Social and Information Networks

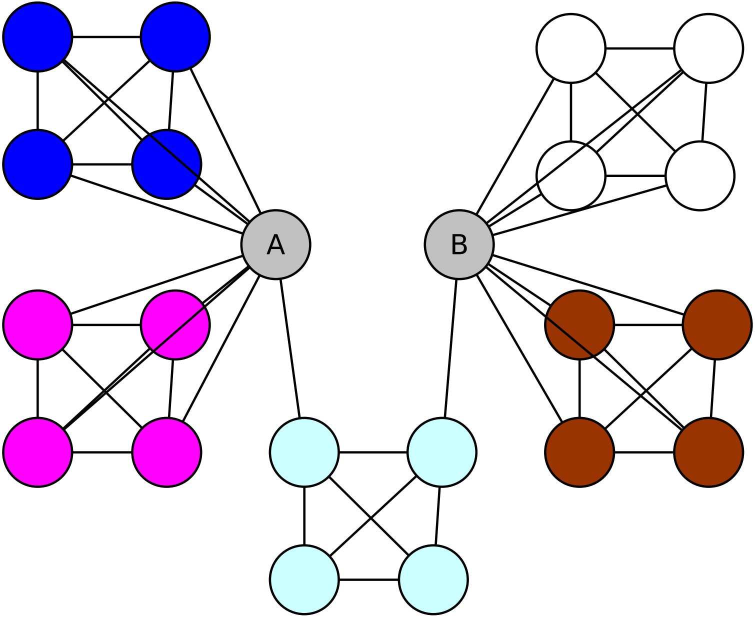
University of Toronto CSC303
Winter/Spring 2021

Week 2: January 18-22 (2021)

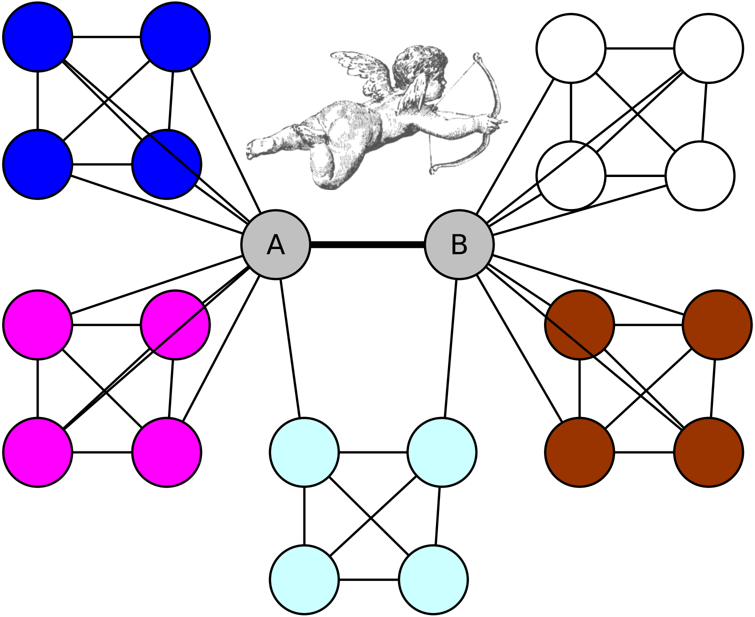
Mon. Jan 18th: Announcements & Corrections

- Anonymized Zoom Chat logs of lectures are now available in Quercus, and will be posted after lectures
- Zoom links for tutorials have been posted. Class division by surname is as follows:
  - A-G: Section 1
  - H-Q: Section 2
  - R-Z: Section 3
- For this week **only**, there is no section 1. So A-C please attend Section 2, and D-G please attend Section 3.
- A quick recap of why dispersion works for identifying romantic relationships
- A review of the survey, and announcements on delivery method & office hours

