lorenzo.dallamico@unito.it

di.unito.it

Lecture 04.ns03

Strong and weak ties

Course: **Complex Networks Analysis and Visualization**
Sub-Module: **NetSci**

-

From local to global

- networks play a powerful role **bridging the local and 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.**
- For example: 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?
- Important example: the famous "**strength of weak ties**" hypothesis from sociology [**Mark Granovetter**, 1973]

Motivating question

As part of his Ph.D. thesis research in the late 1960s, Mark Granovetter interviewed people who had recently changed employers to learn how they discovered their new jobs

He found that:

1. many people learned information leading to their current jobs through **personal contacts** (not surprising…)
2. these personal contacts were often described by interview subjects as **acquaintances** rather than close friends (more strikingly)

Why?

Granovetter's hypothesis

The answer that Granovetter proposed to this question is striking in the way it links two different perspectives on distant friendships:

1. one **structural**, focusing on the way these friendships span different portions of the full network
2. another **interpersonal**, considering the purely local consequences that follow from a friendship between two people being either strong or weak.

The answer transcends the specific setting of job-seeking, and offers a way of thinking about the architecture of social networks more generally.

Strong and weak ties

Triadic Closure

The Strength of Weak Ties

Tie Strength and Network Structure in Large-Scale Data

Tie Strength, Social Media, and Passive Engagement

Closure, Structural Holes, and Social Capital

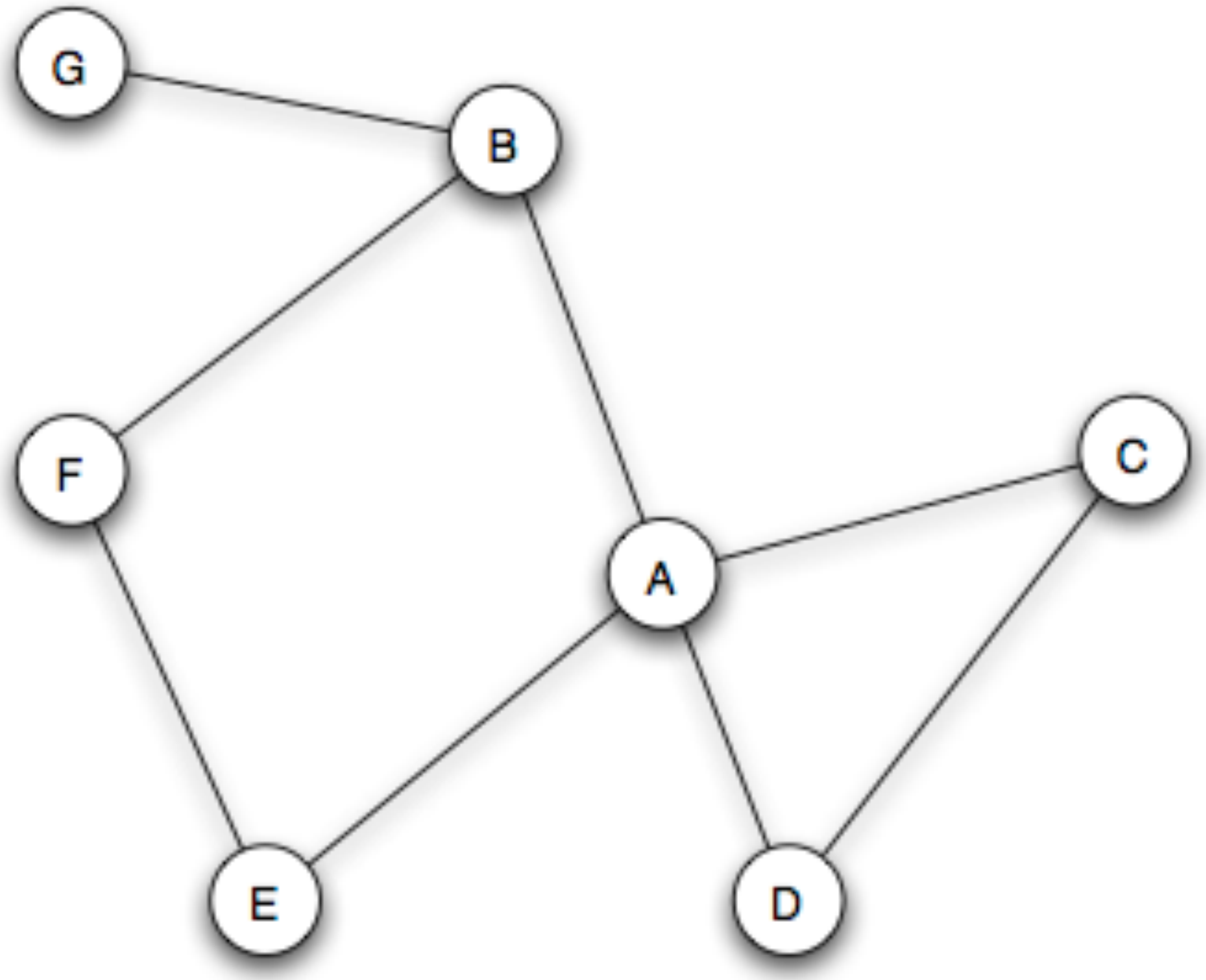
Triadic Closure

Evolution of a network in time

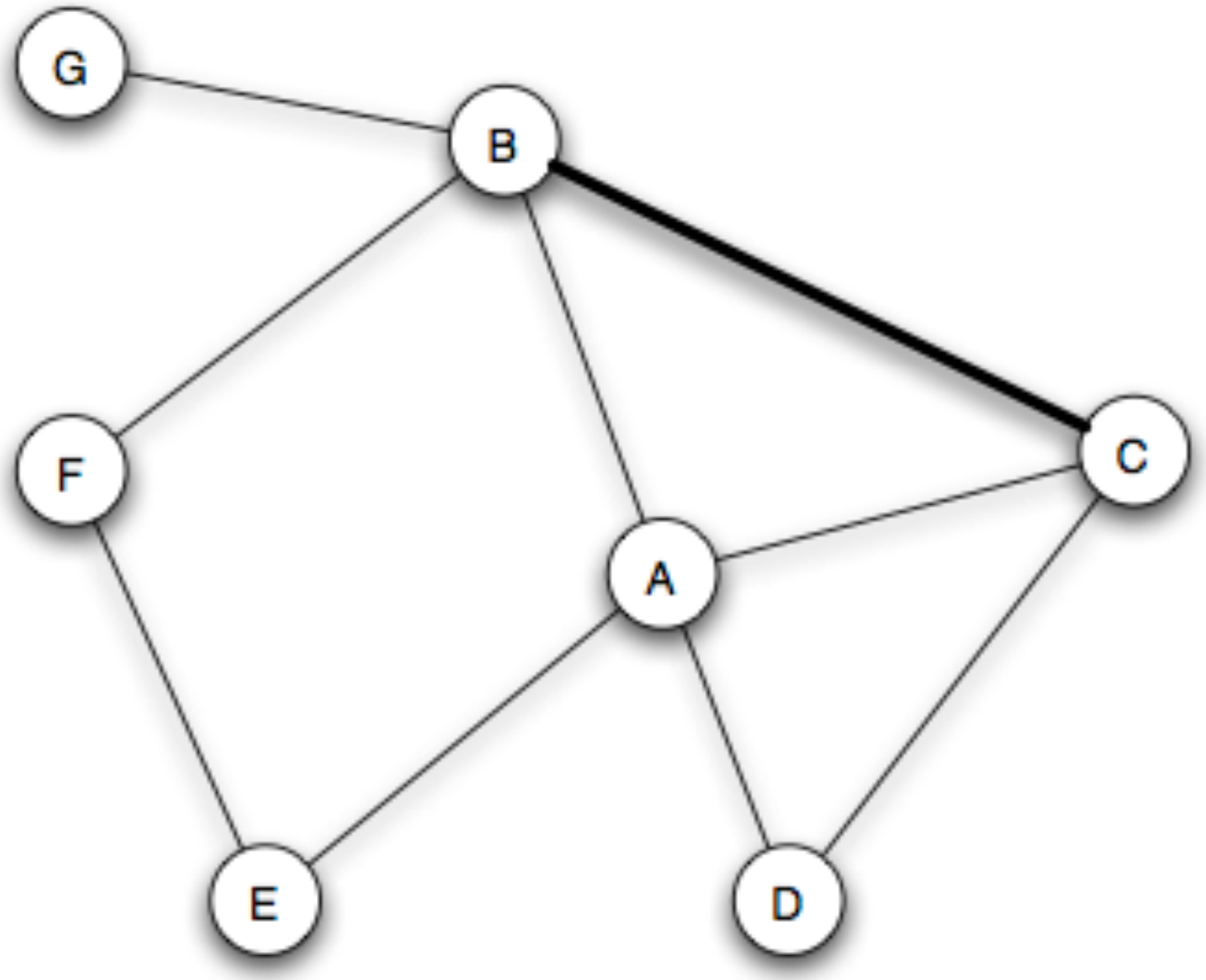
- So far, we took a snapshot of the nodes and edges at a particular moment in time, and then ask about paths, components, distances, and so forth.
- What are the mechanisms by which nodes arrive and depart, and by which edges form and vanish?
- Basic answer:
  If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future
  [**Rapoport**, 1953]

