Social Networks

POLI 100F

Course Plan

- 8/1 – Course introduction, student polls
- **8/3 – Network analysis: basics**
- 8/8 – Network analysis: static networks
- 8/10 – Network analysis: dynamic networks
- 8/15 – Social norms: diffusion
- 8/17 – Social norms: planned change
- 8/22 – Political networks
- 8/24 – Political networks
- 8/29 – Network theory
- 8/31 – Network theory, review

Evaluation

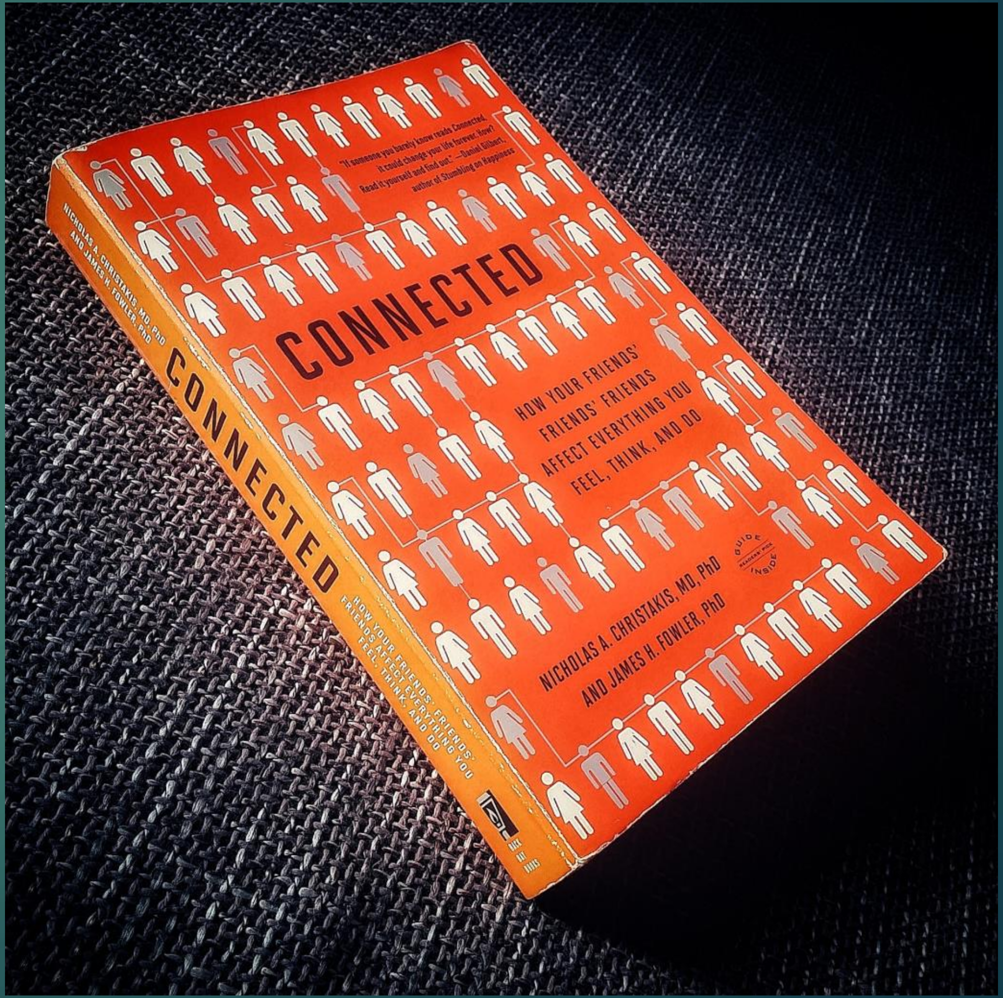
- Here's how your **final grade** will be calculated:
- Problem Set #1 - 30% [due Week 2]
- Problem Set #2 - 30% [due Week 3]
- Research proposal - 40% [due Week 5; no final exam]

- **Attendance** at lecture is not required, but it is recommended because you'll have the opportunity to ask questions. All lectures will be **recorded** and posted on the corresponding Canvas page.

Required Text

- There's only one **required text** for the course, and it's in stock at the UCSD bookstore and available on Amazon for less than $20.

Nicholas A. Christakis and James H. Fowler. 2011. Connected: How Your Friends' Friends' Friends Affect Everything You Feel Think and Do. New York: Little Brown, ISBN: 9780316036139.



Office hours

- I'll be holding **office hours** on Wednesdays from 9-11am. You can sign up at the course Canvas page ("Start Here").
  - If that time's inconvenient or if all the slots are full, we can set something up by appointment. Message me on Canvas or email me at mdraper@ucsd.edu.

What are social networks?

- Networks capture the pattern of interactions between the parts of a system.
- Social networks are networks involving people.
- "[A] social network is a set of actors, or other entities, and a set or sets of relations defined on them" (Knoke & Yang).

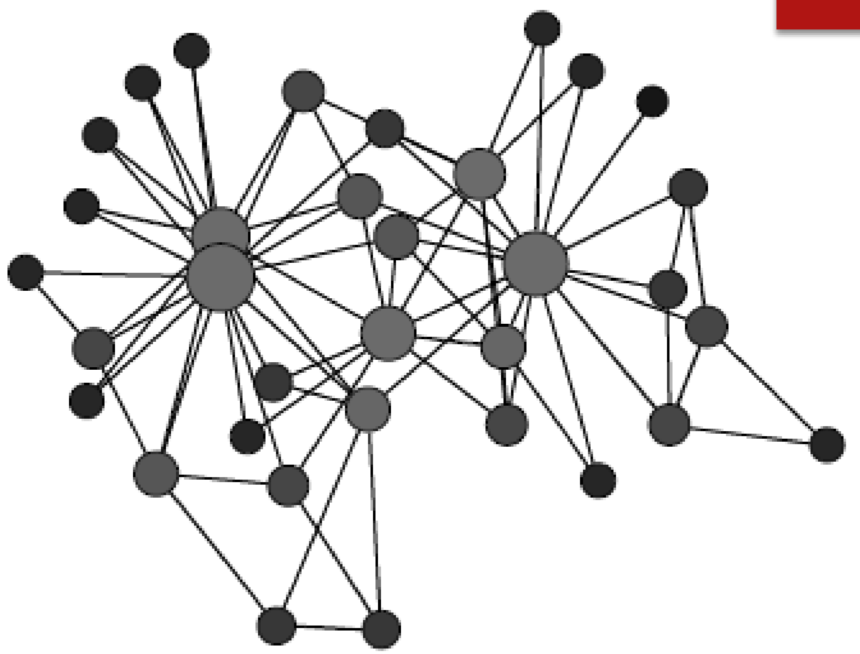
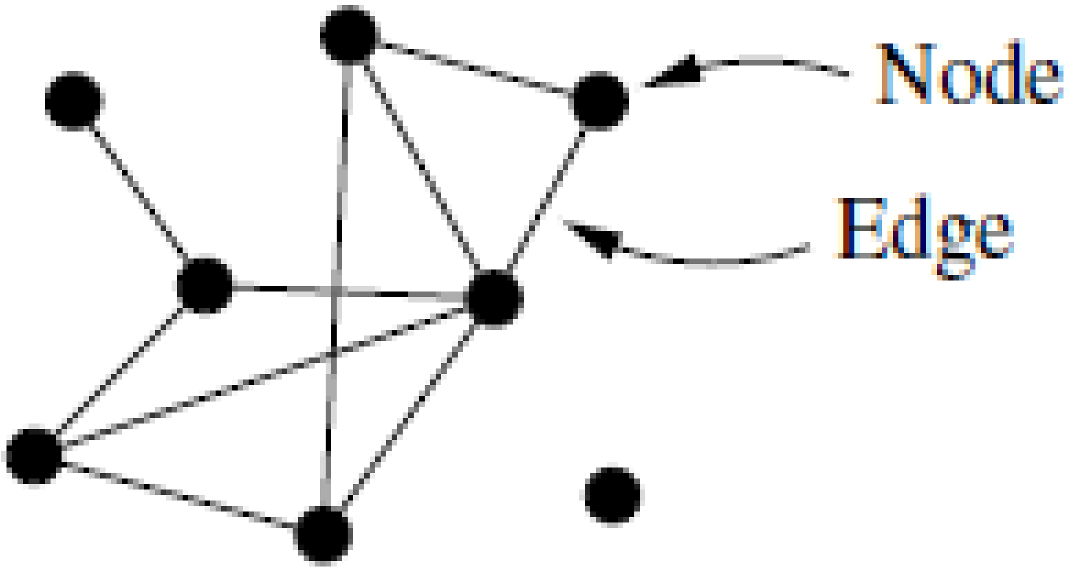


Figure 1.2: Friendship network between members of a club. This social network from a study conducted in the 1970s shows the pattern of friendships between the members of a karate club at an American university. The data were collected and published by Zachary [479].

What are social networks?

- We can call these actors **nodes**, and the relations connecting them **edges**.
  - Nodes are also sometimes called "vertices."
  - Edges are also sometimes called "links"



A small network composed of eight nodes and ten edges.

What are social networks?

- In social networks, nodes represent people, and edges represent social connections (broadly defined).
  - The arrangement of the nodes and edges in graphical space is arbitrary.

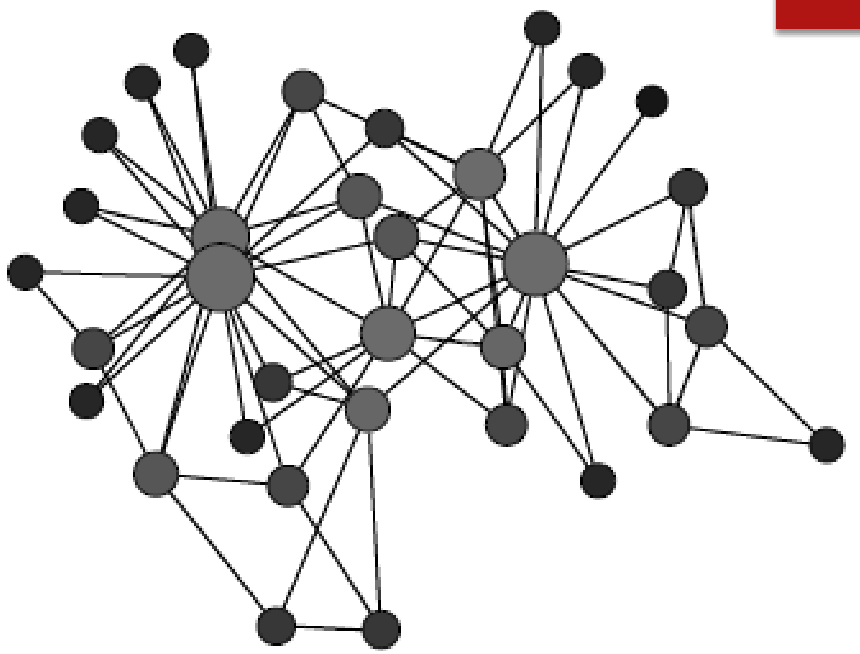


Figure 1.2: Friendship network between members of a club. This social network from a study conducted in the 1970s shows the pattern of friendships between the members of a karate club at an American university. The data were collected and published by Zachary [479].