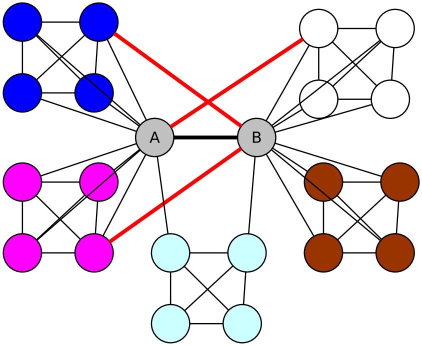
Dispersion Example



Dispersion Example

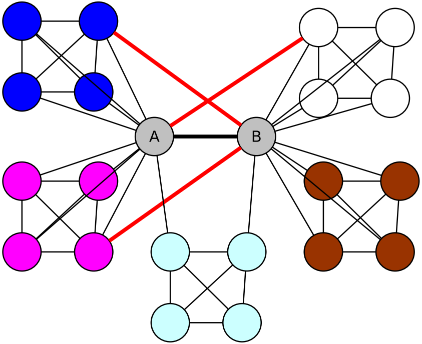


Dispersion Example



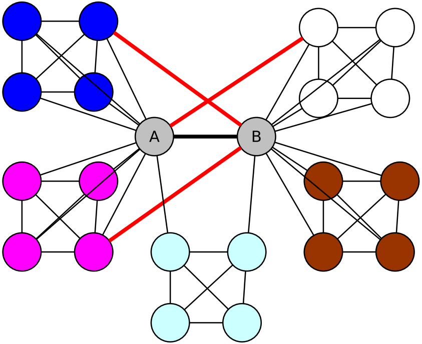
- Embededness of (A, B) is their number of mutual friends
- Easily skewed by dense groups (e.g. white nodes are chess club where everyone is eachothers' friend)

Dispersion Example



- Dispersion is how poorly connected A and B's mutual friends are if A and B are removed from the graph
- Romantic relationship can create pairs of mutual friends in different focii (e.g. A's friend in chess club & B's dark blue friend)

Dispersion Example



- Dispersion is how poorly connected A and B's mutual friends are if A and B are removed from the graph
- Normal friendships are less likely to form these connections (e.g. unlikely to be connections from white nodes to brown nodes if brown nodes are extended family)

Survey Results: Delivery Method

From the following options, what is your preferred delivery method?

Synchronous lecture with recording posted afterwards	55 respondents	**56 %**
Asynchronous pre-recorded full-length lecture that is replayed and discussed during class	10 respondents	10 %
Asynchronous pre-recorded mini-lectures (1 per topic) that is replayed and discussed during class	33 respondents	34 %

Do you **object** to any of the following delivery methods?

Synchronous lecture with recording posted afterwards	10 respondents	**31 %**
Asynchronous pre-recorded full-length lecture that is replayed and discussed during class	22 respondents	69 %
Asynchronous pre-recorded mini-lectures (1 per topic) that is replayed and discussed during class	11 respondents	34 %
No Answer	66 respondents	206 %

- As I also have anecdotal evidence from other instructors that synchronous lectures are better didactically, I've decided that things will continue as they have been in the first week

Survey Results: Office Hours

Monday 16:00-17:00	66 respondents	69 %
Wednesday 16:00-17:00	60 respondents	63 %
Friday 16:00-17:00	67 respondents	70 %
Monday 10:00-11:00	38 respondents	40 %
Wednesday 10:00-11:00	39 respondents	41 %
Friday 10:00-11:00	43 respondents	45 %
Monday 22:00-23:00	56 respondents	58 %
Wednesday 22:00-23:00	59 respondents	61 %
No Answer	2 respondents	2 %

- Weekly office hours will be Fridays, 4-5PM, immediately after class
- I will block off Fridays 10-11AM and Wednesday 10-11PM, **but I won't hold office hours unless I'm emailed ahead of time**
- New link on Quercus, office hours will not be recorded

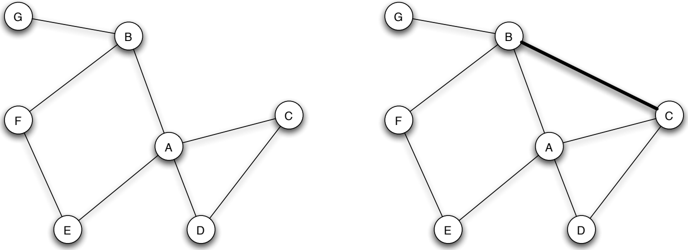
Survey Results: Penguins

- We had ∞% more adorable penguins than expected

Survey Results: Questions/Concerns/Suggestions

- Can we switch to Piazza?
  - Sorry but no, Discourse isn't great, but it's run by the university and it works. I inquired last term about various addons to increase functionality; unfortunately they couldn't get them working for this term but I'm hopeful for the future
- Can we work on the critical review individually?
  - Sorry, but no. In part it's a question of grading resources, in part I want some group work in the course to force you to speak with your peers – knowing someone in the course is valuable for studying and for bouncing ideas off of
- Could slides be released ahead of time?
  - I'll try to do so moving forwards. They might not match up exactly (I sometimes tweak slides, and there won't be announcements as my TARDIS is unreliable) but they core slides will be there
- Can you try to release the recordings soon after?
  - I'll try to getting it out within 3 hours of lecture, but I can't make guarantees and it'll probably be longer for tutorial recordings

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. [E&K Figure 3.1]

- 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?
- Why is a new friendship tie formed? (Friendship ties can range from just knowing someone to a true 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).

