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: evolution
- 8/17 – Social norms: diffusion
- 8/22 – Social norms: planned change
- 8/24 – Political networks
- 8/29 – Political networks
- 8/31 – Network theory, review

Evaluation

- Here's how your **final grade** will be calculated:
- Problem Set #1 - 30% [due August 12 @ 11:59pm]
- Problem Set #2 - 30% [due August 19 @ 11:59pm]
- Research proposal - 40% [due September 2; no final exam]

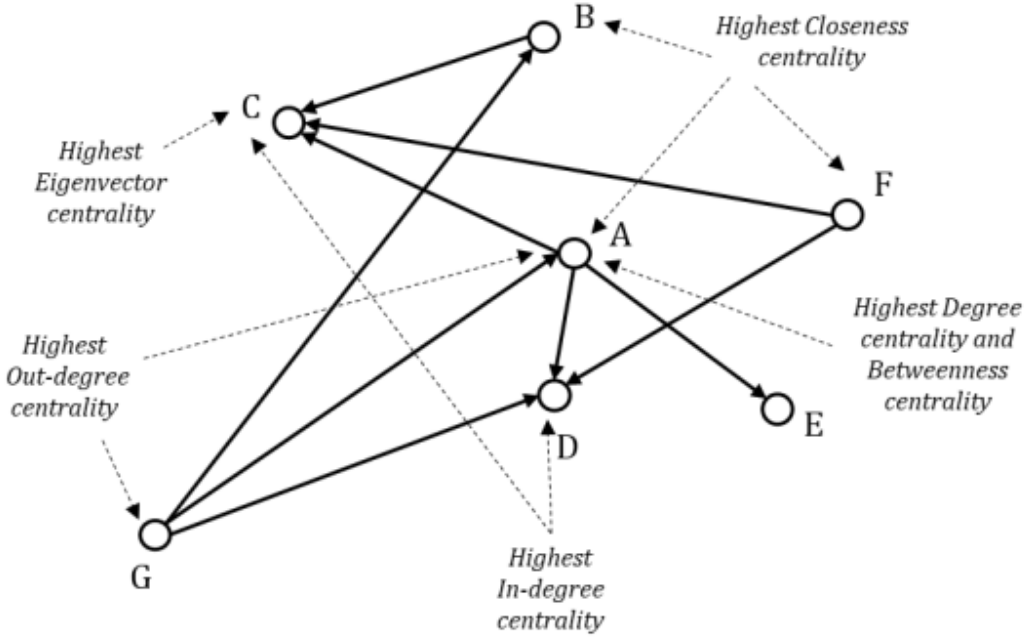
- **Attendance** at lecture is not required, but it is recommended because you'll have the opportunity to ask questions. All lectures will be **recorded** and posted on the corresponding Canvas page.

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.

Social Network Analysis Concepts

- "A **graph** is a way of specifying relationships among a collection of items. A graph consists of a set of objects, called **nodes**, with certain pairs of these objects connected by links called **edges**."
- "We say that two nodes are neighbors if they are connected by an edge."
- "[W]e define a **directed graph** to consist of a set of nodes, as before, together with a set of **directed edges**; each directed edge is a link from one node to another, with the direction being important…When we want to emphasize that a graph is not directed, we can refer to it as an **undirected graph.**"



Social Network Analysis Concepts

- "[A] **path** is…a sequence of nodes with the property that each consecutive pair in the sequence is connected by an edge… a **cycle** is a path with at least three edges, in which the first and last nodes are the same, but otherwise all nodes are distinct."
- "[W]e say that a graph is connected if for every pair of nodes, there is a path between them."
  - "there is no a priori reason to expect graphs in other settings to be connected"

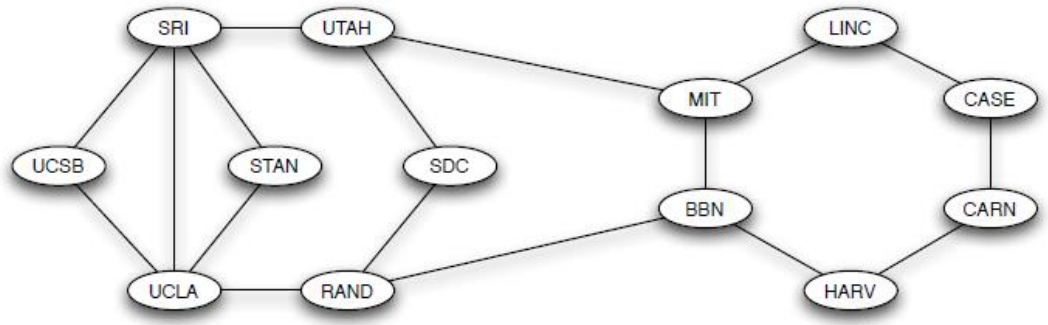


Figure 2.3: An alternate drawing of the 13-node Internet graph from December 1970.

"[E]very edge in the 1970 Arpanet belongs to a cycle, and this was by design"

Social Network Analysis Concepts

- "[A] **connected component** of a graph (often shortened just to the term "component") is a subset of the nodes such that:
  - (i) every node in the subset has a path to every other; and
  - (ii) the subset is not part of some larger set with the property that every node can reach every other.
- "…large, complex networks often have what is called a giant component, a deliberately informal term for a connected component that contains a significant fraction of all the nodes….when a network contains a giant component, it almost always contains only one."

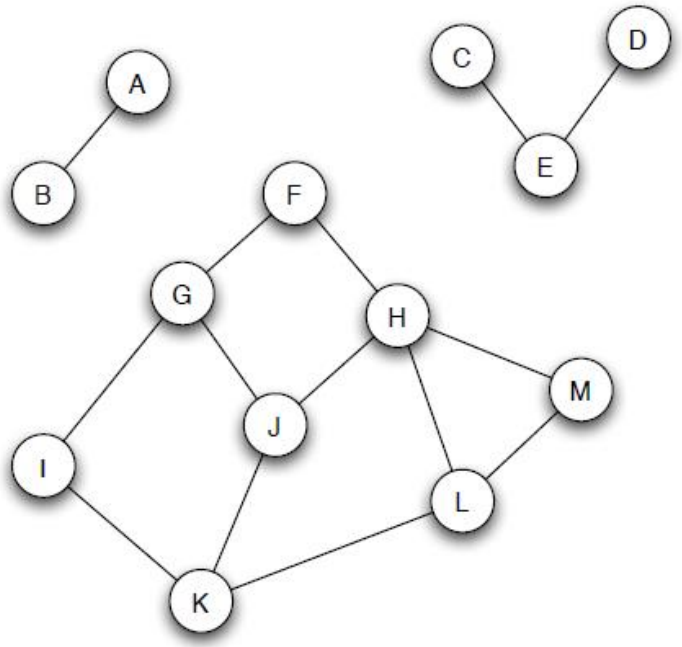


Figure 2.5: A graph with three connected components.

Social Network Analysis Concepts

- "…large, complex networks often have what is called a **giant component**, a deliberately informal term for a connected component that contains a significant fraction of all the nodes….when a network contains a giant component, it almost always contains only one."
- "All it would take is a single edge from someone in the first of these components to someone in the second, and the two giant components would merge into a single component…hence two co-existing giant components are something one almost never sees in real networks."

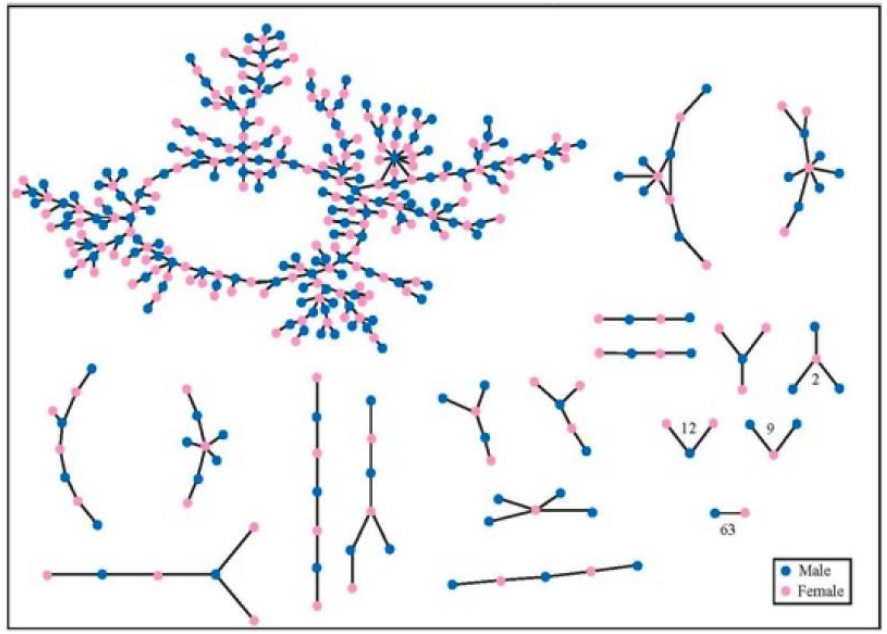


Figure 2.7: A network in which the nodes are students in a large American high school, and an edge joins two who had a romantic relationship at some point during the 18-month period in which the study was conducted [49].