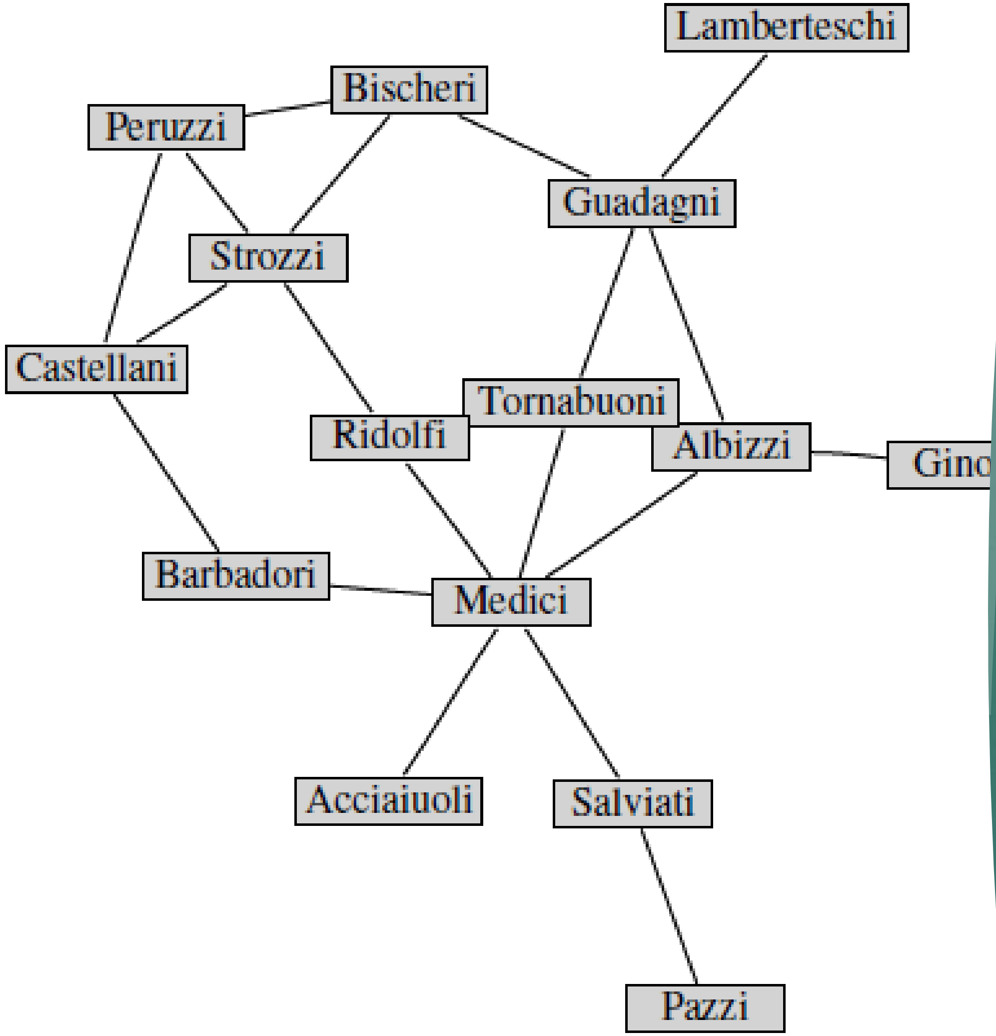
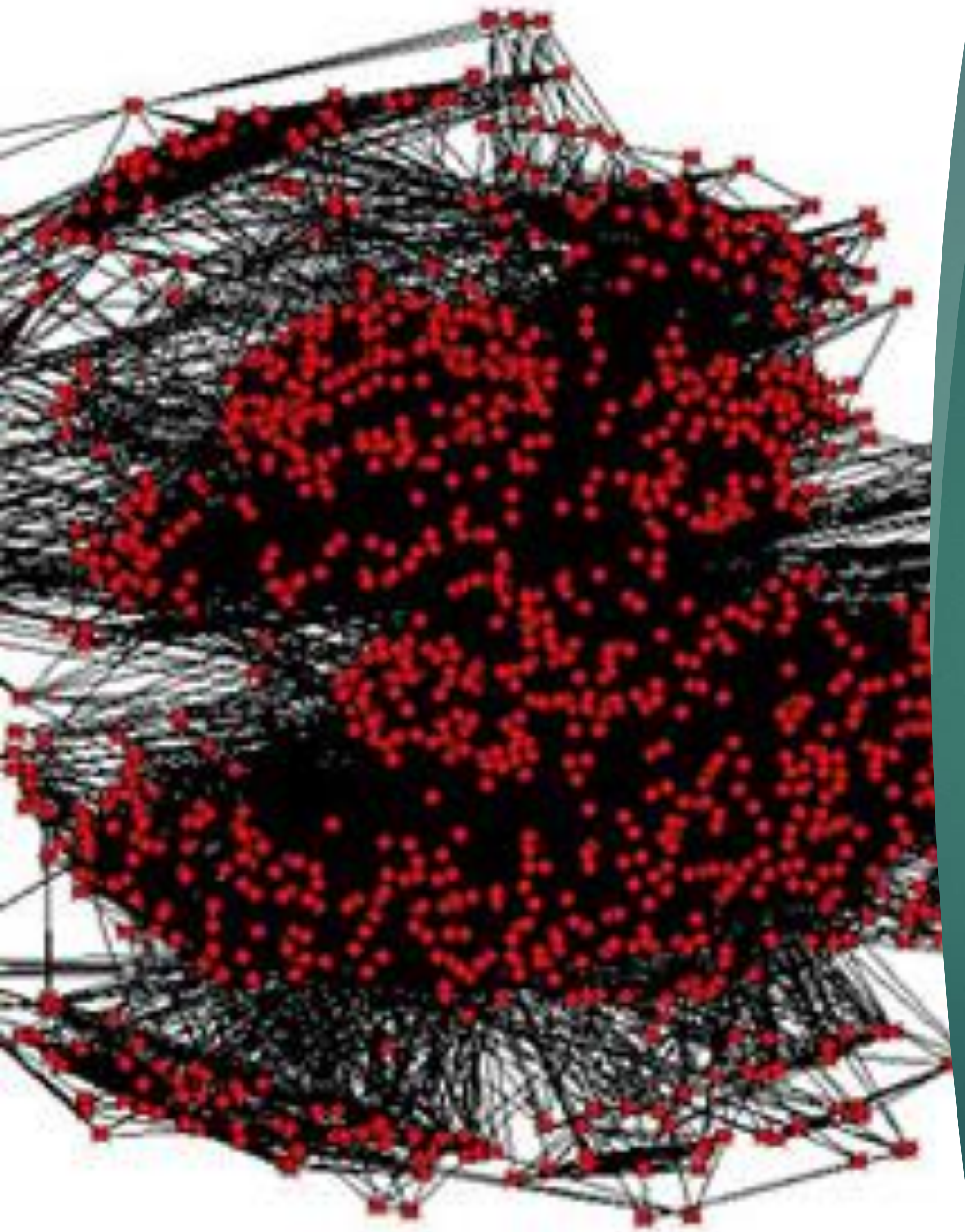


Figure 4.3: Intermarriage network of the ruling famili of Florence in the fifteen century. In this network th nodes represent families and the edges represent ties c marriage between them. After Padgett and Ansell [377].

Why study social networks?

- Recall that an **edge** is defined as the relation connecting two **nodes**.
- Because we can define edges in many different ways, we can use the tools of social network analysis to look at lots of different questions:
  - Friendships, professional relationships, exchanges of value, communication patterns, romantic or sexual relationships, and many others.
  - See *Connected*, Ch. 1.



Why study social networks?

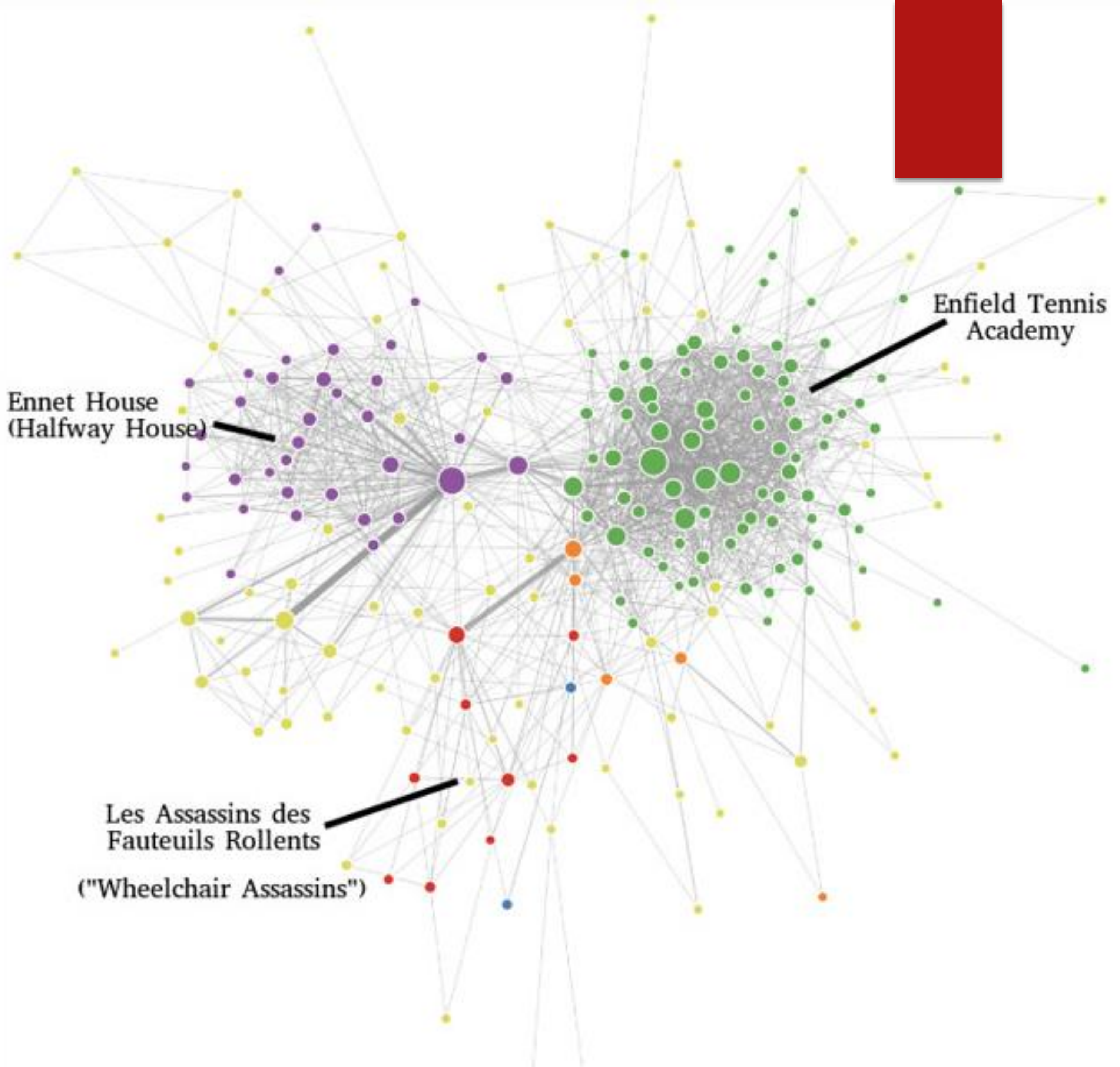
- The choices available to us when we decide to model a network include:
  - How to think about the nodes:
    - Just people?
    - People and something else? [**Multi-modal**]
  - How to understand the links:
    - What do they represent?
    - Diffusion of information, disease?
    - Are they uni-directional (directed)?
      - Ex: Twitter, snail mail
    - Or bi-directional (undirected)?
      - Ex: Facebook, marriage

Abstraction

- When we represent a social network (graphically or with a matrix), we're abstracting from reality and creating a simplified representation of the connection patterns. This is called a **topology**.
- The features we choose to represent depend on the question(s) we're asking. This means that our model is only "right" or "wrong" with respect to some purpose.
- When we're evaluating a network, we should ask: "what features of reality do we think matter to produce the effect we're interested in studying?"

Abstraction

- ▶The boundaries of a network are also related to the question we're asking – ultimately we will have to impose some boundary on it.
- ▶We can imagine this network as one big network, or as three separate networks.



