Social Network Analysis

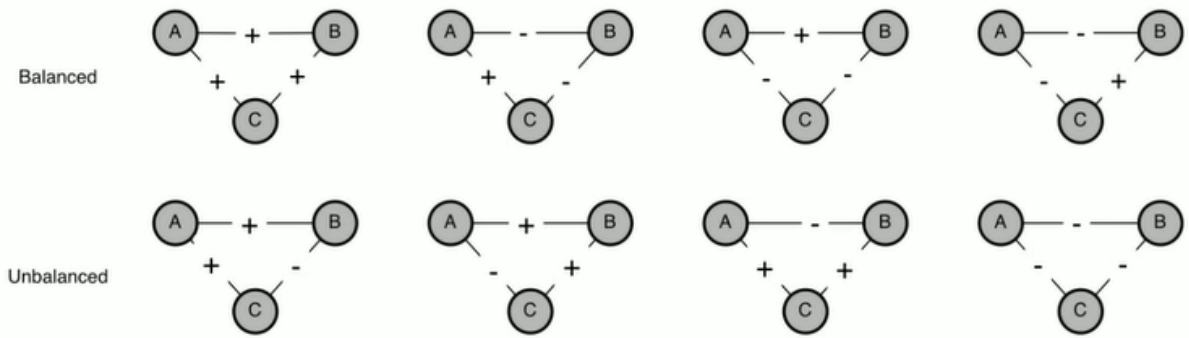
- Classic findings in networks
- Key concepts
- Collecting data
- A history of networks

Classic findings 1 of 3

The strength of weak ties

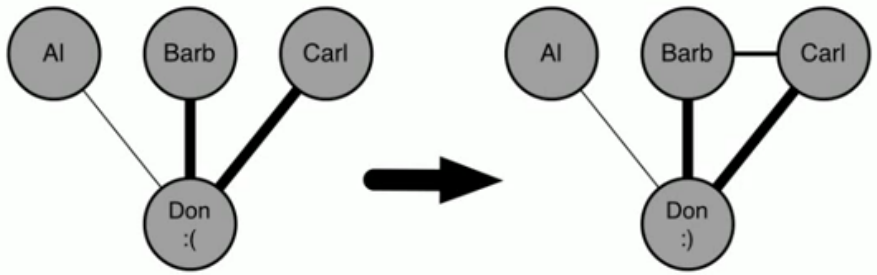
- People tend to want their close friends to know each other.
- As they ‘close these triads’ they create clustering.
- Information circulates faster within clusters.
- So to find novel information, such as a job opening, we can look to the unclosed triads of weak ties rather than the highly clustered strong ties.

The Strength of Weak Ties I: Social Balance Theory



- "The enemy of my enemy is my friend"
- Some networks are 'unbalanced', thus implying they are unstable. Others are 'balanced' and thus, stable.
- This concept came from Heider, a psychologist.

Triadic Closure



- A simplified extension to Heiderian Balance is the idea of triadic closure: The stronger an unclosed triad, the more the focal individual will want to close this connection.
- Heider originally dealt with signed ties, whereas this concept deals with weighted ties.

Classic findings 2 of 3

The six degrees of separation

- Everyone in the world is connected to at least someone else by virtual of birth.
- On average, how many hops would it take to get from one to another.
- Milgram asked people to send a package from Nebraska to Boston.
- The median chain of referrals was six. Hence the six degrees of separation.

How about 3.5 Degrees?

The majority of the people on Facebook have averages between 2.9 and 4.2 degrees of separation. Figure 1 (below) shows the distribution of averages for each person.

