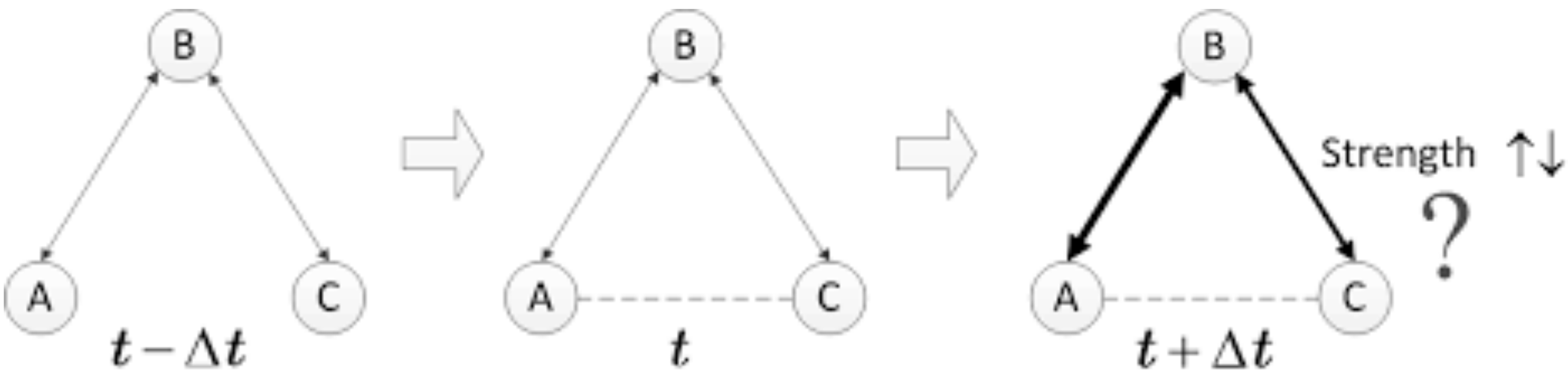
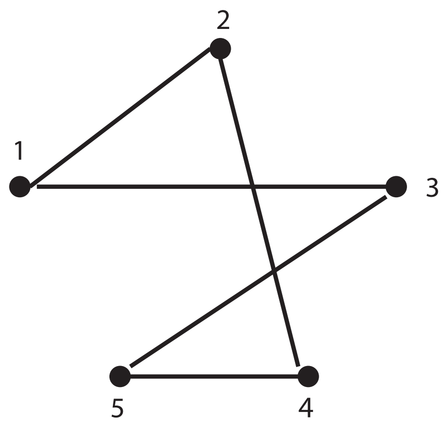
Closure, Clustering and Other Concepts from Sociology

Triadic Closure and Clustering Coefficient



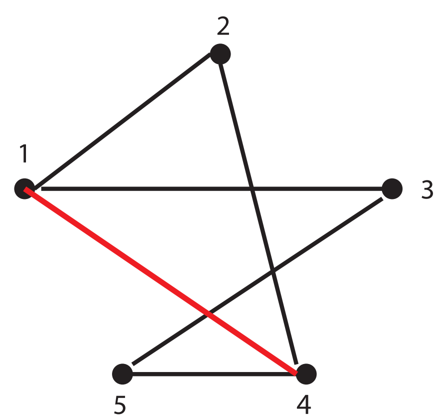
Triadic Closure

Many network concepts have come from sociology, studying interactions of groups of people. Ramifications here are huge for financial interests, the spread of information (and misinformation), and the spread of epidemics.



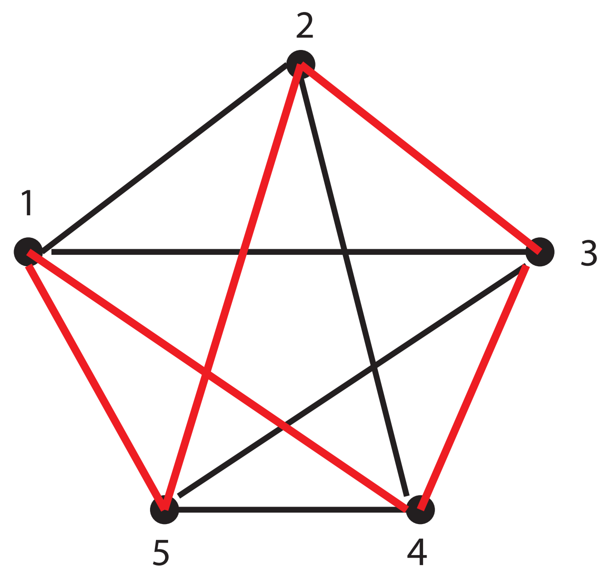
Consider this friendship graph. Person 1 and person 4 have a friend in common, person 2.

So it is likely that 1 and 4 will themselves become friends. This is an example of triadic closure.



Triadic Closure

As a result of this triadic closure, a triangle has been formed in the network. Additional triadic closure will result, over time, in more triangles. In this example, the end result is formation of a complete graph. This would be called a clique if it is a subgraph of a larger friendship network.

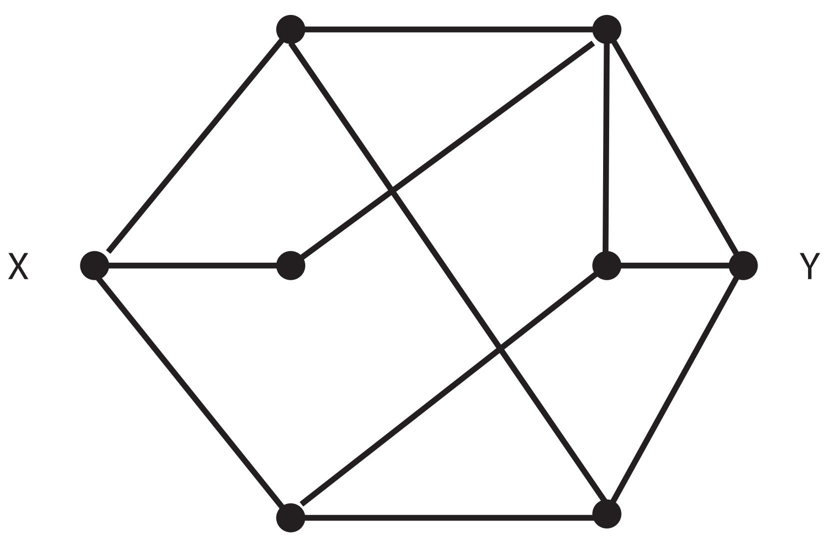


Networks in which nodes/edges are added/deleted over time are called dynamic networks. Triadic closure is a strong dynamic force in social networks.