-> **Triadic Closure**

Triadic Closure



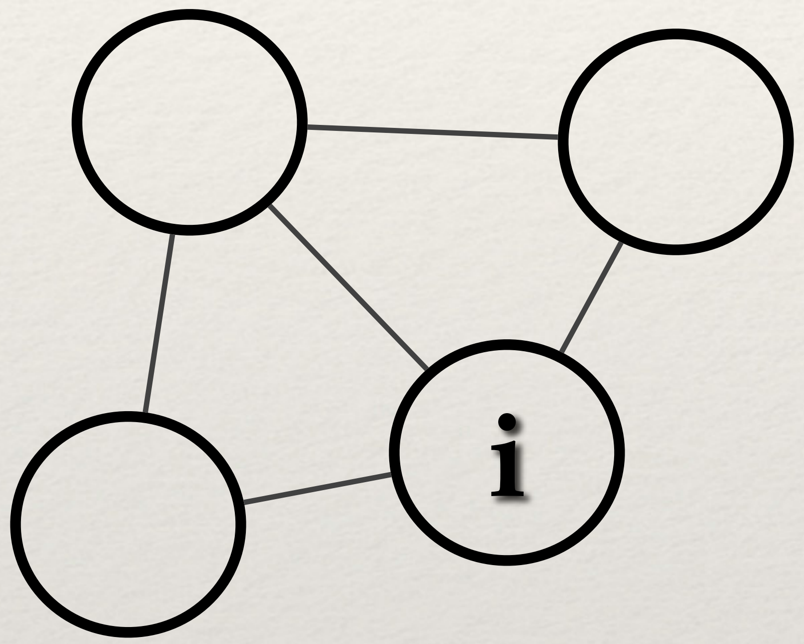
(a) *Before B-C edge forms.*



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

Relation with the clustering coefficient

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 is the **clustering coefficient**



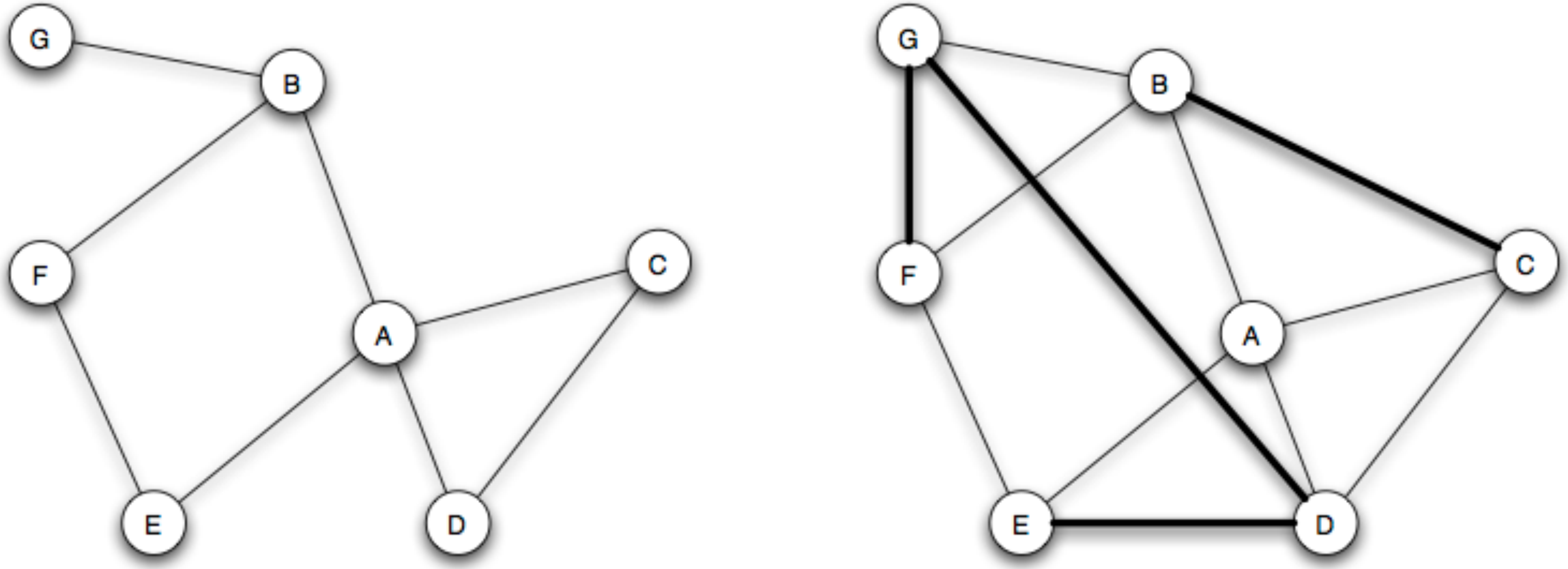
$$C_t(i) = \frac{2\tau(i)}{k_i(k_i - 1)} = \frac{1 \cdot 2}{3 \cdot 2} = \frac{1}{3}$$

$$C_{t'}(i) = \frac{2}{3}$$

The clustering coefficient of a node A can also be defined as the *probability that two randomly selected friends of A are friends with each other*.

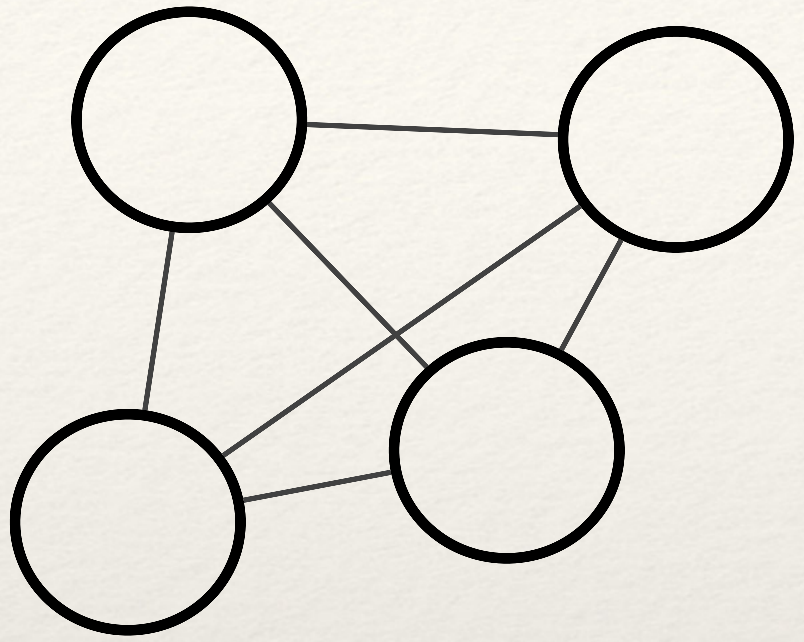
The more strongly triadic closure is operating in the neighborhood of the node, the higher the clustering coefficient will tend to be.

After a given span of time…



(a) *Before new edges form.* (b) *After new edges form.*

Closing triangles



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 more strongly triadic closure is operating in the neighborhood of the node, the higher the clustering coefficient will tend to be.

Side note on clustering coefficient

- high cc: extremely frequent in social networks ("friendship transitivity")
- low cc: if compared with the cc in a randomly generated graph
  - such a random graph must have the same number of nodes, and an equal number of edges reshuffled

$$G, G_R$$ randomization of G

$$CC(G) >> CC(G_R) \Rightarrow$$ "high" clustering coefficient

Reasons for triadic closure

- Intuitively very natural, but experience suggests some of the basic psycho-social reasons why it operates:
  - **Opportunity** - if A spends time with both B and C, then there is an increased chance that they will end up knowing each other and potentially becoming friends
  - **Trusting** - B and C are friends with A: they have a basis for trusting each other that an arbitrary pair of unconnected people might lack.
  - **Incentive** - 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.

Fritz Heider. The Psychology of Interpersonal Relations. John Wiley & Sons, 1958.

Side note on "incentive"

It has empirical reflections that show up in natural but troubling ways in public-health data.

For example**, it has been** found that **teenage girls** who have a **low clustering** coefficient in their network of friends are significantly more likely to **contemplate suicide** than those whose clustering coefficient is high…

