Social Network Analysis

Introduction & Preliminaries

Course Outline

- **Graph Theory and Social Networks**
- Visualizing Social Networks
- Network Dynamics
- Information Networks and the World Wide Web
- Game Theory
- Applications of SNA in various domains

Strong & Weak Ties

SNA – what it does

- Networks bridge the local to the global — to offer explanations for how simple processes at the level of individual nodes and links can have complex effects that ripple through a population as a whole.

Key points

- How information flows through a social network..
- How different nodes can play structurally distinct roles in this process..
- How these structural considerations shape the evolution of the network itself over time...

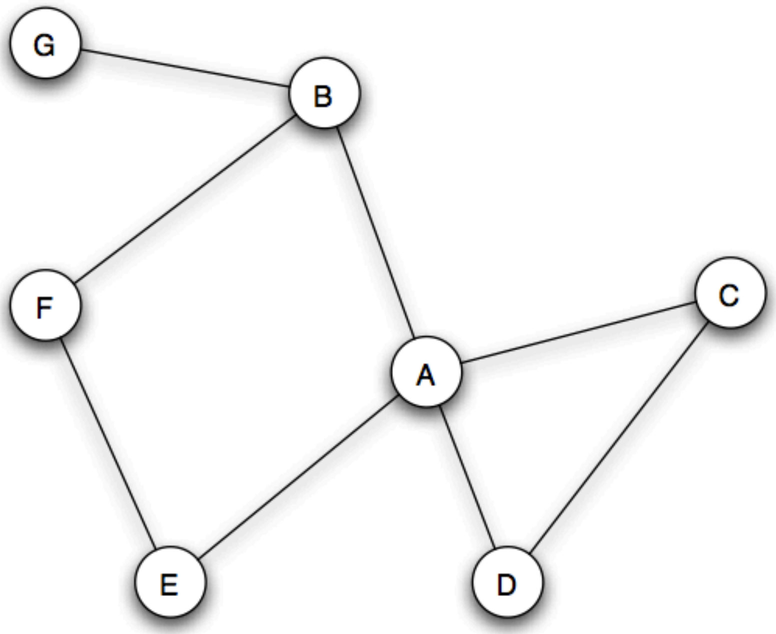
Granovetter's Problem

- How did people find new jobs?
- Through personal contacts..
- acquaintances, not close friends..
- So, how does the graph come into the picture?

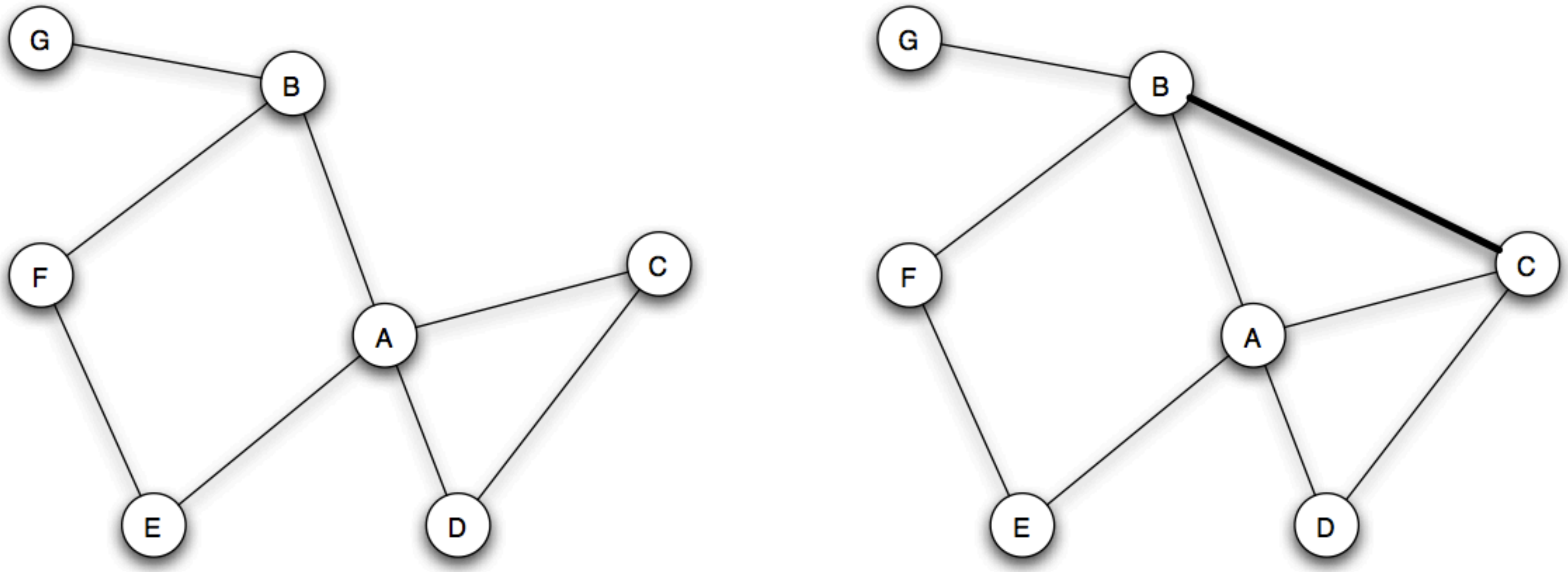
Granovetter's Problem

- Links two different perspectives on distant relationships.
- One interpersonal, using purely local considerations of friendship between two people being strong (friend) and weak (acquaintance).
- The other structural, how friendships span different parts of the entire network.

How are friendships formed?



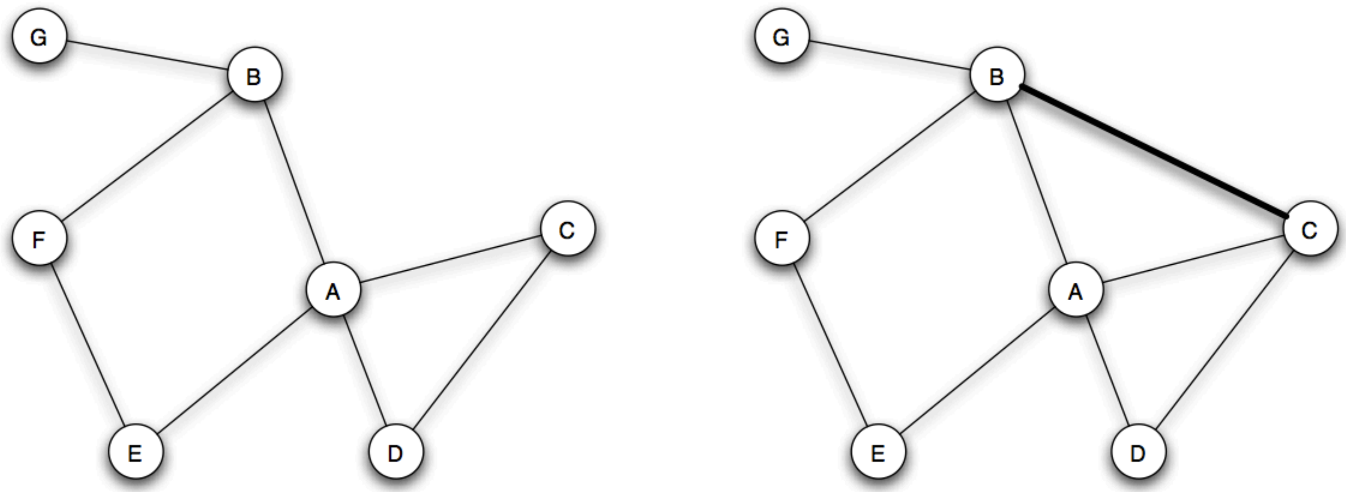
How are friendships formed?