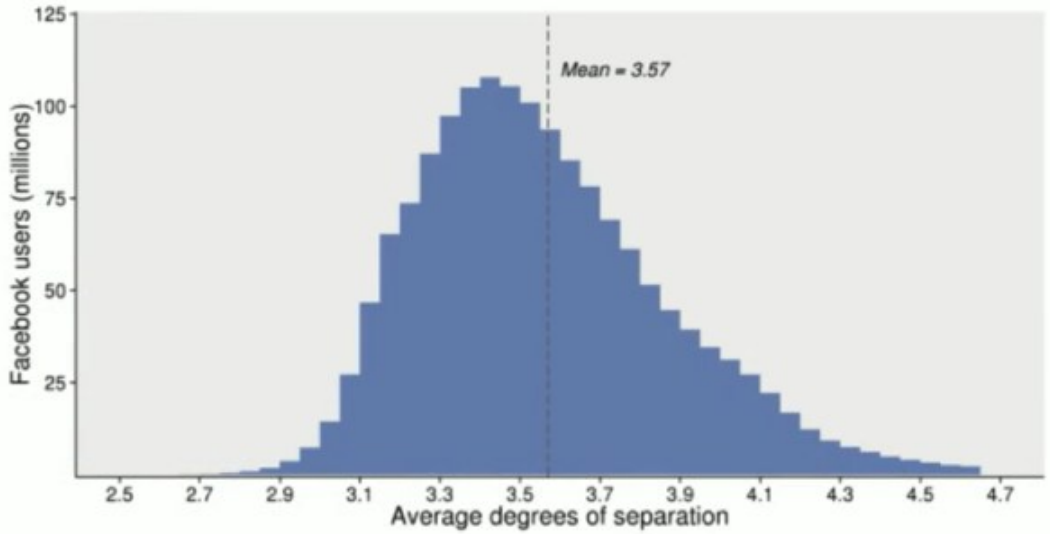


Figure 1. Estimated average degrees of separation between all people on Facebook. The average person is connected to every other person by an average of 3.57 steps. The majority of people have an average between 3 and 4 steps.

Bhagat, S., Burke, M., Diuk, C., Filiz, I.O., Edunov, S. Three and a half degrees of separation. Retrieved from https://research.fb.com/three-and-a-half-degrees-of-separation/

Classic findings 3 of 3

Community lost, Saved, or Transformed?

- With a shift to modernity via cars, telephones and urbanization, there are fears of social disintegration.
- People have asked if ‘community’ is disappearing?
- Wellman studied the personal networks of people in East York. Their networks did not overlap like a large community.
- However, everyone had resilient and supportive relationships: community has transformed to personal communities.

Terminologies in Social Network Analysis

- **Node (Vertex):** Represents an individual or entity in the network.
- **Edge (Link**): Represents a connection between two nodes.
- **Degree of a Node**: The number of connections a node has.
- **Path**: A sequence of edges connecting two nodes.
- **Geodesic Distance**: The shortest path between two nodes.