Abstraction

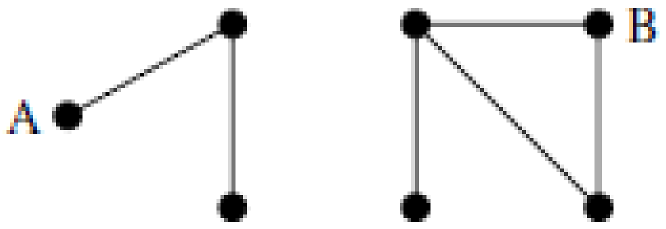


Figure 6.13: A network with two components. There is no path between nodes like A and B that lie in different components.

- Component - a **component** is a subset of the nodes of a network such that there exists at least one path from each member of that subset to each other member.

Social Network Analysis Terminology

- **Multi-mode Network** – Nodes represent more than one type of entity.
  - Contrast with uni-modal network: nodes only represent one type.
  - Example: some nodes may represent voters, others candidates or officeholders.
  - Also called affiliation or bipartite networks.
- **Node Attribute** – features or classifications of a node.
  - Example: whether a node has been exposed to a piece of information (or disease).

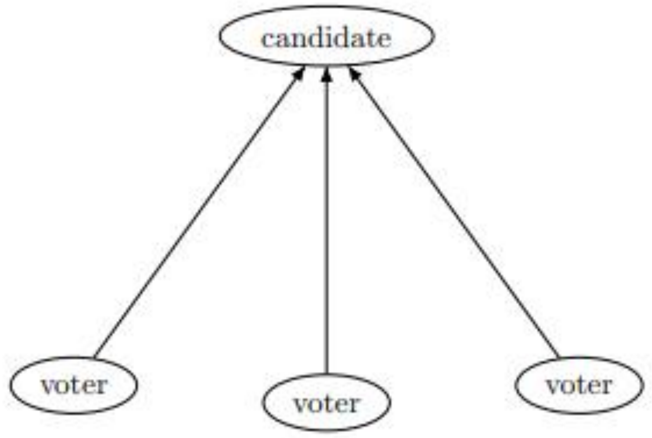
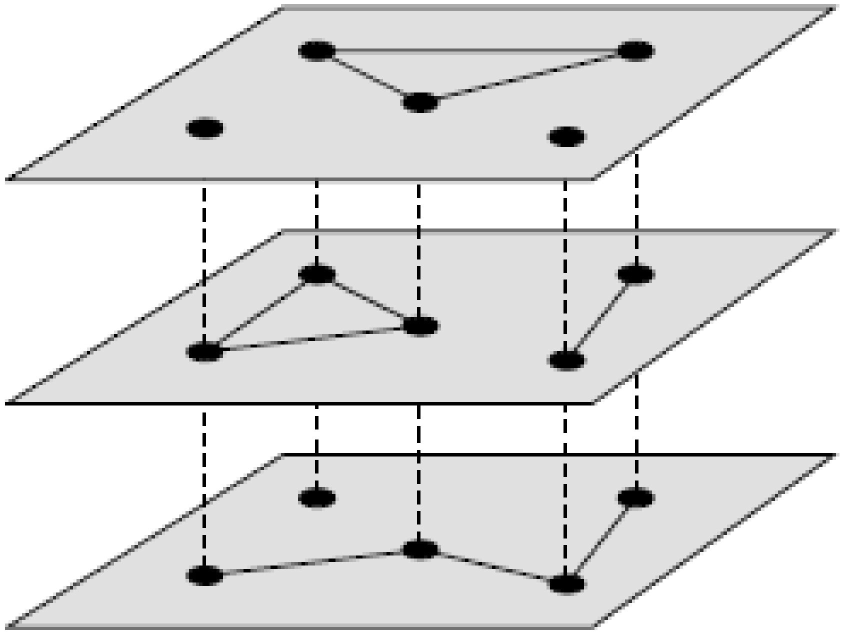


Figure 1: A social network of voters and a candidate.

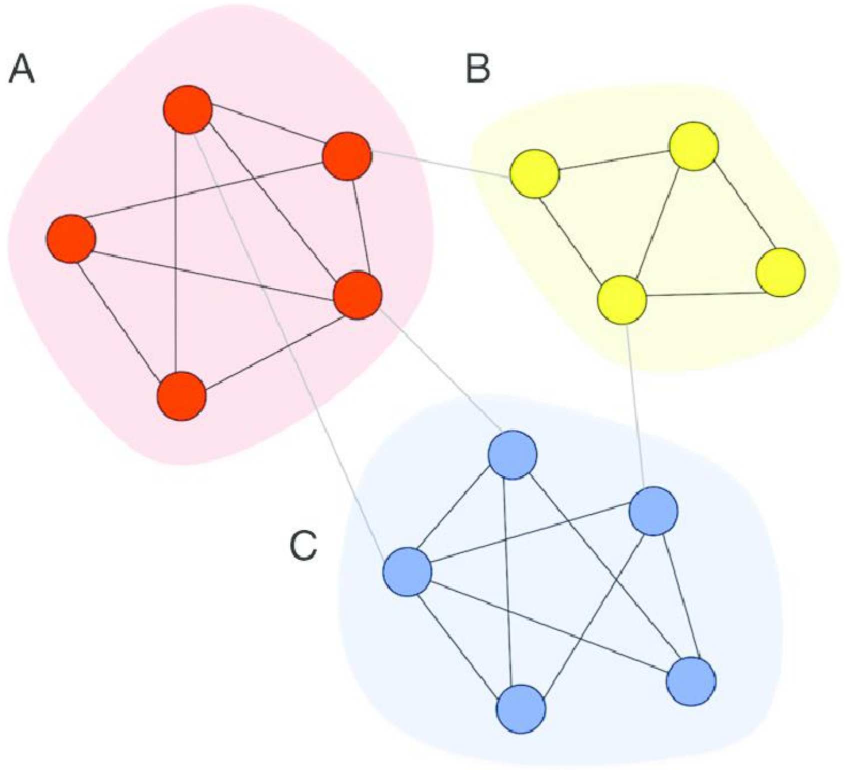
Social Network Analysis Terminology

- **Multiplex network** – more than one set of links; links represent different types of connection.
  - Example: different survey questions (who do you trust, who do others trust).
- Tie strength: the intensity of a link.
- Mark Granovetter - The Strength of Weak Ties (1973)
  - Weak ties are disproportionately likely to link to other parts of a network.



Social Network Analysis Terminology

- **Homophily** – the tendency of nodes to connect to others that are like them in some way. Also called “assortative mixing.”
  - Occasionally reversed: heterophily, “disassortative mixing.”
  - Theories of homophily: path dependency, similarity lowers costs, kin selection, many others.



Social Network Analysis Terminology

- **Contagion** – the tendency for behaviors, emotions or other conditions to spread through a network.
  - Behavioral contagion, emotional contagion.
  - Similar to but distinct from biological contagion.
- **Directionality** – links can be uni-directional or bi-directional.
  - Directed or undirected.
- Remember, where the nodes are located in 2d space is an arbitrary decision made by the analyst.
  - The same network can be visualized in different ways.