Clustering Coefficient (C)

This is a measure that attempts to capture the degree of triadic closure in a network. The clustering coefficient of a node X is defined as the probability that two randomly selected friends of X are friends with each other. If there are q friends, then there are $\binom{q}{2}$ friend pairs.

That is, it is the fraction of pairs of X's friends that are connected to each other by edges.

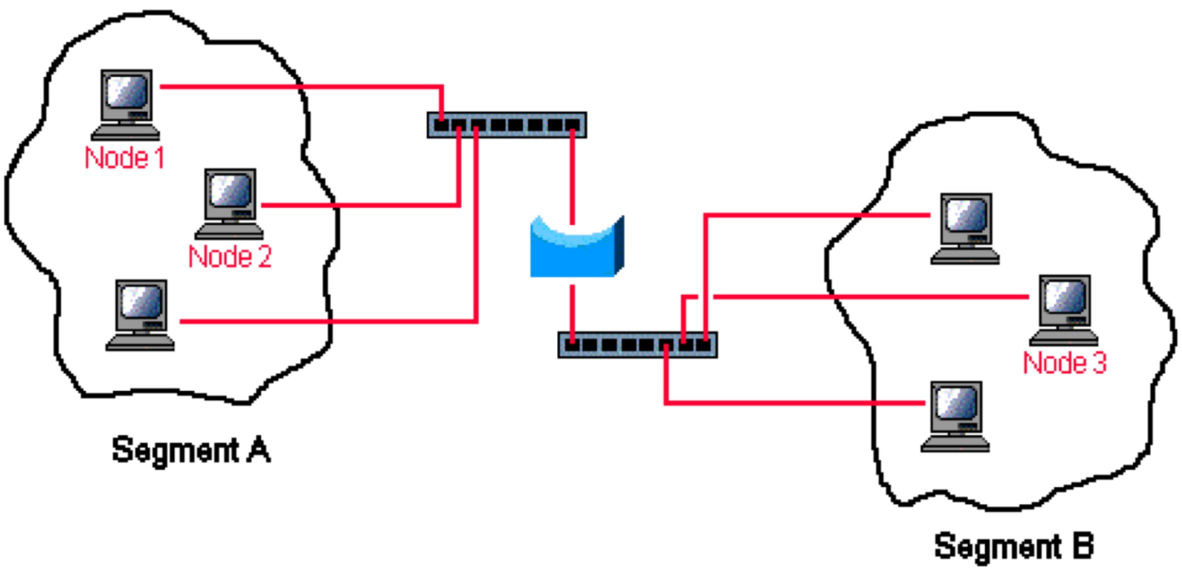


$$C_X = \frac{0}{3} = 0$$

$$C_Y = \frac{1}{3}$$

Nodes in social networks tend to have **high clustering coefficient**. This is due to the effects of triadic closure.

Bridges and Ties



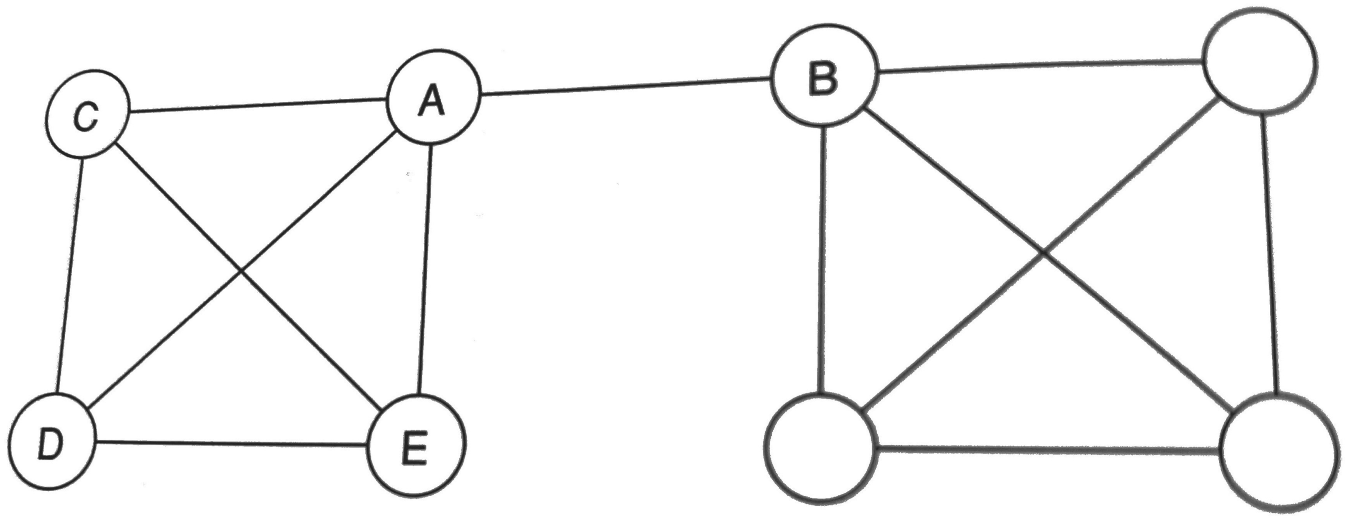
Bridges and Local Bridges

In a study by Mark Granovetter for his PhD, he conducted interviews with a large number of people and determined that many found jobs through acquaintances. That's not surprising. What is surprising is that this number was greater than the number of people who found jobs through close friends.

WHY?

Bridges Between Cliques

Easley and Kleinberg 2010

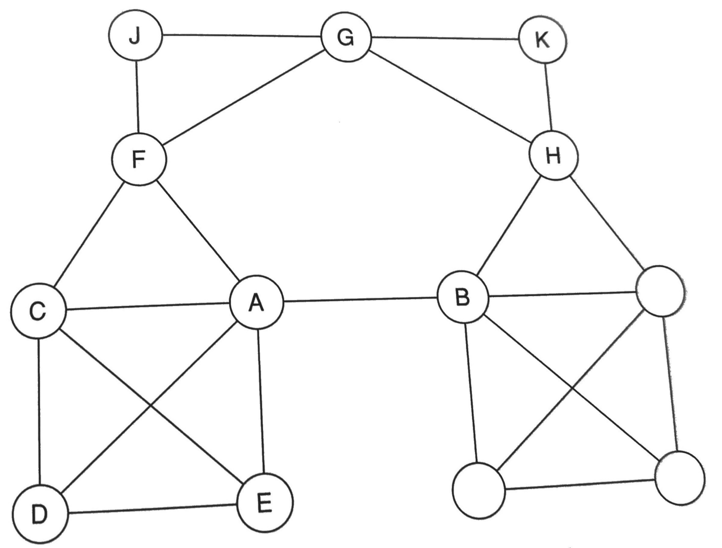


The edge joining A and B in this network is special. If it is removed, then A and B lie in two different components of the graph. Such an edge is called a **bridge**, since it bridges components of the graph.