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

What may be going on here?

Triadic Closure



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

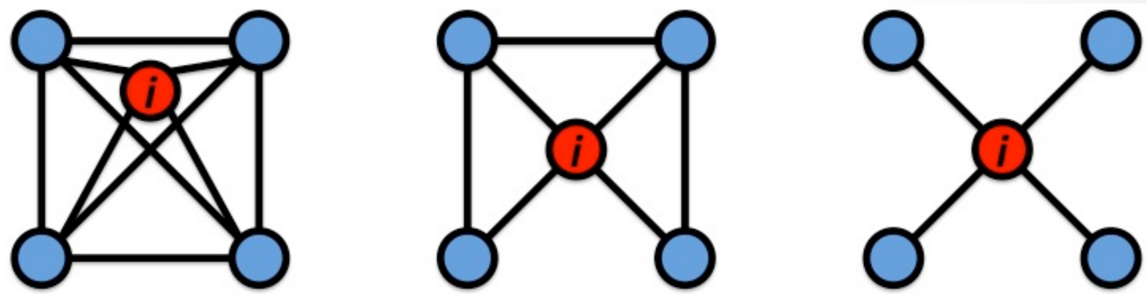
- If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at

Does this remind you of anything you've seen recently?

Clustering Coefficient

- **What fraction of your neighbours are connected?**
  - Node i with degree k_i
  - Clustering Coefficient C_i for a vertex i is in [0,1]

$$\boxed{C_i = \frac{2e_i}{k_i(k_i - 1)}}$$



$C_i = 1$ $\qquad$ $C_i = 1/2$ $\qquad$ $C_i = 0$

Clustering coefficient is a "local" property – each vertex has one.

Watts & Strogatz, Nature 1998.

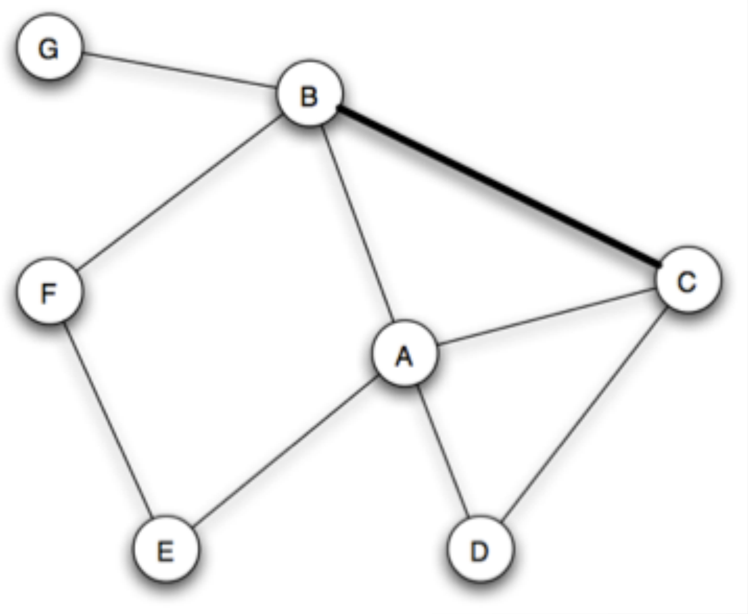
Clustering Coefficient

- ..is the fraction of pairs of A’s friends that are connected to each other by edges.
- ..of a node A can also be defined as the probability that two randomly selected friends of A are friends with each other.
- The more the triadic closure (dynamic) is operating in the neighbourhood of a node, the higher the clustering coefficient (static) will be.

Reasons for Triadic Closure

- When they have a common friend A, there is an increased opportunity for B and C to meet.

- Since each of B and C is friends with A (provided they are mutually aware of this) gives them a basis for trusting each other that an arbitrary pair of unconnected people might lack.

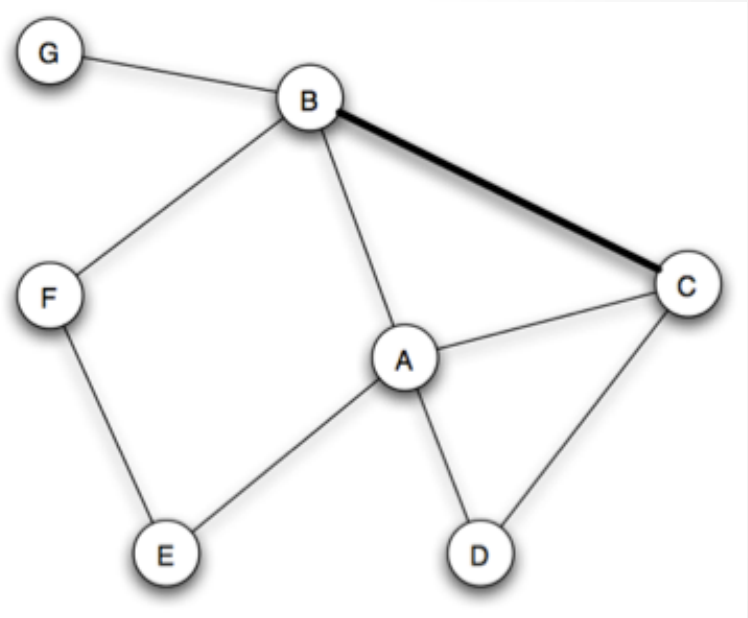


- A third reason is based on the incentive A may have to bring B and C together: if A is friends with B and C, then it becomes a source of latent stress in these relationships if B and C are not friends with each other.

The story so far...

- Acquaintances rather than friends were helpful in finding jobs.

- How friendships may be formed, and the idea of triadic closure ("formation of triangles").



- The reasons for triadic closure.

- What are we trying to do?

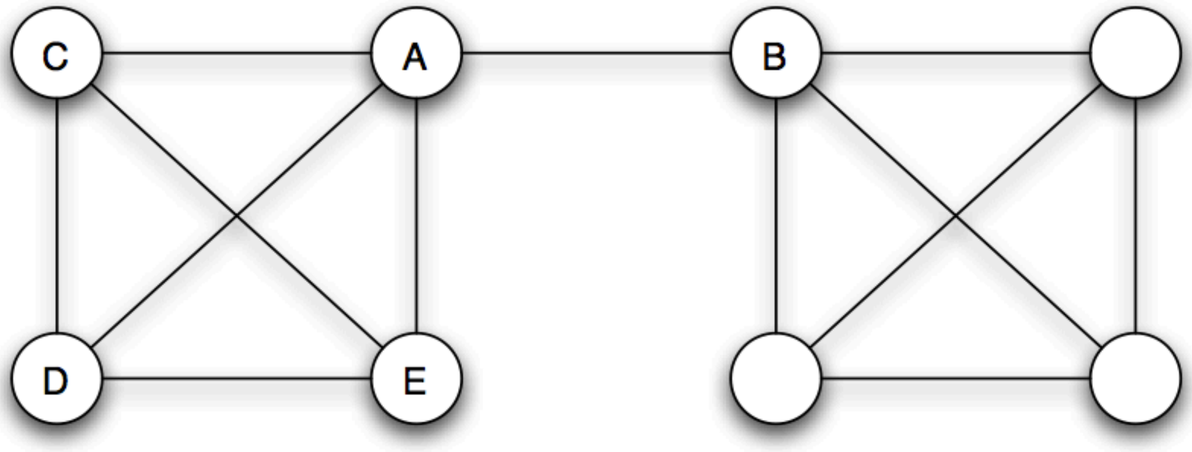
  Differentiate between friends and acquaintances – strong and weak ties.

