CSI 445/660 – Part 2
(Strong and Weak Ties)

Ref: Chapter 3 of [Easley & Kleinberg].

Strong and Weak Ties

Importance: These notions help in understanding how "local" ties and processes in networks impact their "global" functioning.

Background:

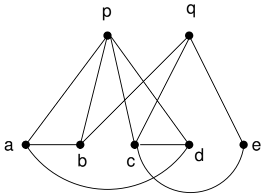
- Mark Granovetter (1943 –)
- Professor of Sociology, Stanford University

- During late 1960's, Granovetter interviewed many people who recently changed jobs.
- Main question: How did you find about the new job?
- Typical answer: Through personal contacts.

Background (continued)

- Many of these contacts were **acquaintances** rather than close friends.
- Granovetter wanted to understand/explain this social phenomenon (without being specific to the "job seeking" domain).
- Led to his work on the "strength of weak ties".

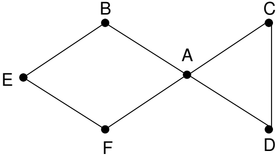
Definition:



- Nodes a, b, c and d are the **neighbors** of p.
- Nodes b, c and e are the **neighbors** of q.

- Applicable to networks that evolve over time.
- Suggested by Georg Simmel (German Sociologist) in 1908 and developed further by Granovetter.

Example:



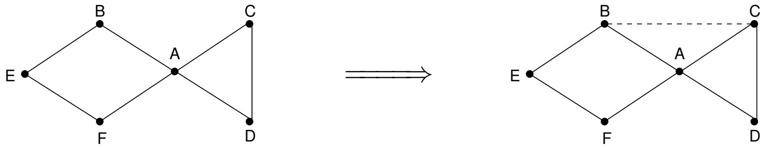
- A friendship network.
- **Question:** Why might this network grow over time?

Basic Principle: **(Triadic Closure)**

If two people have a common friend, then there is an increased likelihood that they will become friends at some point in the future.

Triadic Closure (continued)

Example:

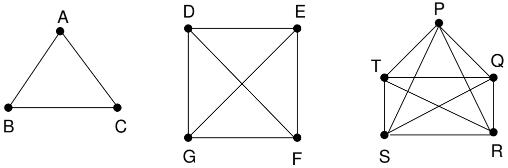


- Network on the left: B and C have a common friend (namely, A).
- By the triadic closure principle, nodes B and C are likely to become friends in the future.
- Nodes A, B and C would then form a **triangle**; edge {B, C} "closes" this triangle (network on the right).
- Examples of other future edges: {F, D} and {B, F}.

Quantifying Triadic Closure

- A common measure: **Clustering Coefficient**.
- Need some preliminaries before defining this measure.

Complete Graph (Clique):



- A clique contains all possible edges between its nodes.
- **Fact:** The number of edges in a clique with k nodes $=$ $k(k-1)/2$.

Definition of Clustering Coefficient

Definition: Suppose the degree of node A is d and the number of edges among the neighbors of A is e. Then, the **clustering coefficient** of A, denoted by $\mathrm{CCF}(A)$, is given by

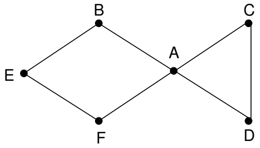
$$\mathrm{CCF}(A) \;=\; \frac{e}{[d(d-1)/2]}$$

Notes:

- The expression $d(d-1)/2$ is the number of edges in a clique with d nodes.
- For any node A, $0 \;\leq\; \mathrm{CCF}(A) \;\leq\; 1$.
- Also called **local clustering coefficient**.

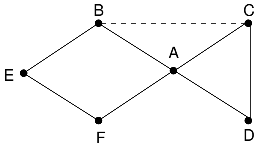
Examples: Clustering Coefficient Calculation

Example 1:



- Degree of A = 4.
- No. of edges among the neighbors of A = 1.
- $\mathrm{CCF}(\mathsf{A}) \;=\; 1/[4(4-1)/2] \;=\; 1/6.$

Example 2:

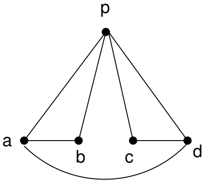


- Degree of A = 4.
- No. of edges among the neighbors of A = 2.
- $\mathrm{CCF}(\mathsf{A}) \;=\; 2/[4(4-1)/2] \;=\; 1/3.$ (Thus, triadic closure increases the clustering coefficient.)

Clustering Coefficient and Triadic Closure

Question: How is the definition of CCF related to triadic closure?