In terms of a social network, a bridge would be a friendship that couples two different clusters of friends. Each such cluster is called a clique.

Local Bridges are More Common

Easley and
Kleinberg 2010



Bridges are very rare in social networks, and usually disappear when larger networks of individuals are considered. Nodes A and B might think that their friendship is all that couples their respective cliques of friends, but actually there are other connections they might not know about. We should therefore define a less restrictive type of bridge.

An edge joining two nodes A and B is a local bridge if its endpoints have no friends in common – in other words, deleting the edge would increase the distance between A and B to a value greater than 2. This distance is called the span of a local bridge between A and B.

Local Bridges and Triadic Closure

If the A-B edge is a local bridge, can it also be part of a triangle in the graph?

No, because if it is part of a triangle, then if removed, the distance between A and B would be 2 (not greater than 2).

Weak and Strong Ties

What's the difference between a friend and an acquaintance? A friend is someone with strong ties, while an acquaintance has weaker ties.

We can weight the edges as W=weak and S=Strong.

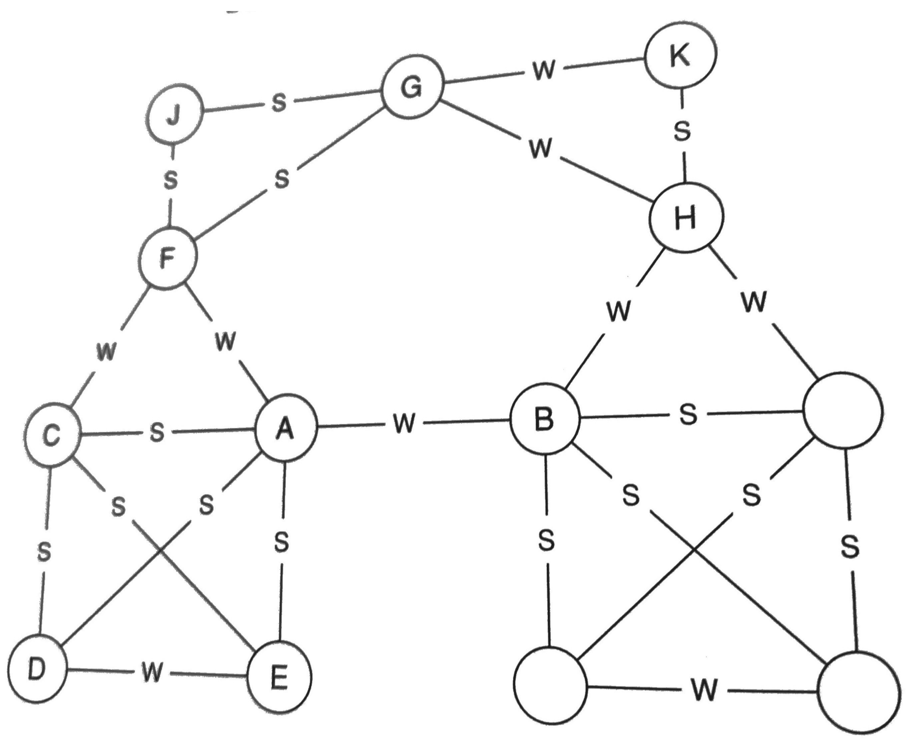
In the last graph, do you think the A-B edge would be weak or strong?

Probably weak, since A and B have no common friends.

In most cases, local bridges are weak ties.

Weak and Strong Ties

Easley and
Kleinberg 2010



Acquaintances are better at recommending jobs since they are in touch with an entirely different clique of people who might have heard something about the job. In contrast, your friends are likely to have heard the same things you have, so will provide no new information.

Generalizing the Notions of Tie Strength and Local Bridges

Rather than assigning edges W and S strengths, one could use a real number. This might correspond to the number of minutes of direct or cell phone conversation per month, for example.

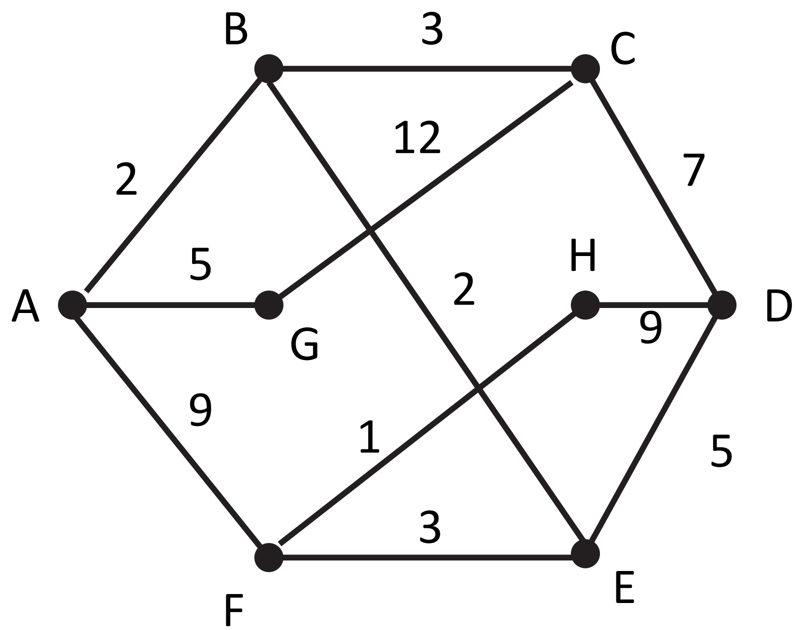
Since there are so few local bridges in most social networks, it makes sense to soften the definition to include “almost” local bridges. To this end, define the neighborhood overlap of an A-B edge as:

$$\boxed{\text{Overlap} = \frac{\text{number of nodes who are neighbors of both A and B}}{\text{number of nodes who are neighbors of at least one of A or B}}}$$

Note that this is the same definition we used earlier in graph partitioning, then called structural equivalence. Same thing, different context.

The neighborhood overlap is 0 when an edge is a true local bridge. If the overlap is near 0, then the edge is almost a local bridge.