Why acquaintances…

- The close friends we have, we talk to them frequently, and share information more frequently with them.
- So, it is less likely that they have information we don't.
- The reasons for triadic closure.
- What are we trying to do?

  Differentiate between friends and acquaintances – strong and weak ties.

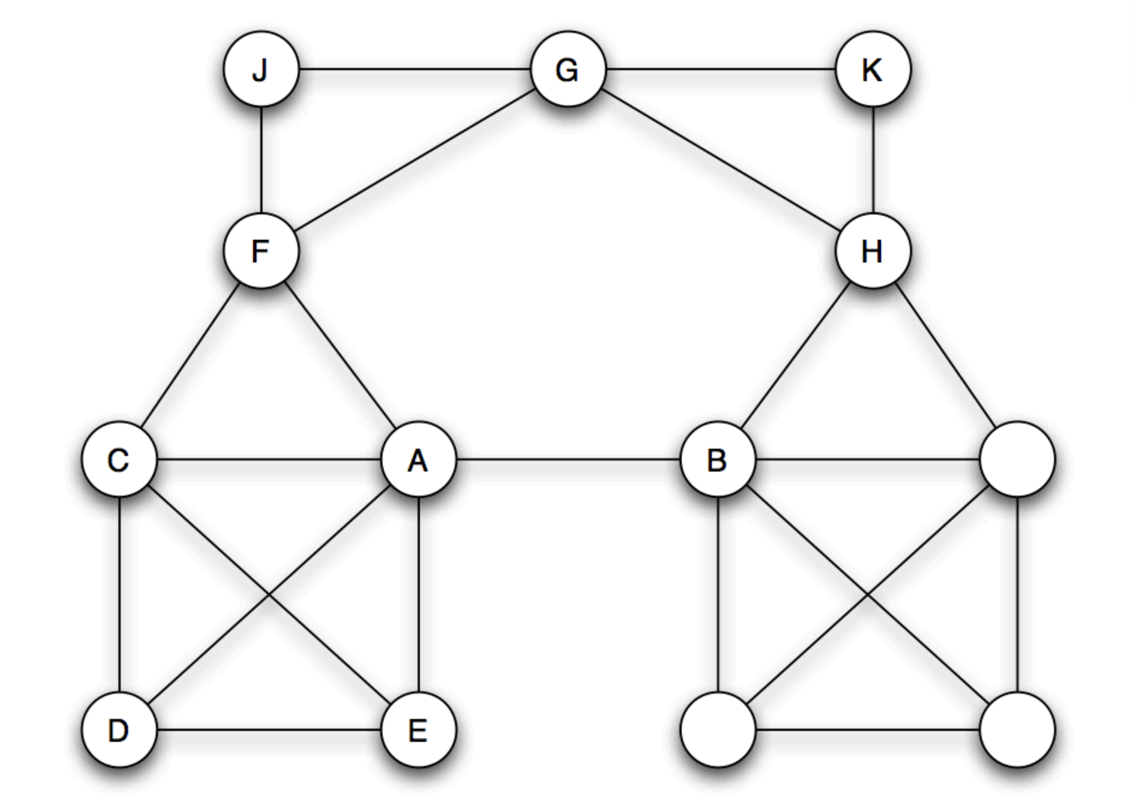
Bridges



A-B is a bridge; it's removal disconnects the graph into 2 components, with A and B lying in different components.

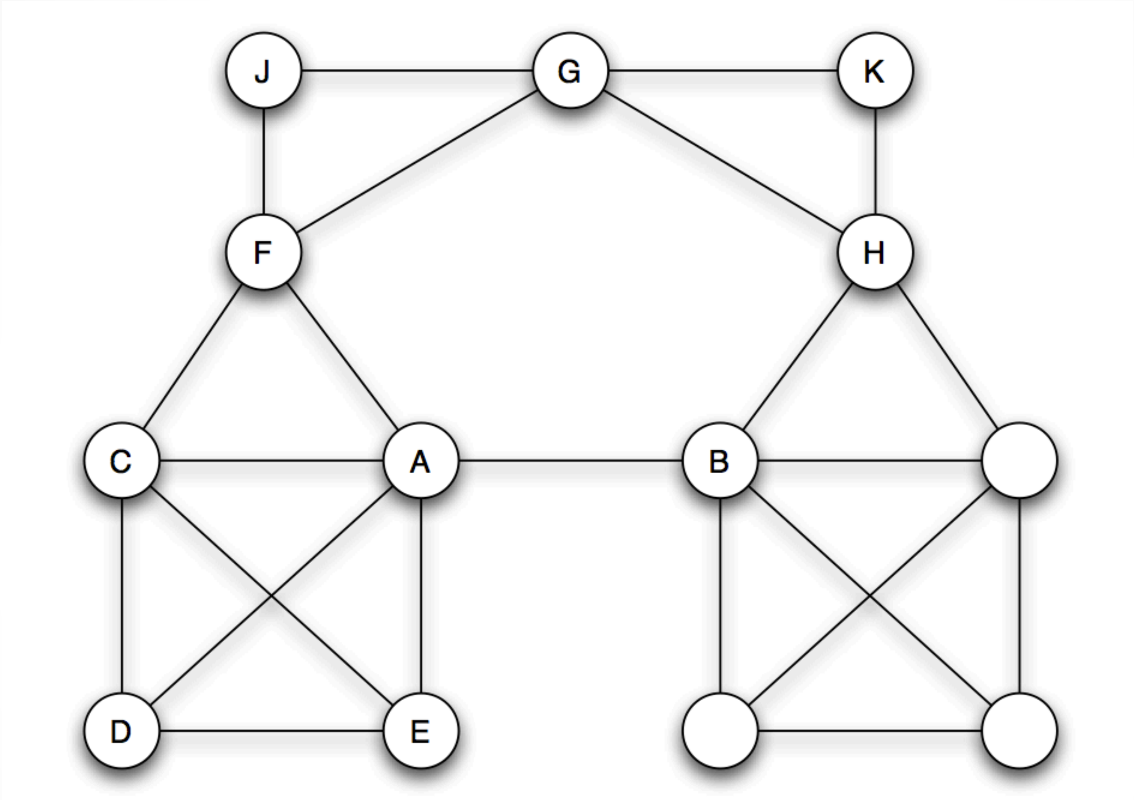
In real social networks, bridges may be rare, so…?

