Granovetter’s principle, homophily

Network science (I606)

Outline

- Strong and weak ties
- Triadic closure
- Measuring homophily
- Selection vs social influence
- Degree assortativity

Strong and weak ties

Mark Granovetter, *The strength of weak ties*, American Journal of Sociology 78 (6), 1360-1380 (1973)

Goal: bridging local and global

Questions:
1) How does information flow over a social network?
2) Which roles do the nodes play in this process?
3) How does the (social) network evolve over time?

Background

Granovetter interviewed people who recently changed jobs and asked them how they discovered their new job

Findings:
1) The job was found through personal contacts
2) The contacts were “acquaintances” instead of “friends”

Question: how come?

Guess: acquaintances may have access to different sources of information than friends

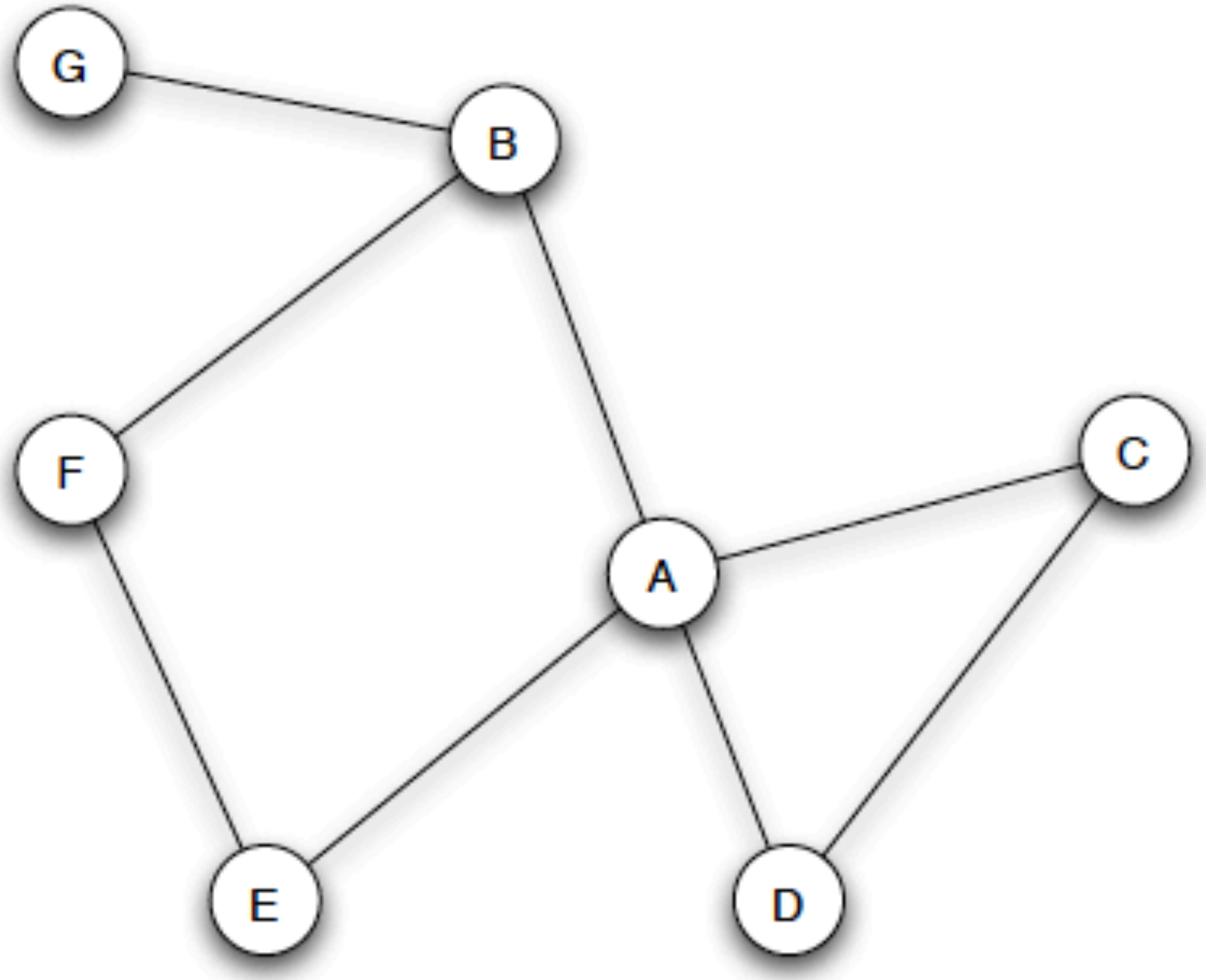
Triadic closure

Question: how are links formed?