Relationship Between Tie Strength and Neighborhood Overlap

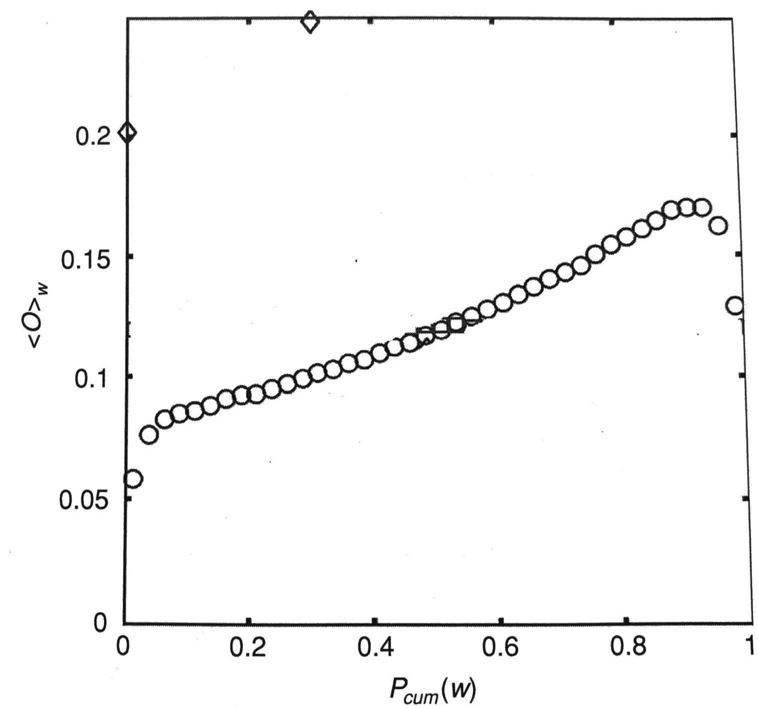


Sort by tie strength:

F-H
A-B
B-E
E-F
B-C
etcetera

Relationship Between Tie Strength and Neighborhood Overlap

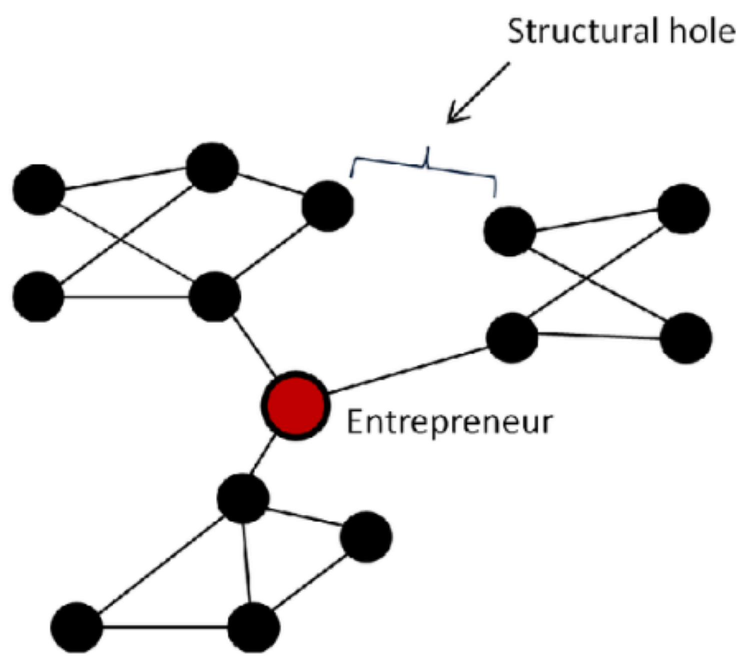
Easley and
Kleinberg 2010



Example of real social network using mobile phone data for tie strength.
X-axis: ordering of edges by tie strength
Y-axis: neighborhood overlap

Edges with stronger tie strength have more neighborhood overlap

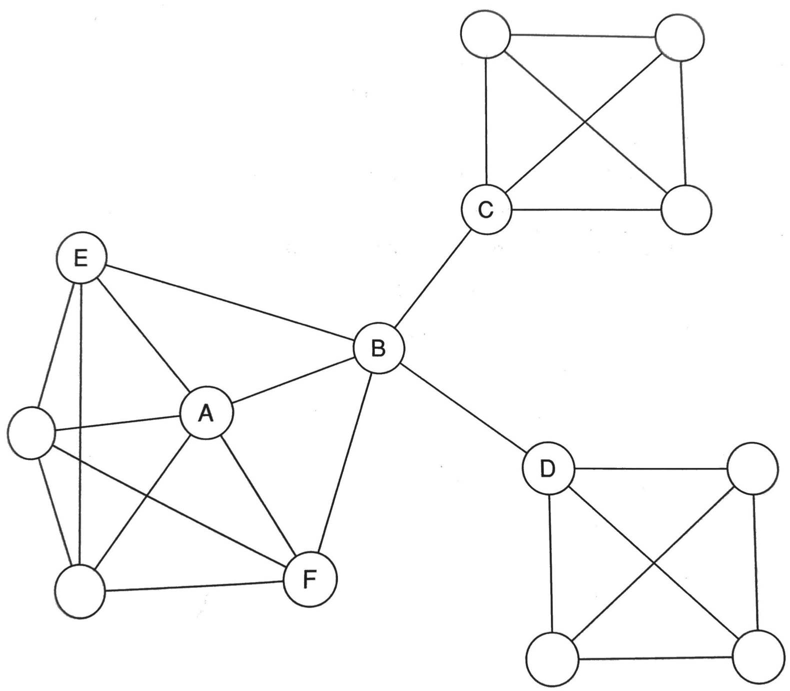
Embeddedness and Structural Holes



Embeddedness

The embeddedness of an edge is the number of common neighbors shared by the two endpoints. This is the numerator in the neighborhood overlap.

Easley and Kleinberg 2010

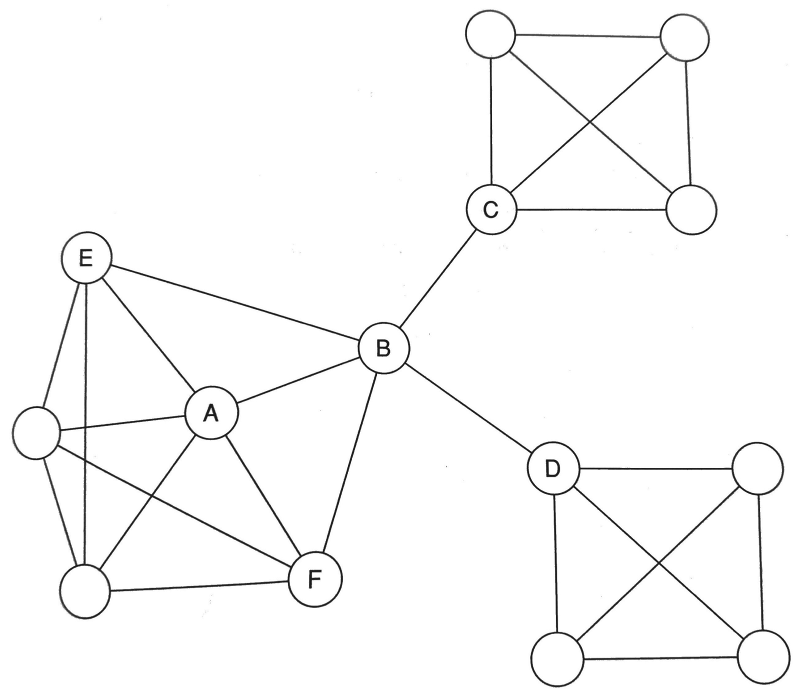


In this example, A-B has embeddedness 2, since A and B have two common neighbors, E and F. Edge B-C is a local bridge, with embeddedness of 0.