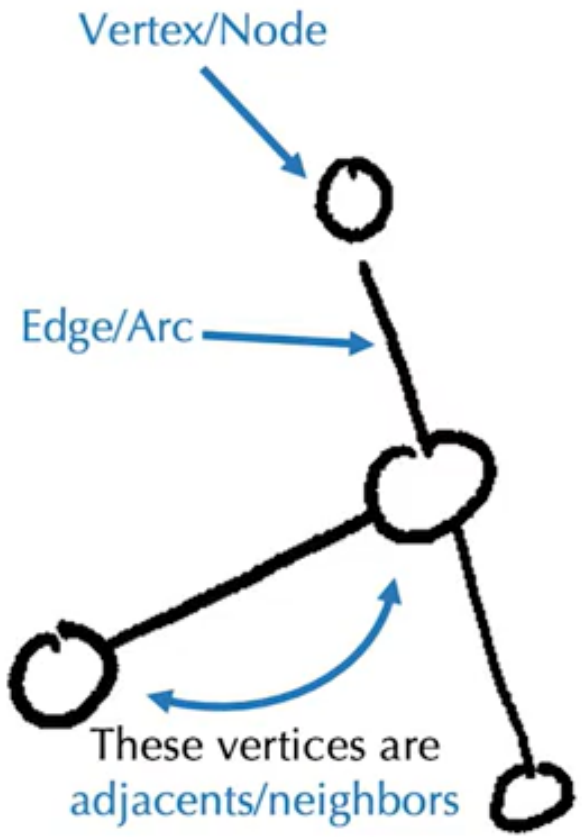
Properties of Nodes in a Social Network

Complete Mutuality

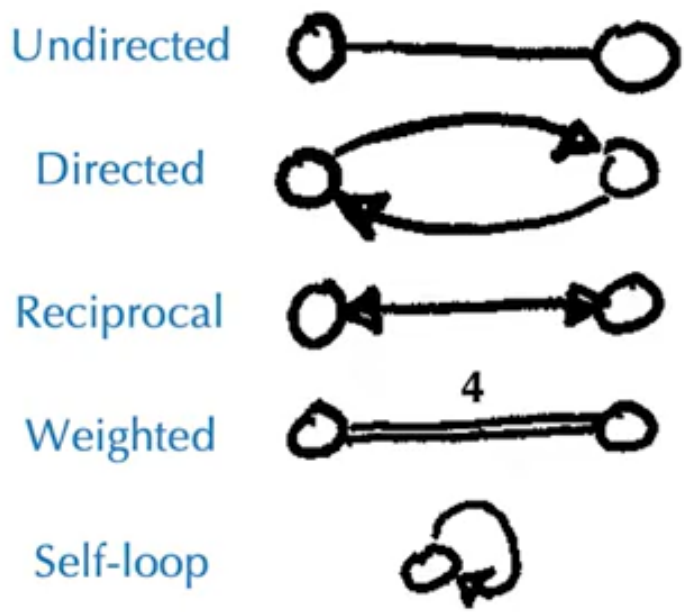
A subset of a network is said to have complete mutuality if every node is **directly connected** to every other node in that subset.

This results in a **clique**—a **maximally** connected subgraph.

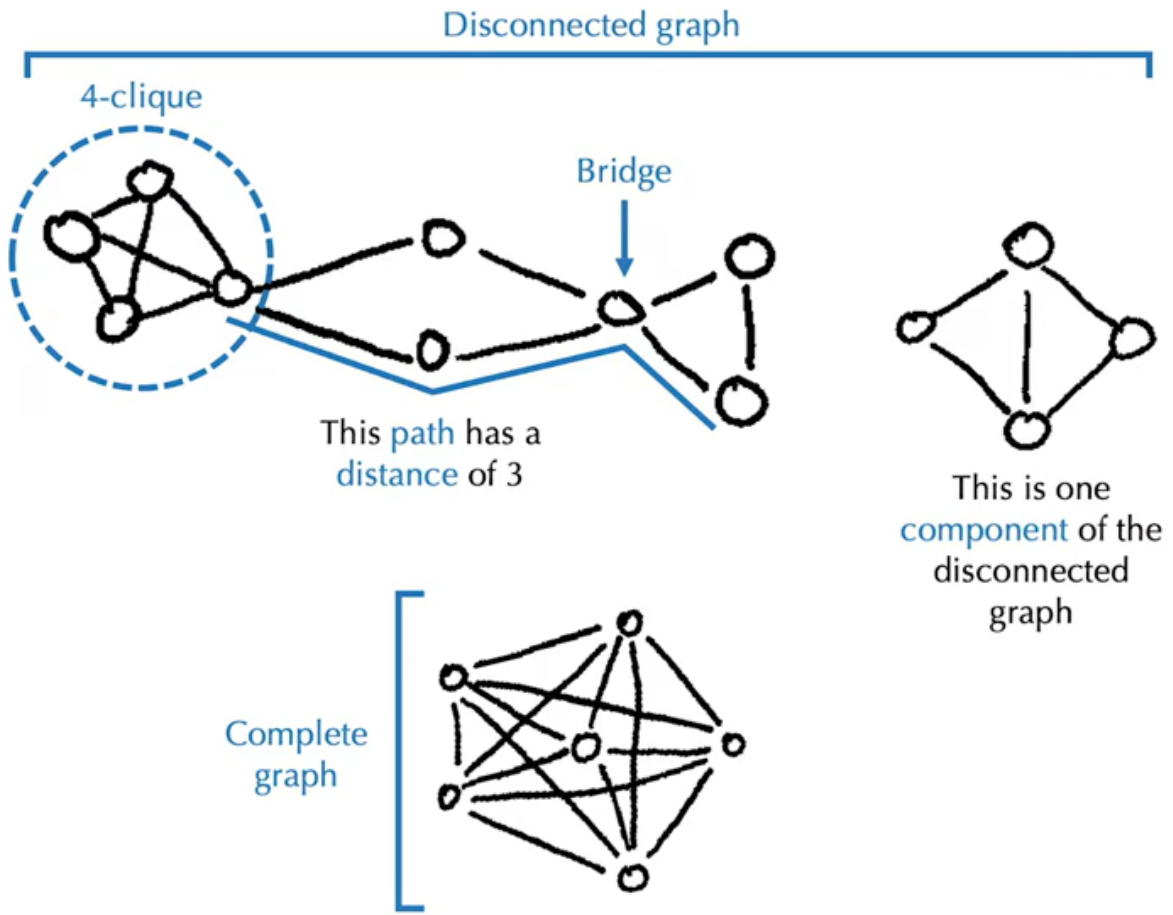
- **Clique:** A set of nodes where **every** node is connected to **every** other node.
- If a group of friends **all** know each other, they form a **clique**.
- **Example:**
  - In a research paper co-authorship network, if three researchers co-author papers **together**, they form a clique.