Local Bridges



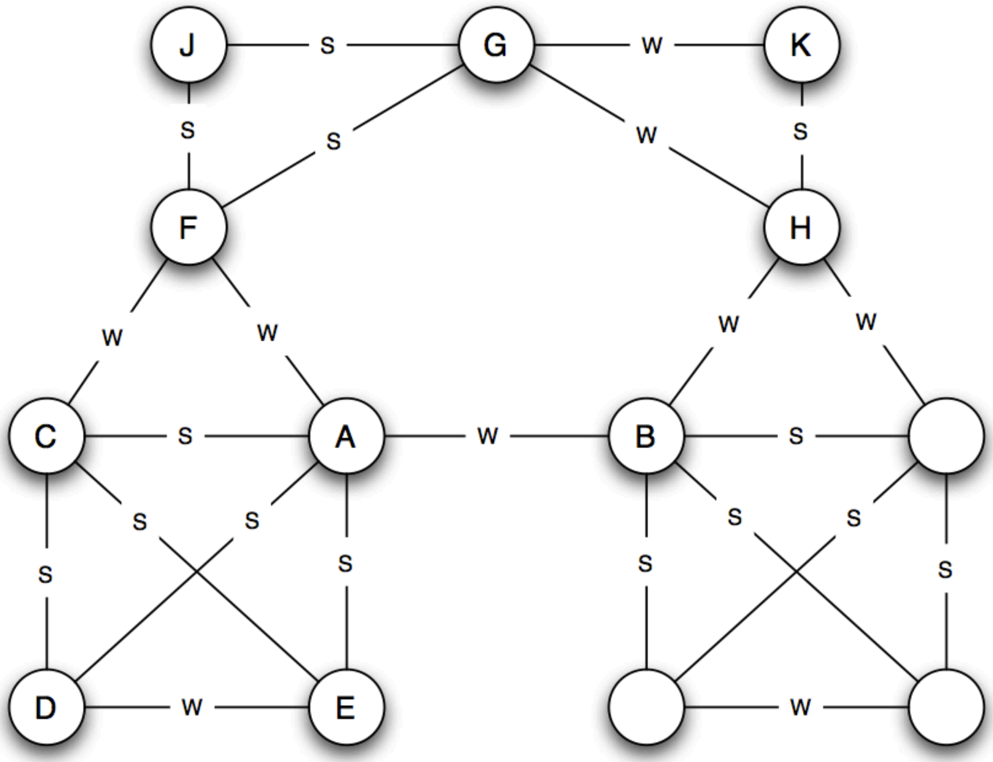
The A-B edge is a local bridge of span 4, since the removal of this edge would increase the distance between A and B to 4.

Local Bridges



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.

Local Bridges

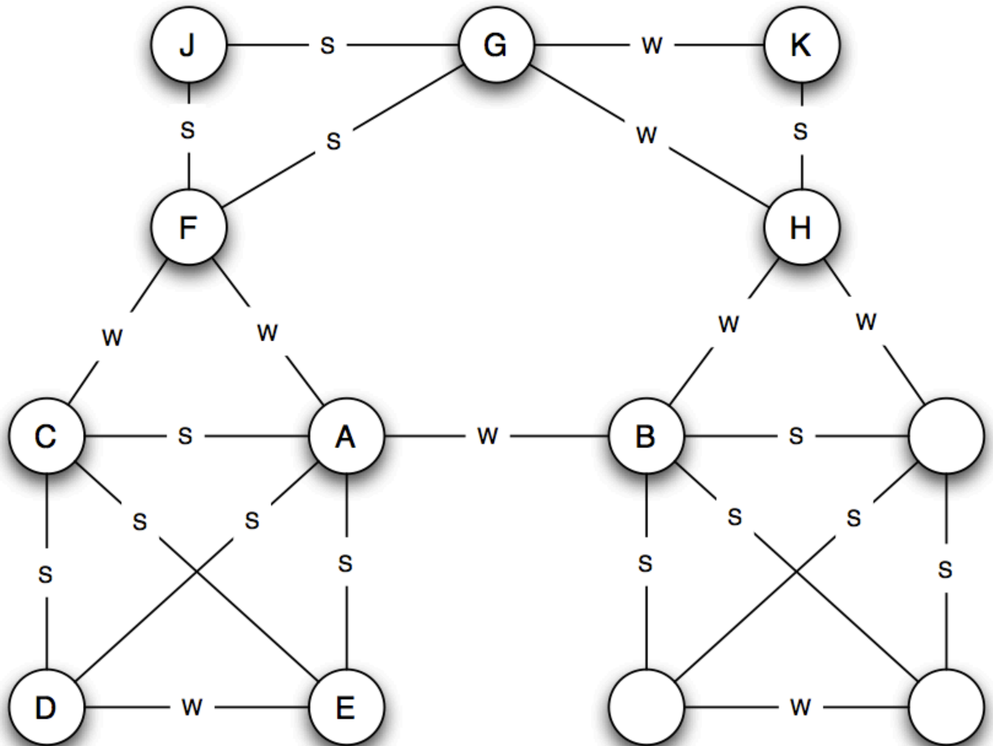


The labelling in the figure satisfies the Strong Triadic Closure Property at each node: if the node has strong ties to two neighbours, then these neighbours must have at least a weak tie between them.

Strong Triadic Closure Property

- If a node A has edges to nodes B and C, then the B-C edge is especially likely to form if A's edges to B and C are both strong ties.
- Granovetter's version:
- We say that a node A violates the Strong Triadic Closure Property if it has strong ties to two other nodes B and C, and there is no edge at all (either a strong or weak tie) between B and C.
- We say that a node A satisfies the Strong Triadic Closure Property if it does not violate it.

Strong Triadic Closure Property



Each edge of the social network is labeled here as either a strong tie (S) or a weak tie (W), to indicate the strength of the relationship.

Local Bridges and Weak Ties

- A purely local, interpersonal distinction between kinds of links — whether they are weak ties or strong ties — as well as a global, structural notion — whether they are local bridges or not.

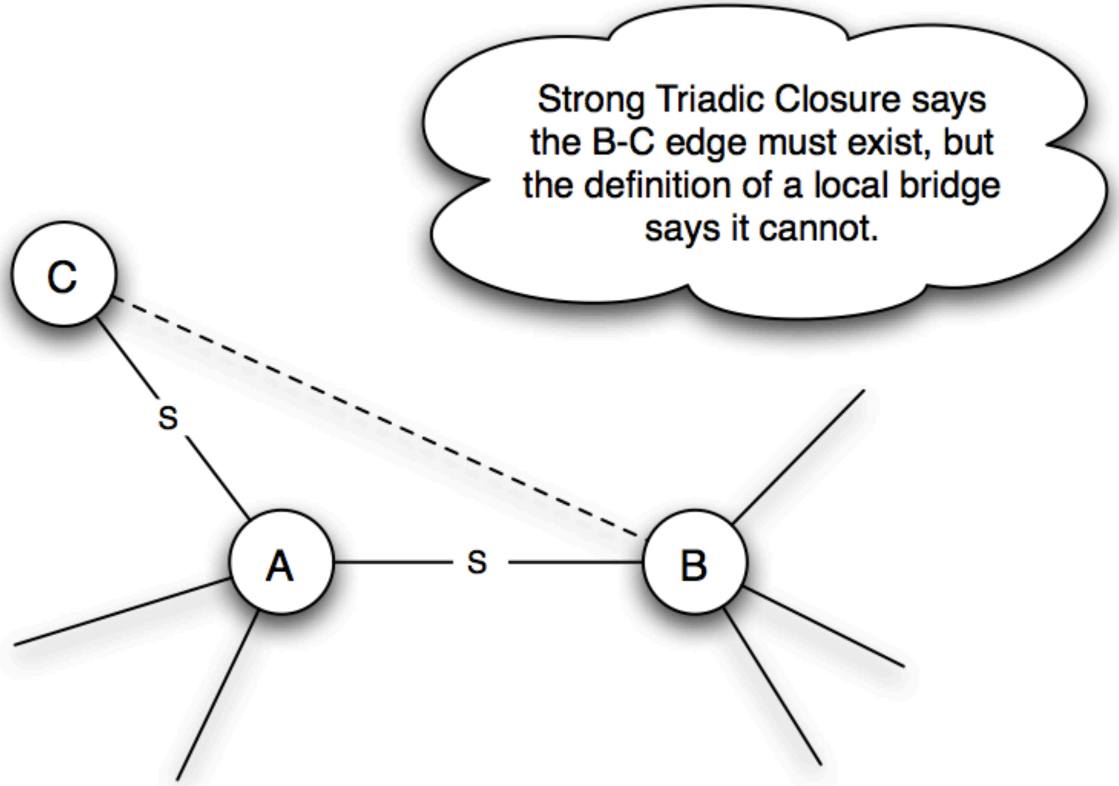
- What is the connection?