=> the structure of social connections is an indicator for catastrophic events

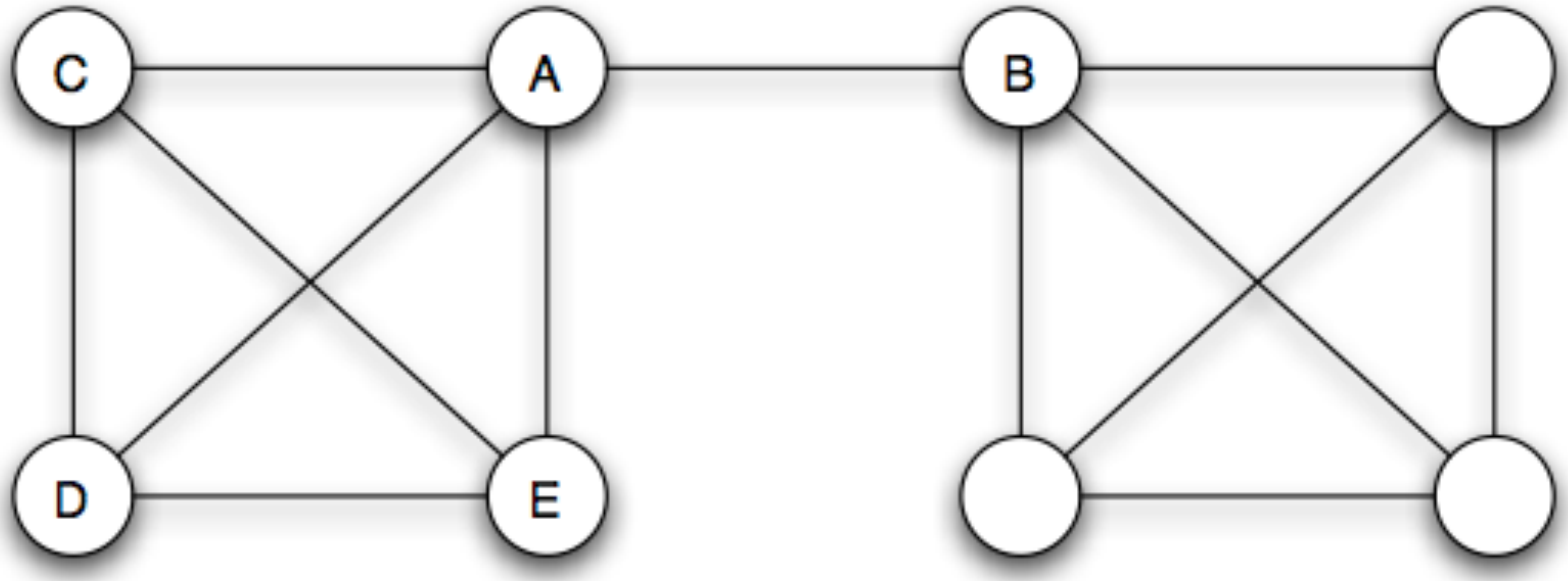
Peter Bearman and James Moody. *Suicide and friendships among American adolescents*. American Journal of Public Health, 94(1):89–95, 2004.

The Strength of Weak Ties

The Strength of Weak Ties

- **Triadic closure** turns out to be one of the **crucial ideas** needed to unravel what's going on
- How links are involved?
  - weak (acquaintance) or
  - strong (friend)?
- Let's start by positing **that information about good jobs is something that is relatively scarce**.
- hearing about a promising job opportunity from someone suggests that *they have access to a source of useful information that you don't.*

Bridge



What about (A,B) relationship?

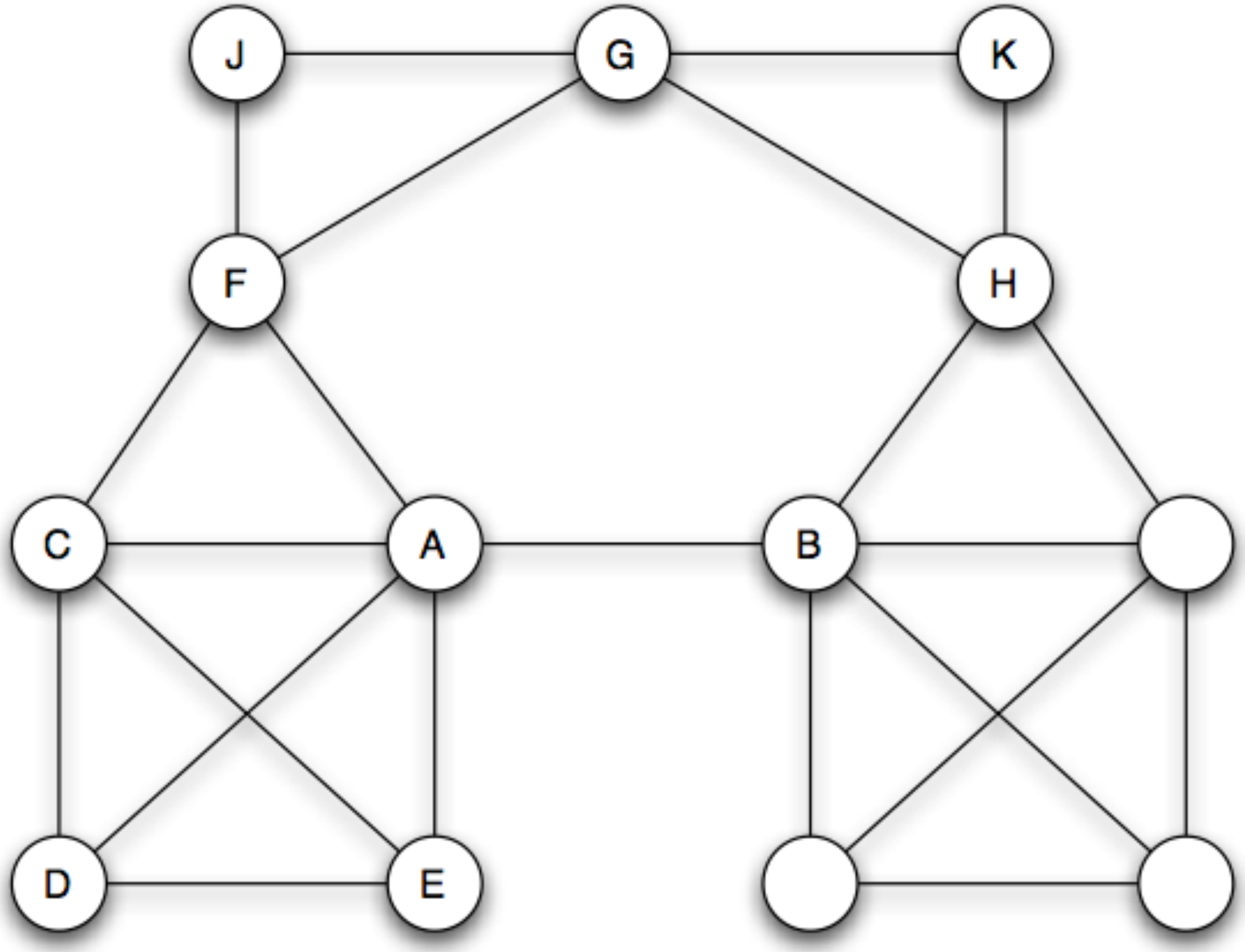
(Strong) definition of a bridge

*We say that an edge joining two nodes A and B in a graph is a **bridge** if deleting the edge would cause A and B to lie in two different components.*

In other words, this edge is literally the **only route** between its endpoints, the nodes A and B.

Bridges are presumably **extremely rare** in real social networks. You may have a friend from a very different background, and it may seem that your friendship is the only thing that bridges your world and his, but one expects **in reality** that **there will be other, hard-to-discover, multi-step paths that also span these worlds**.

Longer paths



Definition of a local bridge

- In previous figure, the A-B edge isn't the only path that connects its two endpoints; though they may not realize it, A and B are also connected by a **longer path** through F, G, and H
- This is much more common than a bridge in real social networks. We need another definition:
  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 > 2*

Span of a local bridge

- We say that the **span** of a local bridge is the **distance its endpoints would be from each other if the edge were deleted**
- In the figure, span(A-B) = 4
- we can also check that no other edge in this graph is a local bridge, since for every other edge in the graph, the endpoints would still be at distance two if the edge were deleted.
- 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.

