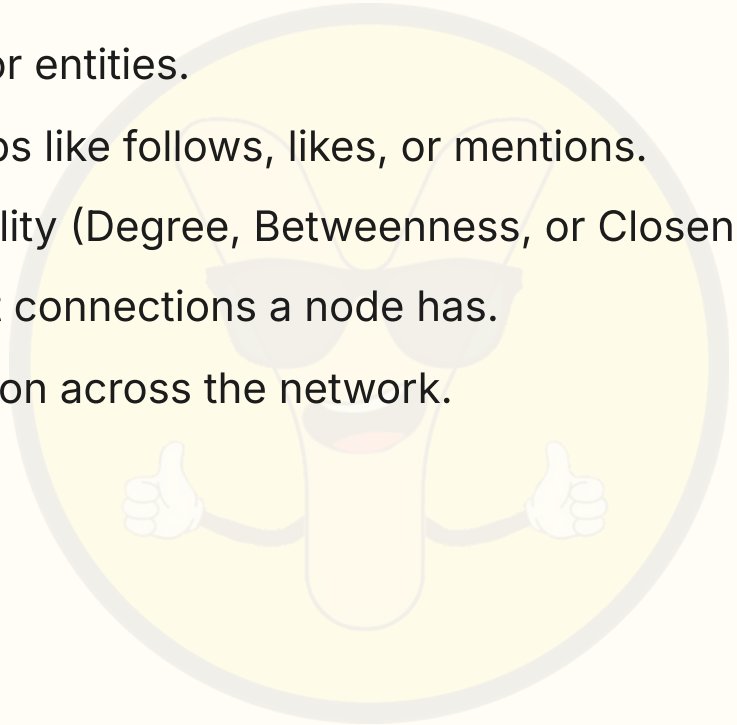
- Common formats: Edge lists, Adjacency lists, and Adjacency matrices.
- Datasets: Twitter (follower graphs), Facebook (friendship networks), LinkedIn (professional connections).
- Characteristics: Sparse, scale-free, and exhibit small-world properties.
- Tools: NetworkX (Python), Gephi, and Pajek are standard for processing these datasets.



Nodes, Ties, and Influencers

Understanding the components of social influence.

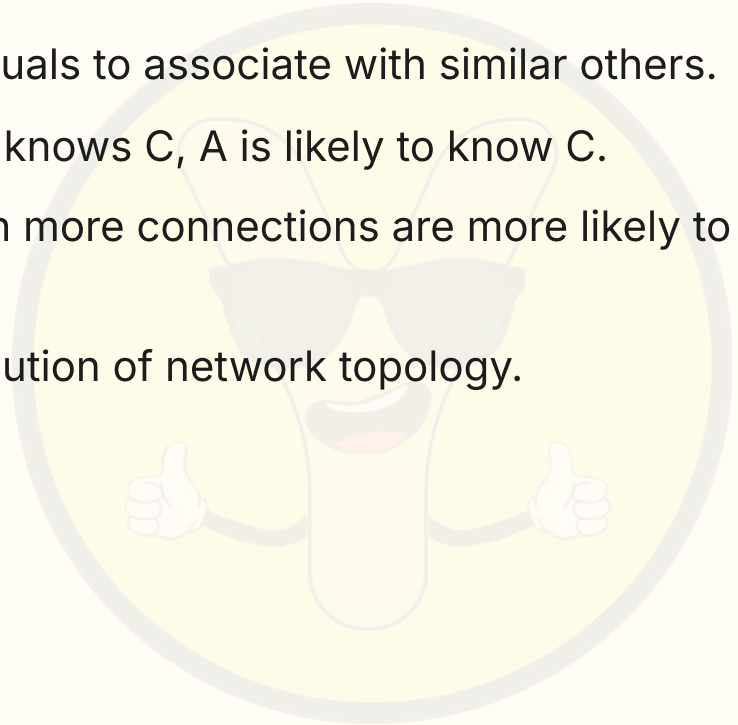
- Nodes: Represent individual users or entities.
- Ties (Edges): Represent relationships like follows, likes, or mentions.
- Influencers: Nodes with high centrality (Degree, Betweenness, or Closeness).
- Degree Centrality: Number of direct connections a node has.
- Influencers drive information diffusion across the network.



Making Connections

Mechanisms through which social networks grow.