Social Network Analysis Concepts

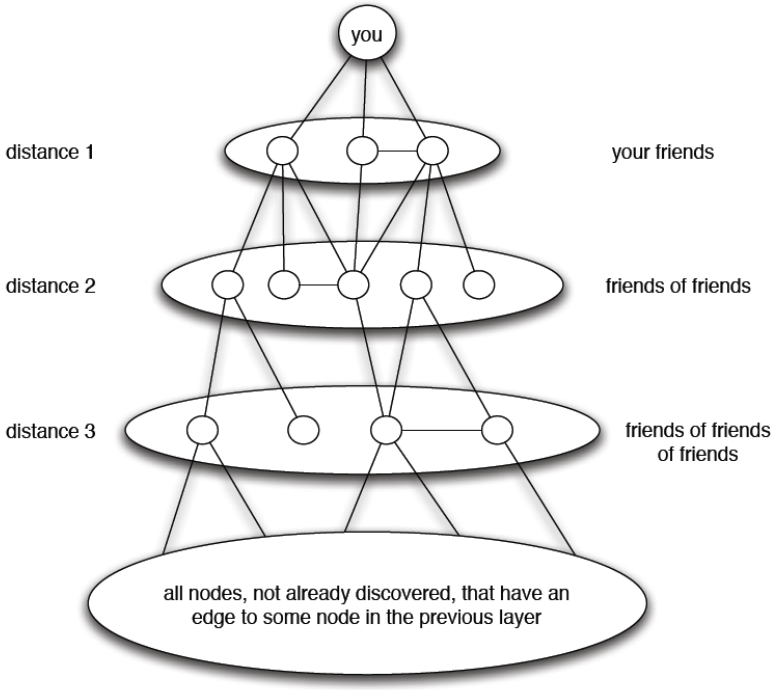
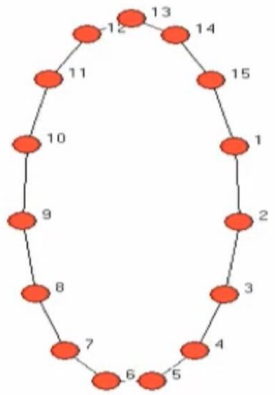


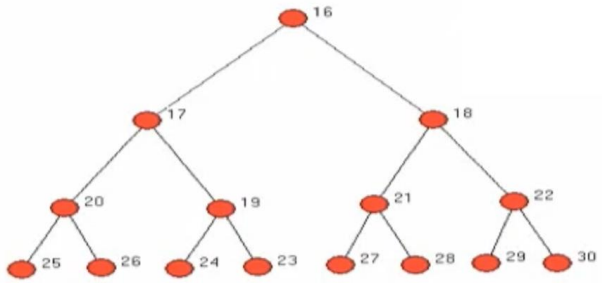
Figure 2.8: Breadth-first search discovers distances to nodes one "layer" at a time; each layer is built of nodes that have an edge to at least one node in the previous layer.

- **Breadth-first search**: categorize nodes by their path distance from some starting node.

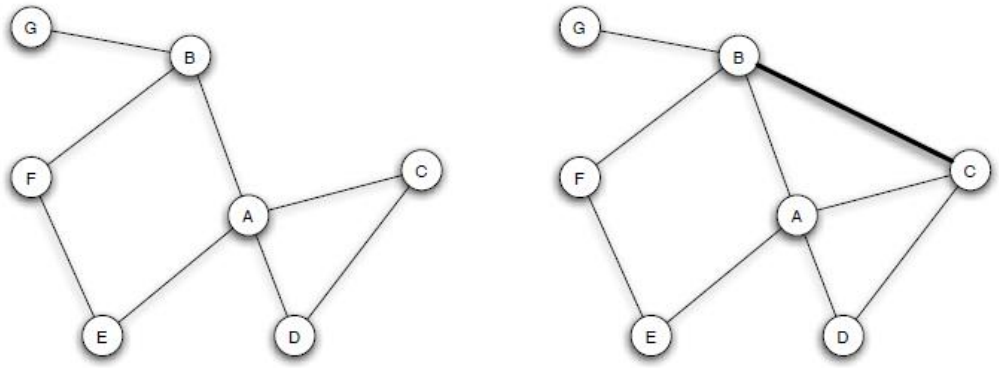
Social Network Analysis Concepts

- Recall that a network's **diameter** is the largest geodesic in the network (or if the network is unconnected, the largest geodesic of the largest component).
- Consider these two networks. Can we calculate their diameter?
- In the ring network, the diameter will scale linearly (n/2). In the tree network, the diameter scales logarithmically ($2\log_2(n+1)$).





Social Network Analysis Concepts



(a) *Before B-C edge forms.* (b) *After B-C edge forms.*

Figure 3.1: The formation of the edge between B and C illustrates the effects of triadic closure, since they have a common neighbor A.

- **Triadic closure**: if nodes B and C have a friend A in common, then the formation of an edge between B and C produces a situation in which all three nodes A, B, and C have edges connecting each other…"
- Recall Granovetter's hypothesis that the so-called forbidden triad was rare. We can interpret this as a claim that triads will tend to close over time.

Social Network Analysis Concepts

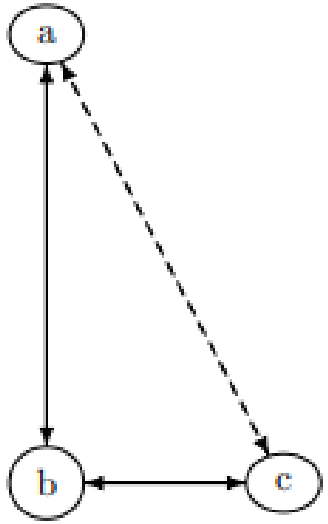


Figure 4: A triad

- Recall the **clustering coefficient** [C].
- "The clustering coefficient of a node A is defined as the probability that two randomly selected friends of A are friends with each other. In other words, it is the fraction of pairs of A's friends that are connected to each other by edges."
- "…the clustering coefficient of a node ranges from 0 (when none of the node's friends are friends with each other) to 1 (when all of the node's friends are friends with each other), and the more strongly triadic closure is operating in the neighborhood of the node, the higher the clustering coefficient will tend to be."

Social Network Analysis Concepts

- "The basic role of triadic closure in social networks has motivated the formulation of simple social network measures to capture its prevalence. One of these measures is the clustering coefficient.
- **The clustering coefficient** of a node A is defined as the probability that two randomly selected friends of A are friends with each other. In other words, it is the fraction of pairs of A's friends that are connected to each other by edges."

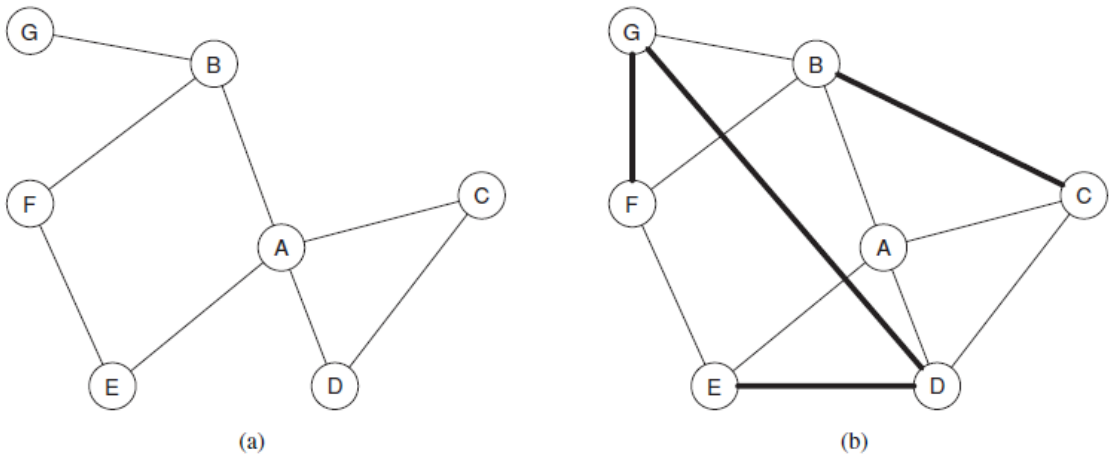


Figure 3.2. If we watch a network for a longer span of time, we can see multiple edges forming; some form through triadic closure while others (such as the D-G edge) form even though the two endpoints have no neighbors in common. The network is illustrated (a) before and (b) after new edges form.

Social Network Analysis Concepts

- "For example, the clustering coefficient of node A in Figure 3.2(a) is 1/6 (because there is only the single C-D edge among the six pairs of friends B-C, B-D, B-E, C-D, C-E, and D-E), and it increases to 1/2 in the second snapshot of the network in Figure 3.2(b) (because there are now the three edges B-C, C-D, and D-E among the same six pairs)."
- "The more strongly the process of triadic closure operates in the neighborhood of the node, the higher the clustering coefficient will tend to be."

