# CSC200: Lecture 4

Allan Borodin

# Announcements

- My apologies for the tutorial room mixup on Wednesday. The room SS 1088 is only reserved for Fridays and I forgot that.
- My office hours: Tuesdays 2-4 (SF 2303B) or by appointment; note that I may have to move office hours some weeks but will notify class. You can also drop in and if I am not busy, I am happy to meet.
- First quiz on Friday, October 2. Next week, I will announce the scope of the quiz. It is timed for 15 minutes and the remaining part of the tutorial will follow. We will be using SS 1069 and SS 1088 so please attend the appropriate tutorial.

# Today's agenda

- Last lecture:
  - We finished our introduction of basic graph concepts
  - Basic graph definitions (see Chapter 2) and some additional concepts:
    - forests and trees for undirected graphs
    - directed paths
    - cycles and trees for directed graphs
    - briefly discussed the use of node and edge weighted graphs.
    - briefly discussed embeddedness and dispersion with regard to the romantic relation prediction problem.
  - We also just began discussing chapter 3 of the text.
- This lecture: Cntinued discussion of "Strong and Weak Ties" (Chapter 3 of textbook).

# Chapter 3: Strong and Weak Ties

There are two themes that run throughout this chapter.

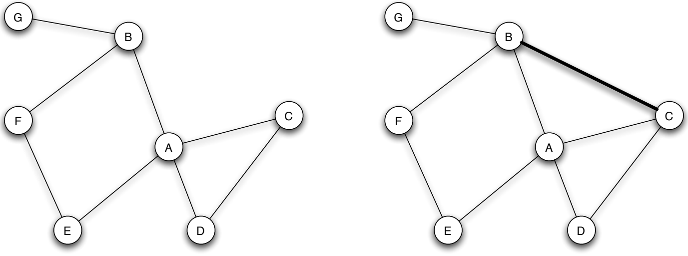
1. Strong vs. weak ties and "the strength of weak ties" is the specific defining theme of the chapter. Also start the discussion of how networks evolve.

2. The larger theme is in some sense "the scientific method".
   - Formalize concepts, construct models of behaviour and relationships, and test hypotheses.
   - Models are not meant to be the same as reality but to abstract the important aspects of a system so that it can be studied and analyzed.
   - See the discussion of the strong triadic closure property on pages 49-50 of textbook.

**Notes**

- strong ties: stronger links, corresponding to friends
- weak ties: weaker links, corresponding to acquaintances

# Triadic closure (undirected graphs)




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

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

- Triadic closure: mutual “friends” of say $A$ are more likely (than “normally”) to become friends over time.
- How do we measure the extent to which triadic closure is occurring?
- How can we know why a new friendship is formed? (Such ties can range from just knowing someone to friendship .)

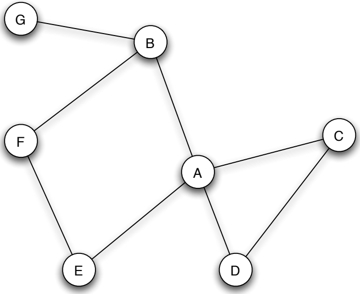
# Measuring the extent of triadic closure

- The clustering coefficient of a node $A$ is a way to measure (over time) the extent of triadic closure (perhaps without understanding why it is occurring).
- Let $E$ be the set of an undirected edges of a network graph. For an node $A$, the clustering coefficient is the following ratio:

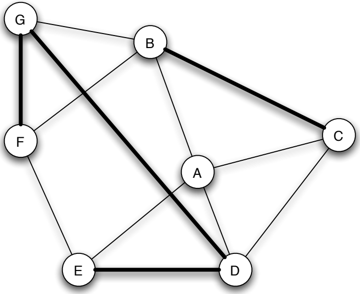
$$\frac{\left|\left\{(B,C) \in E : (B,A) \in E \text{ and } (C,A) \in E\right\}\right|}{\left|\left\{\{B,C\} : (B,A) \in E \text{ and } (C,A) \in E\right\}\right|}$$

- The numerator is the number of all edges $(B, C)$ in the network such that $B$ and $C$ are adjacent to $A$.
- The denominator is the number of all unordered pairs $\{B, C\}$ such that $B$ and $C$ are adjacent to $A$.