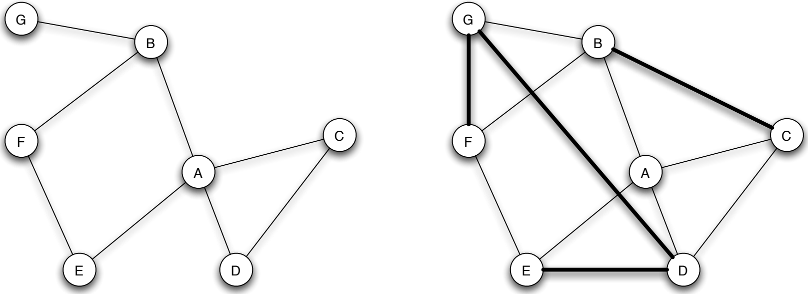
- Let E be the set of an undirected edges of a network graph. (Forgive the abuse of notation where in the previous and next slide E is a node name.) For a node A, the clustering coefficient is the following ratio:

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

- The numerator is the number of all edges (B, C) in the network such that B and C are adjacent to (i.e. mutual friends of) A.

- The denominator is the total 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 $\{D, E\}$). We sometimes refer to a pair of adjacent edges like $(A, B), (A, C)$ as an "open triangle" if (B, C) does not exist.
- 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)).

Driving forces behind Triadic Closure

- Social psychology suggests: Increased opportunity, incentive, and trust

Driving forces behind Triadic Closure

- Social psychology suggests: Increased opportunity, incentive, and trust

- It also predicts that having friends (especially good friends with strong ties) who are not themselves friends causes *latent stress*

Granovetter's thesis: the strength of weak ties

- In 1960s interviews: Many people learn about new jobs from personal contacts (which is not surprising) and often these contacts were acquaintances rather than friends. Is this surprising?

Granovetter's thesis: the strength of weak ties

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

Granovetter's thesis: the strength of weak ties

- In 1960s interviews: Many people learn about new jobs from personal contacts (which is not surprising) and often these contacts were acquaintances rather than friends. Is this 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.

Granovetter's thesis: the strength of weak ties

- In 1960s interviews: Many people learn about new jobs from personal contacts (which is not surprising) and often these contacts were acquaintances rather than friends. Is this 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 component in a graph.
- We expect that large social networks will have a "giant component" and few bridges.

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 component in a 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 introduced in the Backstrom and Kleinberg article regarding Facebook relations.
- A local bridges (A,B) plays 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)

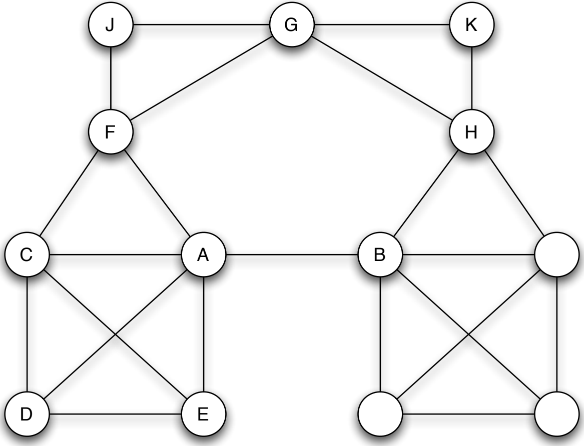


Figure: The edge (A, B) is a local bridge of span 4, since the removal of this edge would increase the distance between A and B to 4. [E&K Figure 3.4]

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.

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.

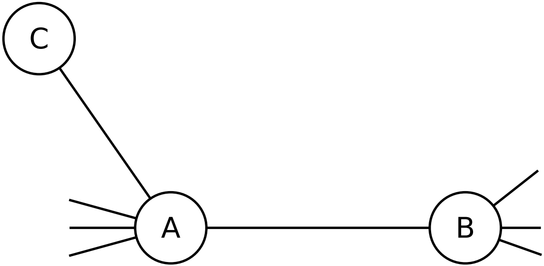
Theorem

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.