Example: Consider the value of $\mathrm{CCF}(\mathsf{p})$ in the following graph.



- Degree of p $=$ 4.
- No. of edges among the neighbors of p $=$ 3.
- $\mathrm{CCF}(\mathsf{p}) = 3/[4(4-1)/2] = 1/2$.

- Each edge between a pair of neighbors of p forms a triangle that includes p.
- The maximum number of triangle that can include p $= 6$.
- So, we can also define $\mathrm{CCF}(\mathsf{p})$ as the ratio

$$\frac{\text{No. of triangles that include p}}{\text{Maximum number of triangles that can include p}}$$

Some Sociological Reasons for Triadic Closure

Assumption: B and C are friends of A.

1. B and C have increased chances of meeting each other and becoming friends.
2. The friendship with A provides a basis for **mutual trust** between B and C.
3. A may have an **incentive** to make B and C friends. (If B and C are not friends, this may be a source of stress for A.)

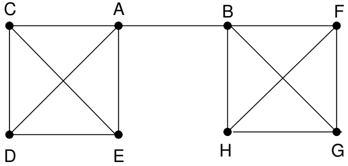
Empirical Evidence for Item 3:

- Bearman & Moody [2004] studied social networks of teenage girls in conjunction with public health records.
- Their finding: Girls whose CCFs are low are more likely to contemplate suicide than those whose CCFs are high.

Bridges and Local Bridges

Definition: A **bridge** is an edge whose removal disconnects a network.

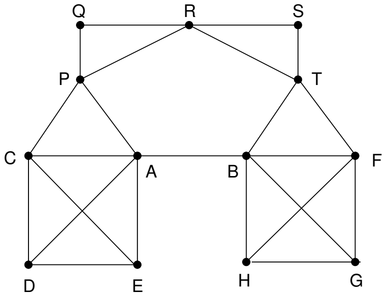
Example:



- Here, {A, B} is a bridge.
- Note that A and B don't have any common neighbors.

- The set of nodes {A, C, D, E} above form a "tightly knit" group.
- Edge {A, B} allows A to "reach a different part" of the network (i.e., it may enable A to get other information that can't be obtained from C, D or E).
- Bridges are rare in social networks. Thus, A and B are likely to be joined through other (longer) paths.

Bridges and Local Bridges (continued)



- Several long paths between A and B.
- This structure is more common in practice.

Definition: An edge {x, y} is a **local bridge** if x and y don't have any common neighbor.

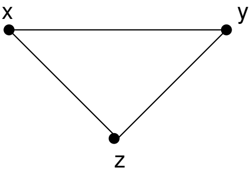
Example: In the above figure, {A, B} is a local bridge.

Observation 1: Every bridge is a local bridge but a local bridge **need not** be a bridge.

Bridges and Local Bridges (continued)

Observation 2: If a local bridge $\{x, y\}$ is removed, then the shortest distance between x and y is **(strictly) larger than 2**.

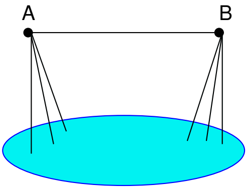
Example:



- Deleting $\{x, y\}$ shortest distance between x and y becomes 2.
- So, $\{x, y\}$ is **not** a local bridge.

Observation 3: An edge is a local bridge only when it **doesn't** form one edge of a triangle. (This is the connection to triadic closure.)

Role of Local Bridges



(Other parts of the network)

- Local bridge {A, B} allows A to get information from B (or vice versa); without the local bridge, A and B will be **far away** from each other.
- All people in the "tightly knit" group that A belongs are likely to have the "same" information.
- So, A is more likely to get new information from a person such as B through a local bridge.

Note: So far, the discussion has not considered whether someone is an "acquaintance" or a "close friend".

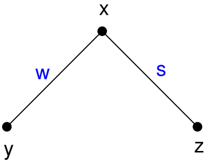
Strong and Weak Ties

- Each edge of the network can be assigned a label "strong" (meaning "close friend") or "weak" (meaning "acquaintance").

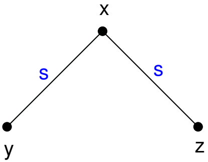
- **Strong Triadic Closure (STC) Condition:** If a node x has **strong** ties to two other nodes y and z, then the graph contains the edge {y, z}.

Note: The STC condition does not specify the label of the edge {y, z}.

Examples:



(Satisfies STC)



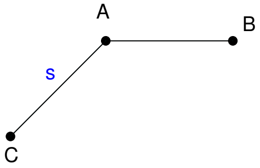
(Violates STC)

Granovetter's Assumption: Every node satisfies STC.