# Example of clustering coefficient




(a) *Before new edges form.*




(b) *After new edges form.*

- The clustering coefficient of node $A$ in Fig. (a) is $1/6$ (since there is only the single edge $(C, D)$ among the six pairs of friends $\{B, C\}$, $\{B, D\}$, $\{B, E\}$, $\{C, D\}$, $\{C, E\}$, and $\{D, E\}$)
- The clustering coefficient of node $A$ in Fig. (b) increased to $1/2$ (because there are three edges $(B, C)$, $(C, D)$, and $(D, E)$).
- Note that another edge $(D, G)$ has also formed for some other reason.

# Interpreting triadic closure

- Does a low clustering coefficient suggest anything?

# Interpreting triadic closure

- Does a low clustering coefficient suggest anything?

- Bearman and Moody reported finding that a low clustering coefficient amongst teenage girls implies a higher probability of suicide (compared to those with high clustering coeficient). How can we understand this finding?

# Interpreting triadic closure

- Does a low clustering coefficient suggest anything?

- Bearman and Moody reported finding that a low clustering coefficient amongst teenage girls implies a higher probability of suicide (compared to those with high clustering coeficient). How can we understand this finding?

- What are the reasons for triadic closure?

# Interpreting triadic closure

- Does a low clustering coefficient suggest anything?

- Bearman and Moody reported finding that a low clustering coefficient amongst teenage girls implies a higher probability of suicide (compared to those with high clustering coeficient). How can we understand this finding?

- What are the reasons for triadic closure?

- Opportunity to meet, trust, incentive ; it can be awkward to have good friends (i.e. with strong ties) who are not themselves friends. The implication is that low clustering coefficient imples few good friends.

# Granovetter's thesis: the strength of weak ties

- In 1960s interviews: Many people learn about new jobs from personal contacts (not surprising) and often these contacts were acquaintances rather than friends (surprising?). Upon a little reflection, this intuitively makes sense.

- The idea is that weak ties link together "tightly knit communities", each containing a large number of strong ties.

- Can we say anything more quantitative about such phenomena?

- To gain some understanding of this phenomena, we need some additional concepts relating to structural properties of a graph.

**Recall**

- strong ties: stronger links, corresponding to friends
- weak ties: weaker links, corresponding to acquaintances

# Bridges and local bridges

- One measure of connectivity is the number of edges (or nodes) that have to be removed to disconnect a graph.
- A bridge (if one exists) is an edge whose removal will disconnect a (connected) graph.
- We expect that large social networks will have a "giant component" and few bridges.
- A local bridge is an edge $(A, B)$ whose removal would cause $A$ and $B$ to have graph distance (called the span of this edge) greater than two. Note: Span is a dispersion measure as informally defined in Lecture 3. Would the span of an edge be useful for the detection of the romantic relation as discussed in Lecture 3?
- Local bridges (A,B) play a role similar to bridges providing access for $A$ and $B$ to parts of the network that would otherwise be (in a useful sense) inaccessible.

# Local bridge $(A, B)$

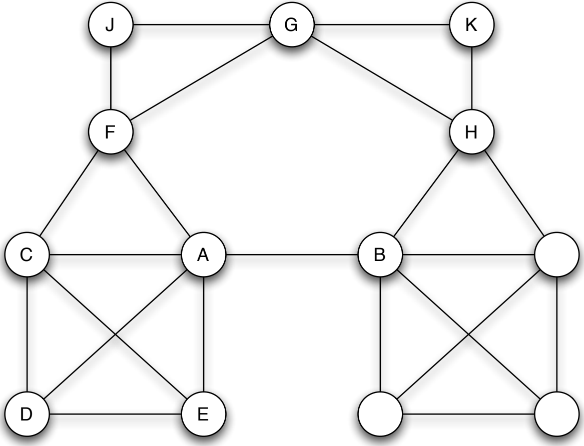



**Figure :** The edge $(A, B)$ is a local bridge of span 4, since the removal of this edge would increase the distance between $A$ and $B$ to 4. [E&K Figure 3.4]

# Strong triadic closure property: connecting tie strength and local bridges

**Strong triadic closure property**

Whenever $(A, B)$ and $(A, C)$ are strong ties, then there will be a tie (possibly only a weak tie) between $B$ and $C$.

- Such a strong property is not likely true in a large social network (that is, holding for every node A)
- However, it is an abstraction that may lend insight.

**Theorem**

*Assuming the strong triadic closure property, for a node involved in at least two strong ties, any local bridge it is part of must be a weak tie.*

Informally, local bridges must be weak ties since otherwise strong triadic closure would produce shorter paths between the end points.