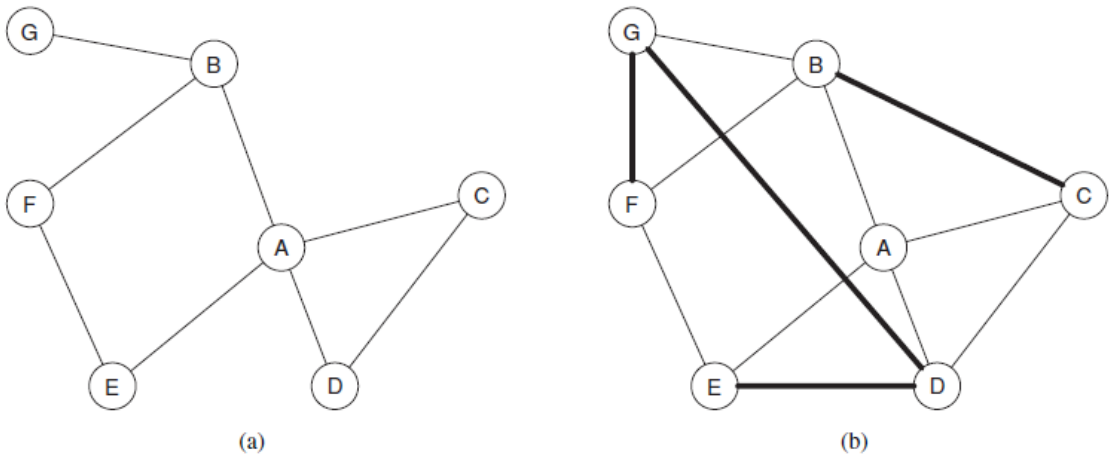


Figure 3.2. If we watch a network for a longer span of time, we can see multiple edges forming; some form through triadic closure while others (such as the D-G edge) form even though the two endpoints have no neighbors in common. The network is illustrated (a) before and (b) after new edges form.

Social Network Analysis Concepts

- "We say that an edge joining two nodes A and B in a graph is a **local bridge** if its endpoints A and B have no friends in common…[in other words] if deleting the edge would increase the distance between A and B to a value strictly more than two."
- "[T]he **span** of a local bridge is the distance its endpoints would be from each other if the edge were deleted."

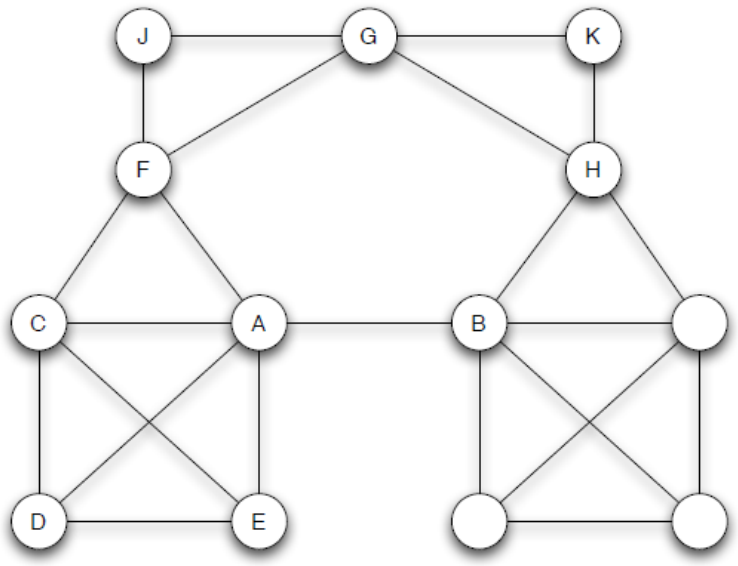


Figure 3.4: The A-B edge is a local bridge of span 4, since the removal of this edge would increase the distance between A and B to 4.

Social Network Analysis Concepts

- "Notice that the definition of a local bridge already makes an implicit connection with triadic closure, in that the two notions form **conceptual opposites**: an edge is a local bridge precisely when it does not form a side of any triangle in the graph."
- "…a local bridge between nodes A and B tends to be a weak tie because if it weren't, triadic closure would tend to produce short-cuts to A and B that would eliminate its role as a local bridge."

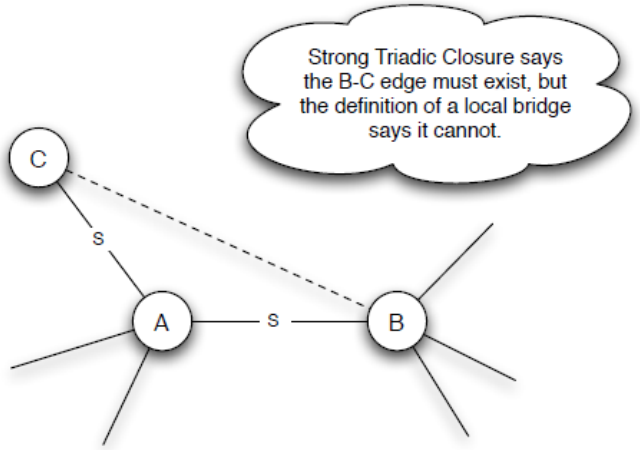


Figure 3.6: If a node satifies Strong Triadic Closure and is involved in at least two strong ties, then any local bridge it is involved in must be a weak tie. The figure illustrates the reason why: if the A-B edge is a strong tie, then there must also be an edge between B and C, meaning that the A-B edge cannot be a local bridge.

Social Network Analysis Concepts

- Global patterns:
  - Path length
  - Degree distribution
- Local patterns:
  - Homophily
  - Clustering
  - Transitivity
- Positional patterns:
  - Centrality
  - Influence

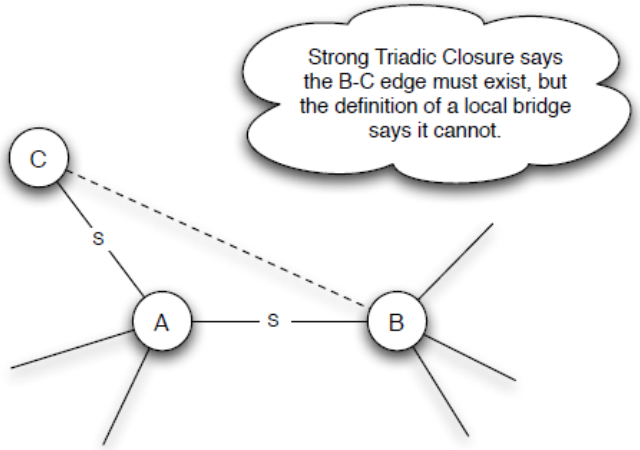
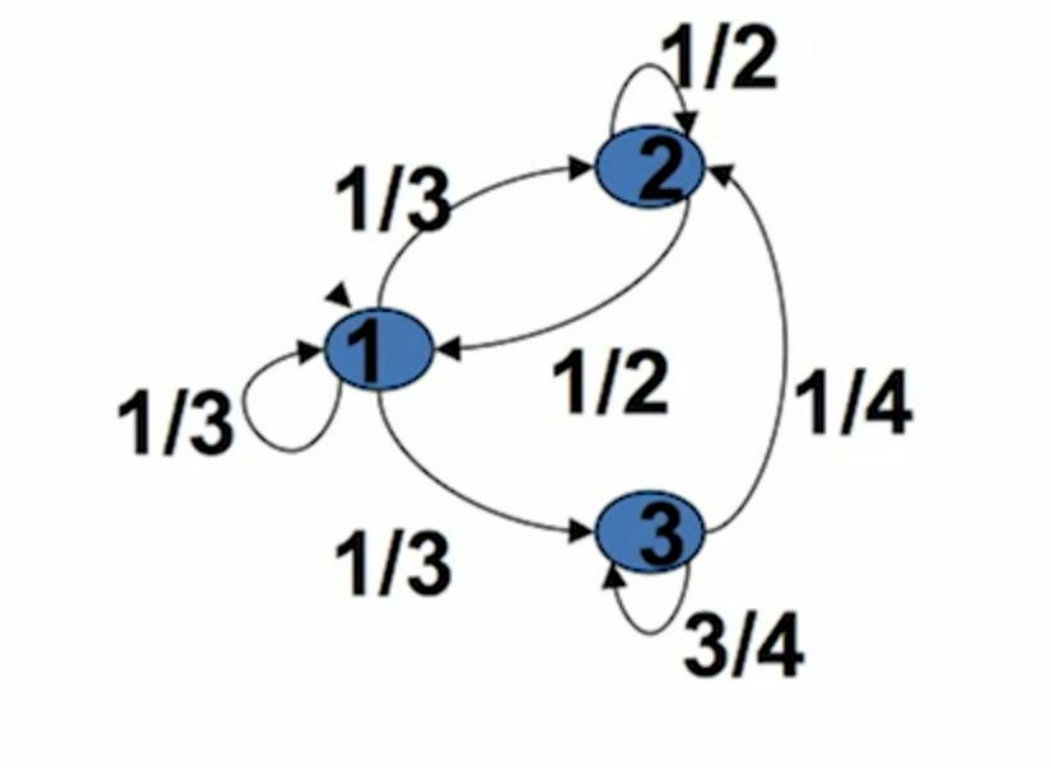


Figure 3.6: If a node satifies Strong Triadic Closure and is involved in at least two strong ties, then any local bridge it is involved in must be a weak tie. The figure illustrates the reason why: if the A-B edge is a strong tie, then there must also be an edge between B and C, meaning that the A-B edge cannot be a local bridge.