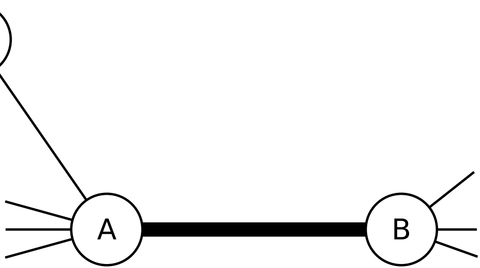
Triadic closure and local bridges

- Let A by any node involved in at least two strong edges and a local bridge. Let (A, B) be a local bridge.



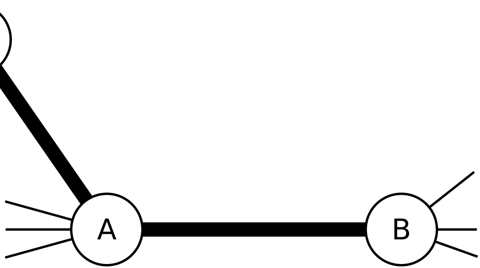
Triadic closure and local bridges

- Let's assume for contradiction that (A, B) is strong



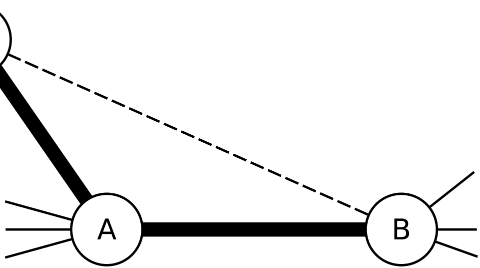
Triadic closure and local bridges

- Let's assume for contradiction that (A, B) is strong



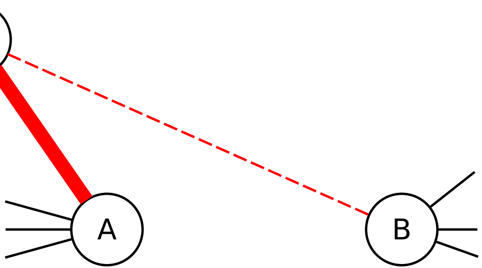
Triadic closure and local bridges

- Let's assume for contradiction that (A, B) is strong



Triadic closure and local bridges

- Let's assume for contradiction that (A, B) is strong



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 (STC) property can help in determining whether a link is a strong or weak tie.

  `www.cs.uoi.gr/~tsap/publications/frp0625-sintos.pdf`

  We will discuss this paper later in the lecture.
- Later we'll discuss Rozenshtein et al [2019]. They assume the existence of *known communities*, and then their goal is to label all edges so as minimize the number of *open triangles violating the STC property* subject to all communities being connected using only strong edges.
  - This work is inspired by the Sintos and Tsaparas [2014] results for inferring the strength of ties, and an earlier [2013] paper by Angluin et al for minimizing the number of edges needed to maintain "communities"

Embeddedness of an edge

Just as there are many specific ways to define the dispersion of an edge, there are different ways to define the embeddedness of an edge.

The general idea is that embeddedness of an edge (u, v) should capture how much the social circles of u and v "overlap". The next slide will use a particular definition for embeddedness.

Why might dispersion be a better discriminator of a romantiic relationship (especially for marriage) than embeddedness?

Large scale experiment relating tie strength and bridges

- 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 end of an edge.
- Closeness to being a local bridge is measured by the neighborhood overlap of an edge (A, B) defined as the ratio

$$\frac{\text{number of nodes adjacent to both } A \text{ and } B}{\text{number of nodes adjacent to at least one of } A \text{ or } B \text{ (excluding } A \text{ \& } B\text{)}}$$

- Question: What does a neighbourhood overlap of zero mean?

Large scale experiment relating tie strength and bridges

- 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 end of an edge.
- Closeness to being a local bridge is measured by the neighborhood overlap of an edge (A, B) defined as the ratio

$$\frac{\text{number of nodes adjacent to both } A \text{ and } B}{\text{number of nodes adjacent to at least one of } A \text{ or } B \text{ (excluding } A \text{ \& } B\text{)}}$$

- Question: What does a neighbourhood overlap of zero mean? Local bridge!