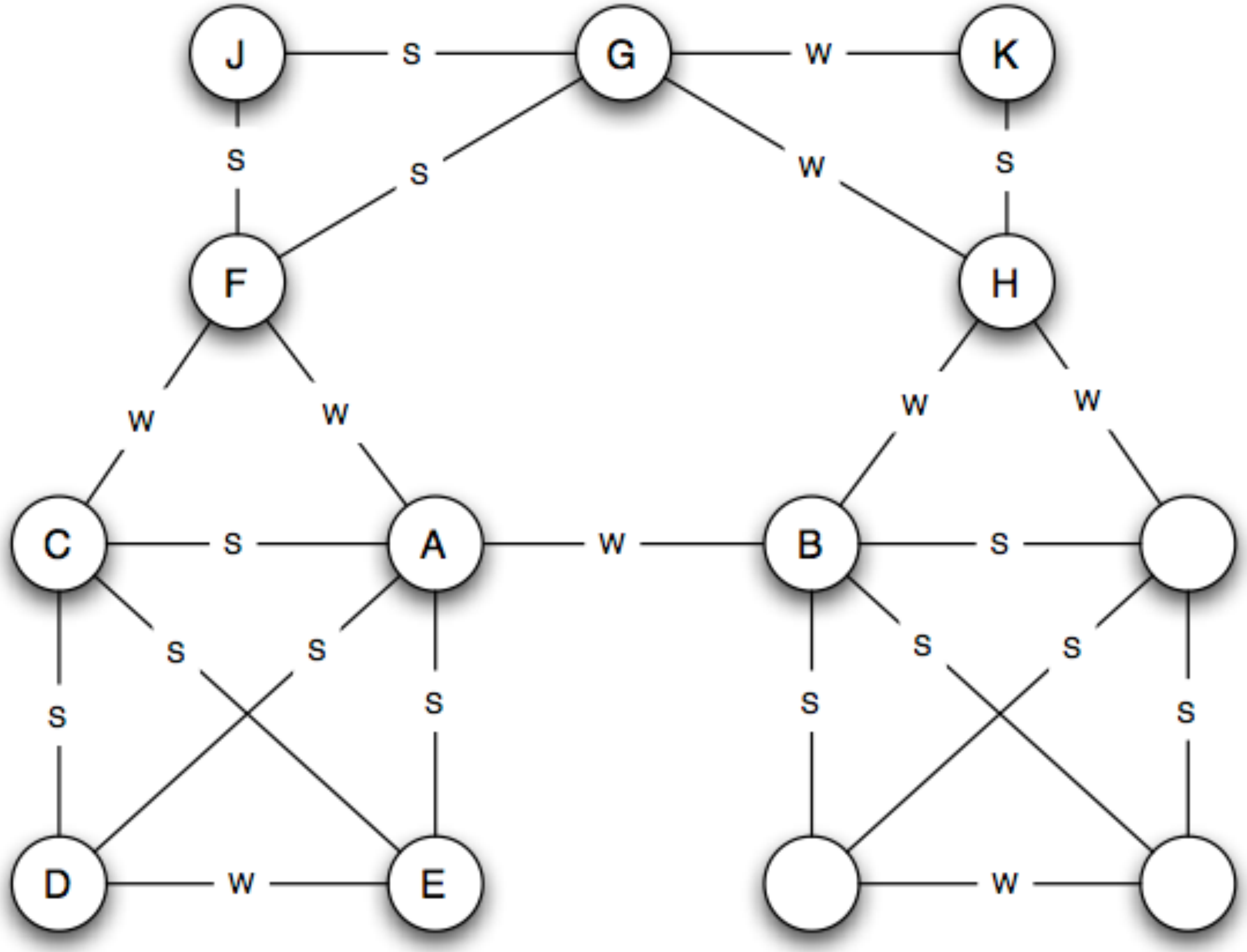
The role of a local bridge

- Local bridges, especially those with reasonably large span, **still play roughly the same role that bridges do**, though in a less extreme way — they provide their endpoints with access to parts of the network, and hence sources of information, that they would otherwise be far away from.
- We might expect that if a node like A is going to get **truly new information**, the kind that leads to a new job, it might come unusually often (though certainly not always) from a friend connected by a local bridge.
  The closely-knit groups that you belong to, though they are filled with people eager to help, are also filled with people who know roughly the same things that you do.

Different level of strength

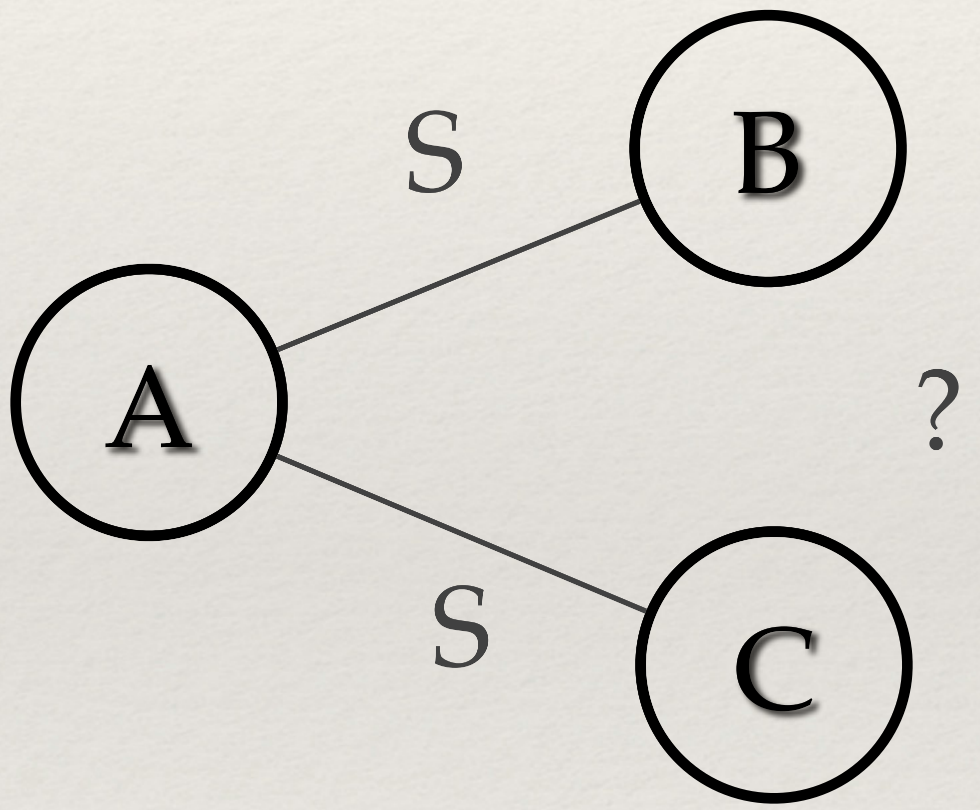
- We deliberately refrain from trying to define "**strength**" precisely, but we mean it to align with the idea that stronger links represent **closer friendship and greater frequency of interaction**.
- For conceptual simplicity we'll categorize all links in the social network as belonging to one of two types: **strong ties** (the stronger links, corresponding to **friends**), and **weak ties** (the weaker links, corresponding to **acquaintances**).
- Now we can take a social network and **annotate** each edge with a designation of it as either strong or weak

Strong and weak ties annotations



Triadic closure and strength of ties

If we recall the arguments supporting **triadic closure**, based on **opportunity, trust, and incentive**, they all act more powerfully when the edges involved are strong ties than when they are weak ties.



Qualitative assumption:

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.

Strong Triadic Closure

To enable some more concrete analysis, Granovetter suggested a more formal (and somewhat more extreme version) of this, as follows:

*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.*

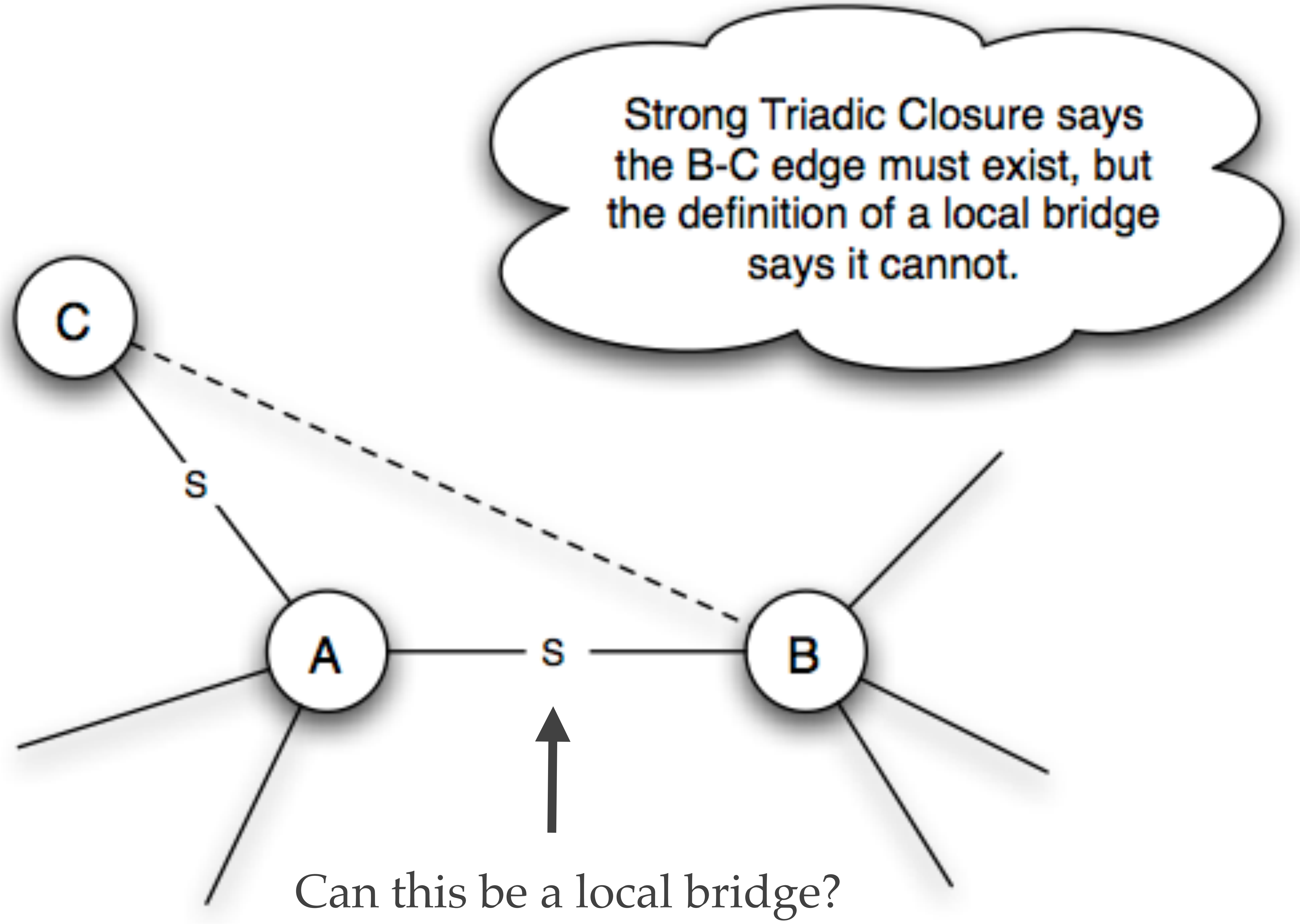
Clearly the Strong Triadic Closure Property is **too extreme** for us to expect it holds across all nodes of a large social network. But it can be useful as an abstraction

Local Bridges and Weak Ties

We now have 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.