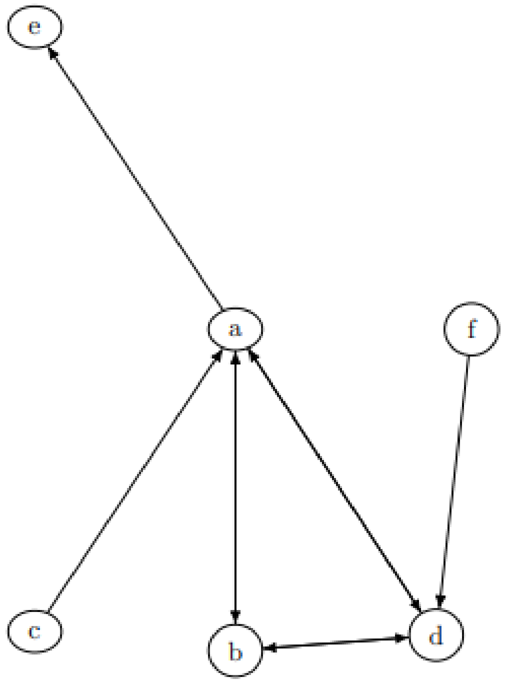
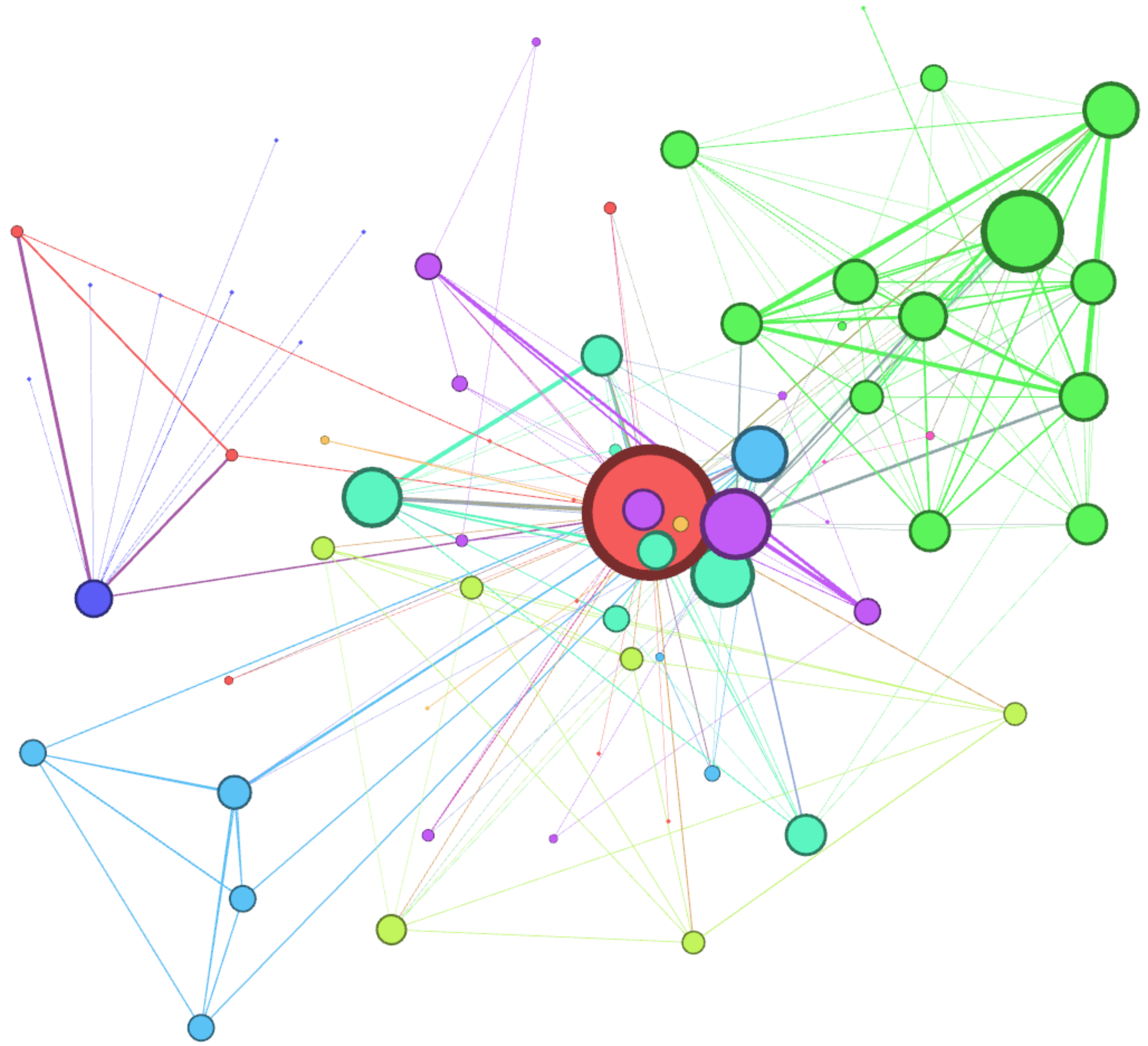
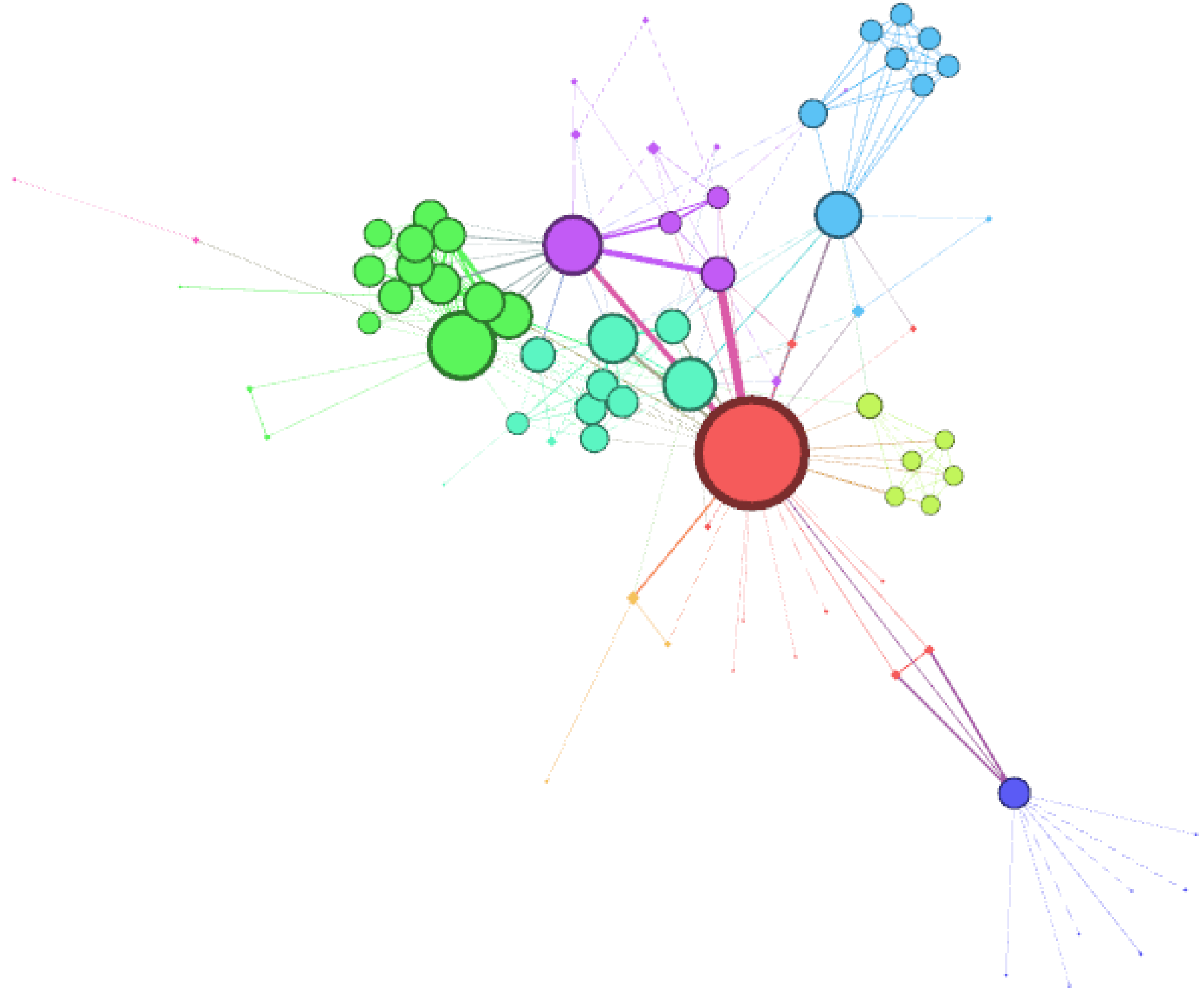
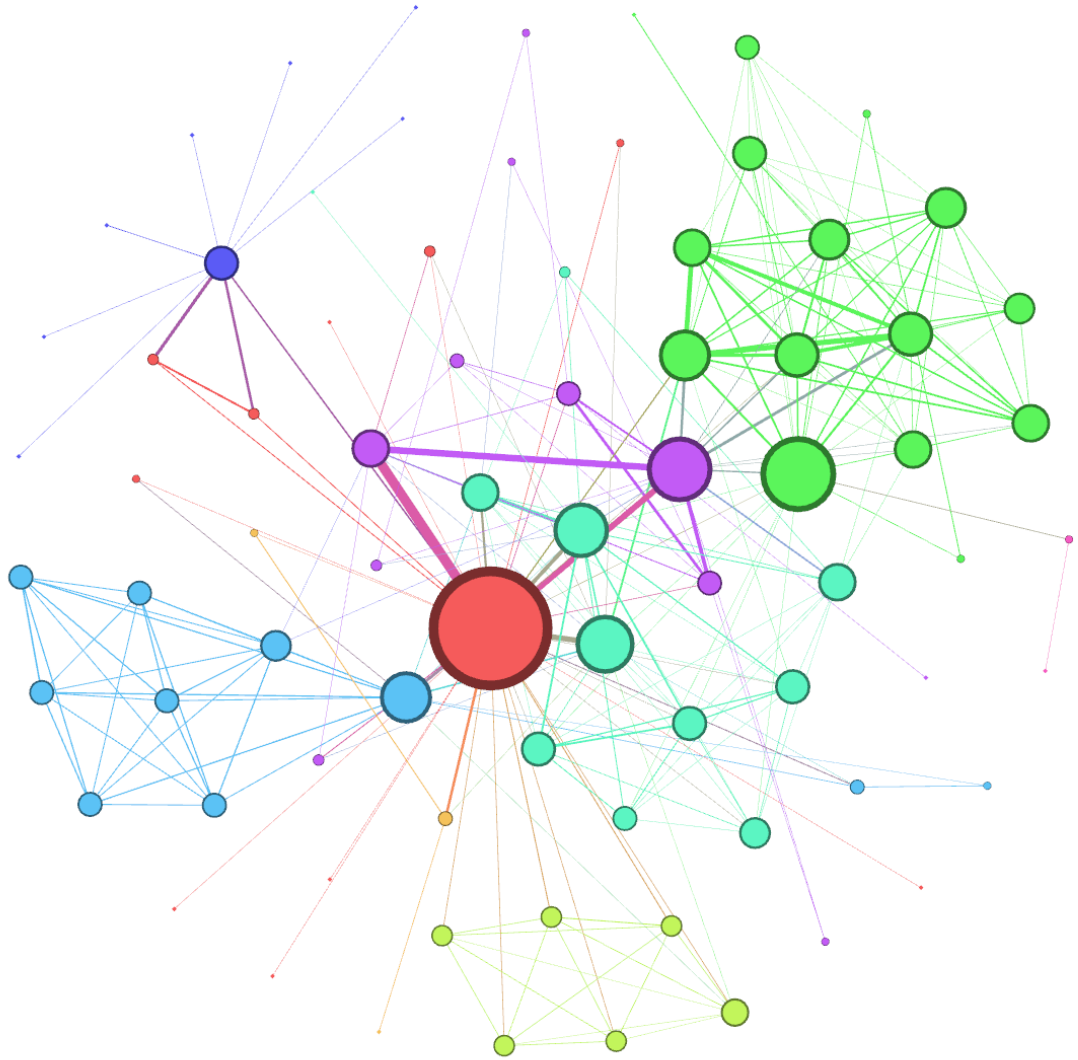
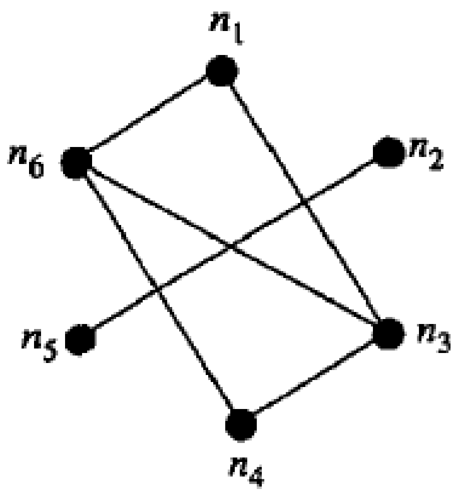


Figure 2: "Who do you go to for advice?"

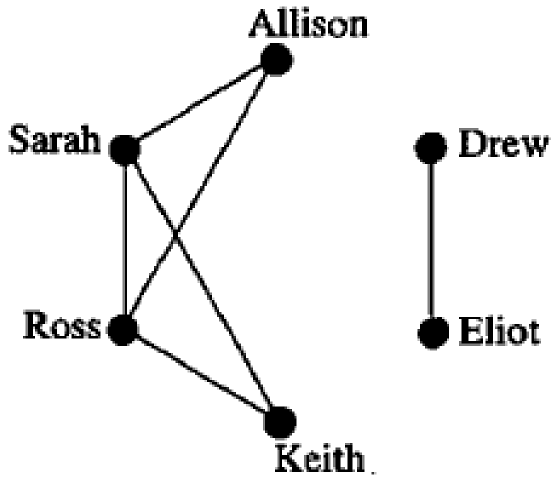








$\mathcal{G}+$



$\mathcal{G}$

$\phi(n_1) = \text{Keith}$
$\phi(n_2) = \text{Eliot}$
$\phi(n_3) = \text{Sarah}$
$\phi(n_4) = \text{Allison}$
$\phi(n_5) = \text{Drew}$
$\phi(n_6) = \text{Ross}$

Fig. 4.13. Isomorphic graphs

Social Network Analysis Terminology

- **Adjacency Matrix**: a dataset representing all the links in the network. [represented by A]
- This is "the fundamental mathematical representation of a network."
- Symmetrical if the network is undirected, but not if it's directed.
- Unless we allow self-links, the main diagonal of the matrix (top left to bottom right) will contain all zeros.

	a	b	c	d	e	f
a	0	1	0	1	1	0
b	1	0	0	1	0	0
c	1	0	0	0	0	0
d	1	1	0	0	0	0
e	0	0	0	0	0	0
f	0	0	0	1	0	0

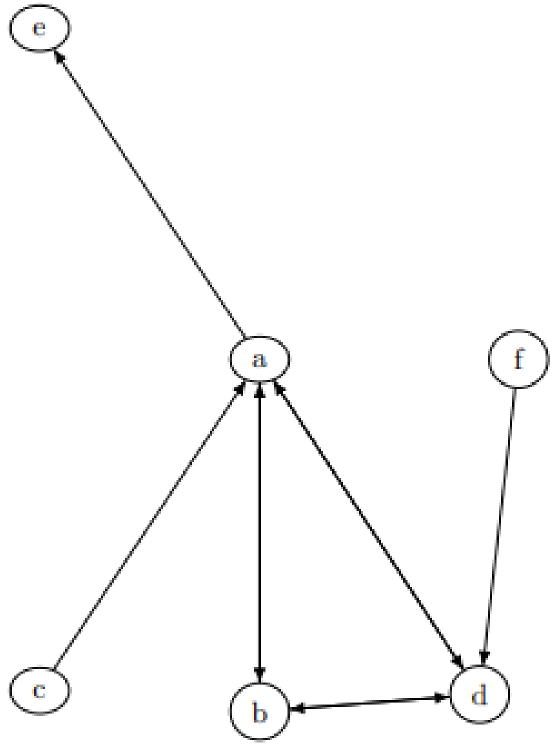


Figure 2: "Who do you go to for advice?"