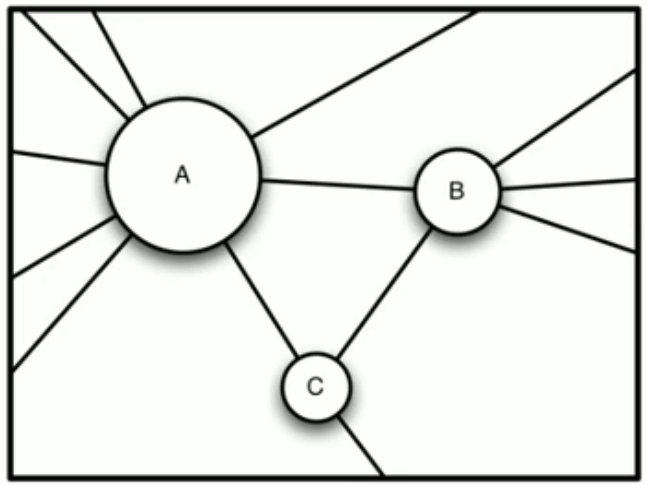
Type of edges:



Terminology



Units of analysis: Position

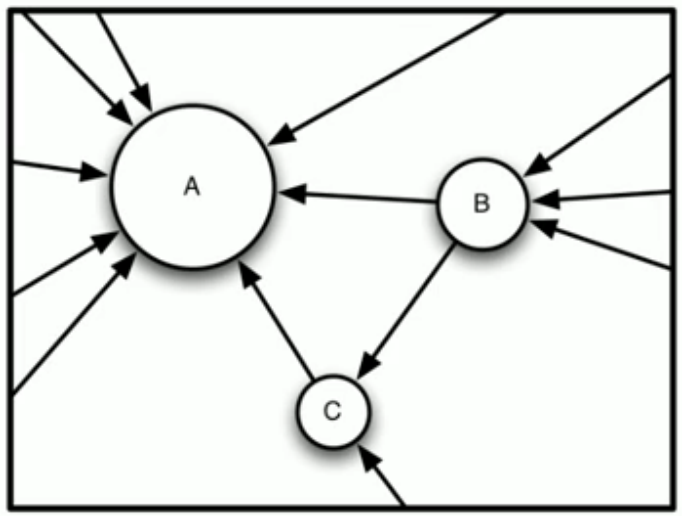


Position:
Looking for a node's position is one of the firsst and most sensible tasks for the analyst.

Nodes with more links are considered more *central.*

Degree centrality is a score for more links, where a link is considered a 'degree'.

Units of analysis: Position



In- and out-degree:
Some links are directed.

When links represent a message, nodes with a higher *in-degree* are more popular. Nodes with more *out-degree* are more authoritative.

The Configuration Model

- Degree has become so important, we normally ‘fix’ this distribution and estimate other parameters around it.
- This requires a means for generating random networks with a fixed degree. This is known as a **configuration model**.