- Claim: If a node A in a network satisfies the Strong Triadic Closure Property and is involved in at least two strong ties, then any local bridge it is involved in must be a weak tie.

- Assuming the Strong Triadic Closure Property and a sufficient number of strong ties, the local bridges in a network are necessarily weak ties.

- *How to prove this?*

Local Bridges and Weak Ties



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.

So, what happened?

- We found the connection between the local property of tie strength and the global property of serving as a local bridge.

- A way to think about the way in which interpersonal properties of social-network links are related to broader considerations about the network's structure.

So, what did we learn?

- Simplifying assumptions (STCP) are useful when they lead to statements that are robust in practice.
  - Informally, 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.

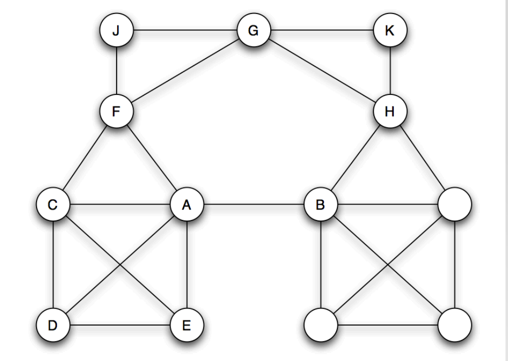
- When the underlying assumptions are stated precisely, as they are here, it becomes possible to test them on real-world data.

- This analysis provides a concrete framework for thinking about the initially surprising fact that life transitions such as a new jobs are often rooted in contact with distant acquaintances.
  - The argument is that these are the social ties that connect us to new sources of information and new opportunities.
  - Their conceptual "span" in the social network (the local bridge property) is directly related to their weakness as social ties.
  - This dual role as weak connections but also valuable conduits to hard-to-reach parts of the network — this is the surprising strength of weak ties.

The story continues...

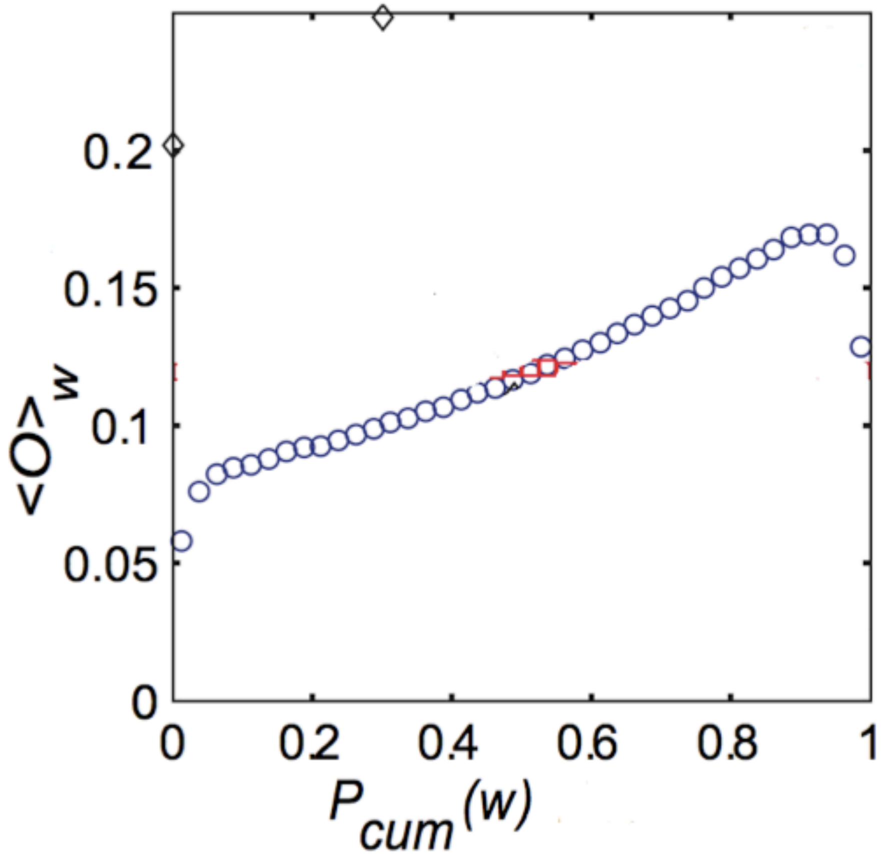
- For many years after Granovetter's work, these predictions remained untested on large social networks..
- Difficulty of finding data that would reliably capture these parameters.
- Enter digital communication data.
- Onnela et al. studied the who-talks-to-whom network maintained by a cell-phone provider that covered roughly 20% of a national population.

The story continues…



- So, how to define the metrics for a practical situation?
- Tie strength: we can make the strength of an edge a numerical quantity, defining it to be the total number of minutes spent on phone calls between the two ends of the edge.
- It is also useful to sort all the edges by tie strength, so that for a given edge we can ask what percentile it occupies this ordering of edges sorted by strength.
- *What about local bridges and "almost" local bridges?*
- Define the neighbourhood overlap of an edge connecting A and B as…
  - $$\frac{\text{number of nodes who are neighbours of both A and B}}{\text{number of nodes who are neighbours of at least one of A and B}}$$

Tie-strength & Network Structure in Large-Scale Data



A plot of the neighbourhood overlap (O) of edges as a function of their percentile (P) in the sorted order of all edges by tie strength. The fact that overlap increases with increasing tie strength is consistent with the theoretical predictions.

Empirical Results on Tie Strength and Neighbourhood Overlap

- First, we can ask how the neighbourhood overlap of an edge depends on its strength; the strength of weak ties predicts that neighbourhood overlap should grow as tie strength grows.

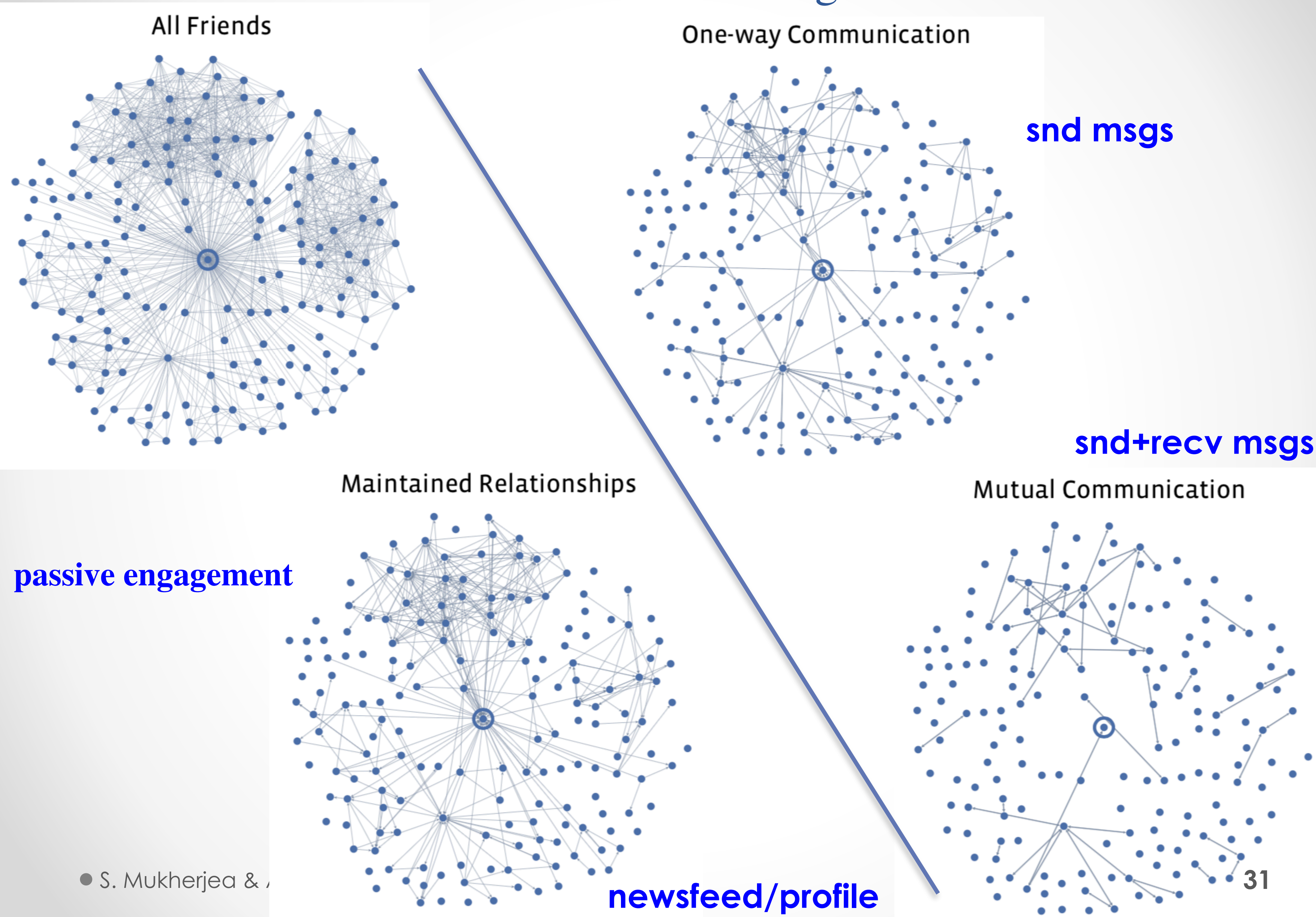
- How can this type of data can be used to evaluate the more global picture suggested by the theoretical framework, that weak ties serve to link together different tightly-knit communities that each contain a large number of stronger ties?