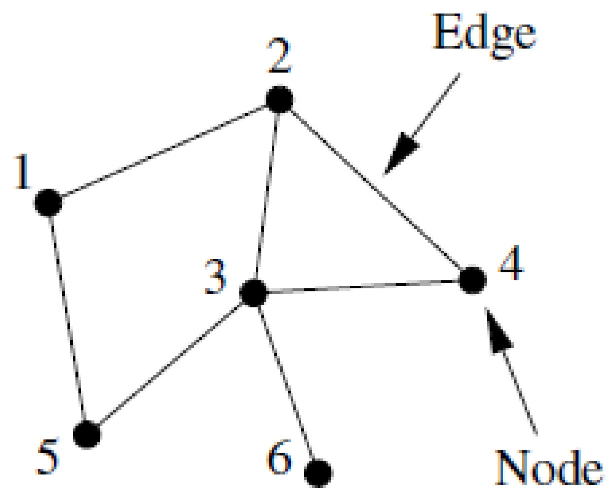
Social Network Analysis Terminology

The adjacency matrix **A** of the network is now defined to be the $n \times n$ matrix with elements A_{ij} such that

$$A_{ij} = \begin{cases} 1 & \text{if there is an edge between nodes } i \text{ and } j, \\ 0 & \text{otherwise.} \end{cases} \tag{6.1}$$

For example, the adjacency matrix of the network in Fig. 6.1a is

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}. \tag{6.2}$$



(a)

Social Network Analysis Terminology

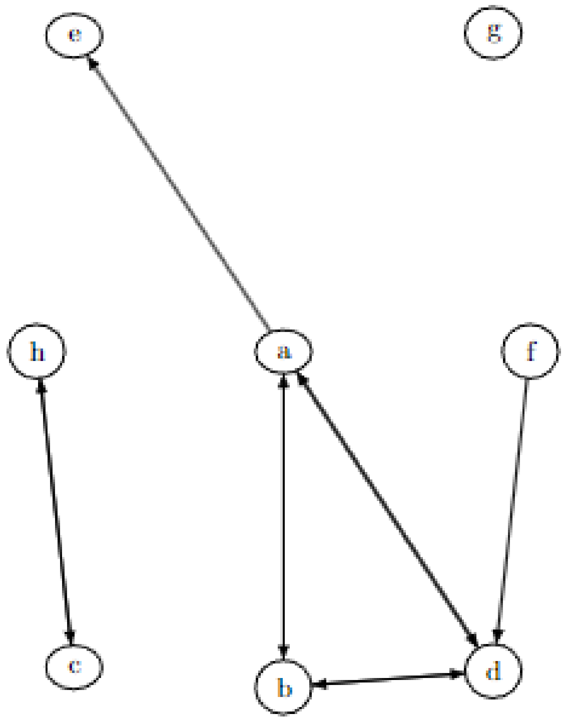


Figure 3: "Who do you go to for advice?"

- **Isolate**: a node with no links to any other nodes
- **Pendant**: a node with a single link to another node
- **Dyad**: two connected nodes
- **Triad**: three connected nodes
- **Neighbor:** a node one link away
- **Bridge:** a node located on the only path between two other nodes

Social Network Analysis Terminology

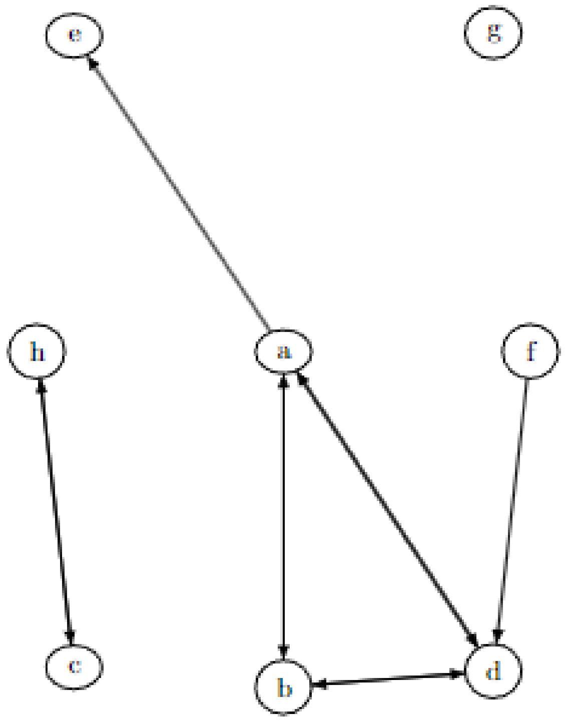


Figure 3: "Who do you go to for advice?"

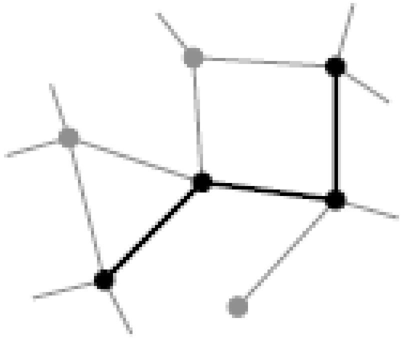
- **Degree**: the number of connections a node has to other nodes. [represented by "k"]
  - In directed networks, each directed link adds one degree.
  - In undirected networks, degree is divisible into in-degree and out-degree.
    - **In-degree**: number of incoming links to a node
    - **Out-degree**: number of outgoing links to a node
  - So in this example, d has an in-degree of 3 and an out-degree of 2. Similarly, f has an in-degree of 0 and an out-degree of 1.

Social Network Analysis Terminology

- We can expand the adjacency matrix to incorporate degrees.
  - The out-degree will be the sum of the rows.
  - The in-degree will be the sum of the columns.
- The mathematics of social network analysis consists mainly of applied matrix algebra.

	a	b	c	d	e	f	Dg
a	0	1	0	1	1	0	3
b	1	0	0	1	0	0	2
c	1	0	0	0	0	0	1
d	1	1	0	0	0	0	2
e	0	0	0	0	0	0	0
f	0	0	0	1	0	0	1
Dg	3	2	0	3	1	0	18

Social Network Analysis Terminology



A walk of length three in an undirected network.

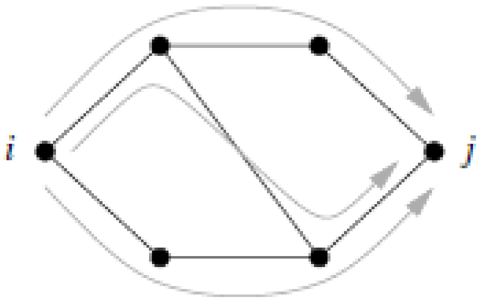
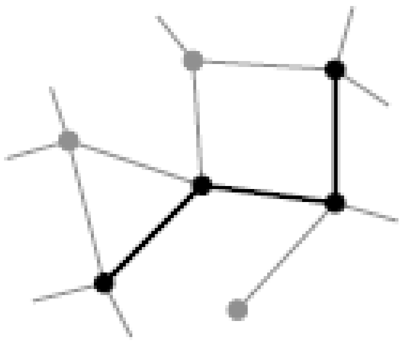


Figure 6.12: There are three shortest paths between nodes *i* and *j* in this network, each of length three.

- A **walk:** a route among the nodes passing along the links.
  - "…any sequence of nodes such that every consecutive pair of nodes in the sequence is connected by an edge."
- A **path**: a walk that never passes through the same node twice.
  - A **geodesic**: the shortest path (fewest steps) between two nodes.
  - The **diameter** of the network is the largest geodesic in the network.

Social Network Analysis Terminology



A walk of length three in an undirected network.

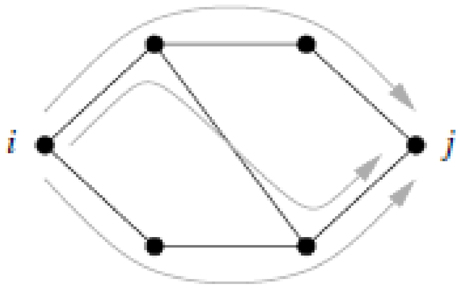


Figure 6.12: There are three shortest paths between nodes *i* and *j* in this network, each of length three.

- A **cycle**: a walk that ends on the same node where it began.
  - "A cycle in a directed network is a closed loop of edges with the arrows on each of the edges pointing the same way around the loop."
- The **average path length** of a network is the average length of all the geodesics.
  - Six degrees of separation (Milgram): average path length of 5.5.

Social Network Analysis Terminology

- Identifying **communities** in a network:
- A **component** is "a subset of the nodes of a network such that there exists at least one path from each member of that subset to each other member, and such that no other node in the network can be added to the subset while preserving this property."

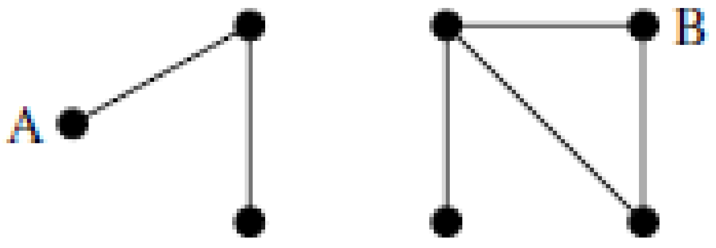
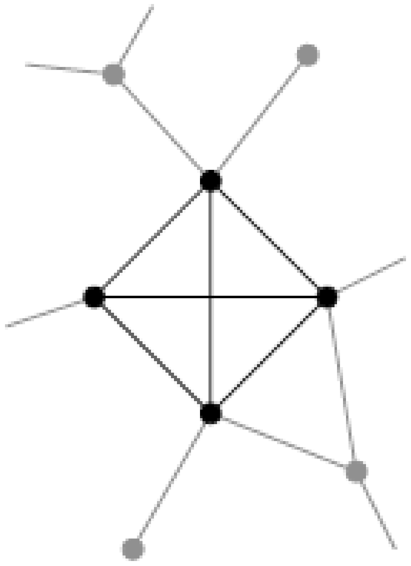


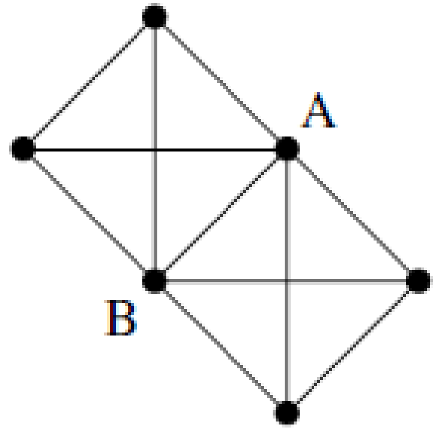
Figure 6.13: A network with two components. There is no path between nodes like A and B that lie in different components.

Social Network Analysis Terminology

- Identifying **communities** in a network:
- A **clique** is a community where every node is connected to every other node. Cliques can overlap.



A clique of four nodes within a network.



Two overlapping cliques. Nodes A and B in this network both belong to two cliques of four nodes.

Social Network Analysis Terminology

- A **weakly-connected** component has a (non-directed) path between every dyad.
- A **strongly-connected** component has a directed path from each node to every other node.