Theorem: If a node A satisfies STC and is involved in at least two strong ties, then every local bridge involving A must be a weak tie.

Proof: To be discussed in class.

An informal explanation:



- A local bridge {A, B} is generally a weak tie.
- If not, STC would produce shortcuts that would eliminate its role as a local bridge (i.e., A and B would become part of the same "tight knit" community).

Summary: Strength of Weak Ties

- Local bridges help in getting information from other parts of the network.
- Under STC, local bridges represent weak ties.
- The formalism relates tie strengths to network structure.
- High level principles from Granovetter's work:
    1. Weak links connect together tightly knit groups.
    2. As tie strength increases, local bridges tend to become edges in tightly knit groups.

Extension of the Study to Large Networks

- Granovetter's study used small (manually constructed) social networks to support the conclusions.
- Other researchers have tested the high level principles resulting from Granovetter's work on large networks.
- Mathematical results require sharp dichotomies:
  - An edge is either a local bridge or not a local bridge.
  - An tie is either weak or strong.
- Such requirements should be relaxed when conducting empirical large studies on practical networks.

Example: A Cell Phone Network Study

Ref: Onnela et al. [2007] (Reference [334] in the text.)

Information About the Network:

- A cell phone network observed over a period of 18 weeks.
- Each node is a user and edge $\{x, y\}$ means that x and y called each other at least once during the observation period.
- 4.6 million nodes and about 7 million edges. (The number of users represents about 20% of a country's population.)
- Giant component had 84% of the nodes.

Relaxing the Notion of Tie Strength:

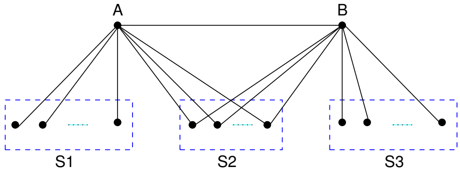
- Tie strength measured by the number of minutes of conversation.
- Edges are sorted by their strengths and the **percentile** values of edges are considered.

Cell Phone Network Study (continued)

Relaxing the Notion of Local Bridge:

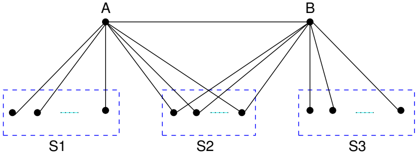
- Number of local bridges in large networks is small.
- So, a slightly relaxed notion ("almost local bridges") is used.

Preliminary Definitions:



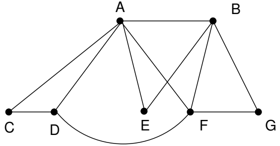
- $S1$: Nodes that are **neighbors of A but not neighbors of B**.
- $S2$: Nodes that are neighbors of **both** A and B.
- $S3$: Nodes that are **neighbors of B but not neighbors of A**.

Cell Phone Network Study (continued)



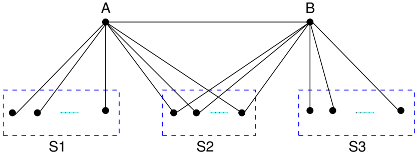
- $S1 \cup S2 \cup S3$ is the set of nodes which are neighbors of **at least one of A and B**. (Neither A nor B is part of $S1 \cup S2 \cup S3$.)
- $S2$ is the set of nodes that are neighbors of **both** A and B. (The quantity $|S2|$ is called the **embeddedness** of the edge {A, B}.)

Example:



Embeddedness of $\{A, B\} = 2$.

Definition: Neighborhood Overlap



Definition: Suppose A and B are nodes in a network G which contains the edge {A, B}. Then the **neighborhood overlap** of edge {A, B}, denoted by $\mathrm{NOV(A, B)}$, is defined by

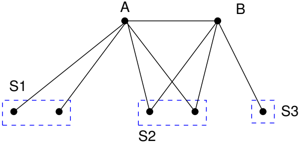
$$\mathrm{NOV(A, B)} \ = \ \frac{\text{\#Nodes which are neighbors of both A and B}}{\text{\#Nodes which are neighbors of at least one of A and B}}$$

An equivalent definition: Suppose sets S1, S2 and S3 for the edge {A, B} are as shown in the above figure. Then the **neighborhood overlap** of edge {A, B} is defined by

$$\mathrm{NOV(A, B)} \ = \ \frac{|\mathrm{S2}|}{|\mathrm{S1}| + |\mathrm{S2}| + |\mathrm{S3}|}$$