CENTRALITY / node position

- Centrality measures help in understanding the structure of social networks by identifying key influencers, intermediaries, and well-connected individuals.
- Depending on the application, different centrality measures are useful in analyzing communication, power dynamics, and the spread of information.

1. Degree Centrality

- Definition: Measures the number of direct connections a node has with other nodes.
- Formula:-

$$C_D(v) = deg(v)$$

where $deg(v)$ is the number of edges connected to node v.

- A node with a high degree centrality has many direct connections and can quickly disseminate information.
- The more connections one node has, the more they are able to get information out.

2. Betweenness Centrality

- **Definition**: Measures how often a node lies on the shortest paths between pairs of nodes.
- **Formula**:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where:

- σ_{st} is the total number of shortest paths from node s to node t,
- $\sigma_{st}(v)$ is the number of those shortest paths that pass through node v.

- **Interpretation**: Nodes with high betweenness centrality act as bridges or intermediaries controlling the flow of information.

3. Closeness Centrality

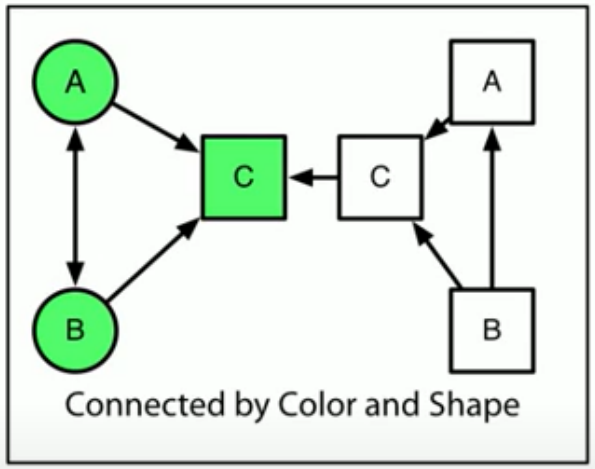
- Measures how close a node is to all other nodes in the network. It is the inverse of the sum of the shortest path distances from the node to all other nodes.
- Formula:-

$$C_C(v) = \frac{1}{\sum_u d(v,u)}$$

where $d(v,u)$ is the shortest distance from node v to node u.

- Nodes with high closeness centrality can quickly access or spread information throughout the network.

Units of analysis: Micro



Homophily:
Individuals of like type are particularly prone to linking to one another.

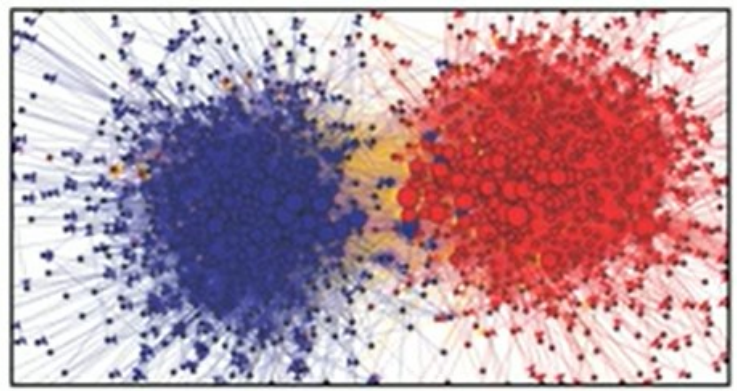
The trick is not finding homophily, but the right kind of homophily.

Baseline homophily is how much you would expect by chance. In-breeding homophily is homophily over and above this value.

Metrics for Calculating Homophily

- EI index: (External ties-Internal ties)/(External +Internal)
- Yule's Q: EI normalized by proportion in the same and different category.
- Assortativity: Newman's measure for considering the propensity of like to link to like. Can also be considered categorical.

Units of analysis: Meso



Communities:
Clusters of nodes tend to group together. Where there are more nodes within a group than between a group we consider this a community.

Community detection is one of the most common tools in network analysis outside of looking for node position. Communities tend to emerge based on the micro level processes previously discussed.