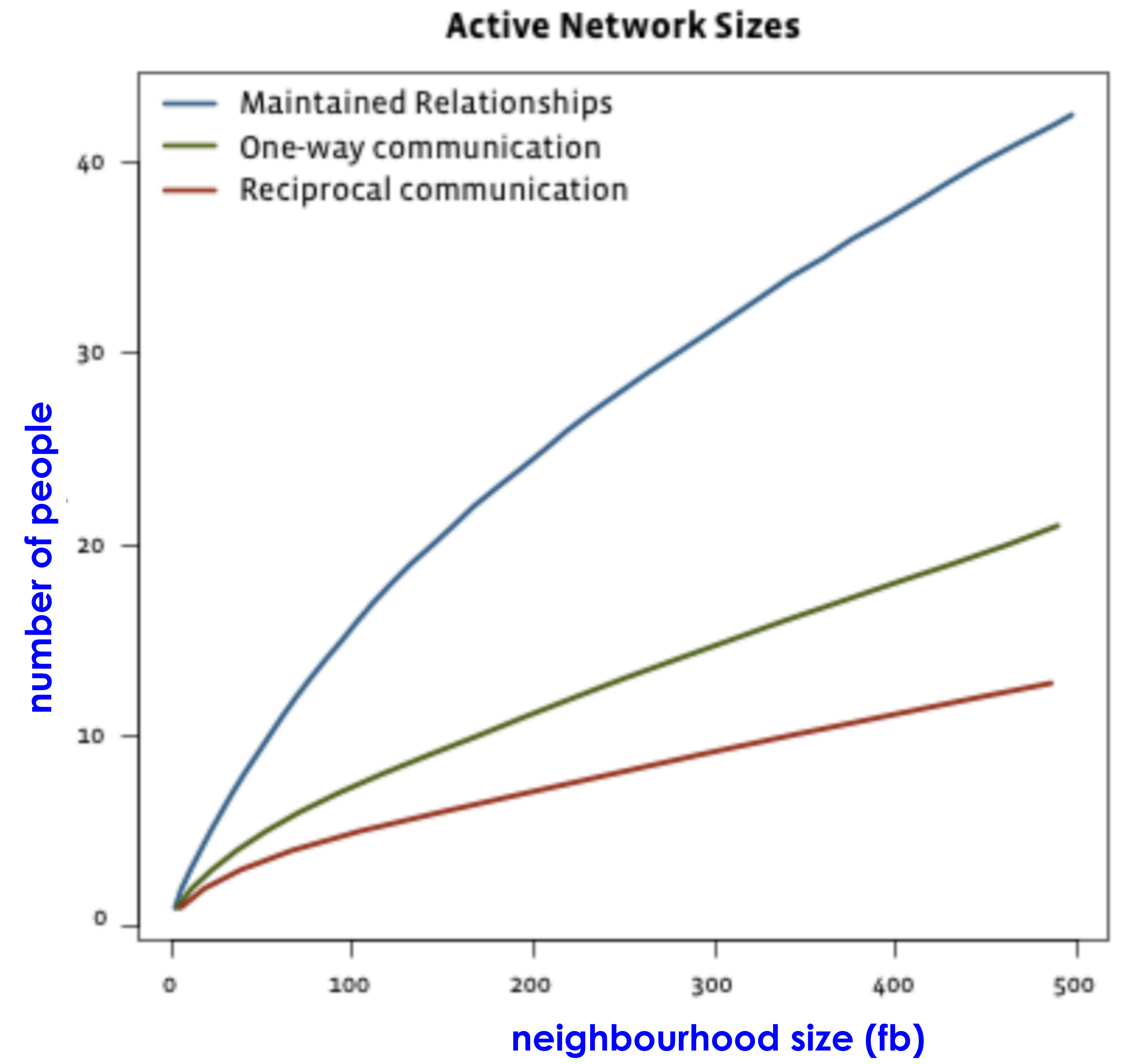
- Onnela et al., first deleted edges from the network one at a time, starting.
  - Starting with the strongest ties and working downward in order of tie strength, the giant component shrank steadily and gradually.
  - Starting from the weakest ties and working upward in order of tie strength. In this case, they found that the giant component shrank more rapidly, and moreover that its remnants broke apart abruptly once a critical number of weak ties had been removed.