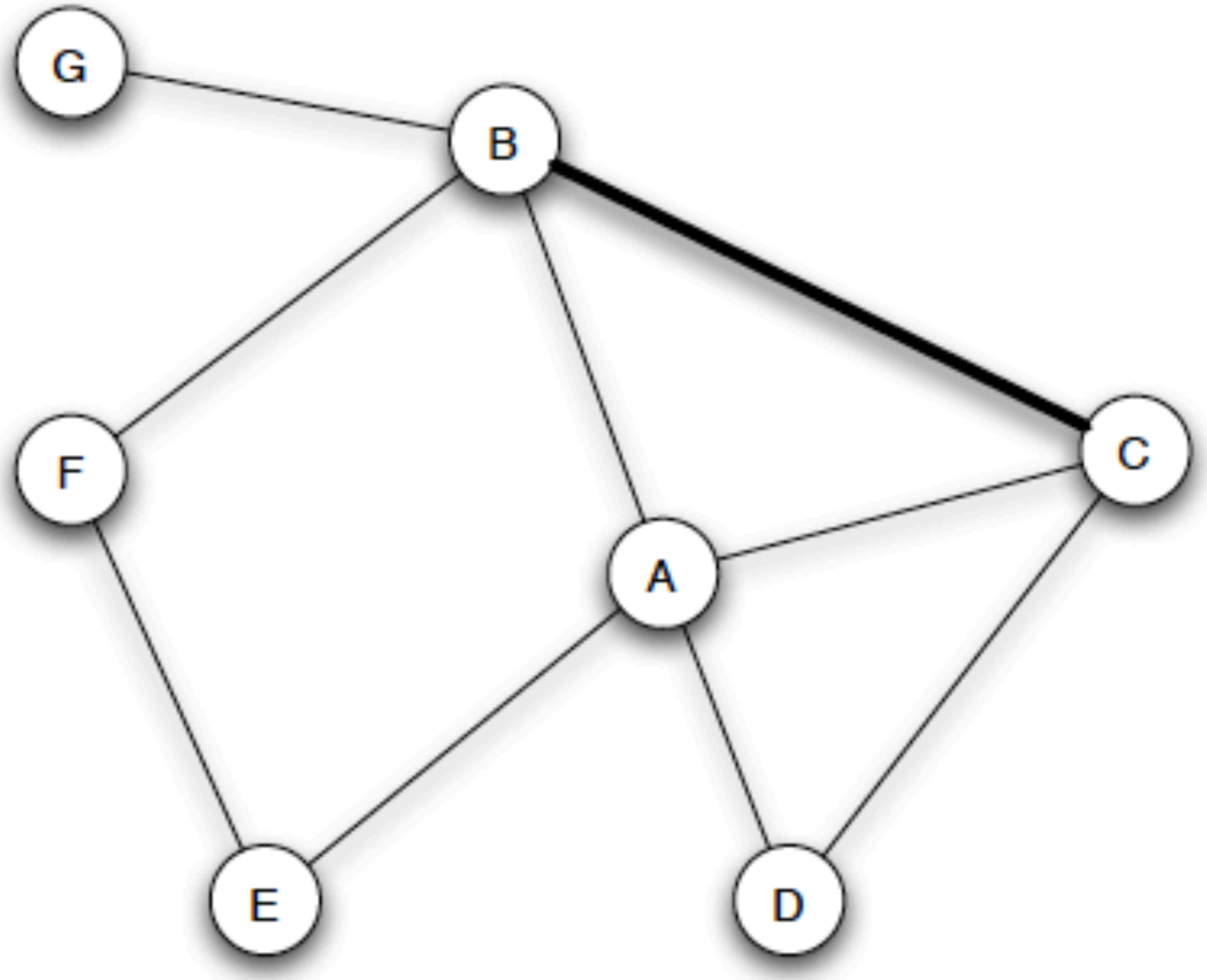
Principle of triadic closure: *If two people in a social network have a friend in common, then there is an increased likelihood that they will become friends themselves at some point in the future* [Rapoport 1953]