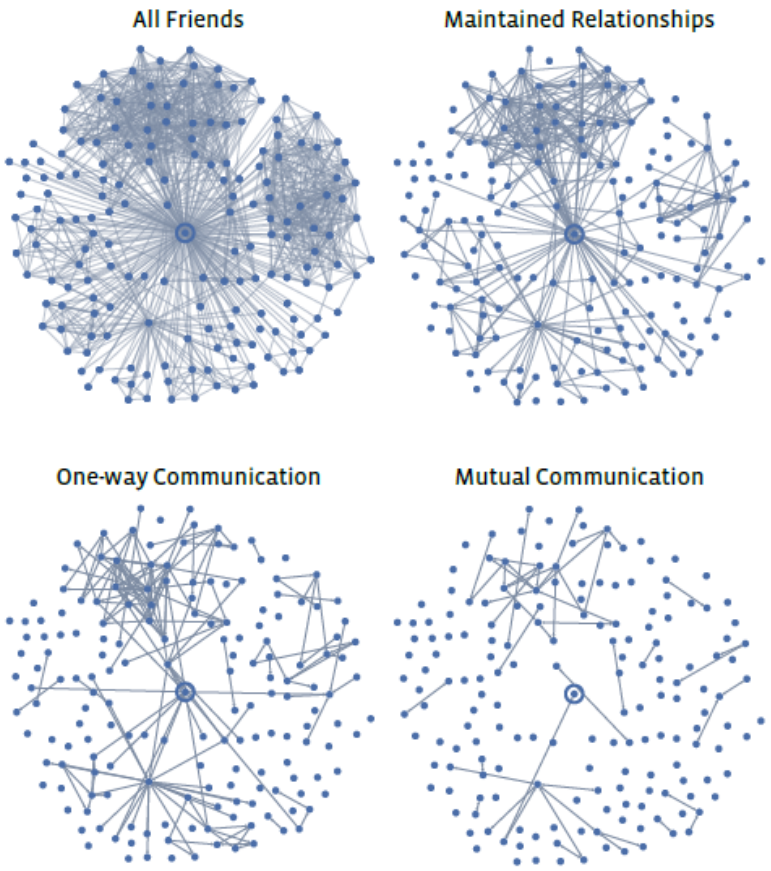
Social Network Analysis Concepts

- Tie strength: we've already seen that we can distinguish between strong and weak ties. We can also **weight** the tie strength using a continuous measure. This can (but need not) sum to 1 depending on the application.
  - Ex. How much time do you spend together each day? [need not sum to 1]
  - Ex. Who influences your decision? [sums to 1]



Social Network Analysis Concepts

- Marlow et al. : "A link represents reciprocal (mutual) communication, if the user both sent messages to the friend at the other end of the link, and also received messages from them during the observation period.
- A link represents one-way communication if the user sent one or more messages to the friend at the other end of the link (whether or not these messages were reciprocated).
- A link represents a maintained relationship if the user followed information about the friend at the other end of the link, whether or not actual communication took place;
- ["following information" here means either clicking on content via Facebook's News Feed service (providing information about the friend) or visiting the friend's prole more than once."]



Social Network Analysis Concepts

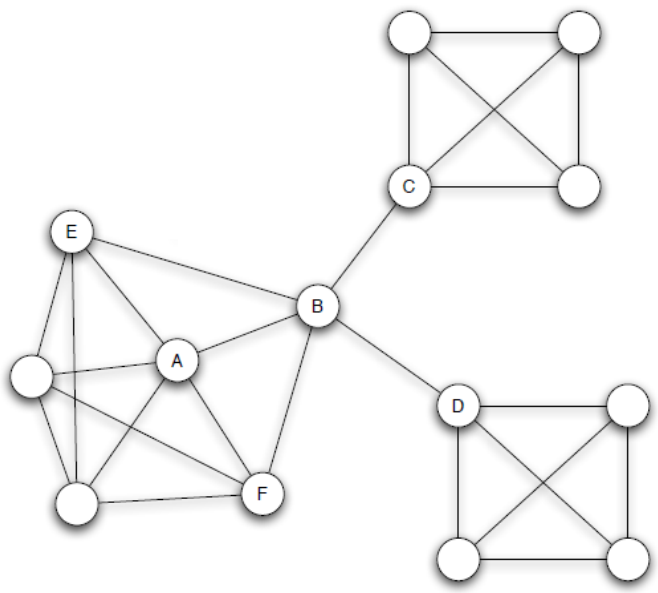


Figure 3.11: The contrast between densely-knit groups and boundary-spanning links is reflected in the different positions of nodes A and B in the underyling social network.

- "We define the **embeddedness** of an edge in a network to be the number of common neighbors the two endpoints have."
  - Node A – high clustering coefficient
  - Links from/to A – high embeddedness
- **Social capital** in networks: "A long line of research in sociology has argued that if two individuals are connected by an embedded edge, then this makes it easier for them to trust one another, and to have confidence in the integrity of the transactions (social, economic, or otherwise) that take place between them."

Social Network Analysis Concepts

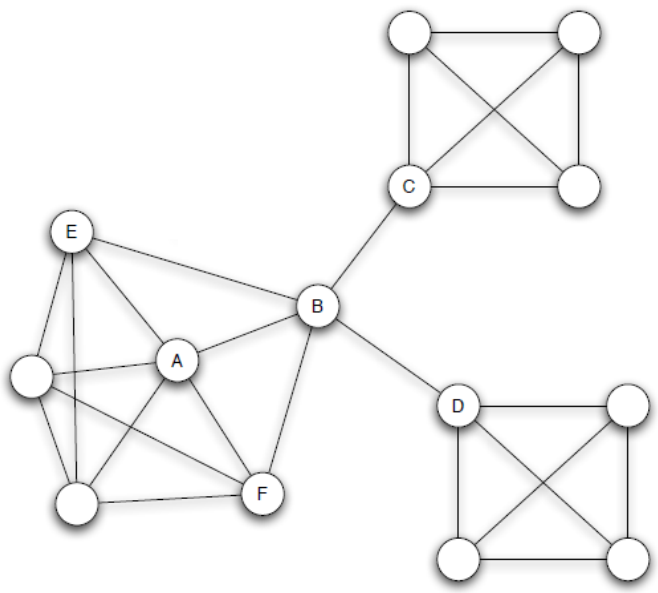
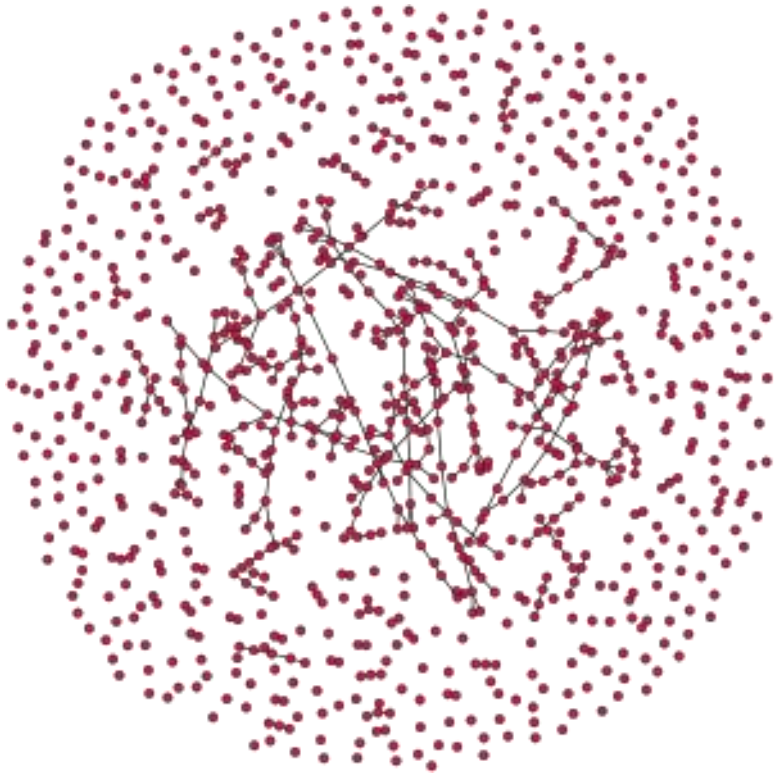


Figure 3.11: The contrast between densely-knit groups and boundary-spanning links is reflected in the different positions of nodes A and B in the underyling social network.

- "No similar kind of deterring threat exists for edges with zero embeddedness, since there is no one who knows both people involved in the interaction. In this respect, the interactions that B has with C and D are much riskier than the embedded interactions that A experiences."
- "[but] B has early access to information originating in multiple, non-interacting parts of the network."

Social Network Analysis Concepts

- **Random networks (random graphs):** We might want to compare the properties of networks we observe in the real world to those of randomly-generated networks.
  - Does our network of interest have parameters that look random, or is it non-random in some systematic and interesting way?
- For example, do new nodes connect to existing nodes at random, or according to some pattern?