On the surface, there is no direct connection between the two notions, but in fact using triadic closure we can establish a connection, in the following **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.



Can this be a local bridge?

Mathematical proof

The proof proceeds by contradiction.

Take some network, and consider a node A that satisfies the Strong Triadic Closure Property and is involved in at least two strong ties.

Now suppose A is involved in a local bridge — say, to a node B — that is a strong tie.

First, since A is involved in at least two strong ties, and the edge to B is only one of them, it must have a strong tie to some other node, which we'll call C.

Proof (cont'd)

Is there an edge connecting B and C?

Since the edge from A to B is a local bridge, A and B must have no friends in common, and so the B-C edge must not exist.

This **contradicts** Strong Triadic Closure: since the A-B and A-C edges are both strong ties, the B-C edge must exist.

This contradiction shows that our initial premise**, the existence of a local bridge that is a strong tie, cannot hold**. QED

Interpersonal and structural properties

This argument completes the connection we’ve been looking for between the local property of tie strength and the global property of serving as a local bridge.

As such, it gives us 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.**

Assumption too strong?

- Strong Triadic Closure is very strong as an assumption.
- Simplifying assumptions are useful when they lead to statements that are robust in practice, making sense as qualitative conclusions that hold in approximate forms even when the assumptions are relaxed.
- In fact, the mathematical argument can be summarized more informally and approximately as saying that in real life, **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

Test on real data

- When the underlying assumptions are stated precisely it becomes possible to test them on **real-world data**.
- In the past few years researchers have studied the relationship of tie strength and network structure quantitatively across **large populations**, and have shown that the conclusions described here in fact hold in an **approximate** form.
- We describe some of this empirical research later.

Concrete framework

- 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, and 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.

Tie Strength and Network Structure in Large-Scale Data

Tie Strength and Network Structure in Large-Scale Data

- For many years after Granovetter's initial work these predictions **remained relatively untested** on large social networks, due to the difficulty in finding data that reliably captured the strengths of edges in large-scale, realistic settings.
- This changed rapidly once detailed traces of **digital communication** became available.
- Such "**who-talks-to-whom**" data exhibits the two ingredients we need for empirical evaluation: the **network structure** of communication among pairs of people, and the **total time** that two people spend talking to each other — the more time spent communicating during the course of an observation period, the stronger we declare the tie to be.

$$(w, s) \to (0,1) \to [0, +\infty]$$

The case of cell-phone network

- In 2007 the who-talks-to-whom network maintained by a cell-phone provider that covered roughly **20% of a national population** has been studied.
- The **nodes** correspond to cell-phone users, and there is an **edge** joining two nodes if they made phone calls to each other in both directions over an **18-week** observation period.
- The network can be viewed as a reasonable sampling of the conversations occurring within a social network representing a significant fraction of one country's population..
- The data exhibits many of the broad structural features of typical large social networks, including a **giant component** — a single connected component containing most (in this case **84%**) of the individuals in the network.

J.-P. Onnela, J. Saramaki, J. Hyvonen, G. Szabo, D. Lazer, K. Kaski, J. Kertesz, and A.-L. Barabasi. *Structure and tie strengths in mobile communication networks*. Proc. Natl. Acad. Sci. USA, 104:7332–7336, 2007.

Generalizing the Notions of Weak Ties and Local Bridges

- The theoretical formulation is based on two definitions that impose sharp dichotomies on the network:
  - an edge is either a strong tie or a weak tie, and $\forall e \in L : w(e) \in [0, +\infty]$
  - it is either a local bridge or it isn't.
- For both of these definitions, it is useful to have versions that exhibit smoother gradations when we go to examine real data at a large scale.
- 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.

Neighborhood overlap

Since a very small fraction of the edges in the cell-phone data are local bridges, it makes sense to soften this definition as well, so that we can view certain edges as being "almost" local bridges.

Hence, we define the **neighborhood overlap** of an edge connecting A and B to be the ratio:

$$O_{AB} = \frac{|N(A) \cap N(B)|}{|N(A) \cup N(B)|}$$

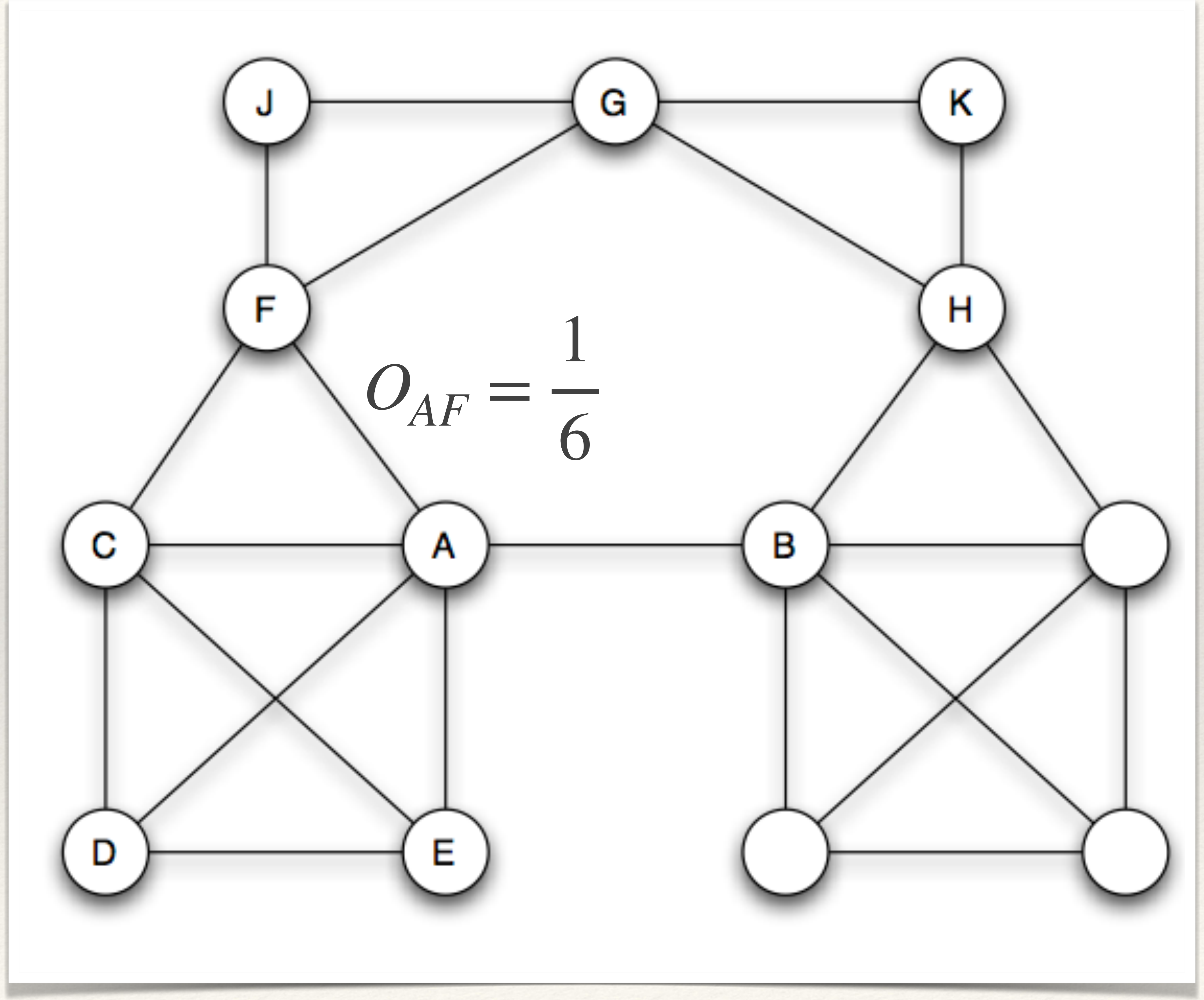
where in the denominator we don't count A or B themselves.

Example

- Consider the edge A-F.

- The denominator of the neighborhood overlap for A-F is determined by the nodes B, C, D, E, G, and J, since these are the ones that are a neighbor of at least one of A or F.

- Of these, only C is a neighbor of both A and F, so the **neighborhood overlap is 1/6**.



Key feature

- This ratio in question is 0 precisely when the numerator is 0, and hence when the edge is a local bridge.