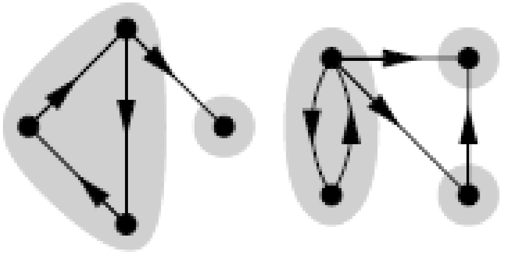
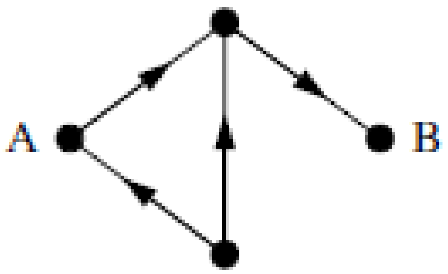


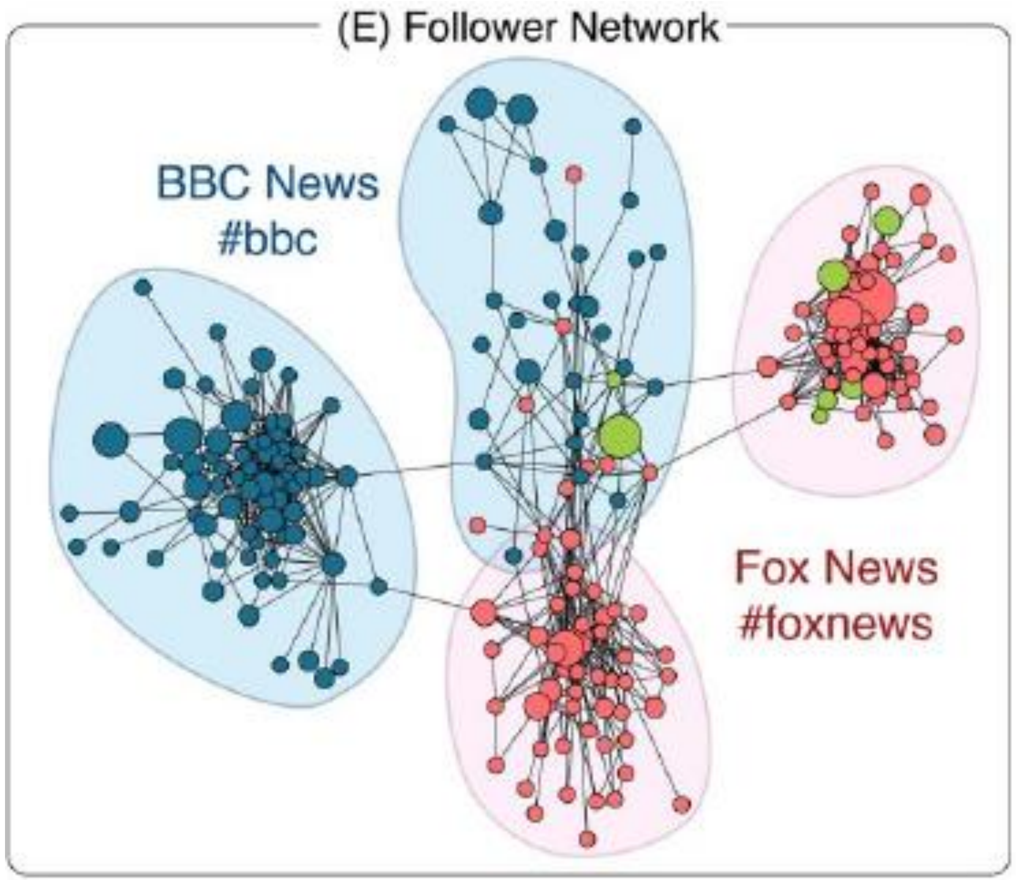
Figure 6.14: Components in a directed network. This network has two weakly connected components of four nodes each, and five strongly connected components (shaded).



There is a directed path from A to B in this network, but none from B to A.

Social Network Analysis Terminology

- Example: community structure among Twitter users who post particular hashtags.
- Weng et al. 2013: Community structure among Twitter users sharing the hashtags #BBC and #FoxNews. Blue nodes represent #BBC users, red nodes are #FoxNews users, and users who have used both hashtags are green. Node size is proportional to usage (tweet) activity, links represent mutual following relations.



Social Network Analysis Terminology

- In mathematics a relation “” is said to be **transitive** if a b and b c together imply a c.
- In social network theory, nodes are connected by edges. Networks aren’t perfectly transitive, because a<->b and b<->c doesn’t mean c->a or a->c.
- However, in social networks, if we know that a->b and b->c, there’s a greater likelihood that c->a.
- “the friend of my friend is more likely than average to be my friend.”
- This is what we mean by **network transitivity**.

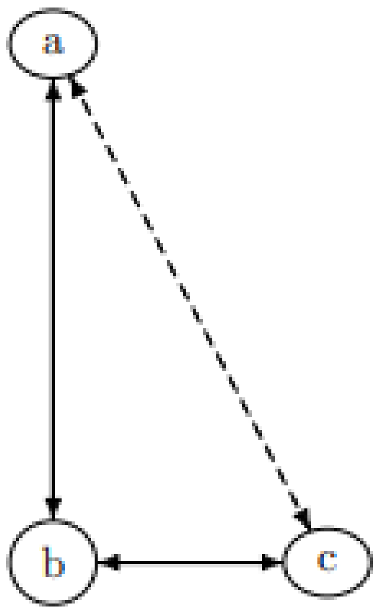
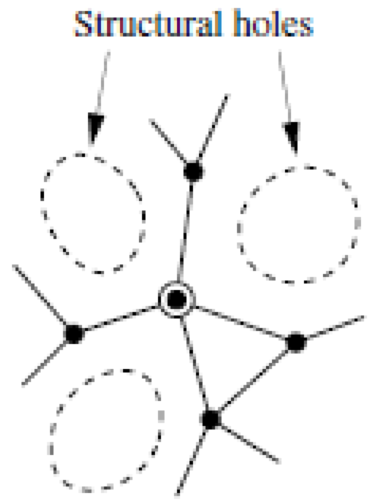


Figure 4: A triad

Social Network Analysis Terminology

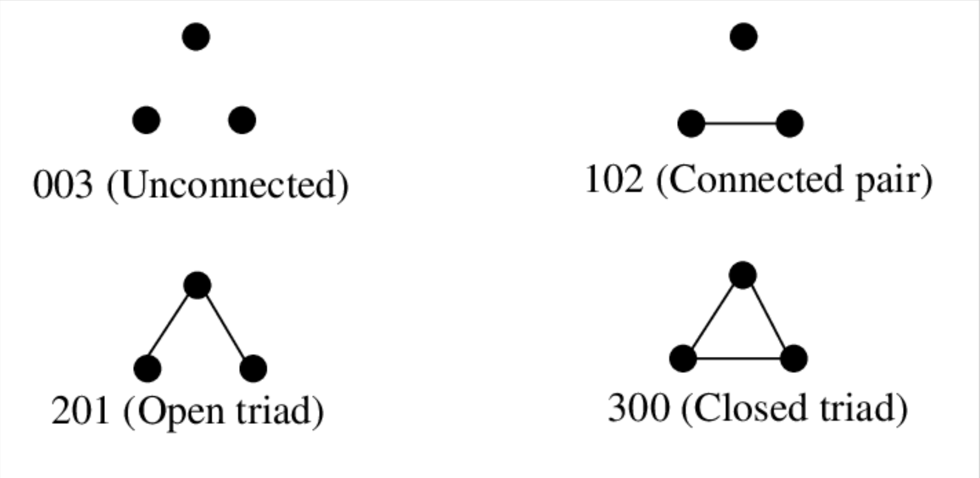
- While it is common in social networks, for the neighbors of a node to be connected among themselves, it often happens that these expected connections between neighbors are missing. The missing links are called **structural holes**. These can be desirable (ex: epidemic prevention).
- A node whose role in the network is enhanced by structural holes is known as a **broker**, like the central node here:

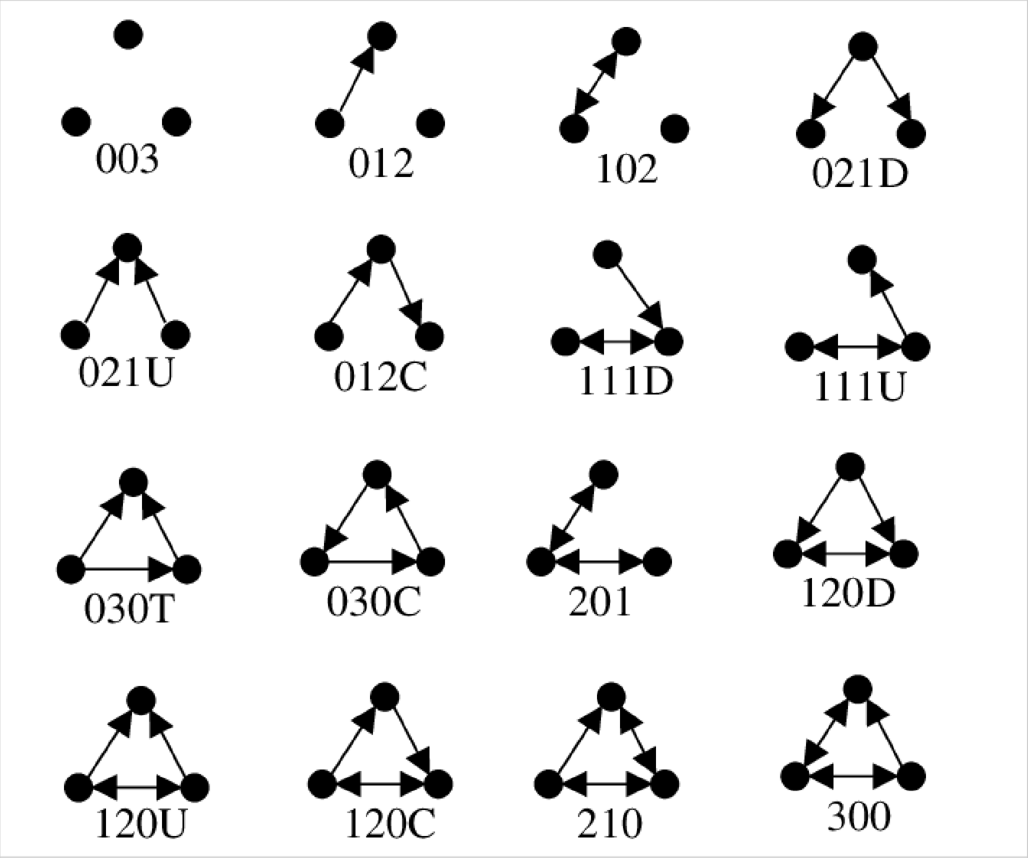


When the neighbors of a node are not connected to one another we say the network contains "structural holes."

Social Network Analysis Terminology

- There are sixteen possible **triad types** that we can encounter in a directed social network.
  - Which ones are cliques?
- In an undirected network, there are just four:





Social Network Analysis Terminology

- We can quantify the level of transitivity in the network as the fraction of paths of length two that are closed, or (more commonly) as the number of connected triads divided by the number of triads. This is called the **clustering coefficient** [C].
- No perfectly transitive components (or networks) exist that are not also cliques. Another way to put this is that all cliques are completely transitive.
- Think of it as "the mean probability that two people with a common friend are themselves friends."