Social Network Analysis Concepts

- **Erdos-Renyi graphs**: G(n,p) / G(n,M)
  - "In the G(n,M) model, a graph is chosen uniformly at random from the collection of all graphs which have n nodes and M edges."
  - "In the G(n,p) model, a graph is constructed by connecting labeled nodes randomly. Each edge is included in the graph with probability p, independently from every other edge."
- G = graph; n = number of nodes
- p = probability of forming an edge (uniform attachment)
- M = number of independent edges

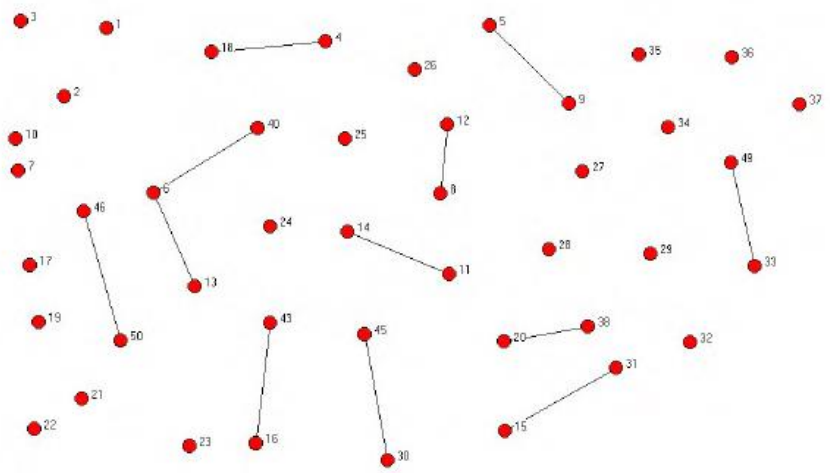


Figure 4.2.2. A First Component with More than Two Nodes: A Random Network on 50 Nodes with p=.01

Social Network Analysis Concepts

- Properties of Erdos-Renyi random graphs:
- **Average degree**: k = (n-1)*p
- **Clustering coefficient**: C = c/(n-1)
- **Number of edges**: M = (n choose 2)p
  - From combinatorics: "choose 2 elements out of n elements."

$$\binom{n}{k} = \frac{n!}{k!(n-k)!},$$

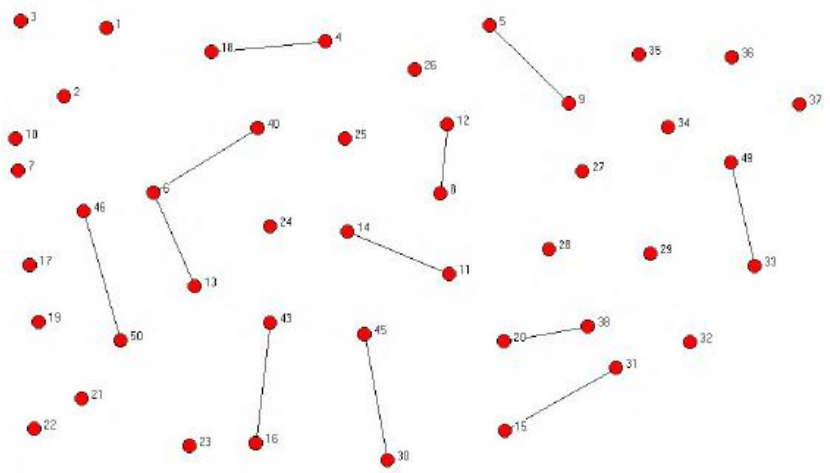


Figure 4.2.2. A First Component with More than Two Nodes: A Random Network on 50 Nodes with p=.01

Social Network Analysis Concepts

- "… the graph in Figure 4.2 shows the friendship network of a (small) hypothetical classroom in which the three shaded nodes are girls and the six unshaded nodes are boys."
- "Consider a given edge in this network. If we independently assign each node the gender male with probability p and the gender female with probability q, then both ends of the edge will be male with probability p^2, and both ends will be female with probability q^2."
- "On the other hand, if the first end of the edge is male and the second end is female, or vice versa, then we have a cross-gender edge, so this happens with probability 2pq. So we can summarize the test for homophily…as follows: "

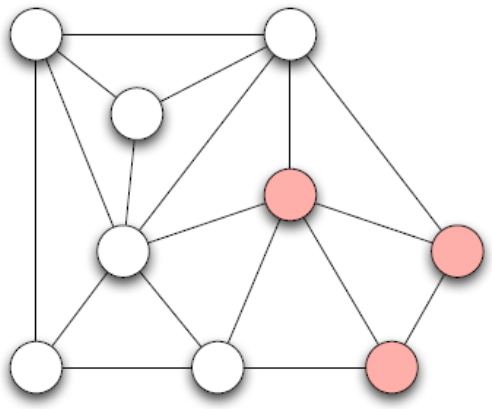


Figure 4.2: Using a numerical measure, one can determine whether small networks such as this one (with nodes divided into two types) exhibit homophily.

Social Network Analysis Concepts

- "**Homophily Test**: If the fraction of [heterogeneous] edges is significantly less than 2pq, then there is evidence for homophily.""
- "In Figure 4.2, for example, 5 of the 18 edges in the graph are cross-gender."
- "Since p = 2/3 and q = 1/3 in this example, we should be comparing the fraction of cross-gender edges to the quantity 2pq = 4/9 = 8/18."

"In other words, with no homophily, one should expect to see 8 cross-gender edges rather than 5, and so this example shows some evidence of homophily."

  - But is it statistically significant?

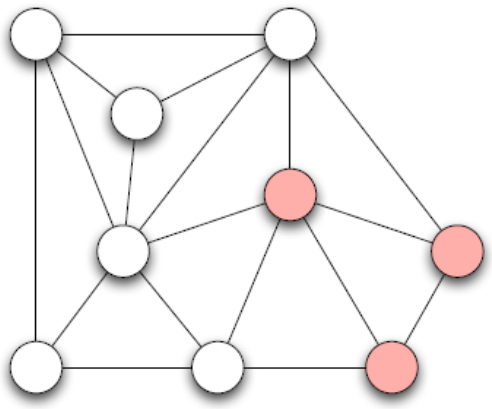
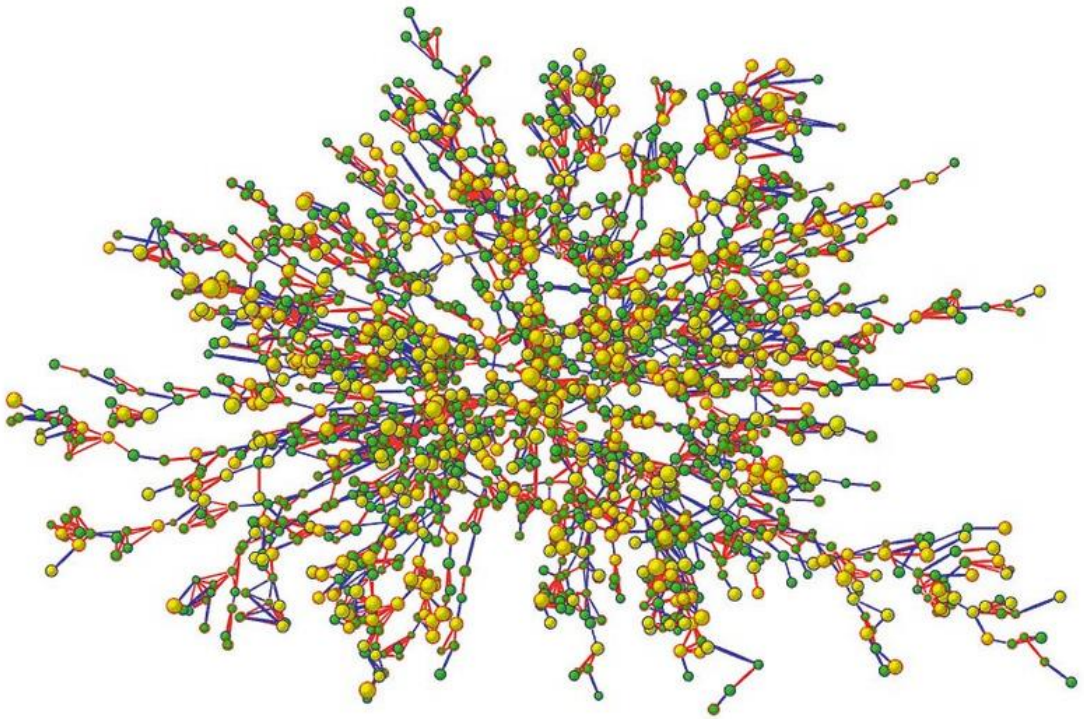


Figure 4.2: Using a numerical measure, one can determine whether small networks such as this one (with nodes divided into two types) exhibit homophily.