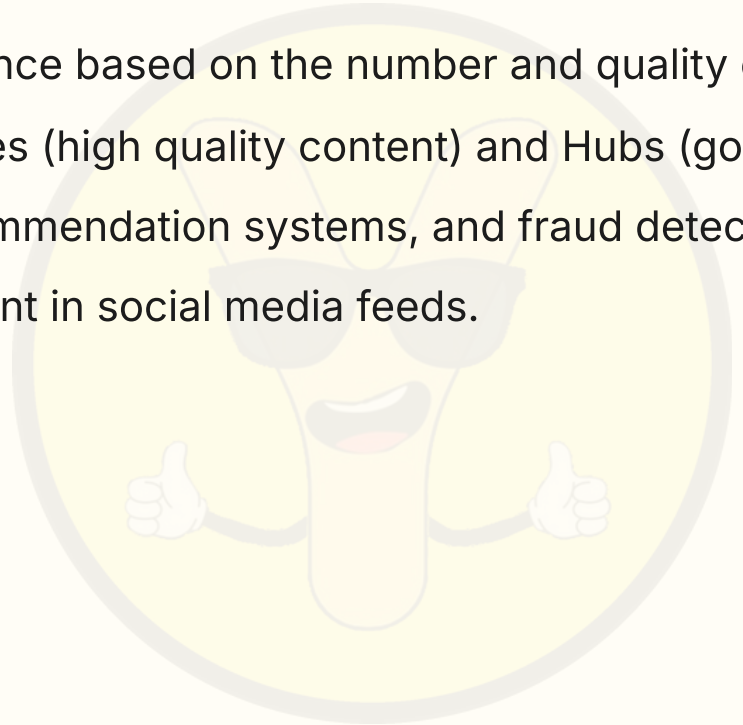
- Homophily: The tendency of individuals to associate with similar others.
- Triadic Closure: If A knows B and B knows C, A is likely to know C.
- Preferential Attachment: Nodes with more connections are more likely to receive new connections (Rich-get-richer).
- These mechanisms explain the evolution of network topology.



Link Analysis

Techniques to evaluate the importance of nodes based on their links.

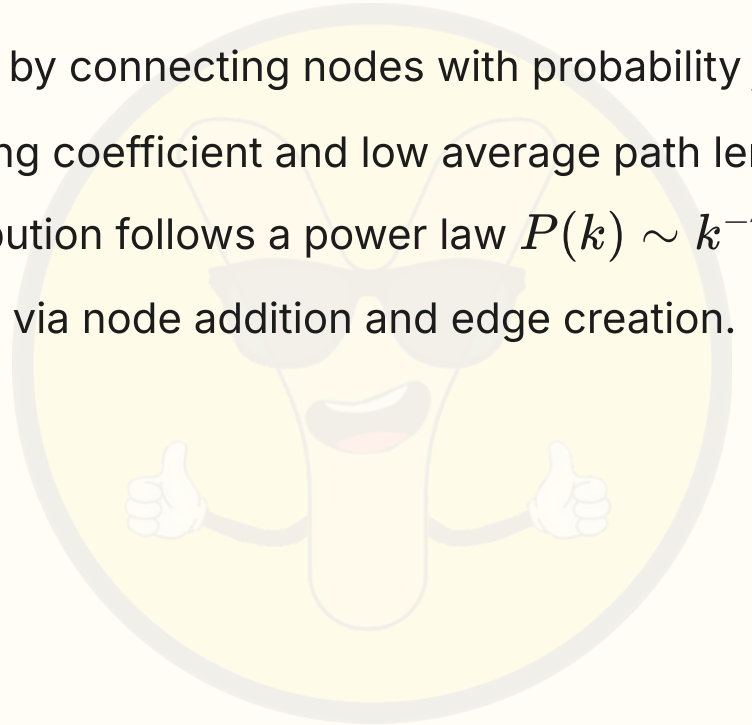
- PageRank: Measures node importance based on the number and quality of incoming links.
- HITS Algorithm: Identifies Authorities (high quality content) and Hubs (good directories).
- Applications: Search engines, recommendation systems, and fraud detection.
- Link analysis helps in ranking content in social media feeds.



Random Graphs and Network Evolution

Mathematical models to understand network growth.

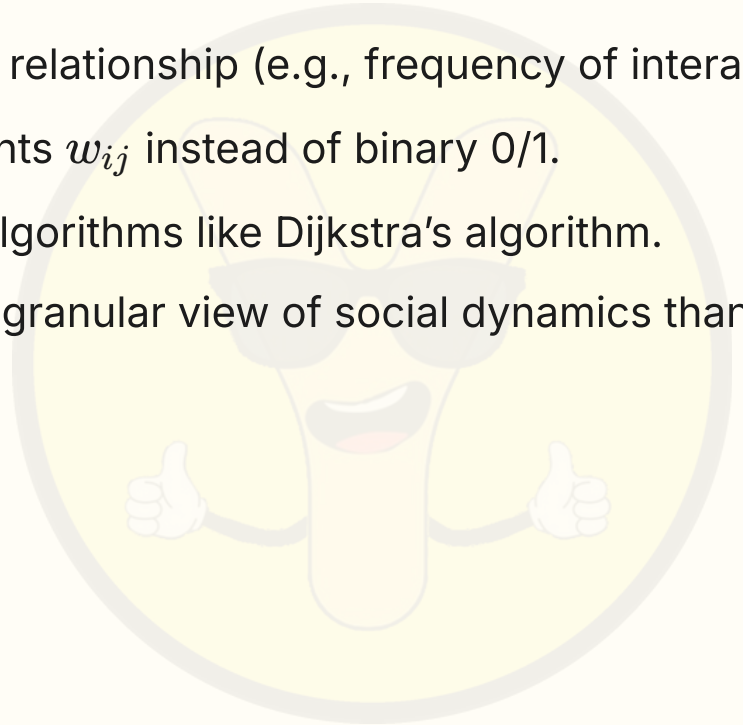
- Erdős–Rényi Model: Graphs formed by connecting nodes with probability p.
- Small-world property: High clustering coefficient and low average path length.
- Scale-free networks: Degree distribution follows a power law $P(k) \sim k^{-\gamma}$.
- Evolution: Networks grow over time via node addition and edge creation.