Embeddedness

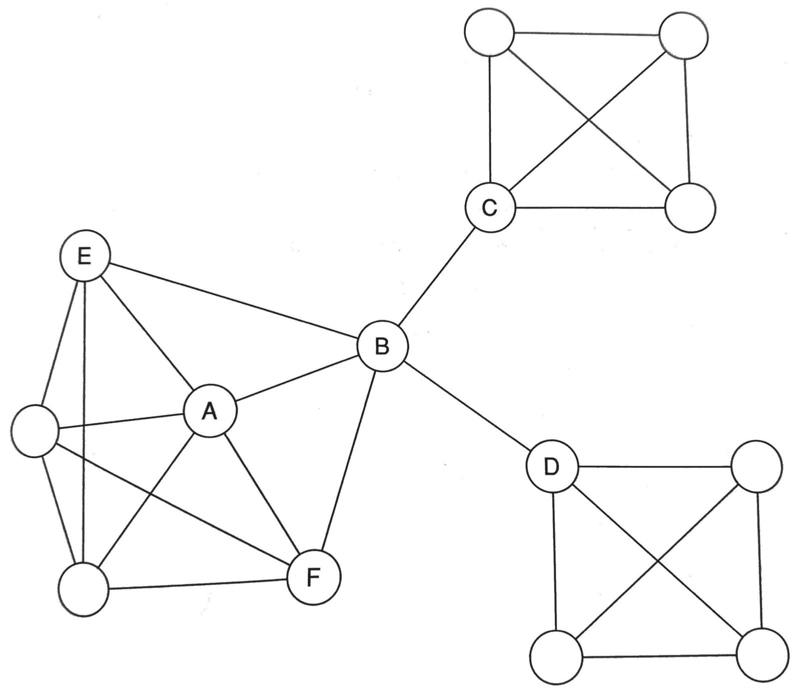
Node A is clearly part of a clique. This is quantified by the fact that each of its edges has high embeddedness. This is in stark contrast to node B, whose edges have lower (or 0) embeddedness.

Easley and Kleinberg 2010



A great deal of sociological research shows that if two individuals are connected by an embedded edge, then it is easier for them to trust each other and any potential transactions they may have together.

Structural Holes

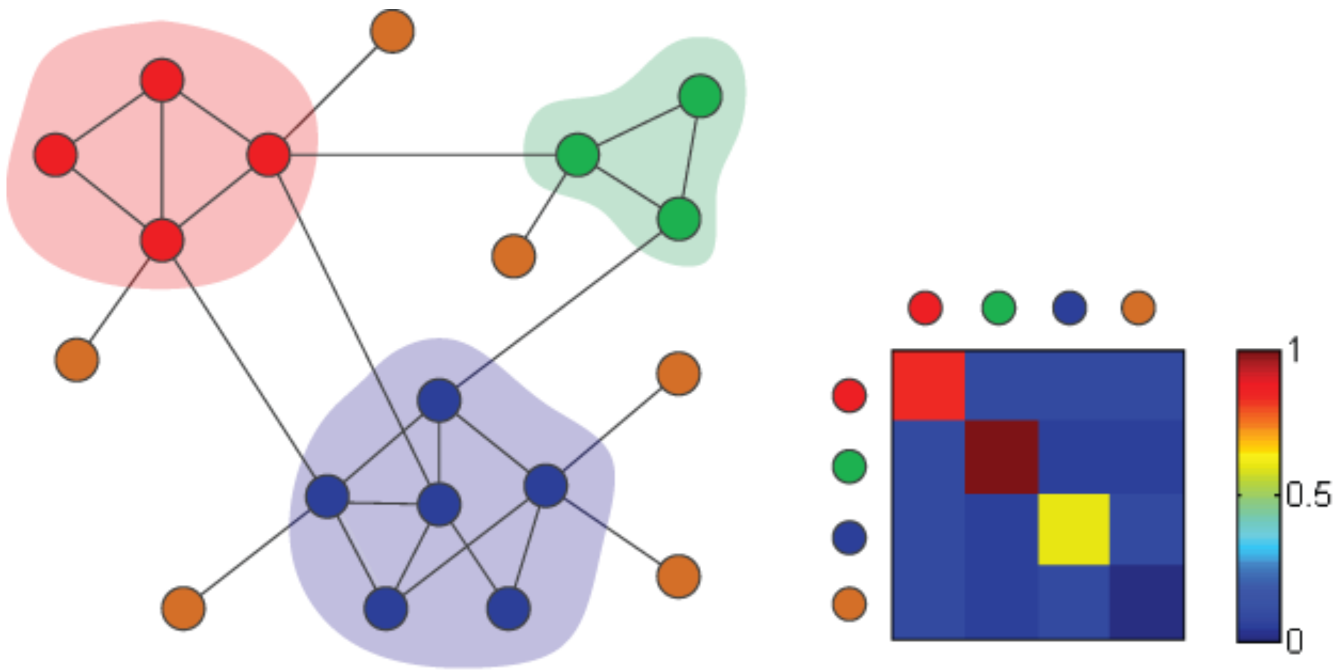


Easley and Kleinberg 2010

There are disconnections or empty spaces between the three clusters of nodes. These are examples of structural holes. Node B fills those holes. Why might that be advantageous for this individual?

B has access to disparate information that can be combined in novel ways. B is also the gatekeeper between the different groups.

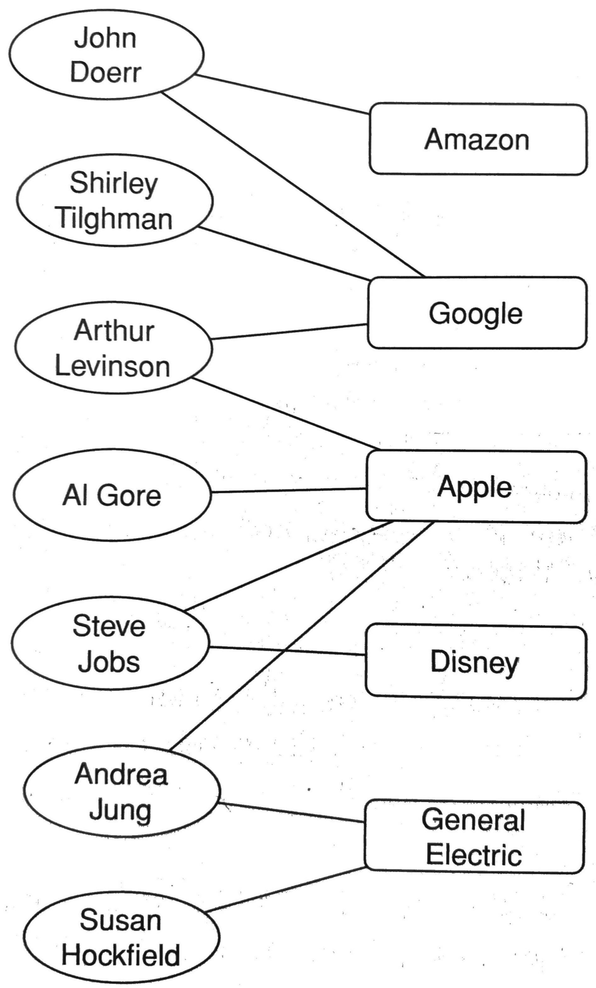
Link Formation in a Social Network



(a) (b)