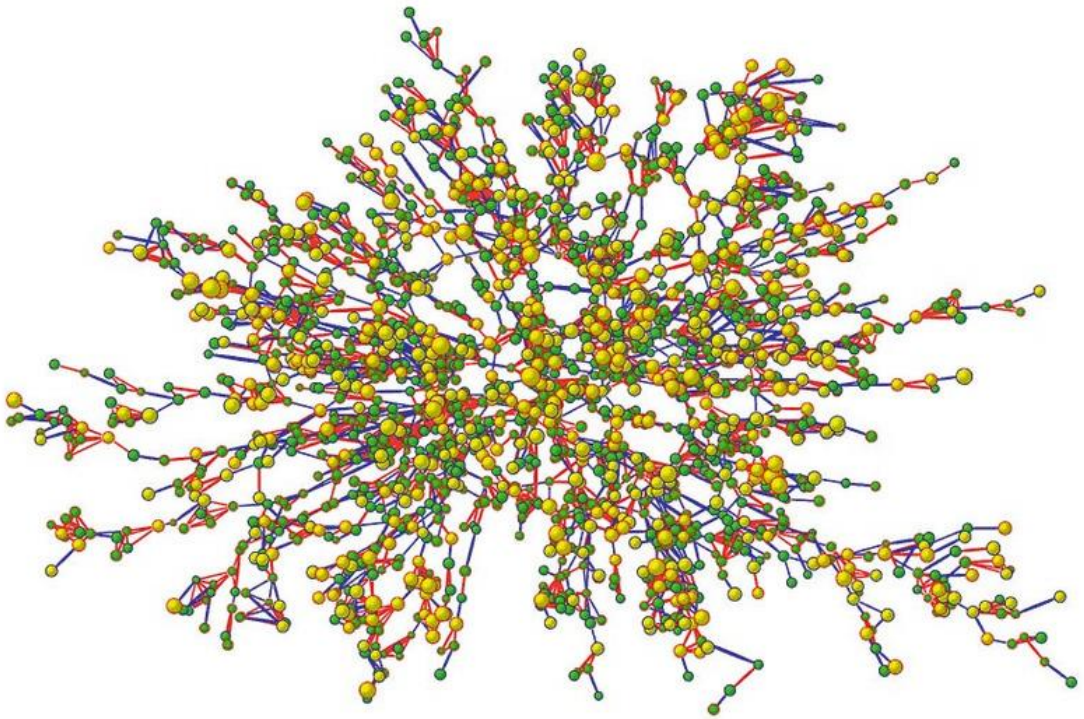
Social Network Analysis Concepts

- "Christakis and Fowler (2014)...found that obese and non-obese people clustered in the network in a fashion consistent with homophily…people tend to be more similar in obesity status to their network neighbors than in a version of the same network where obesity status is assigned randomly."
- "The problem is then to distinguish among several hypotheses for why this clustering is present:"



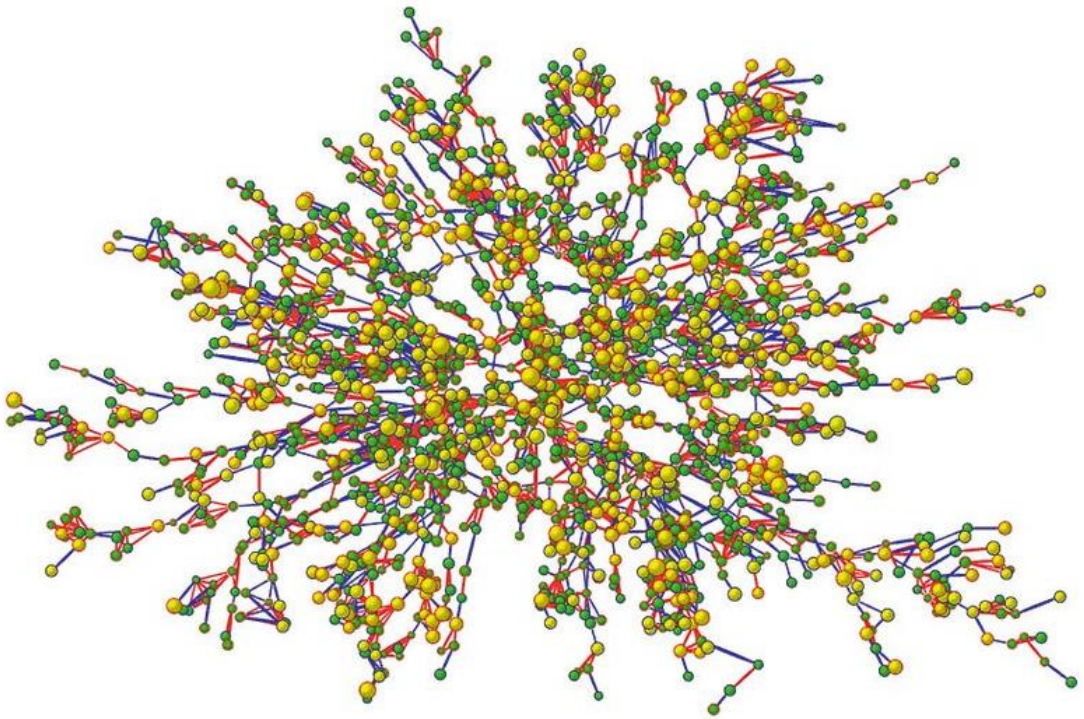
Social Network Analysis Concepts

- (i) because of selection effects, in which people are choosing to form friendships with others of similar obesity status?
- (ii) because of the confounding effects of homophily according to other characteristics, in which the network structure indicates existing patterns of similarity in other dimensions that correlate with obesity status? or
- (iii) because changes in the obesity status of a person's friends was exerting a (presumably behavioral) influence that affected his or her future obesity status?"

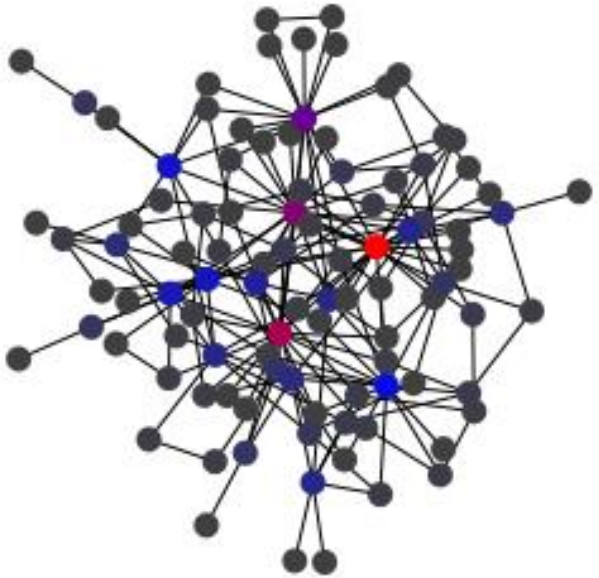


Social Network Analysis Concepts

- “Statistical analysis in Christakis and Fowler's paper argues that, even accounting for effects of types (i) and (ii), there is significant evidence for an effect of type (iii) as well: that obesity is a health condition displaying a form of social influence, with changes in your friends' obesity status in turn having a subsequent effect on you.”
- “This suggests the intriguing prospect that obesity (and perhaps other health conditions with a strong behavioral aspect) may exhibit some amount of "contagion" in a social sense: you don't necessarily catch it from your friends the way you catch the flu, but it nonetheless can spread through the underlying social network via the mechanism of social influence.”



Social Network Analysis Concepts



- In the random networks we just considered, new nodes are linked to existing nodes at random (p). What if we make the probability of attachment proportional to the number of links a node already has?
- Preferential attachment: a model of network formation that makes the probability of linking to an existing node proportional to the number of nodes that link already has.
  - "The rich get richer."
- Emergent properties: emerge at particular scales.
  - Example: giant component, disease transmission

Social Network Analysis Concepts

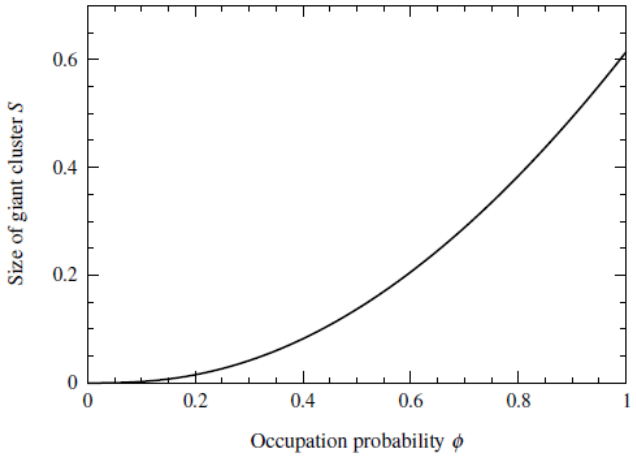


Figure 15.5: Size of the giant cluster for a scale-free network. The size of the giant cluster for a configuration model network with a power-law degree distribution and exponent $\alpha = 2.5$, a typical value for real-world networks. Note the non-linear form of the curve near $\phi = 0$, which means that S, while technically non-zero, becomes very small in this regime. Contrast this figure with Fig. 15.4 for the giant cluster size in a network with an exponential degree distribution.