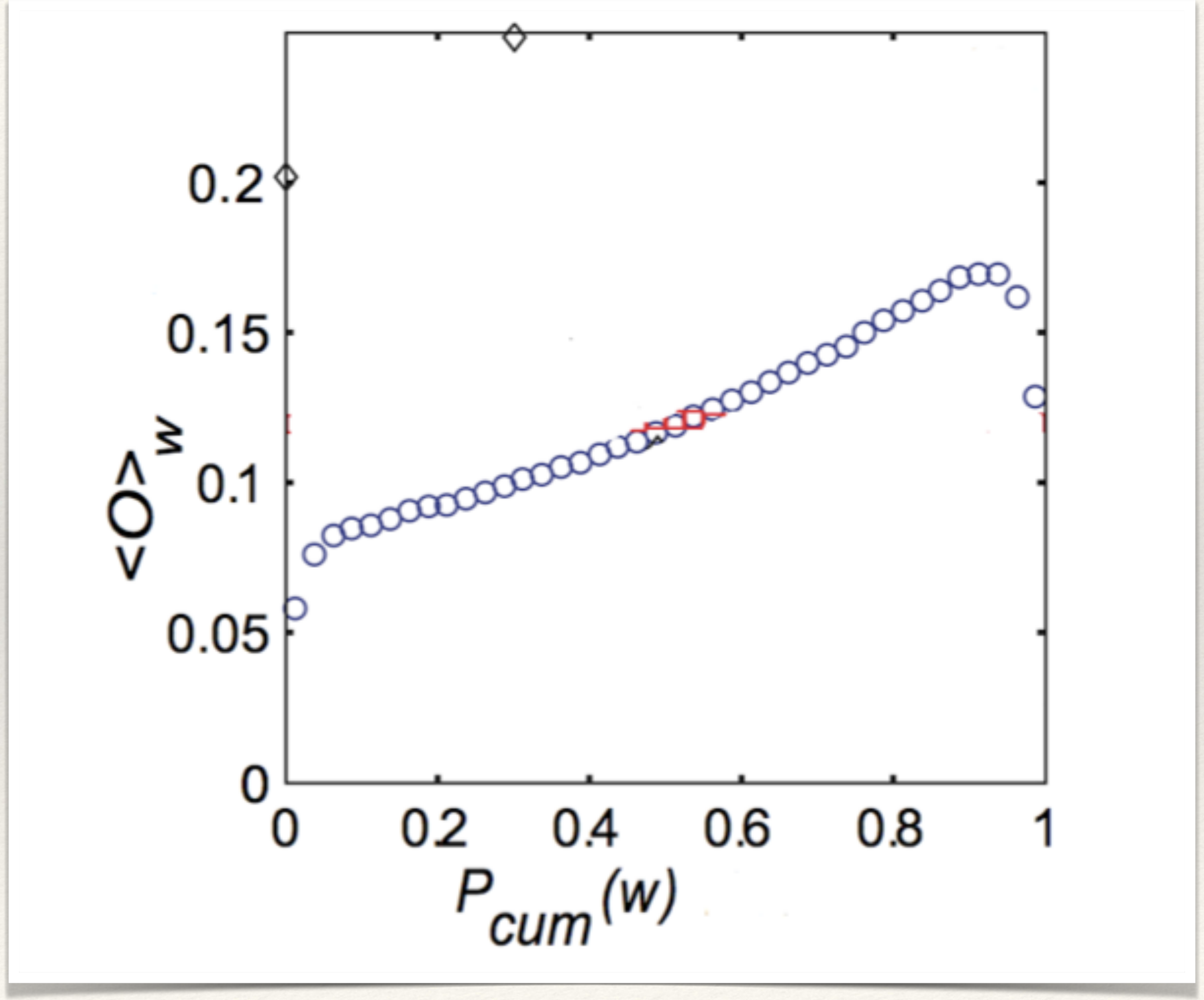
$$O_{A,B} = 0 \Leftrightarrow \text{(A,B) is a local bridge}$$

- So the notion of a **local bridge is contained within this definition** — local bridges are the edges of neighborhood overlap 0 — and hence we can think of edges with very small neighborhood overlap as being "almost" local bridges.
- Intuitively, edges with very small neighborhood overlap consist of nodes that travel in "social circles" having almost no one in common.

Plotting neighborhood overlap

A plot of the neighborhood overlap of edges as a function of their percentile 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



Indirect analysis of Onnela et al.

- They first **deleted edges** from the network one at a time, **starting with the strongest ties** and working downward in order of tie strength.
- The **giant component shrank steadily** as they did this, its size going down gradually due to the elimination of connections among the nodes.
- They then tried the same thing, but **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.
- This is consistent with a picture in which **the weak ties provide the more crucial connective structure** for holding together disparate communities, and for **keeping the global structure of the giant component intact**.

Important observation

- This is just a first step toward evaluating theories of tie strength on network data of this scale.
- It illustrates some of the inherent challenges: given the size and complexity of the network, we cannot simply look at the structure and "see what's there."
- Indirect measures must generally be used, and *since 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 in the way that one can on small datasets.*

Tie Strength, Social Media, and Passive Engagement

Tie Strength, Social Media, and Passive Engagement

- As it is well-known to users of **social-networking tools**, people maintain large explicit lists of friends in their profiles — in contrast to the ways in which such friendship circles were once much more implicit, and in fact relatively difficult for individuals even to enumerate or mentally access.
- Which is the effect?
- 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.
- <u>When we see people maintaining hundreds of friendship links on a social-networking site, we can ask how many of these correspond to strong ties</u> that involve frequent contact, and how many of these correspond to weak ties that are activated relatively rarely.

Tie Strength on Facebook

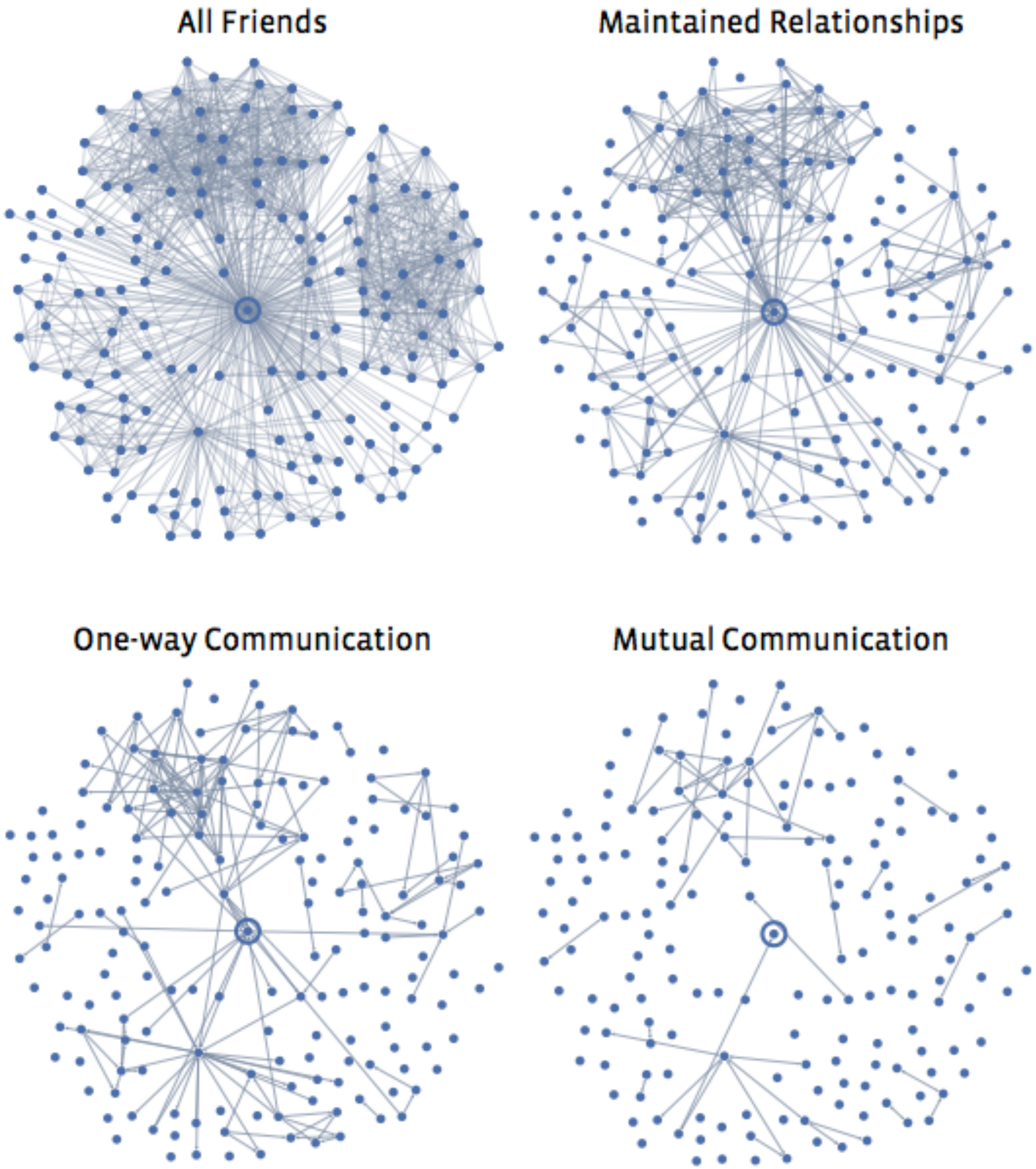
At Facebook, Cameron Marlow and his colleagues analyzed the friendship links reported in each user's profile, asking to what extent each link was actually used for social interaction, beyond simply being reported in the profile.