Closure is Influenced by Affiliation

Easley and Kleinberg 2010

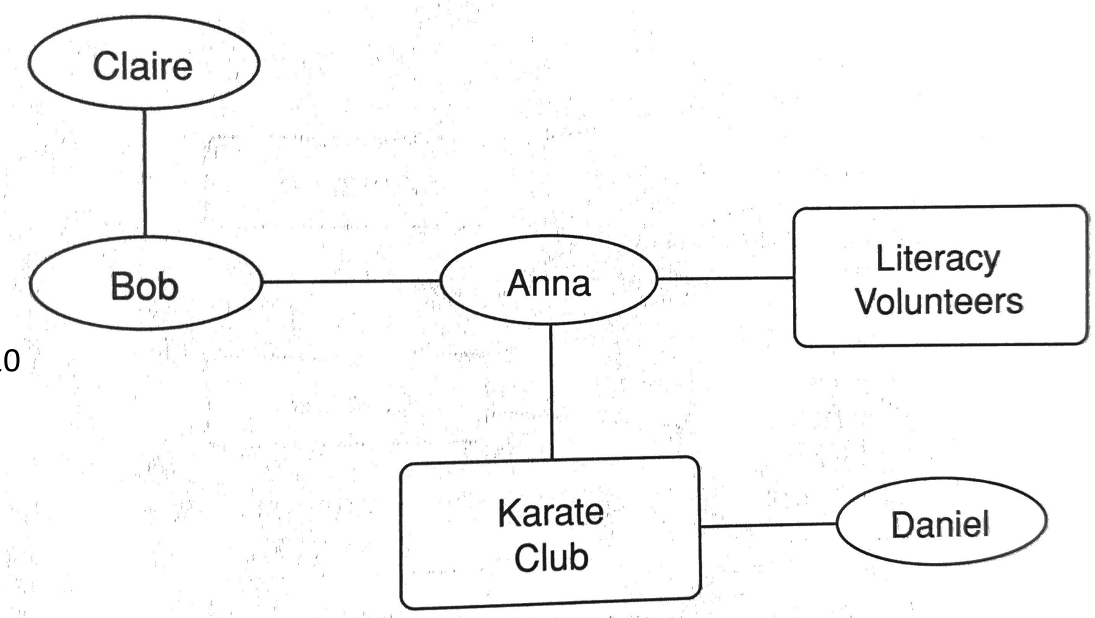


People on the left were on the board of directors of companies on the right.

What type of network is this? It is a bipartite affiliation network.

Having common affiliations will promote interactions of the "actors"

Closure is Influenced by Affiliation

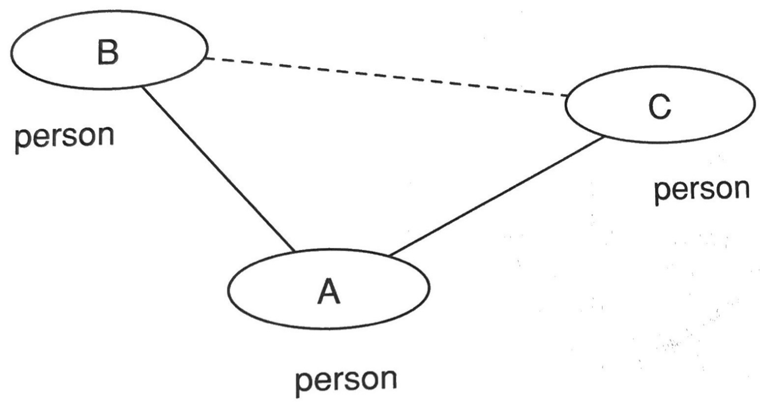


Easley and Kleinberg 2010

Simple example of a social-affiliation network. There are now edges between actors, as well as between actors and "foci" (something that brings people together)