# Strong triadic closure property continued

- Again we emphasize (as the text states) that "Clearly the strong triadic closure property is too extreme to expect to hold across all nodes ... But it is a useful step as an abstraction to reality, ..."
- Sintos and Tsaparas give evidence that assuming the strong triadic closure property can help in determining whether a link is a strong or weak tie. (www.cs.uoi.gr/ tsap/publications/frp0625-sintos.pdf)

# Strong triadic closure property continued

- Again we emphasize (as the text states) that "Clearly the strong triadic closure property is too extreme to expect to hold across all nodes ... But it is a useful step as an abstraction to reality, ..."
- Sintos and Tsaparas give evidence that assuming the strong triadic closure property can help in determining whether a link is a strong or weak tie. (www.cs.uoi.gr/ tsap/publications/frp0625-sintos.pdf)
- More specifically, for a social network where the edges are not labelled they define the following two computational problems: Label the graph edges (by strong and weak) so as to satisfy the strong triadic closure property and either
  1. maximize the number of strong edges or
  2. minimize the number of weak edges
- For computational reasons, it is not usually possible to optimize and it is best to approximately minimize the number of weak edges.
- Their computational results (labeling the edges) are validated with 5 network data sets for which the strength of ties can be determined.

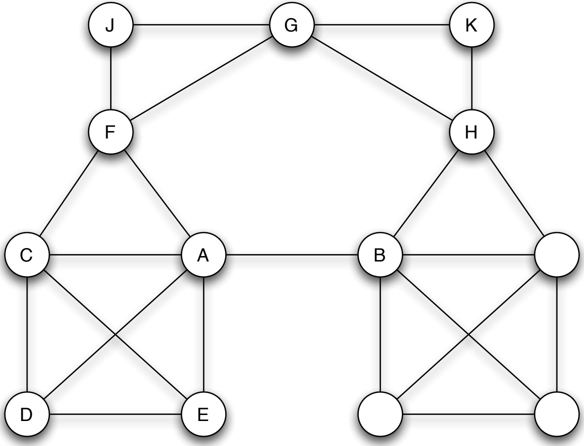
# Large scale experiment supporting strength of weak ties and triadic closure

- Onnela et al. [2007] study of who-talks-to-whom network maintained by a cell phone provider. Large network of cell users where an edge exists if there existed calls in both directions in 18 weeks.
- First observation: a giant component with 84% of nodes.
- Need to quantify the tie strength and the closeness to being a local bridge.
- Tie strength is measured in terms of the total number of minutes spent on phone calls between the two ends of an edge.
- Closeness to being a local bridge is measured by the neighborhood overlap of an edge $(A, B)$ defined as the ratio

$$\frac{\text{number of nodes adjacent to both } A \text{ and } B}{\text{number of nodes } C \neq A, B \text{ adjacent to at least one of } A \text{ or } B}$$

- Local bridges are precisely edges having overlap 0.

# Example: Embeddedness and neighborhood overlap




- The edge $(A, B)$ has embeddedness 0 and hence is a local bridge of span 4, since the removal of this edge would increase the distance between $A$ and $B$ to 4.
- The edge $(B, H)$ has embeddedness 1 and neighborhood overlap $\frac{1}{6}$.

# Onnela et al. study continued

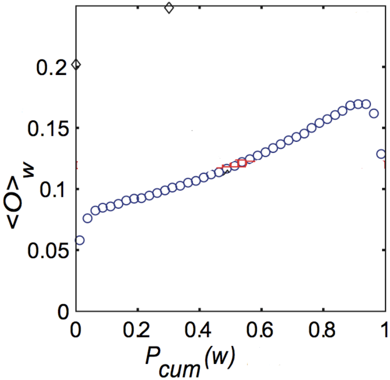



**Figure :** A plot of the neighborhood overlap of edges as a function of their percentile in the sorted order of all edges by tie strength. [E&K Fig 3.7]

- The figure shows the relation between tie strength and overlap.
- Quantitative evidence supporting the theorem: as tie strength decreases, the overlap decreases ⇒ weak ties becoming "almost local bridges".

# Onnela et al. study continued

To support the hypothesis that weak ties tend to link together more tightly knit communities, Onnela et al. perform two simulations:

1. Removing edges in decreasing order of tie strength, the giant component shrank gradually.
2. Removing edges in increasing order of tie strength, the giant component shrank more rapidly and at some point then started fragmenting into several components.