- Preferential attachment means that degrees in the network will be distributed according to a power law.
- Distribution of degrees according to a power law means that the network will be scale free (Pareto distribution) [looks the same at any scale].
  - "Exponentially few nodes with many connections."
  - "Exponentially many nodes with few connections."
- Each new node i adds m edges to existing nodes.
- So at any time t we have t*m edges.
- And the probability of attaching to some node j is:
  - Degree of node j/2*t*m [each edge connects 2 nodes]

Social Network Analysis Concepts

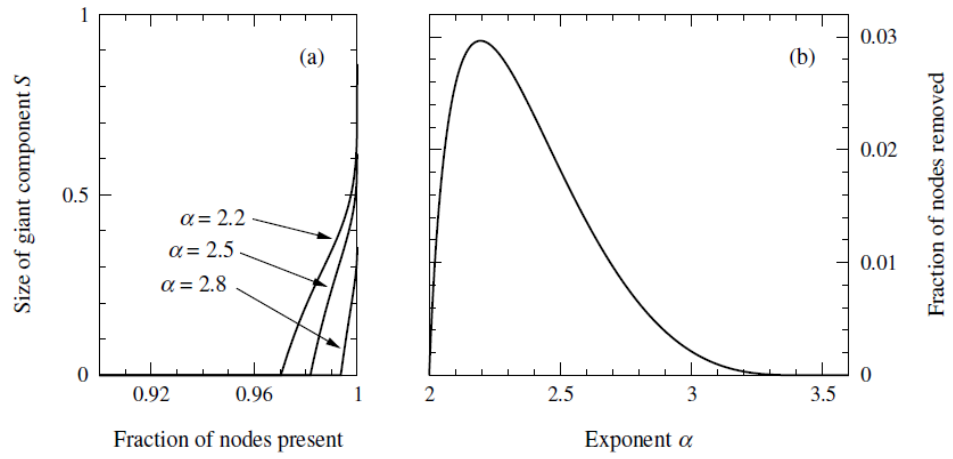


Figure 15.7: Removal of the highest-degree nodes in a scale-free network. (a) The size of the giant cluster in a configuration model network with a power-law degree distribution as nodes are removed in order of their degree, starting with the highest-degree nodes. Only a small fraction of the nodes need be removed to destroy the giant cluster completely. (b) The fraction of nodes that must be removed to destroy the giant cluster as a function of the exponent α of the power-law distribution. For no value of α does the fraction required exceed 3%.

- "Scale-free networks are paradoxically both robust and fragile…On the one hand, they are remarkably robust to the random failure of their nodes, with the giant cluster persisting no matter how many nodes we remove…"
- "On the other hand, scale-free networks are very fragile to attacks targeted specifically at their highest-degree nodes. We need remove only the tiniest fraction of the high-degree hubs in such a network to entirely destroy the giant cluster."

Social Network Analysis Concepts

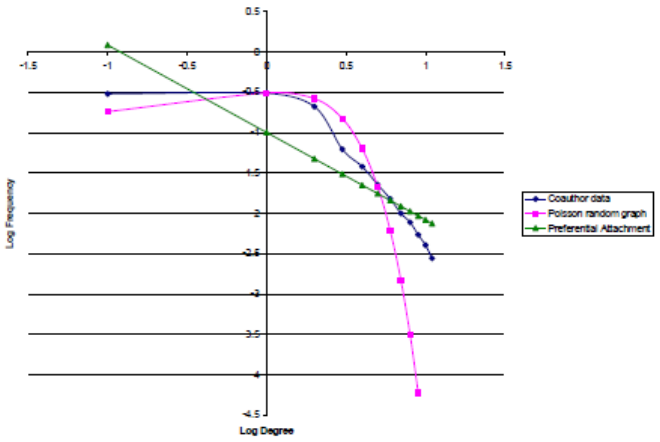
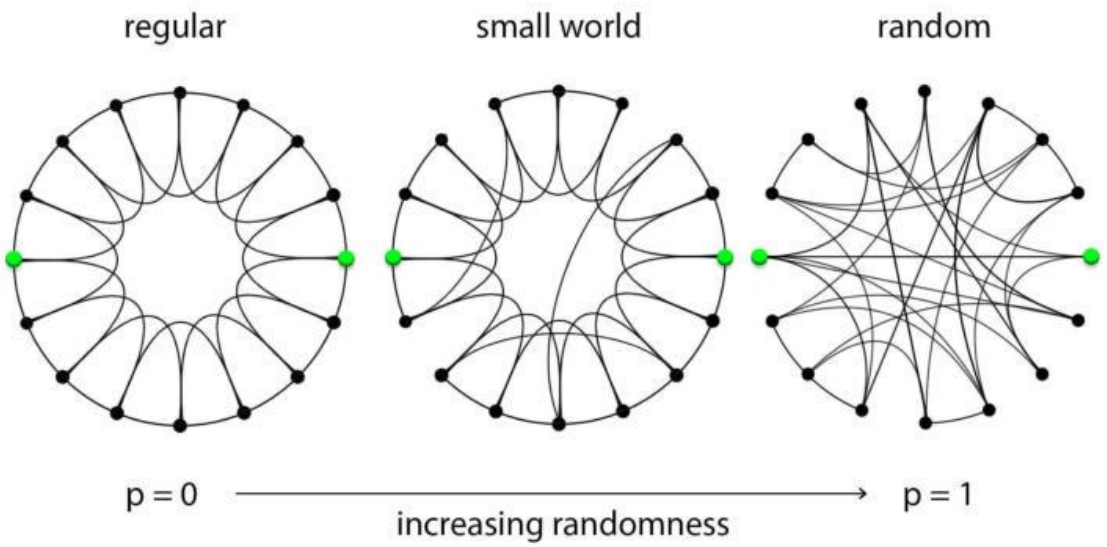


Figure 5.3. A Degree Distribution of A Co-Authorship Network that Fits Between a Uniformly Random Network and One formed via Preferential Attachment

- Hybrid models: we can blend uniform random and preferential attachment in varying proportions.
  - Fitted to data.
- Which resulting network types are most "efficient" at transmitting information?

Social Network Analysis Concepts



- Small-world networks: combine a high clustering coefficient with enough weak ties to give the network a small average path length. [Milgram]
- "…high clustering stems from a distance-based cost structure. Nodes that are closer (or more similar) find it cheaper to maintain links to each other and this generates high clustering."
- "Short overall path length then comes from the fact that if there were no short enough paths between two given nodes, then even if there were a high cost to adding a link, that link would bridge distant parts of the network and bring high benets.to that pair of nodes."

Social Network Analysis Concepts

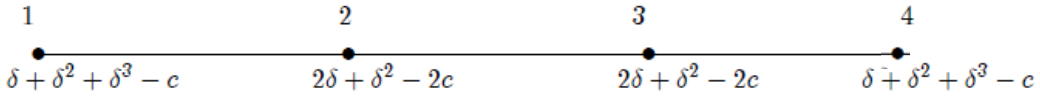


Figure 1.2.4 *The utilities to the players in a three-link four-player network in the symmetric connections model.*

The Symmetric Connections Model:

It is a simple model of social connections that was developed by Jackson and Wolinsky [343]. In this model, links represent social relationships, for instance friendships, between players. These relationships offer benefits in terms of favors, information, etc., and also involve some costs. Moreover, players also benefit from indirect relationships. A "friend of a friend" also results in some indirect benefits, although of a lesser value than the direct benefits that come from a "friend." The same is true of "friends of a friend of a friend," and so forth. The benefit deteriorates with the "distance" of the relationship. This is represented by a factor δ that lies between 0 and 1, which indicates the benefit from a direct relationship and is raised to higher powers for more distant relationships. For instance, in the network where player 1 is linked to 2, 2 is linked to 3, and 3 is linked to 4: player 1 gets a benefit of δ from the direct connection with player 2, an indirect benefit of δ^2 from the indirect connection with player 3, and an indirect benefit of δ^3 from the indirect connection with player 4. The payoffs to this four players in a three-link network is pictured in Figure 1.2.4.

For $\delta < 1$ this leads to a lower benefit from an indirect connection than a direct one. Players only pay costs, however, for maintaining their direct relationships.[16]

Social Network Analysis Concepts

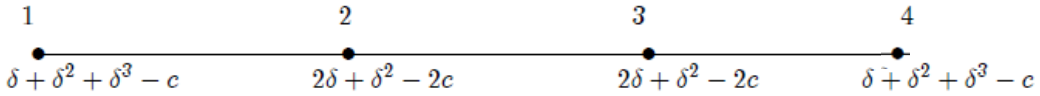


Figure 1.2.4 *The utilities to the players in a three-link four-player network in the symmetric connections model.*

Let us define a network to be *efficient* if it maximizes the total utility to all players in the society. That is, g is efficient if it maximizes $\sum_i u_i(g)$.[18]

It is clear that if costs are very low, it will be efficient to include all links in the network. In particular, if $c < \delta - \delta^2$, then adding a link between any two agents i and j will always increase total welfare. This follows because they are each getting at most δ^2 of value from having any sort of indirect connection between them, and since $\delta^2 < \delta - c$, the extra value of a direct connection between them increases their utilities (and might also increase, and cannot decrease, the utilities of other agents).