Closure is Influenced by Affiliation

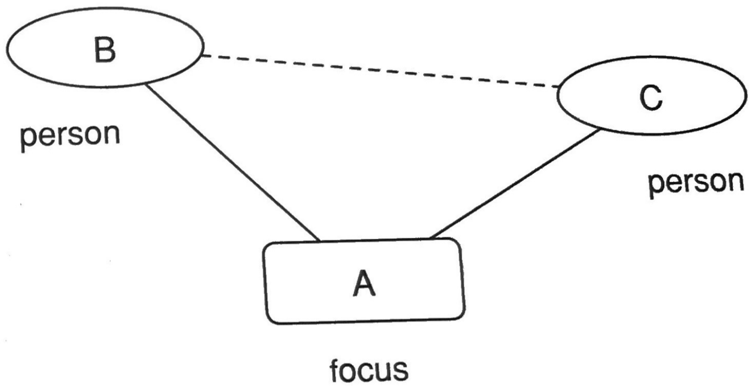
Easley and
Kleinberg 2010



Triadic closure

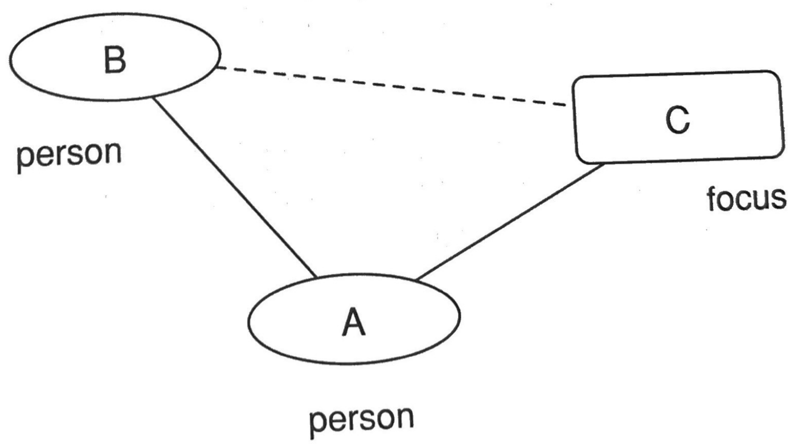
Closure is Influenced by Affiliation

Easley and
Kleinberg 2010



Focal closure

Closure is Influenced by Affiliation

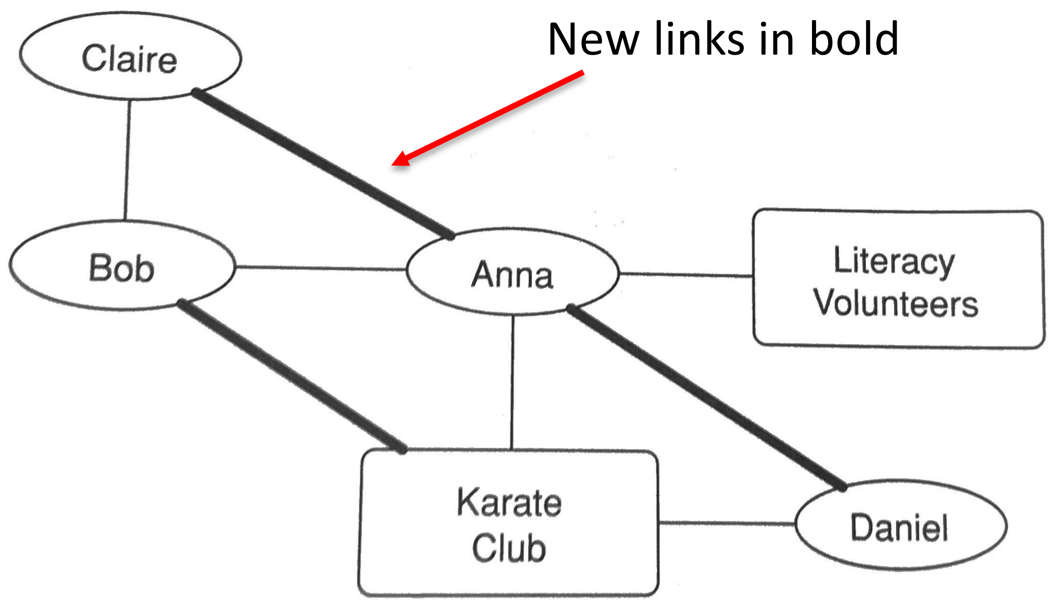


Easley and
Kleinberg 2010

Membership closure

Closure is Influenced by Affiliation

Easley and Kleinberg 2010



In real life, all of these connections can form social links.

Testing for Closure in a Real Social Network

The rationale for links in real social networks is rarely documented, and many people have trouble even remembering how they formed links with their friends and acquaintances. This type of information is stored for online interactions!

Strategy: Using online data,

(1) Take two snapshots of the network at different times

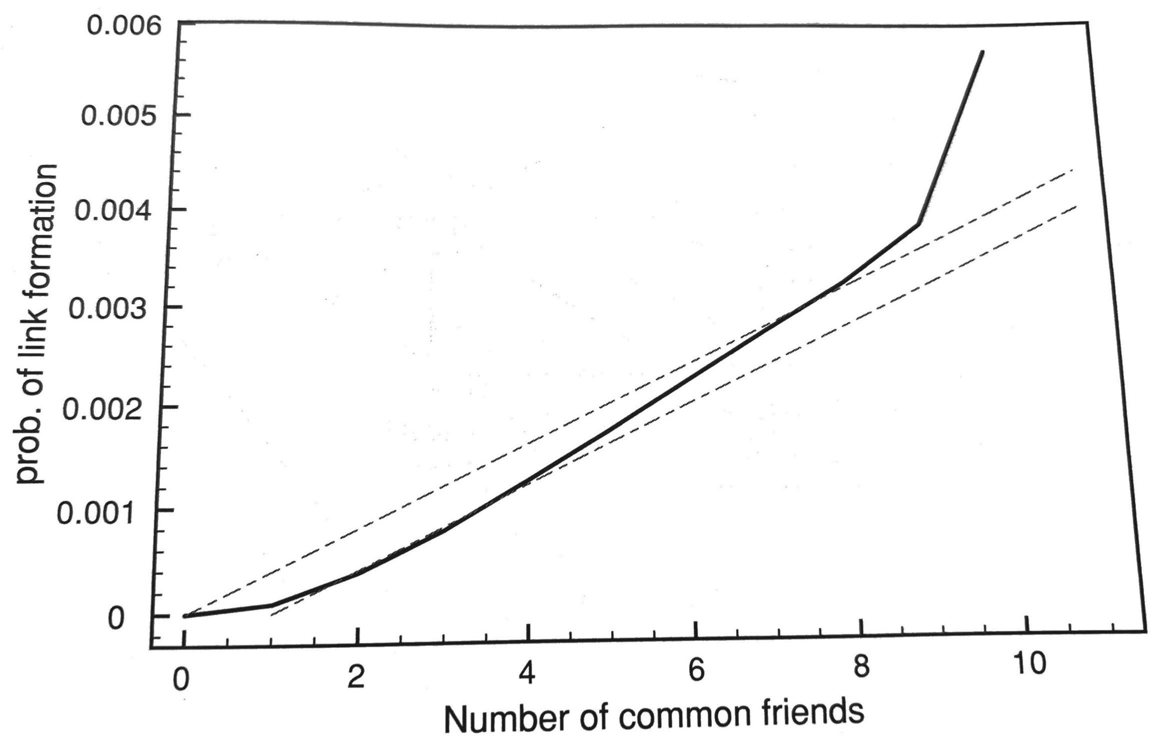
(2) For each k, identify all pairs of nodes who have exactly k friends in common in the first snapshot, but who are not directly connected

(3) Define $T(k)$ to be the fraction of these pairs that have formed an edge by the time of the second snapshot. This approximates the probability that a link will form between two people with k common friends

(4) Plot $T(k)$ as a function of k to illustrate the effect of common friends on the formation of links