Weighted Networks

Networks where edges have associated values (weights).

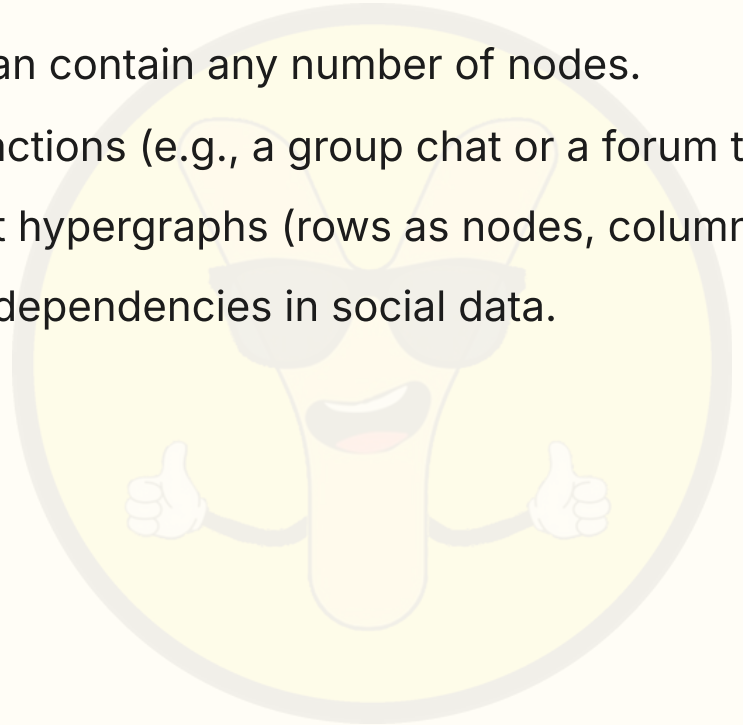
- Weight represents the strength of a relationship (e.g., frequency of interaction).
- Adjacency matrix W contains weights w_{ij} instead of binary 0/1.
- Shortest path calculation requires algorithms like Dijkstra's algorithm.
- Weighted networks provide a more granular view of social dynamics than binary graphs.



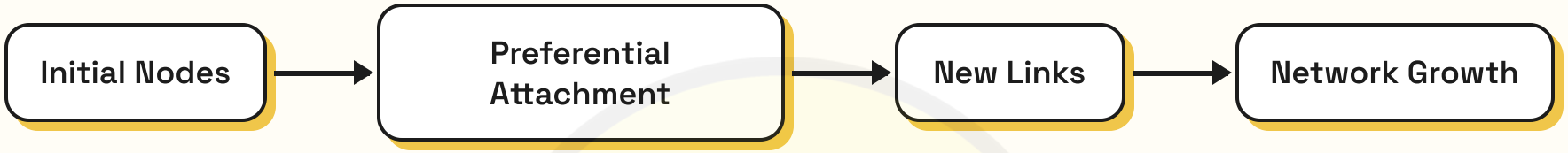
Hypergraphs

Generalization of graphs where edges can connect more than two nodes.

- Hyperedge: A set of vertices that can contain any number of nodes.
- Useful for representing group interactions (e.g., a group chat or a forum thread).
- Incidence Matrix: Used to represent hypergraphs (rows as nodes, columns as hyperedges).
- Hypergraphs capture higher-order dependencies in social data.



Network Evolution Process



The cycle of network evolution through preferential attachment.

Exam corner

- **Q1** Define Adjacency Matrix. How is it used to find the number of paths of length k?
- **Q2** Explain the difference between BFS and DFS in the context of social network traversal.
- **Q3** What is the significance of 'Triadic Closure' in network evolution?
- **Q4** Differentiate between Degree Centrality and PageRank.
- **Q5** Explain the concept of Scale-Free networks and the Power Law distribution.
- **Q6** How does a Weighted Network differ from an Unweighted Network in analysis?
- **Q7** What is a Hypergraph? Provide an example of its application in social media.

Unit 2 done. You are exam-ready.

Revise the Exam corner once more, then open Unit 3 on theycotes.com. Every AKTU subject, unit by unit, free. Notes by Chirag Jain, Ycotes.