In other words, **where are the strong ties among a user's friends?**

Cameron Marlow, Lee Byron, Tom Lento, and Itamar Rosenn. *Maintained relationships on Facebook*, 2009. On-line at *http://overstated.net/2009/03/09/maintained-relationships-on-facebook*

Categories of links

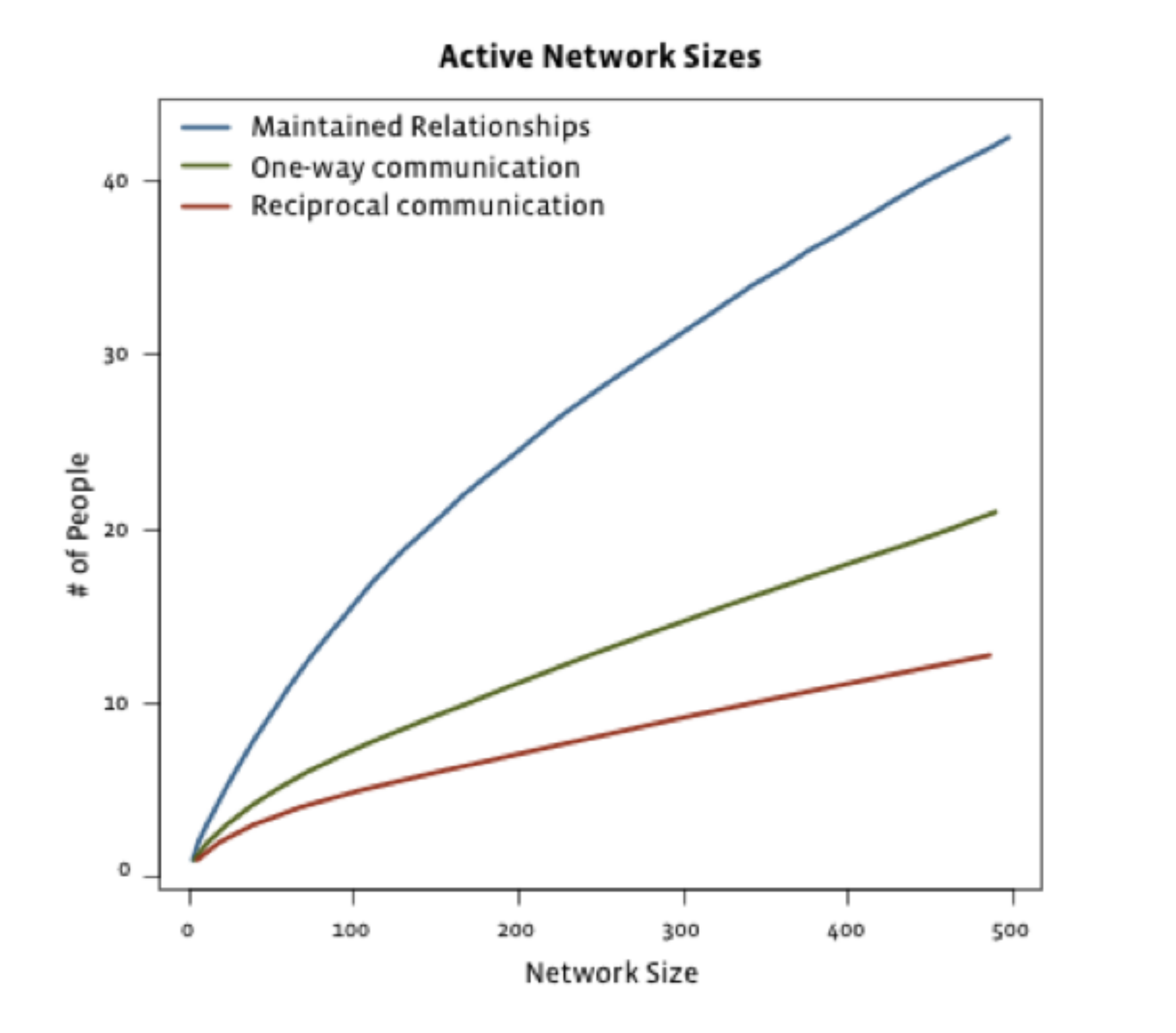
- **reciprocal (mutual) communication**, if the user both sent messages to the friend at the other end of the link, and also received messages from them during the observation period.
- **one-way communication**, if the user sent one or more messages to the friend at the other end of the link (whether or not these messages were reciprocated).
- **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 profile more than once.


All Friends
Maintained Relationships
One-way Communication
Mutual Communication

Qualitative Discussion

- To get a sense of the relative volumes of these different kinds of interaction through an example, previous figure shows the network neighborhood of a sample Facebook user — consisting of all his friends, and all links among his friends.
- The picture in the upper-left shows the set of all declared friendships in this user's profile; the other three pictures show how the set of links becomes sparser once we consider only maintained relationships, one-way communication, or reciprocal communication.
- Moreover, as we restrict to stronger ties, certain parts of the network neighborhood thin out much faster than others.

Plotting different links



Quantitative Discussion

- On the x-axis is the total number of friends a user declares, and the curves then show the (smaller) numbers of other link types as a function of this total.
- There are several interesting conclusions to be drawn from this.
    - it confirms that even for users who report very large numbers of friends on their profile pages (on the order of 500), the number with whom they actually communicate is generally between 10 and 20, and the number they follow even passively (e.g. by reading about them) is under 50.
    - Online social media are very powerful, because they enable **passive engagement**: one keeps up with friends by reading news about them even in the absence of communication.

Tie Strength on Twitter

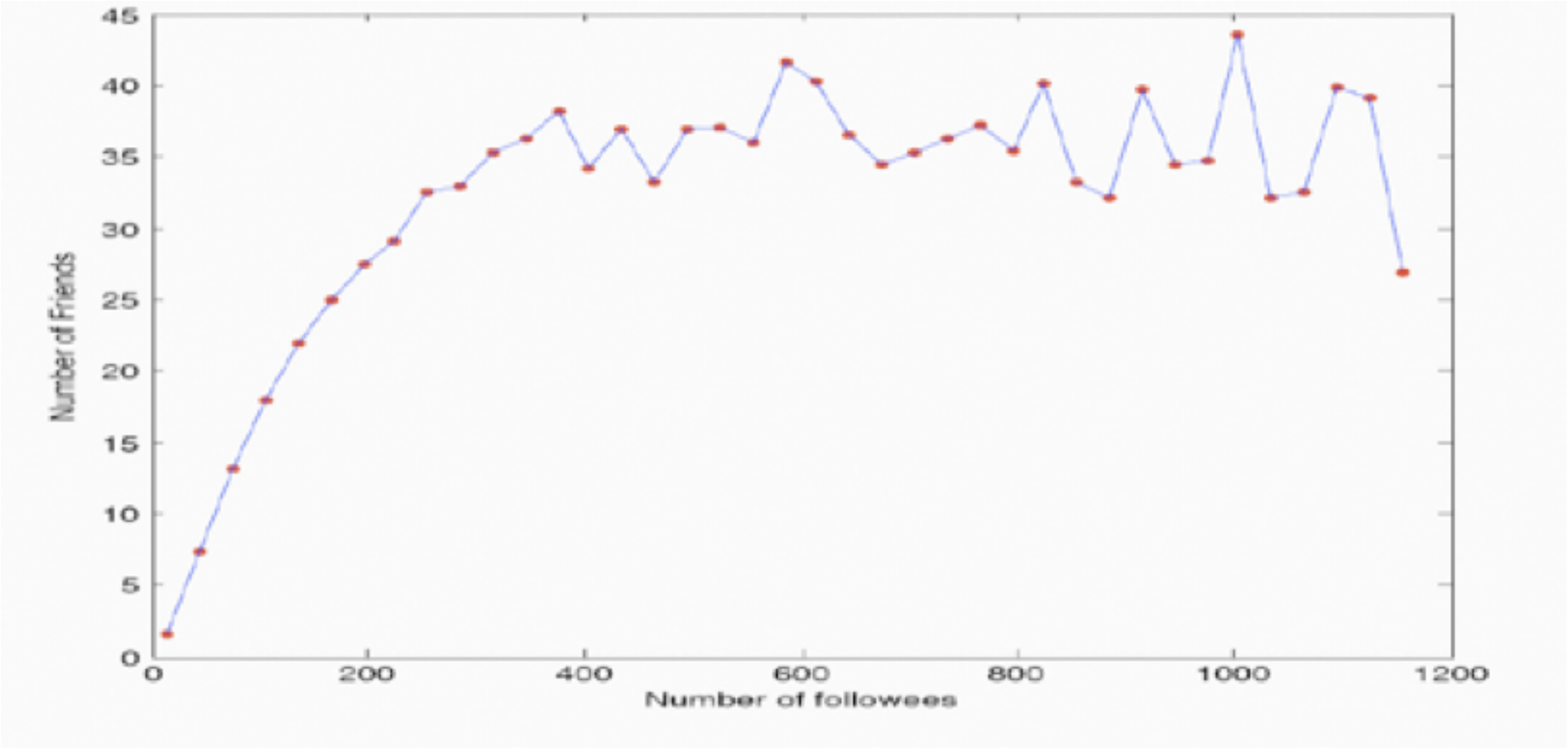
- Twitter also includes social-network features, and these enable one to distinguish between stronger and weaker ties: each user can specify a set of other users whose messages he or she will follow, and each user can also direct messages specifically to another user.
- The former kind of interaction defines a social network based on more passive, **weak ties** — it is very easy for a user to follow many people's messages without ever directly communicating with any of them.
- The latter kind of interaction — especially when we look at users directing multiple messages to others — corresponds to a **stronger** kind of direct interaction.

Follower and followees

- Huberman, Romero, and Wu analyzed the relative abundance of these two kinds of links on Twitter.
- Specifically, for each user they considered the number of users whose messages she followed (her "followees"), and then defined her strong ties to consist of the users to whom she had directed at least two messages over the course of an observation period.