Triadic Closure with E-mail Data

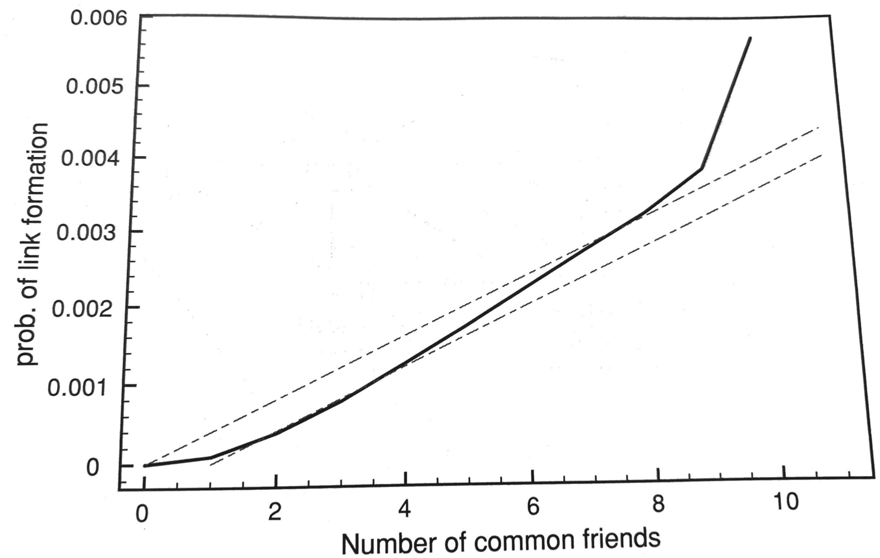
A study by Kossinets and Watts did this, using email communications among 22,000 students at a U.S. university over a one-year period. Two people were considered friends if one sent an email to the other during the preceding 60 days. They took multiple snapshots at different days throughout the year and averaged them to compute *T(k)*. The result is below.



Science, 311, (2006)

Solid curve is actual data

Triadic Closure with E-mail Data



Science, 311, (2006)

Let *p*=Prob[two people with a common friend forms a link]

Assume that each common friend gives this probability, independent of any other common friends.

Then probability of having *k* common friends and not forming a link is $(1-p)^k$

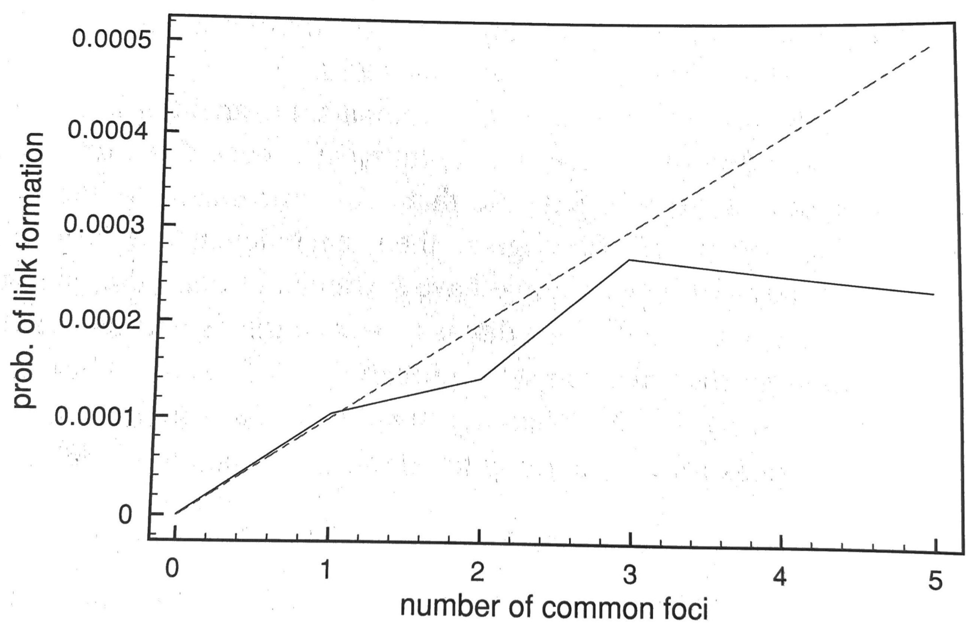
The probability that a link does form is then $T_{ind} = 1-(1-p)^k$. This is plotted as the upper dashed curve in the figure above.

The actual increase in link formation is greater than that from the model, so the probability of link formation with two common friends is more than expected from independent effects of the two common friends. There is a synergy.

Focal Closure with E-mail Data

In the study by Kossinets and Watts, they examined focal closure using classes as foci for the students (two students share a focus if they are in the same class).

Now *T(k)* is the probability they form a link (become email friends) if they have *k* classes together. Result is below.

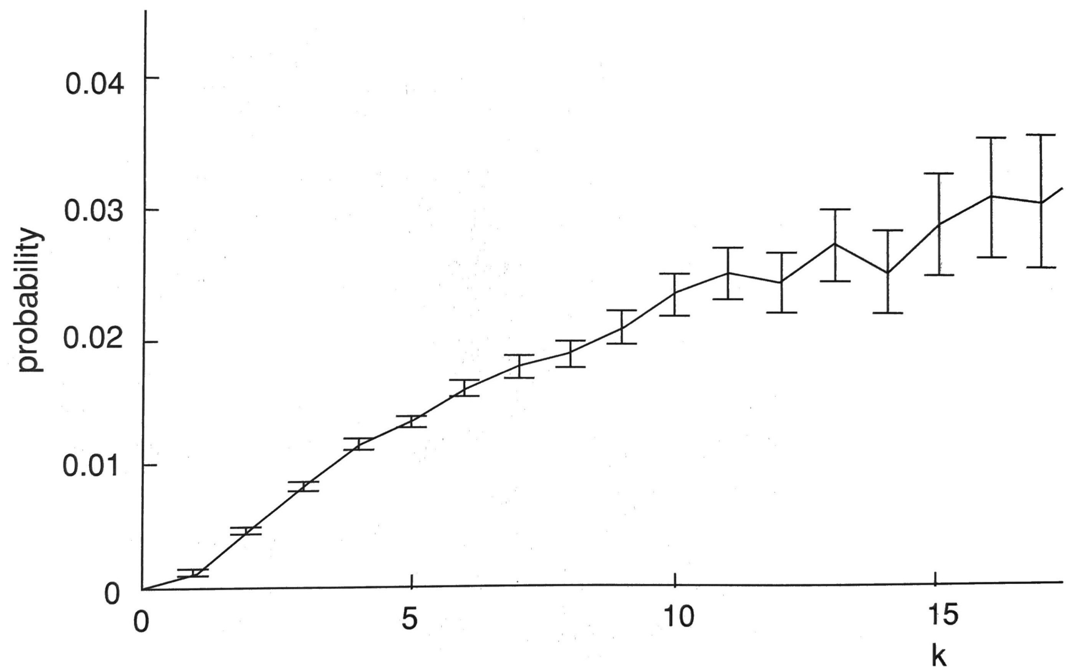


Science, 311, (2006)

While focal closure is evident for the first few values of *k*, it drops below what would be expected from the model that assumes independence. That is, there is a “diminishing returns” effect.

Membership Closure with Wikipedia Data

In a study on membership closure, a Wikipedia article is defined as a focus. A person is associated with that focus if he/she has edited it. If someone is friends with (i.e., has e-mail communication with) *k* editors of a Wikipedia article, what is the probability (*T(k)*) that that person becomes an editor of the article?



It is clear that membership closure occurs here, even for large *k*.

The End