# Word of caution in text regarding such studies

Easley and Kleinberg (end of Section 3.3):

> *Given the size and complexity of the (who calls whom) network, we cannot simply look at the structure... Indirect measures must generally be used and, because one knows relatively little about the meaning or significance of any particular node or edge, it remains an ongoing research challenge to draw richer and more detailed conclusions...*

# Word of caution in text regarding such studies

Easley and Kleinberg (end of Section 3.3):

> *Given the size and complexity of the (who calls whom) network, we cannot simply look at the structure... Indirect measures must generally be used and, because one knows relatively little about the meaning or significance of any particular node or edge, it remains an ongoing research challenge to draw richer and more detailed conclusions...*

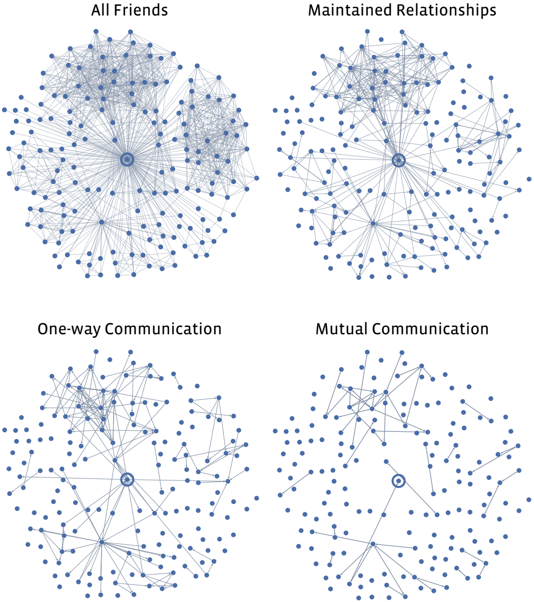
Yogi Berra(1925-2015):

> *In theory there is no difference between theory and practice. In practice there is.*

# Strong vs. weak ties in large online social networks (Facebook and Twitter)

- The meaning of "friend" as in Facebook is not the same as one might have traditionally interpreted the word "friend".
- Online social networks give us the ability to qualify the strength of ties in a useful way.
- For an observation period of one month, Marlow et al. (2009) consider 4 Facebook networks defined by (in increasing order of strength): all friends, maintained (passive) relations of following a user, one-way communication, and reciprocal communication.
  1. These networks thin out when links represent stronger ties.
  2. As the number of total friends increases, the number of reciprocal communication links levels out at slightly more than 10.
  3. How many Facebook friends did you have for which you had a reciprocal communication in the last month?

# Different Types of Facebook Friendships



# A limit to the number of strong ties

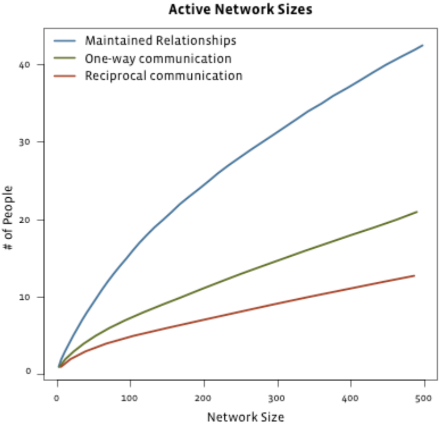



**Figure :** The number of links corresponding to maintained relationships, one-way communication, and reciprocal communication as a function of the total neighborhood size for users on Facebook. [Figure 3.9, textbook]

# Twitter:Limited Strong Ties vs Followers

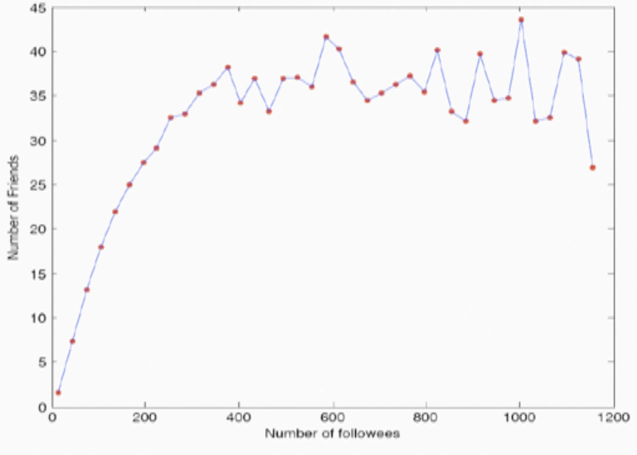



**Figure :** The total number of a user's strong ties (defined by multiple directed messages) as a function of the number of followees he or she has on Twitter. [Figure 3.10, textbook]