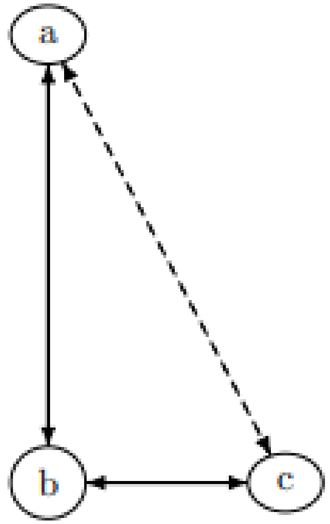
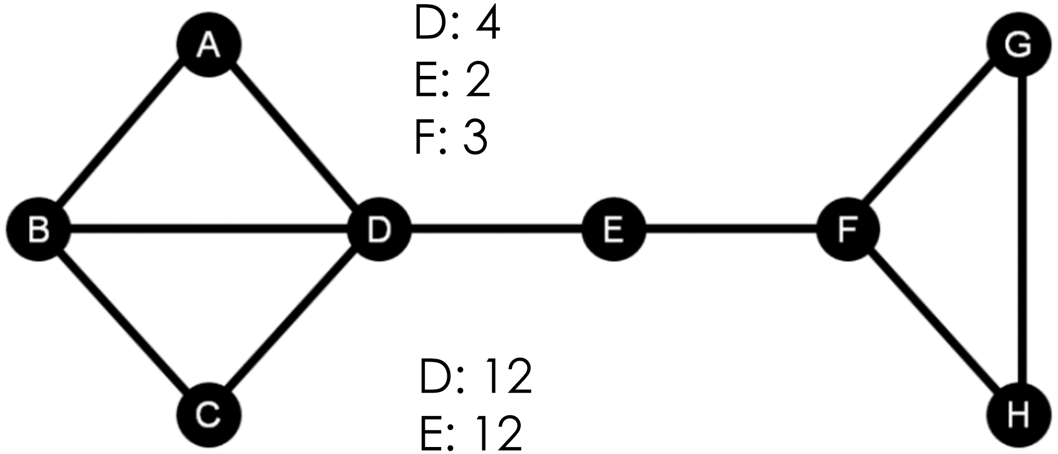


Figure 4: A triad

$$C = \frac{(\text{number of triangles}) \times 3}{(\text{number of connected triples})}.$$

Social Network Analysis Terminology

- There are several different types of **network centrality**: which nodes are most central to the network?
- **Degree Centrality**: which node has the most connections?
  - Can be divided into in-degree centrality and out-degree centrality.
- **Closeness Centrality:** the node that is closest to all the other nodes (the fewest steps in a walk).

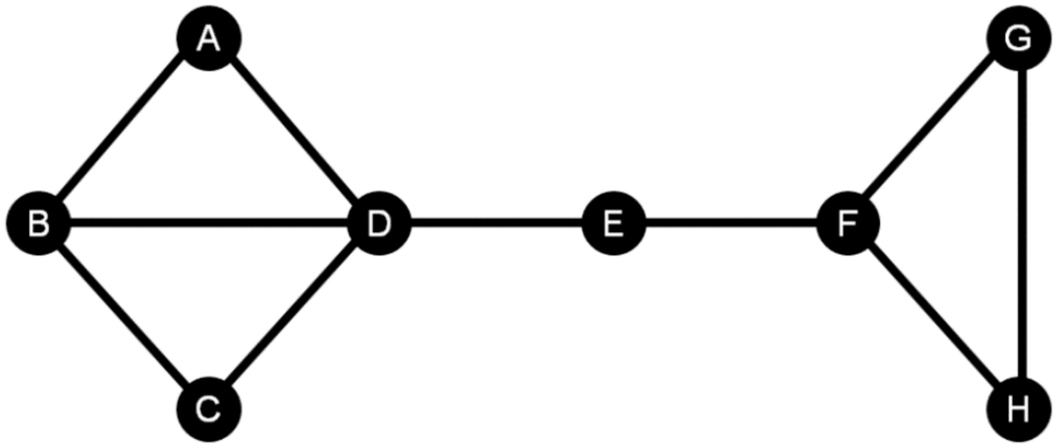


D: 4
E: 2
F: 3

D: 12
E: 12
F: 14

Social Network Analysis Terminology

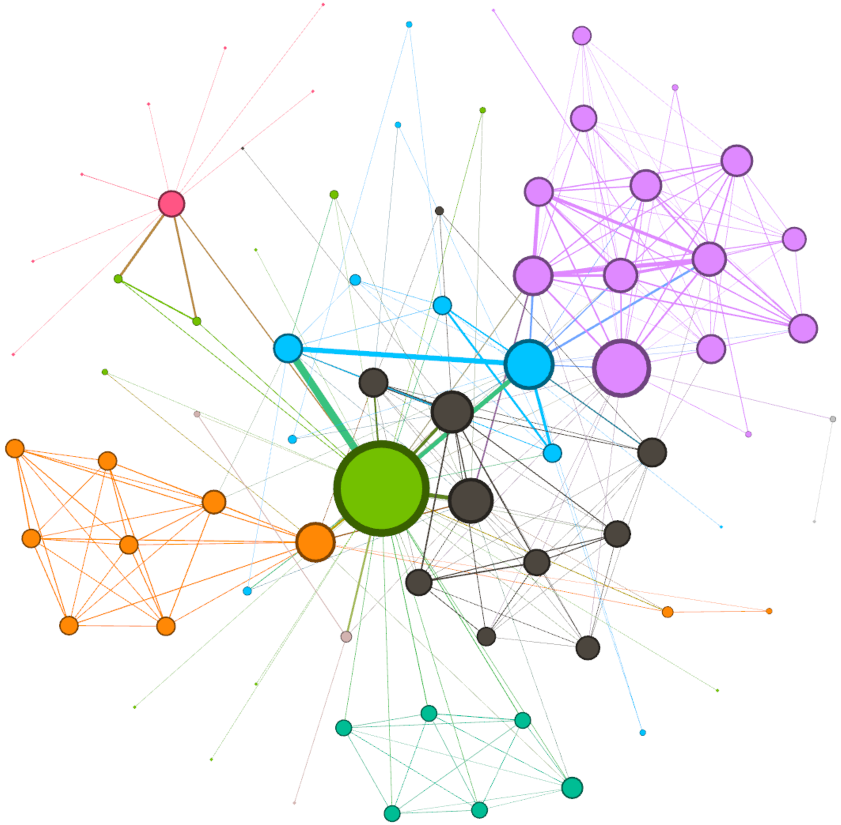
- There are several different types of **network centrality**: which nodes are most central to the network?
- **Betweenness Centrality:** which node is on the most paths connecting other nodes?
  - Ex: A has to go through D to get to E, F, G and H (4). Same for B and C (4+4). E has to go through D to get to A, B and C (3), and same for F, G and H (3+3+3). So the betweenness centrality of D is 4+4+4+3+3+3+3=24.



Social Network Analysis Terminology

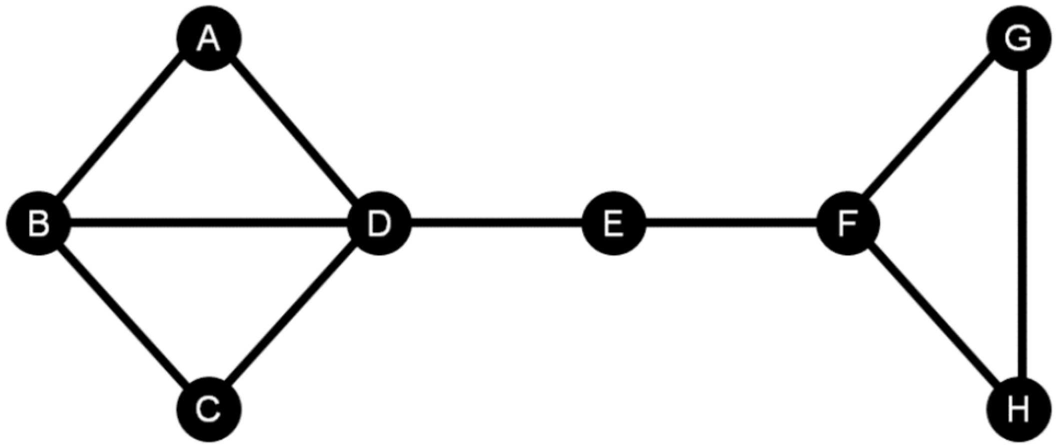
- Using betweenness centrality to determine network subcomponents – the **Girvan-Newman algorithm**.
- 1. Calculate betweenness for all ties.
- 2. Remove the tie with the highest betweenness.
- 3. Repeat until components emerge, and modularity is maximized.

- **Modularity**: high when connections are between nodes of the "same type."



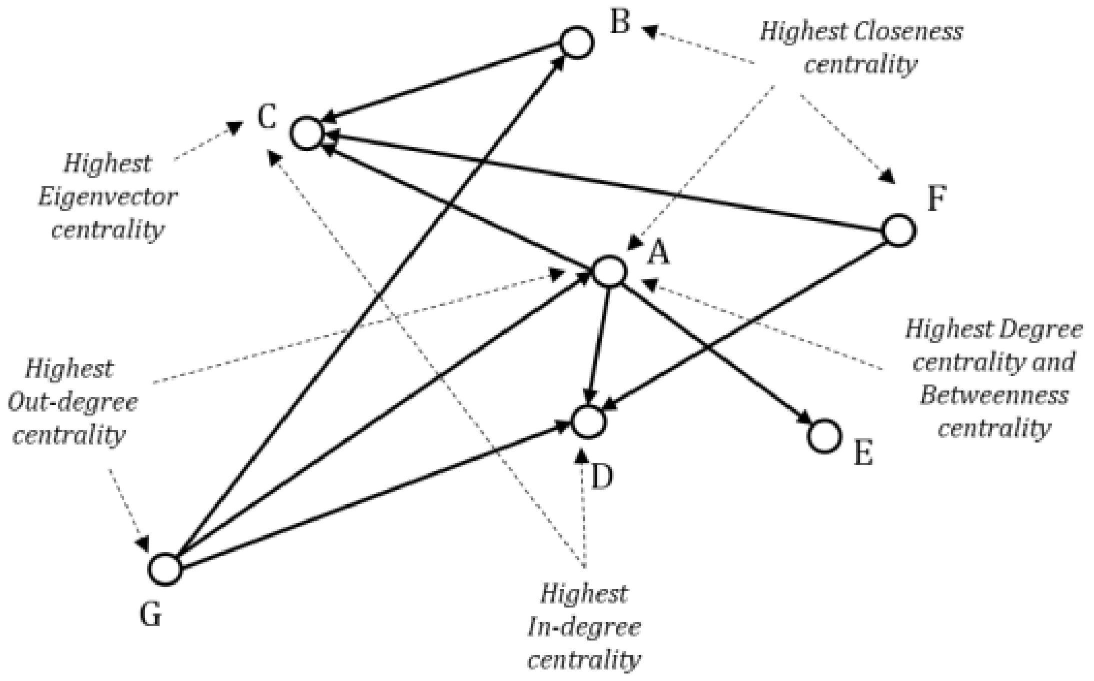
Social Network Analysis Terminology

- There are several different types of **network centrality**: which nodes are most central to the network?
- **Eigenvector Centrality:** the average centrality score of a node's direct links. A node's eigenvector centrality score is proportional to the centrality score of linked nodes.
- Sound familiar? This is (more or less) the concept of centrality behind Google's PageRank algorithm.



Social Network Analysis Terminology

- An overview of centrality measures:
- **Degree Centrality**: how many connections does the node have?
  - Can be divided into **in-degree** centrality and **out-degree** centrality.
- **Closeness Centrality:** how close is the node to all the other nodes?
- **Betweenness Centrality:** how many paths go through this node?
- **Eigenvector Centrality:** how many links do the nodes linked to this link have?



Literature: Sociology

- Mark Granovetter: The Strength of Weak Ties (1973)
- "A fundamental weakness of current sociological theory is that it does not relate micro-level interactions to macro-level patterns in any convincing way…the analysis of processes in interpersonal networks provides the most fruitful micro-macro bridge."
- **Tie strength**: "the strength of a tie is a (probably linear) combination of the amount of time, the emotional intensity, the intimacy (mutual confiding), and the reciprocal services which characterize the tie."