Example of Neighborhood Overlap Computation

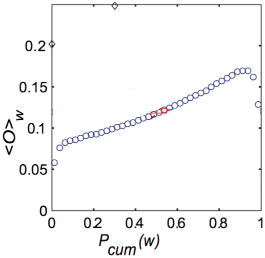
Example:



$$\begin{aligned} \mathrm{NOV(A,B)} &= |S2|/(|S1|+|S2|+|S3|) \\ &= 2/(2+2+1) = 0.4 \end{aligned}$$

- For any local bridge {A, B}, NOV(A,B) = 0.
- So, edges with small NOV values can be considered **"almost local bridges"**.
- One should expect NOV to increase with tie strength. (This is a consequence of second of the high level principles from Granovetter's work.)
- This is supported by the study of Onnela et al.

Results from the Study of Onnela et al.



- Tie strength is along the X-axis and the NOV values are along the Y-axis.
- As the tie strength increases, the NOV value also increases.

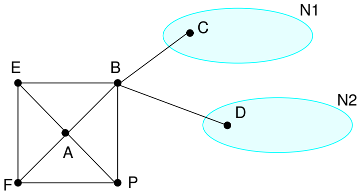
Evidence for Weak links Between Tightly Knit Groups:

- The evidence is indirect. (It is based on two experiments.)
- **Experiment I:** Edges are deleted from the network starting from the strongest edges.
  - Here, the size of the giant component shrank gradually.

Evidence for Weak Links ... (continued)

- **Experiment II:** Edges are deleted from the network starting from the weakest edges.
  - Here, the size of the giant component shrank much more rapidly.

Roles of Nodes:

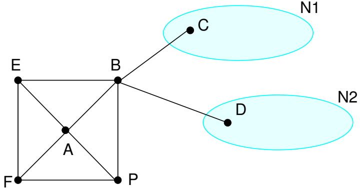


- Node A located in the "middle" of a tightly knit group.
- Node B located at the "interface" between multiple groups.

Question: What is the difference between the experiences of A and B?

Roles of Nodes (continued)

- Recall that the **embeddedness** of an edge {x,y} is the number of neighbors that are common to x and y.

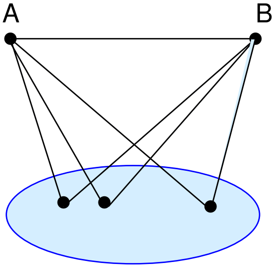


- Embeddedness of edge {A, E} = 2.
- So, {A, E} is **not** a local bridge.

Importance of Embeddedness: If two people are joined by an edge with large embeddedness, it is easier for them to trust each other (and have more confidence in transactions between them).

Roles of Nodes (continued)

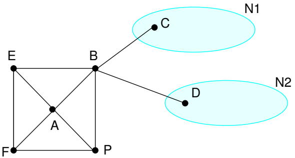
Reason:



(Common neighbors)

- If A "misbehaves", a large number of (common) friends will find out about it.
- As a consequence, A's reputation is likely to suffer.
- So, nodes involved in edges of high embeddedness can trust their friends.

Roles of Nodes (continued)

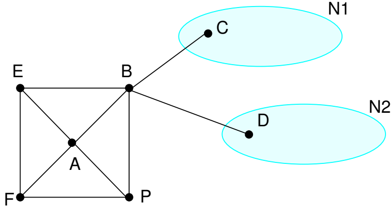


- Node B's situation is different from that of A.
- B has several **bridges** incident on it.
- B is called a **structural hole** (or **articulation point**).

- Ronald Burt (1949 –)
- School of Business, University of Chicago

- Imagine the above graph as representing interactions among managers in a company.

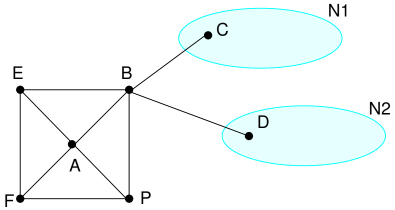
Roles of Nodes (continued)



Why B enjoys a position of power:

- B may have early access to information that originates at different parts of the network.
- B has the opportunity to combine knowledge from disparate sources, thus having more opportunity for creativity.
- B can serve as a "gate keeper" regulating the access of C and D to the group containing A, E, F and P. (If there is an edge from D to P, then that would diminish B's power as "gate keeper".)

Roles of Nodes (continued)



Why B's interests may not be aligned with those of the company:

- To hold on to "power", B may want to control the flow of information among the various groups.
- For the organization to function effectively, information needs to flow readily between the various groups.