Large scale experiment relating tie strength and bridges

- 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 end of an edge.
- Closeness to being a local bridge is measured by the neighborhood overlap of an edge (A, B) defined as the ratio

$$\frac{\text{number of nodes adjacent to both } A \text{ and } B}{\text{number of nodes adjacent to at least one of } A \text{ or } B \text{ (excluding } A \text{ \& } B\text{)}}$$

- Question: What does a neighbourhood overlap of zero mean? Local bridge!
- Question: What relationship would we expect between tie-strength & neighbourhood overlap?

Onnela et al. experiment

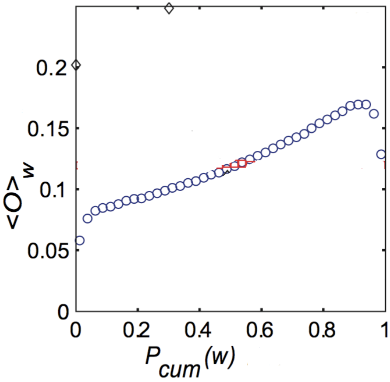


Figure: A plot of the neighborhood overlap of edges as a function of their percentile in the sorted order of all edges by tie strength. [E&K Fig 3.7]

- The figure shows the relation between tie strength and overlap.
- Quantitative evidence supporting the theorem: as tie strength decreases, the overlap decreases; that is, weak ties are becoming "almost local bridges" having ovelap almost equal to 0.

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. . .*

Fri. Jan 22th: Announcements & Corrections

- I hope you enjoyed the first tutorial!
  - Appears that attendance could be better
- Don't forget the first participation quiz is due Tuesday at midnight
- Apologies that yesterday's Tutorial slides were not up in advance, I'll aim to have a draft on the course website moving forwards
- I'm aiming to have lecture recordings on the course website within 3 hours, and tutorial recordings up by midnight the day of
- First office hours today after class, link on Quercus

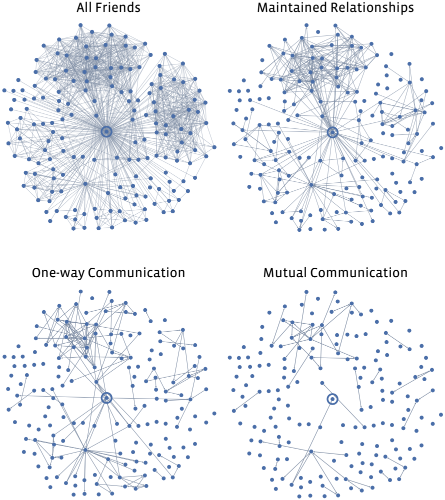
Fri. Jan 22th: Announcements & Corrections

[image from nweismuller]

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 Facebook networks defined by 4 criteria (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 Friendships: The neighbourhood network of a sample Facebook individual



A limit to the number of strong ties

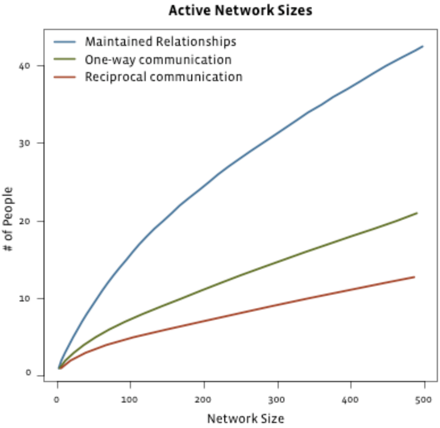


Figure: The number of links corresponding to maintained relationships, one-way communication, and reciprocal communication as a function of the total neighborhood size for users on Facebook. [Figure 3.9, textbook]

Twitter:Limited Strong Ties vs Followers

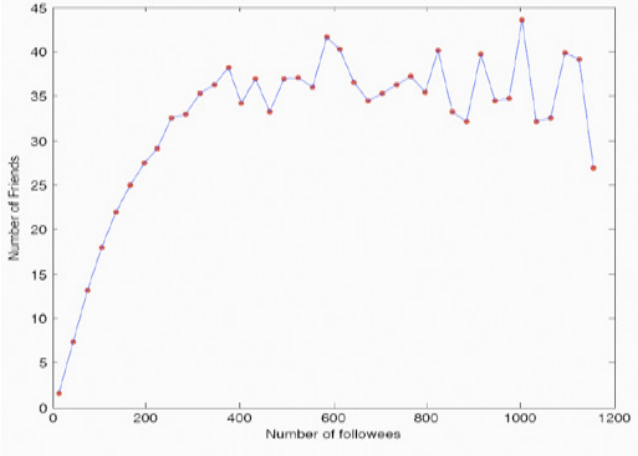


Figure: The total number of a user's strong ties (defined by multiple directed messages) as a function of the number of followees he or she has on Twitter. [Figure 3.10, textbook]