Tie Strength, Social Media, and Passive Engagement

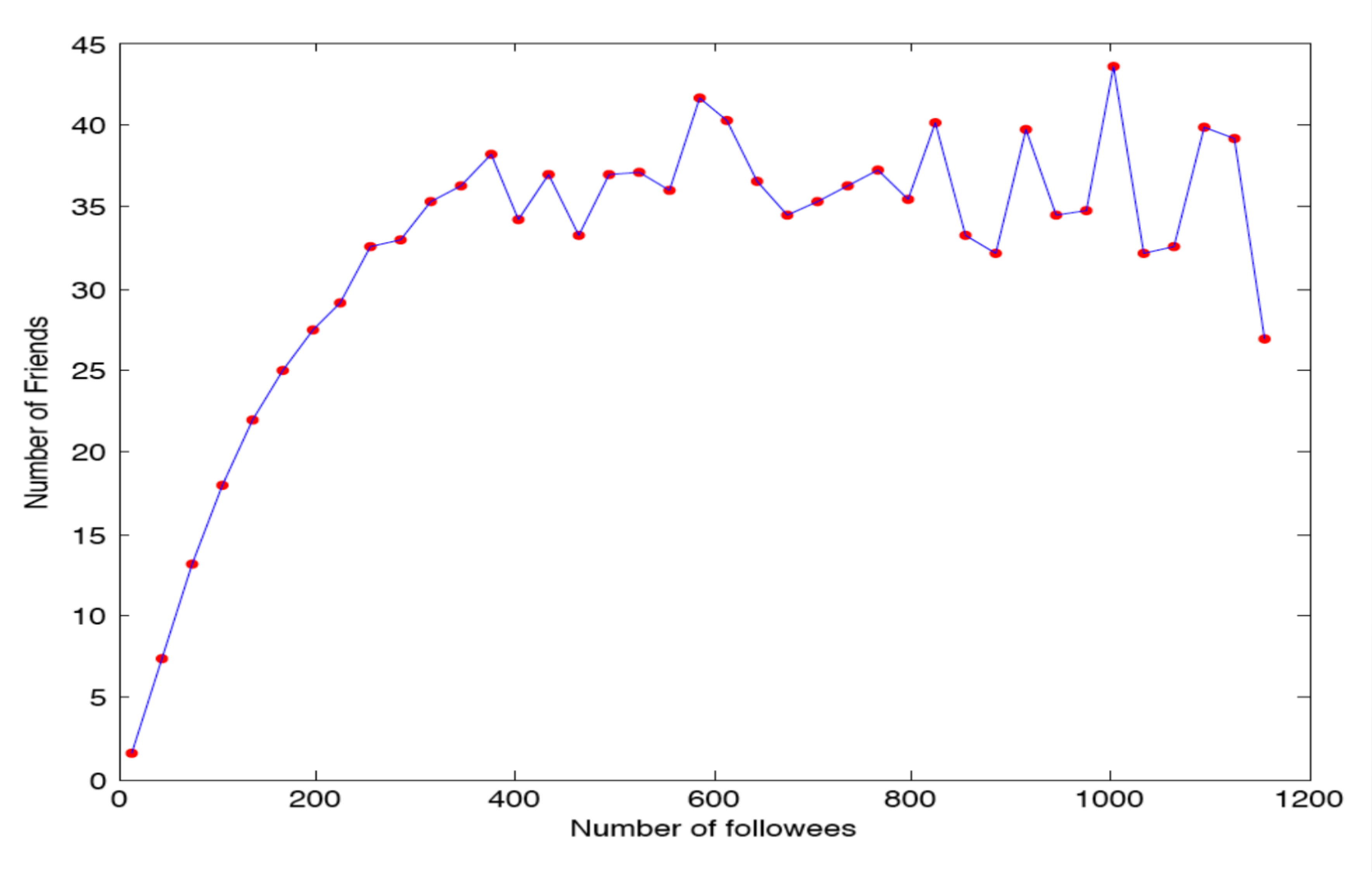
- As is well-known to users of social-networking tools, people maintain large explicit lists of friends.
- In the past, these lists were much more implicit.
- What effect does this have on social network structure more broadly?
- Tie strength can provide an important perspective on such questions, providing a language for asking how on-line social activity is distributed across different kinds of links — and in particular, how it is distributed across links of different strengths.

Facebook Tie Strength




Active Network Sizes
Maintained Relationships
One-way communication
Reciprocal communication
0
10
20
30
40
0
100
200
300
400
500
number of people
neighbourhood size (fb)

Twitter Tie Strength



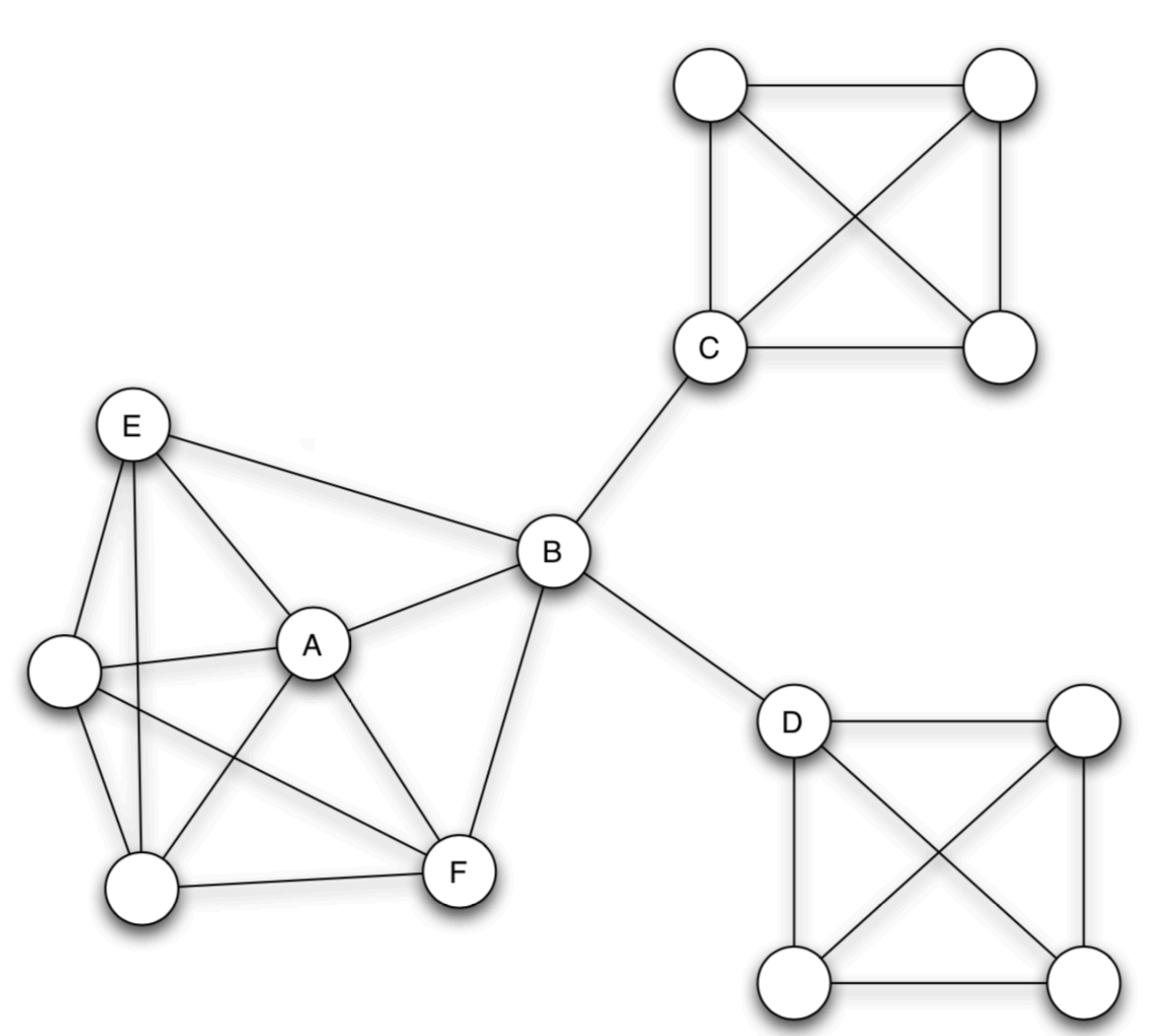
at least 2 directed messages

Tie Strength, Social Media, and Passive Engagement

- As is well-known to users of social-networking tools, people maintain large explicit lists of friends.
- In the past, these lists were much more implicit.
- What effect does this have on social network structure more broadly?
- Tie strength can provide an important perspective on such questions, providing a language for asking how on-line social activity is distributed across different kinds of links — and in particular, how it is distributed across links of different strengths.