When the cost rises above this level, so that $c > \delta - \delta^2$ but c is not too high (see Exercise 1.3), it turns out that the unique efficient network structure is to have all players arranged in a "star" network. That is, there should be some central player who is connected to each other player, so that one player has $n-1$ links and each of the other players has 1 link. The idea behind why a star among all players is the unique efficient structure in this middle cost range, is as follows. A star involves the minimum number

Social Network Analysis Concepts

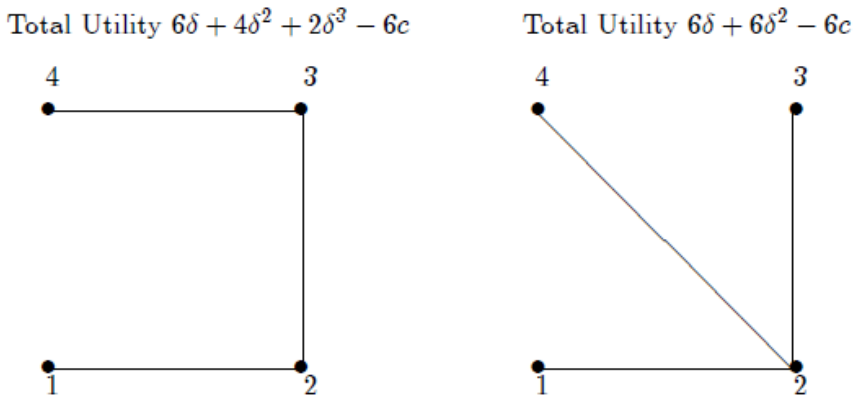


Figure 1.2.4 *The Gain in Total Utility from Changing a "Line" into a "Star".*

structure in this middle cost range, is as follows. A star involves the minimum number of links needed to ensure that all pairs of players are path connected, and it has each player within two links of every other player. The intuition behind this dominating other structures is then easy to see. Suppose for instance we have a network with links between 1 and 2, 2 and 3, and 3 and 4. If we change the link between 3 and 4 to be one between 2 and 4, we end up with a star network. The star network has the same number of links as our starting network, and thus the same cost and payoffs from direct connections. However, now all agents are within two links of each other whereas before some of the indirect connections involved paths of length three. This is pictured in Figure 1.2.4.

As we shall see, this is the key to the set of efficient networks having a remarkably simple characterization: either costs are so low that it makes sense to add links, and then it makes sense to add all links, or costs are so high that no links make sense, or costs are in a middle range and the unique efficient architecture is a star network.

Literature: Psychology

The Surprising Power of Neighborly Advice

Daniel T. Gilbert,[1*] Matthew A. Killingsworth,[1] Rebecca N. Eyre,[1] Timothy D. Wilson[2]

Two experiments revealed that (i) people can more accurately predict their affective reactions to a future event when they know how a neighbor in their social network reacted to the event than when they know about the event itself and (ii) people do not believe this. Undergraduates made more accurate predictions about their affective reactions to a 5-minute speed date ($n = 25$) and to a peer evaluation ($n = 88$) when they knew only how another undergraduate had reacted to these events than when they had information about the events themselves. Both participants and independent judges mistakenly believed that predictions based on information about the event would be more accurate than predictions based on information about how another person had reacted to it.

- ""Before we set our hearts too much upon anything," he wrote, "let us first examine how happy those are who already possess it"
- "La Rochefoucauld was essentially suggesting that forecasters should use other people as surrogates for themselves, and the advantages of his "surrogation strategy" are clear: Because surrogation does not rely on mental simulation, it is immune to the many errors that inaccurate simulations produce."
- "The disadvantages of surrogation are also clear: Individuals differ, and thus, one person's affective reaction is almost certainly an imperfect predictor of another's. But there are at least two reasons to suspect that affective reactions are not as different as people may believe."

Literature: Psychology

The Surprising Power of Neighborly Advice

Daniel T. Gilbert,[1*] Matthew A. Killingsworth,[1] Rebecca N. Eyre,[1] Timothy D. Wilson[2]

Two experiments revealed that (i) people can more accurately predict their affective reactions to a future event when they know how a neighbor in their social network reacted to the event than when they know about the event itself and (ii) people do not believe this. Undergraduates made more accurate predictions about their affective reactions to a 5-minute speed date ($n = 25$) and to a peer evaluation ($n = 88$) when they knew only how another undergraduate had reacted to these events than when they had information about the events themselves. Both participants and independent judges mistakenly believed that predictions based on information about the event would be more accurate than predictions based on information about how another person had reacted to it.

- "First, affective reactions are produced in large part by physiological mechanisms that are evolutionarily ancient, which is why people the world over have very different beliefs and opinions but very similar affective reactions to a wide range of stimuli, preferring warm to cold, satiety to hunger, friends to enemies, winning to losing, and so on."
- "Second, people tend to marry, befriend, work with, and live near those who share their preferences and personality traits, and thus the people from whom they are especially likely to receive surrogation information—the neighbors in their social networks—are especially likely to share their affective reactions."

Literature: Psychology

The Surprising Power of Neighborly Advice

Daniel T. Gilbert,[1*] Matthew A. Killingsworth,[1] Rebecca N. Eyre,[1] Timothy D. Wilson[2]

Two experiments revealed that (i) people can more accurately predict their affective reactions to a future event when they know how a neighbor in their social network reacted to the event than when they know about the event itself and (ii) people do not believe this. Undergraduates made more accurate predictions about their affective reactions to a 5-minute speed date ($n = 25$) and to a peer evaluation ($n = 88$) when they knew only how another undergraduate had reacted to these events than when they had information about the events themselves. Both participants and independent judges mistakenly believed that predictions based on information about the event would be more accurate than predictions based on information about how another person had reacted to it.

- "In short, there is little disagreement among people about the sources of pleasure and pain, and even less disagreement among neighbors. These facts suggest that surrogation may be more powerful than people realize."
- "We tested this hypothesis in two experiments. The events we studied were (i) speed dating, in which undergraduate women predicted how much they would enjoy a 5-min speed date with an undergraduate man, and (ii) peer-evaluation, in which undergraduates predicted how they would feel after being evaluated by a peer (38)."

Literature: Psychology

The Surprising Power of Neighborly Advice

Daniel T. Gilbert,[1*] Matthew A. Killingsworth,[1] Rebecca N. Eyre,[1] Timothy D. Wilson[2]

Two experiments revealed that (i) people can more accurately predict their affective reactions to a future event when they know how a neighbor in their social network reacted to the event than when they know about the event itself and (ii) people do not believe this. Undergraduates made more accurate predictions about their affective reactions to a 5-minute speed date ($n = 25$) and to a peer evaluation ($n = 88$) when they knew only how another undergraduate had reacted to these events than when they had information about the events themselves. Both participants and independent judges mistakenly believed that predictions based on information about the event would be more accurate than predictions based on information about how another person had reacted to it.

- "In both experiments, we gave participants either information that allowed them to simulate the future event (simulation information) or information about the affective reaction of a fellow undergraduate who had experienced the same event in the past (surrogation information). We predicted that participants would make more accurate affective forecasts when they knew nothing about the future event and knew only how someone in their social network had reacted to it."

Literature: Psychology

- "[F]orecasters were considerably more accurate when they used surrogation information (12.50 T 14.10 mm) than when they used simulation information (33.75 T 22.01 mm) [t(86) = 5.38, P < 0.001]. Relative to simulation, surrogation reduced the size of the affective forecasting error by 63%."
- "…although our experiments demonstrate the power of surrogation, they also suggest that people may not normally take advantage of this power. Our participants mistakenly believed that simulation was the superior strategy even after it had failed them."
- "When we want to know our emotional futures, it is difficult to believe that a neighbor's experience can provide greater insight than our own best guess."

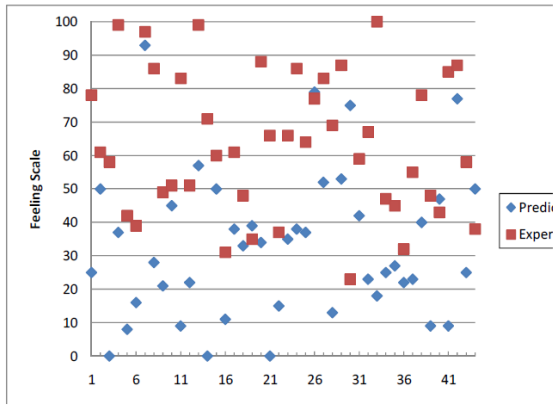


Figure S2a: Affective forecasts (prediction) and affective reports (experience) for each participant in the simulation condition of Experiment 2

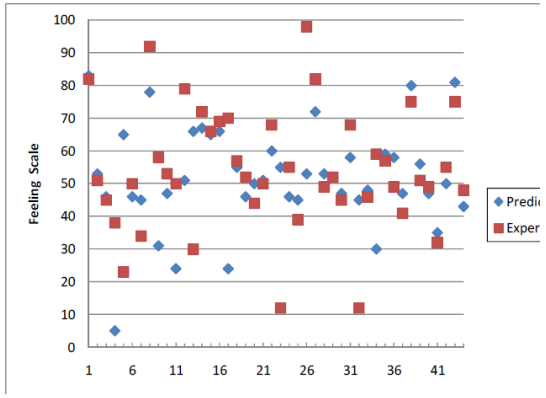


Figure S2b: Affective forecasts (prediction) and affective reports (experience) for each participant in the surrogation condition of Experiment 2