Information spread in a passive network

- The maintained or passive relation network (as in the Facebook network on slide 24) is said to occupy a middle ground between
  1. strong tie network (in which individuals actively communicate), and
  2. very weak tie networks (all "friends") with many old (and inactive) relations.
- "Moving to an environment where everyone is passively engaged with each other, some event, such as a new baby or engagement can propagate very quickly through this highly connect neighborhood."
- We can add that an event might be a political demonstration.

Social capital (as discussed in section 3.5 of EK text)

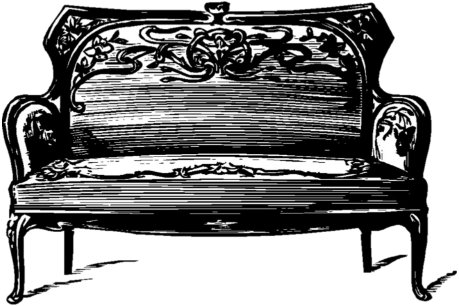
Social capital is a term in increasingly widespread use, but it is a famously difficult one to define.

The term social capital is designed to suggest its role as part of an array of different forms of capital (e.g. economic, cultural, physical etc...) all of which serve as tangible or intangible resources that can be mobilized to accomplish tasks.

Social capital (as discussed in section 3.5 of EK text)

Social capital is a term in increasingly widespread use, but it is a famously difficult one to define.

The term social capital is designed to suggest its role as part of an array of different forms of capital (e.g. economic, cultural, physical etc...) all of which serve as tangible or intangible resources that can be mobilized to accomplish tasks.



Social capital (as discussed in section 3.5 of EK text)

A source of terminological variation is based on whether social capital is a property that is purely intrinsic to a group based only on the social interactions among the groups members or whether it is also based on the interactions of the group with the outside world.

A person can have more or less social capital depending on his or her position in the underlying social structure or network.

“Tightly knit communities” connected by weak ties

- The intuitive concept of tightly knit communities occurs several times in Chapter 3 but is deliberately left undefined.

- In a small network we can sometimes visualize the tightly knit communities but one cannot expect to do this is a large network. That is, we need algorithms and this is the topic of the advanced material in Section 3.6.

“Tightly knit communities” connected by weak ties

- The intuitive concept of tightly knit communities occurs several times in Chapter 3 but is deliberately left undefined.
- In a small network we can sometimes visualize the tightly knit communities but one cannot expect to do this is a large network. That is, we need algorithms and this is the topic of the advanced material in Section 3.6.
- Recalling the relation to weak ties, the text calls attention to how nodes at the end of one (or especially more) local bridges can play a pivotal role in a social network.
- These “gatekeeper nodes” between communities stand in contrast to nodes which sit at the center of a tightly knit community.

Central nodes vs. gatekeepers

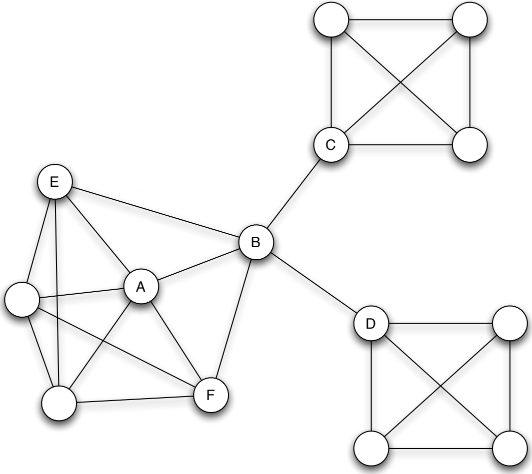


Figure: The contrast between densely-knit groups and boundary-spanning links is reflected in the different positions of central node *A* and gatekeeper node *B* in the underlying social network. [Fig 3.11, textbook]

Social capital of nodes A and B

- The edges adjacent to node A all have high embeddedness. Visually one sees node A as a central node in a tightly-knit cluster. As such, the social capital that A enjoys is its "bonding capital" in that the actions of A can (for example) induce norms of behaviour because of the trust in A.
- In contrast, node B is a bridge to other parts of the network. As such, its social capital is in the form of "brokerage" or "bridging capital" as B can play the role of a "gatekeeper" (of information and ideas) between different parts of the network. Furthermore, being such a gatekeeper can lead to creativity stemming from the synthesis of ideas.
- Some nodes can have both bonding capital and bridging capital.