Closure, Structural Holes, and Social Capital

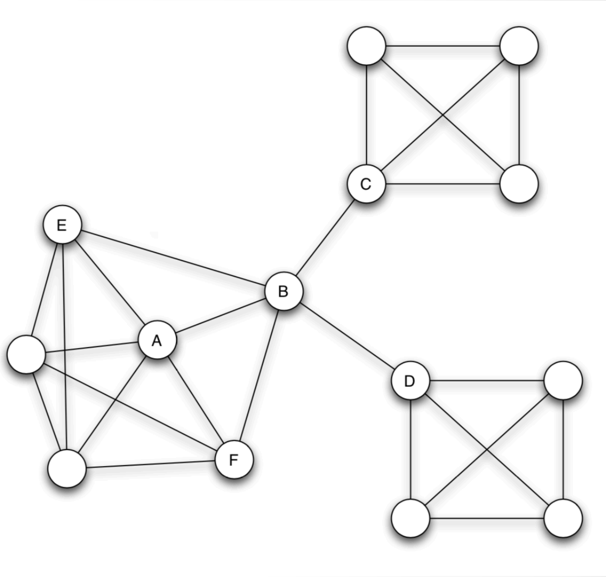
- So far, we have a general view of social networks in terms of tightly-knit groups and the weak ties that link them.
- The analysis has focused primarily on the roles that different kinds of edges of a network play in this structure — with a few edges spanning different groups while most are surrounded by dense patterns of connections.
- There is a lot of further insight to be gained by asking about the roles that different **nodes** play in this structure as well.
- In social networks, access to edges that span different groups is not equally distributed across all nodes:
  - some nodes are positioned at the interface between multiple groups, with access to boundary-spanning edges,
  - while others are positioned in the middle of a single group.



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

Closure, Structural Holes, and Social Capital

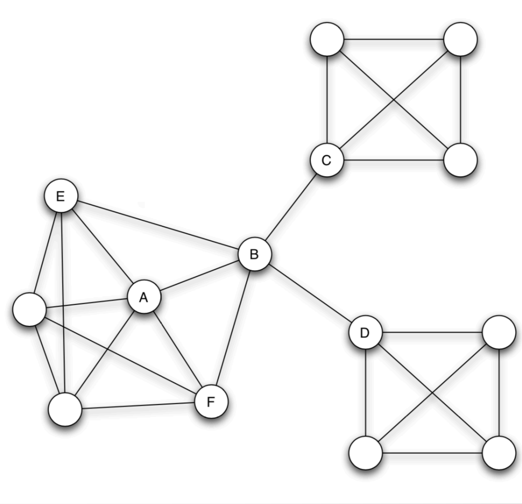
- Embeddedness
  - Node A's set of network neighbours has been subject to considerable triadic closure; A has a high clustering coefficient.
  - We define the *embeddedness* of an edge in a network to be the number of common neighbours the two endpoints have. (embed(A,B) = 2).
  - This is the numerator in the neighbourhood overlap definition; also, a local bridge has embeddedness = 0.



- Sociological research 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.

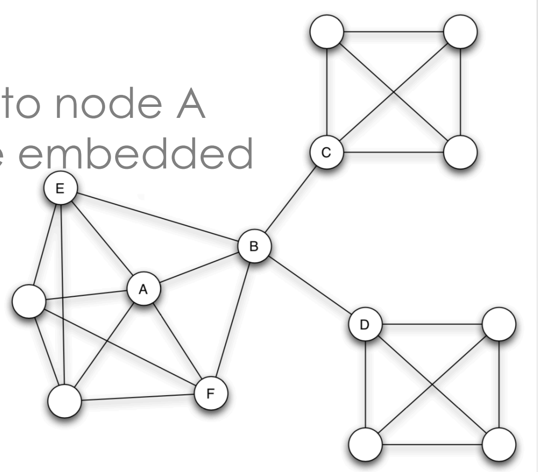
Closure, Structural Holes, and Social Capital

- The presence of mutual friends puts the interactions between two people "on display" in a social sense, even when they are carried out in private;
- In the event of misbehaviour by one of the two parties to the interaction, there is the potential for social sanctions and reputational consequences from their mutual friends.
- No such 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.
  - Moreover, the constraints on B's behaviour are made complicated by the fact that she is subject to potentially contradictory norms and expectations from the different groups she associates with.



Closure, Structural Holes, and Social Capital

- Structural Holes
  - We have been discussing the advantages that accrue to node A from the closure in his network neighbourhood, and the embedded edges that result from this.
  - However, network positions such as node B's, at the ends of multiple local bridges, confer a **distinct** set of equally fundamental advantages (Burt).



- The canonical setting for this argument is the social network within an organisation or company.
  - Rather than the formal organisational hierarchy, we're interested in the more informal network of who knows whom, and who talks to whom on a regular basis.
  - Empirical studies of managers in large corporations has correlated an individual's success within a company to their access to *local bridges*.

Closure, Structural Holes, and Social Capital

- In this network of managers, Node B, with her multiple local bridges, spans a structural hole in the organisation — the "empty space" in the network between two sets of nodes that do not otherwise interact closely.

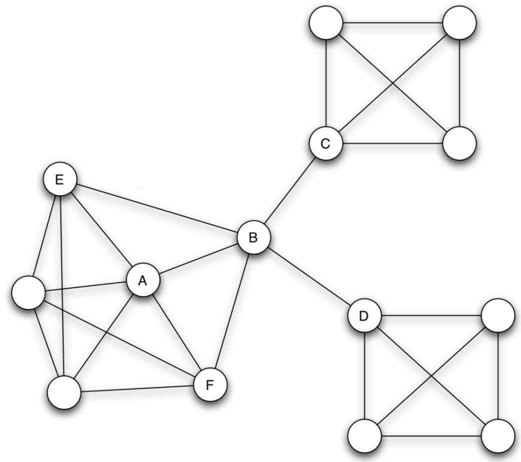
- The argument is that B's position offers advantages in several dimensions relative to A's.

  - The first kind of advantage is an *informational* one: B has early access to information originating in multiple, non-interacting parts of the network.

  - A second, related kind of advantage is based on the way in which standing at one end of a local bridge can be an amplifier for creativity.

    (Experience from many domains suggests that innovations often arise from the unexpected synthesis of multiple ideas, each of them on their own perhaps well-known, but well-known in distinct and unrelated bodies of expertise)

Closure, Struc ocial Capital



- The argument is that B’s position offers advantages in several dimensions relative to A’s.
    - The first kind of advantage is an *informational* one: B has early access to information originating in multiple, non-interacting parts of the network.
    - A second, related kind of advantage is based on the way in which standing at one end of a local bridge can be an amplifier for *creativity*.
    - Third, B’s position in the network provides an opportunity for a kind of social “gate- keeping” — she regulates the access of both C and D to the tightly-knit group she belongs to, and she controls the ways in which her own group learns about information coming from C’s and D’s groups. This becomes a source of *power* for B.

Closure, Structural Holes, and Social Capital

- All of these arguments are framed in terms of individuals and groups deriving benefits from an underlying social structure or social network; as such, they are naturally related to the notion of *social capital*.

- First, social capital is sometimes viewed as a property of a group, with some groups functioning more effectively than others because of favourable properties of their social structures or networks.

- Alternately, it has also been considered as a property of an individual; used in this sense, a person can have more or less social capital depending on his or her position in the underlying social structure or network.

References

1. https://www.cs.cornell.edu/home/kleinber/networks-book/
2. Structure and tie strengths in mobile communication networks by JP Onnela, J Saramäki, J Hyvönen, G Szabó, D Lazer, K Kaski, J Kertész, and AL Barabási. *PNAS* 104, 7332 (2007).

Network Data Sources

1. http://www-personal.umich.edu/~mejn/netdata/
2. https://snap.stanford.edu/data/
3. https://networkdata.ics.uci.edu/index.php