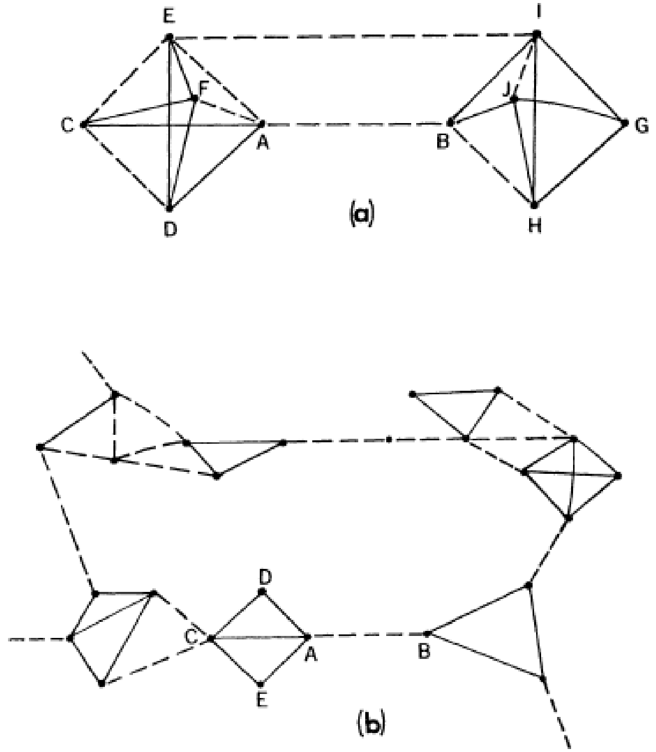


FIG. 2.—Local bridges. *a*, Degree 3; *b*, Degree 13. ——— = strong tie; – – – = weak tie.

Literature: Sociology

- Mark Granovetter: The Strength of Weak Ties (1973)
- Hypothesis: "the stronger the tie between A and B, the larger the proportion of individuals in [the set] to whom they will both be tied, that is, connected by a weak or strong tie."
- "…the triad which is most unlikely to occur, under the hypothesis stated above, is that in which A and B are strongly linked, A has a strong tie to some friend C, but the tie between C and B is absent."

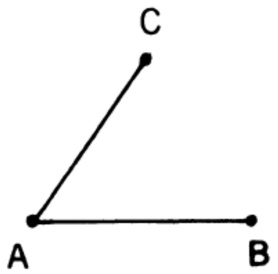


FIG. 1.—Forbidden triad

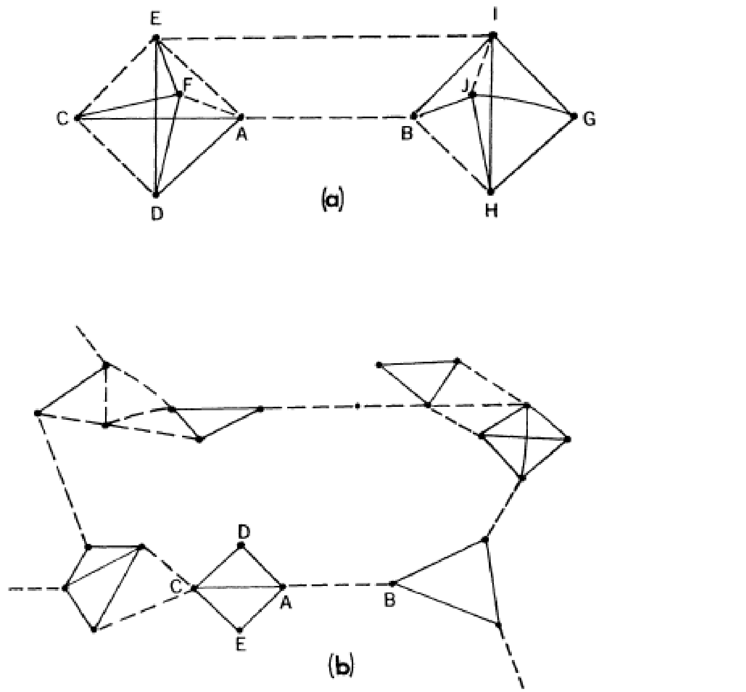


FIG. 2.—Local bridges. *a*, Degree 3; *b*, Degree 13. ——— = strong tie; – – – = weak tie.

Literature: Sociology

- Mark Granovetter: The Strength of Weak Ties (1973)
- "The significance of this triad's absence can be shown by using the concept of a "**bridge**"; this is a line in a network which provides the only path between two points."
- "if the stipulated triad is absent, it follows that, except under unlikely conditions, no strong tie is a bridge…all bridges are weak ties."
- "In large networks it probably happens only rarely, in practice, that a specific tie provides the only path between two points. The bridging function may nevertheless be served locally…[a] **local bridge**…"

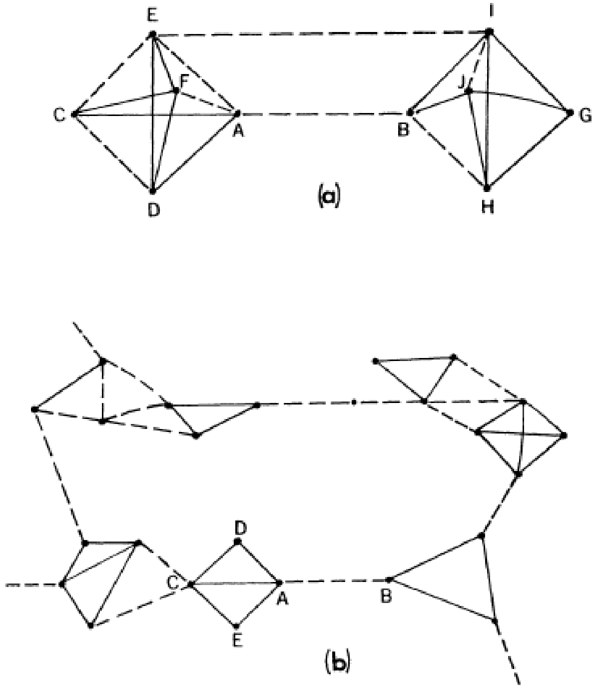


FIG. 2.—Local bridges. *a*, Degree 3; *b*, Degree 13. ——— = strong tie; – – – = weak tie.

Literature: Sociology

- Mark Granovetter: The Strength of Weak Ties (1973)
- “I will refer to a tie as a "local bridge of degree n" if n represents the shortest path between its two points (other than itself), and n > 2. In figure 2a, A-B is a local bridge of degree 3, in 2b, of degree 13.”
- “The significance of weak ties, then, would be that those which are local bridges create more, and shorter, paths”
- “The contention here is that removal of the average weak tie would do more "damage" to transmission probabilities than would that of the average strong one.”

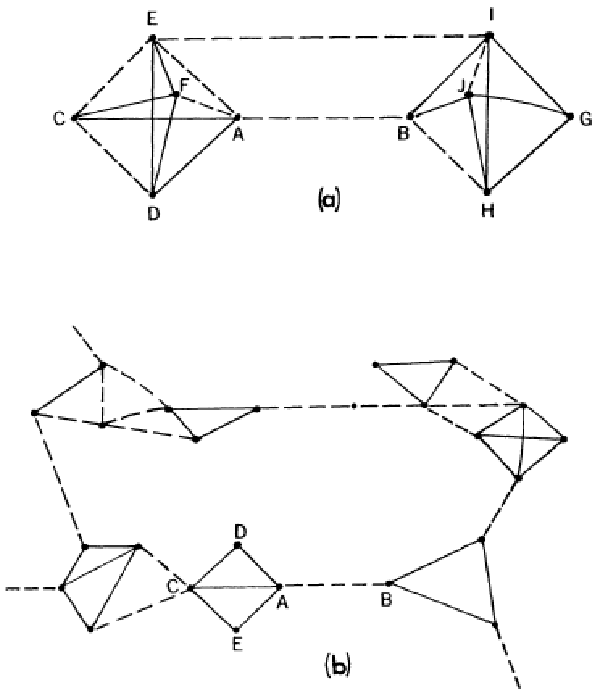


FIG. 2.—Local bridges. *a*, Degree 3; *b*, Degree 13. ——— = strong tie; – – – = weak tie.

Literature: Sociology

- Mark Granovetter: The Strength of Weak Ties (1973)
- "Intuitively speaking, this means that whatever is to be diffused can reach a larger number of people, and traverse greater social distance (i.e., path length),' when passed through weak ties rather than strong."
- "the rumor moving through strong ties is much more likely to be limited to a few cliques than that going via weak ones; bridges will not be crossed."
- Experimental evidence: "The smallest total number of people were reached through the networks generated by first and second choices-presumably the strongest ties-and the largest number through seventh and eighth choices."

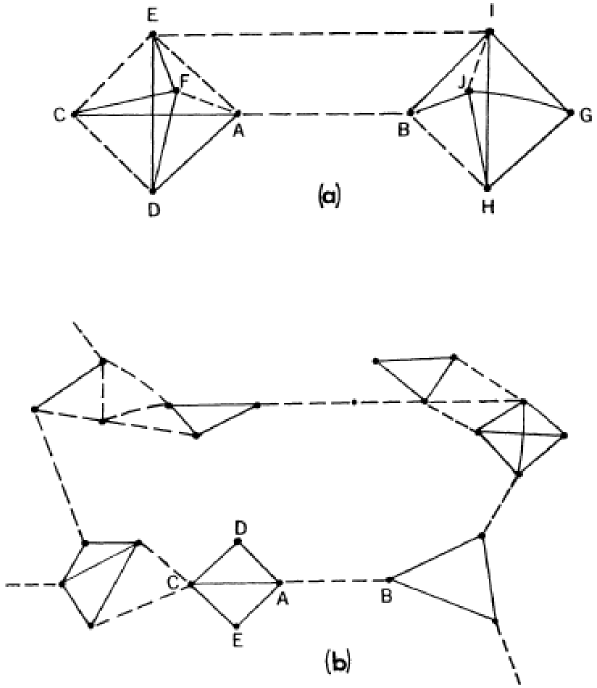


FIG. 2.—Local bridges. *a*, Degree 3; *b*, Degree 13. ——— = strong tie; – – – = weak tie.

Literature: Sociology

- Mark Granovetter: The Strength of Weak Ties (1973)
- "The fewer indirect contacts one has the more encapsulated he will be in terms of knowledge of the world beyond his own friendship circle; thus, bridging weak ties (and the consequent indirect contacts) are important in both ways."
- "It is suggested…that for a community to have many weak ties which bridge, there must be several distinct ways or contexts in which people may form them."
- "the more local bridges (per person?) in a community and the greater their degree, the more cohesive the community and the more capable of acting in concert."

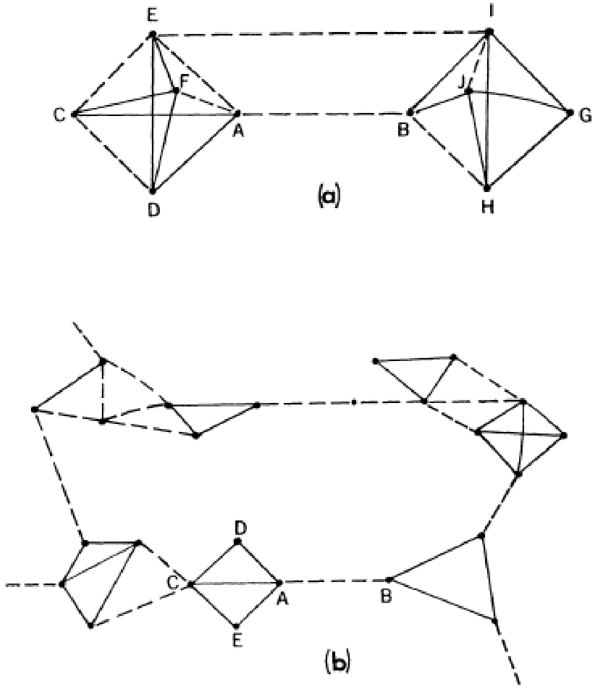


FIG. 2.—Local bridges. *a*, Degree 3; *b*, Degree 13. ——— = strong tie; - - - = weak tie.