Community

- A **community** is a group of people who share common interests, values, beliefs, or goals and interact with each other in a structured way.
- Communities can form based on **geography, profession, social interests, digital platforms, or personal identity.**

Key Characteristics of a Community:

- **Shared Identity:** Members have a common purpose, belief, or interest.
- **Interaction:** People communicate, collaborate, and influence one another.
- **Boundaries:** Some communities have clear membership rules, while others are open.
- **Continuity:** Communities exist over time, either growing, changing, or dissolving.
- **Engagement:** Participation may be active (leaders, contributors) or passive (observers).

Communities

- In **social computing**, communities extend beyond physical locations into **online spaces**, where people interact via social media, gaming platforms, forums, and other digital tools

Communities can be understood from different **perspectives**, each providing unique insights into how people interact, connect, and engage.

1. (a) Systems Perspective

A community functions like a living system, with different interconnected parts working together to maintain stability.

How It Works:

Just like a human body has organs that perform different functions, communities have organizations, leaders, and roles that work together.

If one part fails (e.g., a breakdown in communication), the whole system is affected.

Systems have feedback loops, meaning a small change can cause big shifts (e.g., a viral post changing public opinion).

System Perspective cont....

Examples in Social Computing:

- Social Media Algorithms: Platforms like YouTube and TikTok function as systems where content engagement influences what users see next.
- Misinformation Spread: Fake news can disrupt a social system by influencing elections, causing panic, or shaping public behavior.
- Echo Chambers: Online groups with similar views reinforce each other, making it hard for new ideas to enter.

2. Social Perspective

- Communities are built on **social relationships** and **networks between individuals, organizations**, and **leaders**.

How It Works:

- People form connections (strong ties with close friends, weak ties with acquaintances).
- Influence flows through network centrality—people with more connections hold power.
- Collective behavior emerges as groups shape trends, activism, and movements.

Social Perspective Examples

- Twitter & Activism: Hashtags like #MeToo or #BlackLivesMatter spread globally because of highly connected influencers.
- LinkedIn & Professional Networks: People find jobs, mentors, and opportunities through social ties.
- Online Friendships: Platforms like Discord and Reddit form social groups that replace traditional friendships.

3. Virtual Perspective

- Communities today are not just physical but also exist in virtual spaces where people connect through digital platforms.

How It Works:

- Some communities map to real-world locations (city-based WhatsApp groups).
- Others are fully virtual (gaming communities, crypto groups, Metaverse spaces).
- People form digital identities (avatars, usernames, pseudonyms).

- ***Case Study: Compare Physical vs. Virtual Communities*** *– Uni clubs vs Whatsapp groups*

4. Individual Perspective

People define their own sense of community membership based on interests, identity, and experiences.

How It Works:

- Some people feel stronger connections to online groups than to their local community.
- Multiple identities exist—someone may be a programmer on GitHub, a gamer on Twitch, and a student in class.
- Identity-driven communities form (cultural communities).

Individual Perspective Examples:

- Social Media Profiles: A person's Instagram identity may be different from their LinkedIn presence.
- Gaming Communities: People adopt in-game personas different from real life.
- Digital Activism: People engage in social causes they feel connected to, even if they are not physically affected.
- *why do some communities feel stronger than others?*

Metrics for calculating clustering

- Transitivity. The proportion of all triads which are closed.
- Local clustering coefficient. The proportion of all triads around a specific individual node that are closed.
- Clustering coefficient. The average of the LCC across all nodes.

Network Visualization Software

iGraph
(for R and Python)

Tulip

Less Complex /
Less coding

More Complex /
More coding

A History of networks

- 1800s - birth of graph theory through the konisberg bridge problem (is it possible to cross all bridges without backtracking).
- 1930s - sociometry in small groups.
- 1950s - Bott and networks within couples; Barnes, Harary and graph theory
- 1960s - First egonet studies, Small worlds, name generator surveys
- 1970s - SOWT, blockmodeling, centrality
- 1980s - P1/P2 modeling, software emerges, institutionalization [INSNA]
- 1990s - Visualization techniques, P*, SOAMs, collaborative filtering, pagerank
- 2000s - Community detection; ERGMs, SNS apis.
- 2010 - Big data, neo4j, interactive networks