Florentine marriages: Bridging capital of the Medici

- The Medici are connected to more families, but not by much.
- More importantly: Four of the six edges adjacent to the Medici are bridges or local bridges and the Medici lie on the shortest paths between most pairs of families.

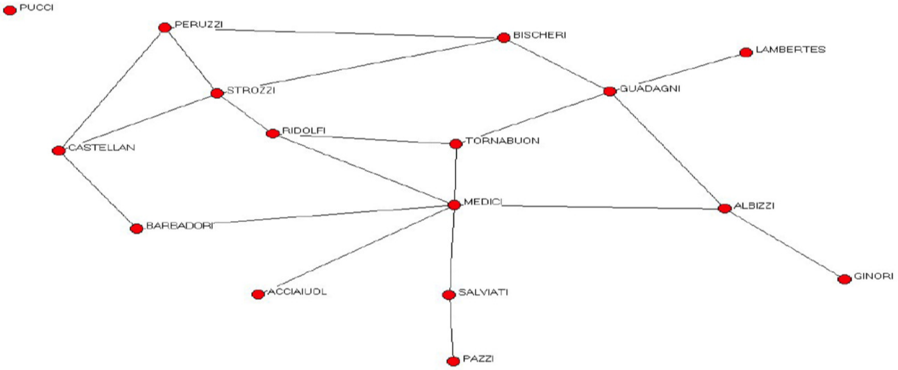
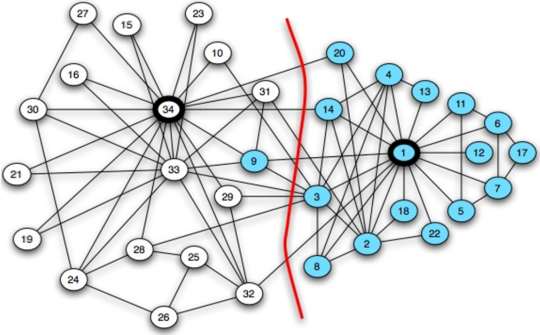


Figure: see [Jackson, Ch 1]

A Balanced Min Cut in Graph: Bonding capital of nodes 1 and 34



- Note that node 34 also seems to have bridging capital.
- Wayne Zachary's Ph.D. work (1970-72): observed social ties and rivalries in a university karate club.
- During his observation, conflicts intensified and group split.
- Could the club boundaries be predicted from the network structure?
- Split could almost be explained by minimum cut in social network.

The Sintos and Tsaparas Study

In their study of the strong triadic closure (STC) property, Sintos and Tsaparas study 5 small networks. They give evidence as to how the STC assumption can help determine weak vs strong ties, and how weak ties act as bridges to different communities.

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

1. Either maximize the number of strong edges, or equivalently
2. minimize the number of weak edges

The computational problem in identifying strong vs weak ties

- For computational reasons (i.e., assuming $P \neq NP$ and showing NP hardness by reducing the max clique problem to the above maximization problem), it is not possible to efficiently optimize and hence they settle for approximations.
- Note that even for the small Karate Club network having only $m = 78$ edges, a brute force search would require trying 2^{78} solutions. Of course, there may be better methods for any specific network.
- The reduction preserves the approximation ratio, so it is also NP-hard to approximate the maximization problem with a factor of $n^{1-\epsilon}$. However, the minimization problem can be reduced (preserving approximations) to the vertex cover problem which can be approximated within a factor of 2.
- Their computational results are validated against the 5 networks where the strength of ties is known from the given data. Notably their worst case approximation algorithm (via the reduction) lead to reasonably good results achieved for the 5 real data networks.

The vertex cover algorithms and the 5 data sets

While there are uncovered edges, the (vertex) greedy algorithm selects a vertex for the vertex cover with maximum current degree. It has worst case $O(\log n)$ approximation ratio. The maximal matching algorithm is a 2-approximation online algorithm that finds an uncovered edge and takes both endpoints of that edge.

Table 1: Datasets Statistics.