Bernardo A. Huberman, Daniel M. Romero, and Fang Wu. *Social networks that matter: Twitter under the microscope*. First Monday, 14(1), January 2009.

Strong ties in Twitter



Result

As we saw for Facebook, even for users who maintain very large numbers of weak ties on-line, the number of strong ties remains relatively modest, in this case stabilizing at a value below 50 even for users with over 1000 followees.

Discussion

By definition, each strong tie requires the continuous investment of time and effort to maintain, and so even people who devote a lot of their energy to building strong ties will eventually reach a limit — imposed simply by the hours available in a day — on the number of ties that they can maintain in this way.

The formation of weak ties is governed by much milder constraints — they need to be established at their outset but not necessarily maintained continuously — and so it is easier for someone to accumulate them in large numbers.

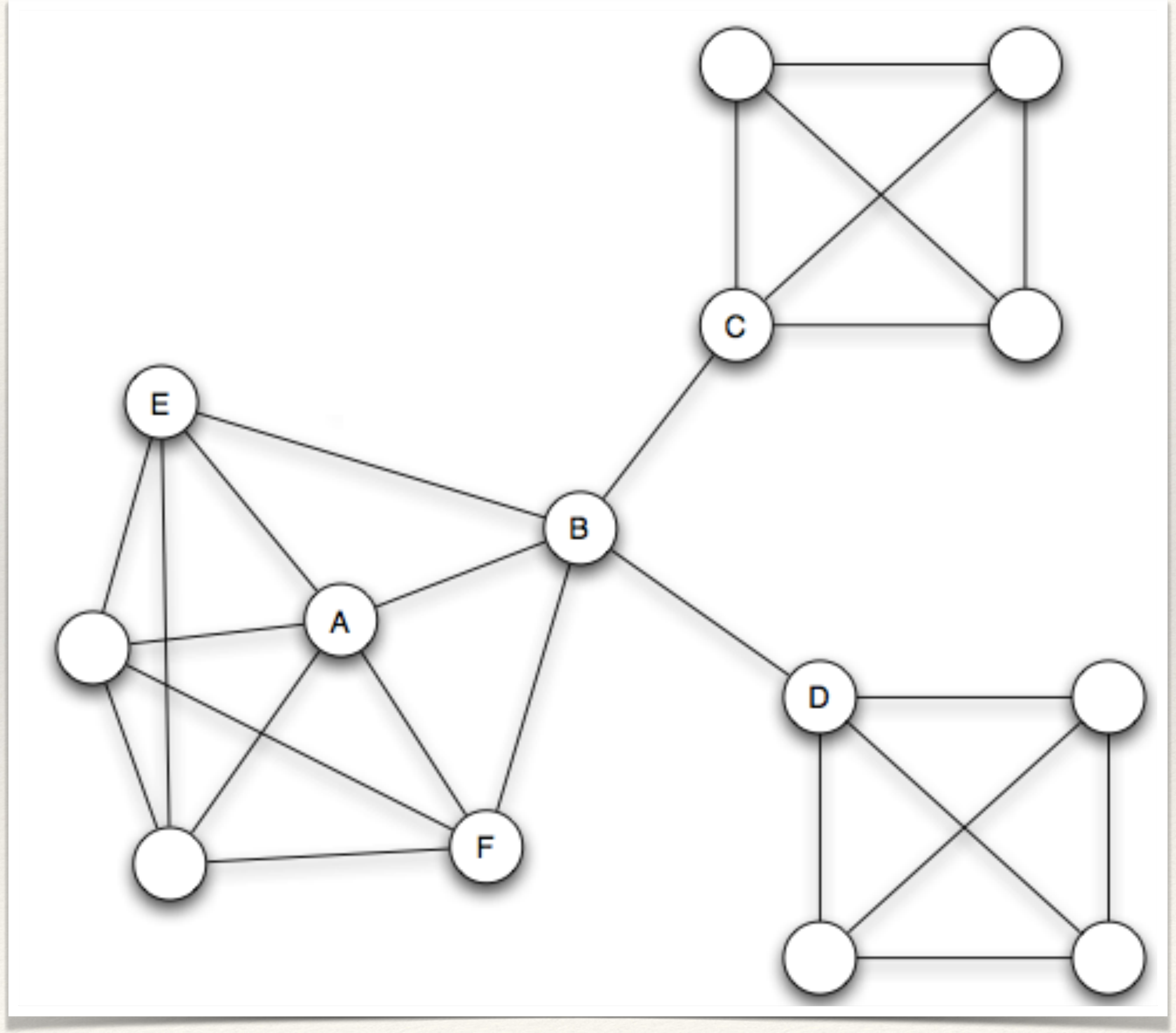
Closure, Structural Holes, and Social Capital

Closure, Structural Holes, and Social Capital

- We made a general view of social networks in terms of **tightly-knit groups** and the **weak ties** that link them
- There is **a lot of further insight** to be gained by asking about **<u>the roles that different nodes</u>** 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 betwee multiple groups**, with access to boundary-spanning edges, while others are positioned in the middle of a single group.
- What is the effect of this **heterogeneity**?

Contrast between nodes

- We can formulate an answer to this question as a story about the **different experiences** that nodes have in a network.
- For example, focus on the contrast between the experience of a node such as A, who sits at the **center of a single tightly-knit group**, and node B, who sits at the **interface between several groups**.



Ronald S. Burt. The network structure of social capital. Research in Organizational Studies, 22:345–423, 2000.

Embeddedness

- Node A is at the **center of a single tightly-knit group.**
- Node A's set of network neighbors has been subject to considerable **triadic closure**.
- A has a **high clustering coefficient**.
- Let's define the ***embeddedness*** *of an edge in a network to be the number of common neighbors the two endpoints have.*

Neighborhood overlap and Local bridges

- Embeddedness = the **numerator of the neighborhood overlap**
- **Local bridges** = edges whose embeddedness equal to 0
- 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)

Misbehavior

- 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 **misbehavior** by one of the two parties to the interaction, there is *the potential for social sanctions and reputational consequences from their mutual friends*.
- 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.

Structural Holes

- A related line of research in sociology has argued that network positions at the ends of **multiple local bridges**, confer a distinct set of equally fundamental advantages.
- Empirical studies of managers in **large corporations** has correlated an individual's success within a company to their access to local bridges.
- At a more abstract level, the central arguments behind these studies are also supported by the **network principles** we have been discussing, as we now explore further .

Benefits of structural holes

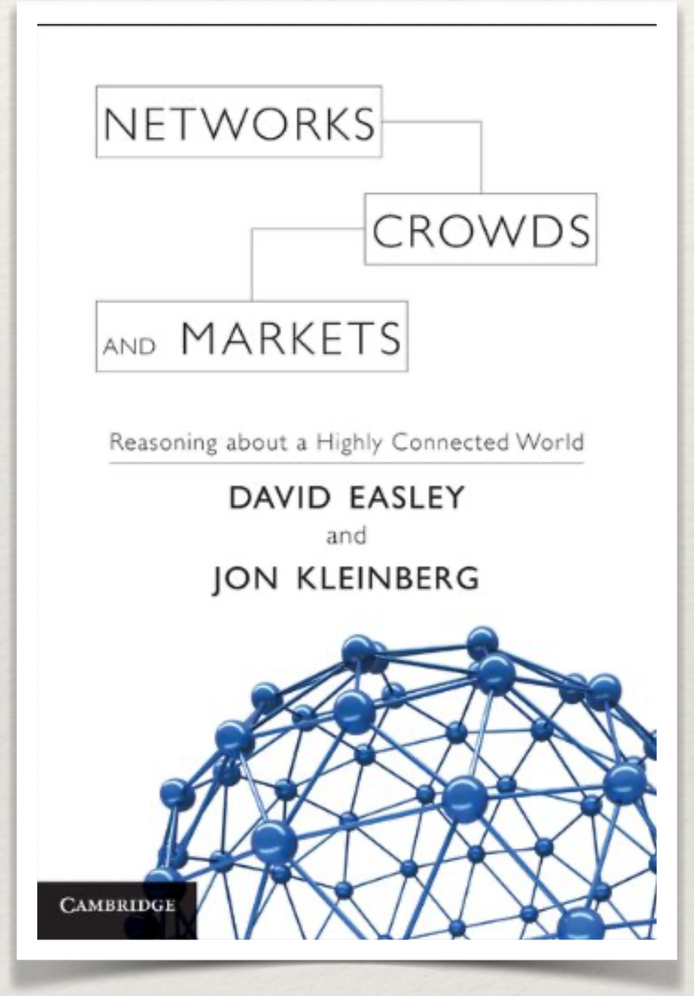
- In Burt's language, a node B with multiple local bridges, spans a **structural hole** in the organization — the "empty space" in the network between two sets of nodes that do not otherwise interact closely.
- A node in this position **offers advantages** in several dimensions to relative nodes with high embeddedness.
- **Informational** advantage: such a node has early access to information originating in multiple, non-interacting parts of the network. This node is investing her energy efficiently by reaching out to different groups rather than basing all her contacts in the same group.

Other advantages

Standing at one end of a local bridge:

- can be an amplifier for **creativity.** 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.
- provides an opportunity for a kind of social “**gate-keeping**” — regulates the access 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.

References



[ns2] Chapter 3 (3.1 - 3.5) —->

https://www.cs.cornell.edu/home/kleinber/networks-book/networks-book-ch03.pdf