Dataset	Nodes	Edges	Weights	Community structure
Actors	1,986	103,121	Yes	No
Authors	3,418	9,908	Yes	No
Les Miserables	77	254	Yes	No
Karate Club	34	78	No	Yes
Amazon Books	105	441	No	Yes

Figure: Weights (respectively, community structure) indicates when explicit edge weights (resp. a community structure) are known.

Tie strength results in detecting strong and weak ties

Table 2: Number of strong and weak edges for Greedy and MaximalMatching algorithms.

	Greedy		MaximalMatching	
	Strong	Weak	Strong	Weak
Actors	11,184	91,937	8,581	94,540
Authors	3,608	6,300	2,676	7,232
Les Miserables	128	126	106	148
Karate Club	25	53	14	64
Amazon Books	114	327	71	370

Figure: The number of labeled links.

Although the Greedy algorithm has an inferior (worst case) approximation ratio, here the greedy algorithm has better performance than Maximal Matching. (Recall, the goal is to maximize the number of strong ties, or equivalently minimize the number of weak ties.)

Results for detecting strong and weak ties

Table 3: Mean count weight for strong and weak edges for Greedy and MaximalMatching algorithms.

	Greedy		MaximalMatching	
	S	W	S	W
Actors	1.4	1.1	1.3	1.1
Authors	1.341	1.150	1.362	1.167
Les Miserables	3.83	2.61	3.87	2.76

Figure: The average link weight.

Question: Is there a problem with average edge strength?

Results for detecting strong and weak ties

Table 3: Mean count weight for strong and weak edges for Greedy and MaximalMatching algorithms.

	Greedy		MaximalMatching	
	S	W	S	W
Actors	1.4	1.1	1.3	1.1
Authors	1.341	1.150	1.362	1.167
Les Miserables	3.83	2.61	3.87	2.76

Figure: The average link weight.

Question: Is there a problem with average edge strength? Easy to skew average if weights have high variance

Tie strength results in detecting strong and weak ties normalized by amount of activity

Table 4: Mean Jaccard similarity for strong and weak edges for Greedy and MaximalMatching algorithms.

	Greedy		MaximalMatching	
	S	W	S	W
Actors	0.06	0.04	0.06	0.04
Authors	0.145	0.084	0.155	0.088

Figure: Using a normalized edge weight based on activity

$$w((a,b)) = \frac{\text{works}(a) \cap \text{works}(b)}{\text{works}(a) \cup \text{works}(b)} \in [0,1]$$

Results for strong and weak ties with respect to known communities

Table 5: Precision and Recall for strong and weak edges for Greedy and MaximalMatching algorithms.

Greedy				
	P_S	R_S	P_W	R_W
Karate Club	1	0.37	0.19	1
Amazon Books	0.81	0.25	0.15	0.69
MaximalMatching				
	P_S	R_S	P_W	R_W
Karate Club	1	0.2	0.16	1
Amazon Books	0.73	0.14	0.14	0.73

Figure: Precision and recall with respect to the known communities.

The meaning of the precision-recall table

The precision and recall for the weak edges are defined as follows:
$P_W = \frac{|W \cap E_{inter}|}{|W|}$ and $R_W = \frac{|W \cap E_{inter}|}{|E_{inter}|}$
$P_S = \frac{|S \cap E_{intra}|}{|S|}$ and $R_S = \frac{|S \cap E_{intra}|}{|E_{intra}|}$

- Ideally, we want $R_W = 1$ indicating that all edges between communities are weak; and we want $P_S = 1$ indicating that strong edges are all within a community.
- For the Karate Club data set, all the strong links are within one of the two known communities and hence all links between the communities are all weak links.
- For the Amazon Books data set, edges are co-purchases, and there are three communities corresponding to liberal, neutral, conservative viewpoints. Of the strong edges predicted, only 22 cross communities:
  - 20 cross-community strong edges have one node labeled as neutral.
  - the rest are between books dealing with the same issue.

Strong and weak ties in the karate club network

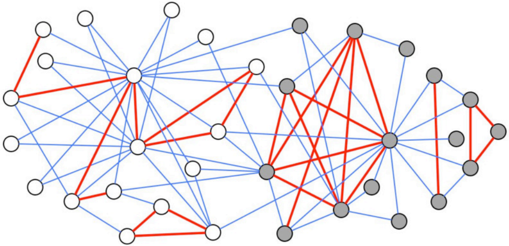


Figure 1: Karate Club graph. Blue light edges represent the weak edges, while red thick edges represent the strong edges.

- Note that all the strong links are within one of the two known communities and hence all links between the communities are weak links.