Anatole Rapoport, *Spread of information through a population with socio-structural bias I: Assumption of transitivity*, Bulletin of Mathematical Biophysics 15 (6), 523-533 (1953)

Triadic closure



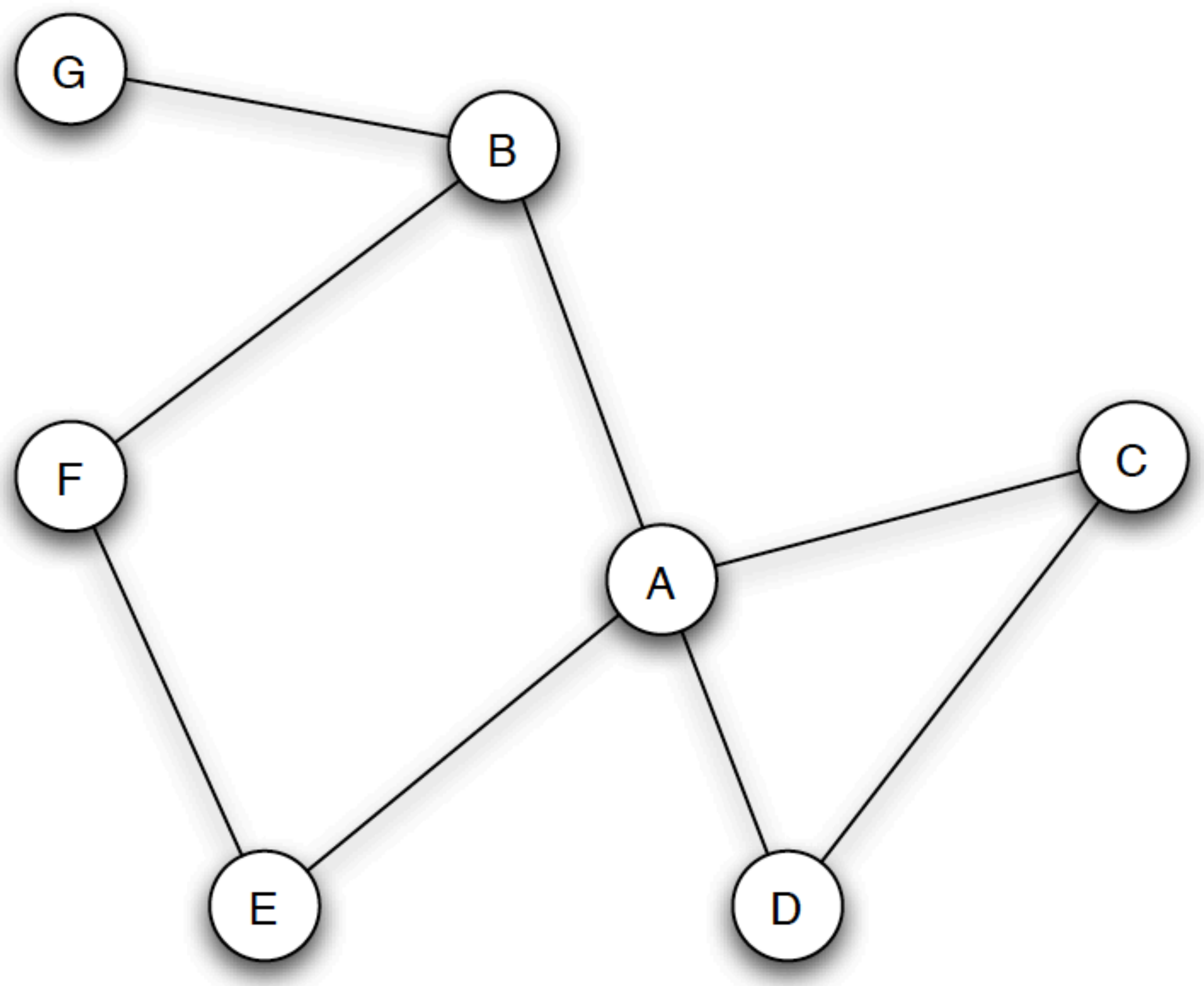
(a) *Before B-C edge forms.*



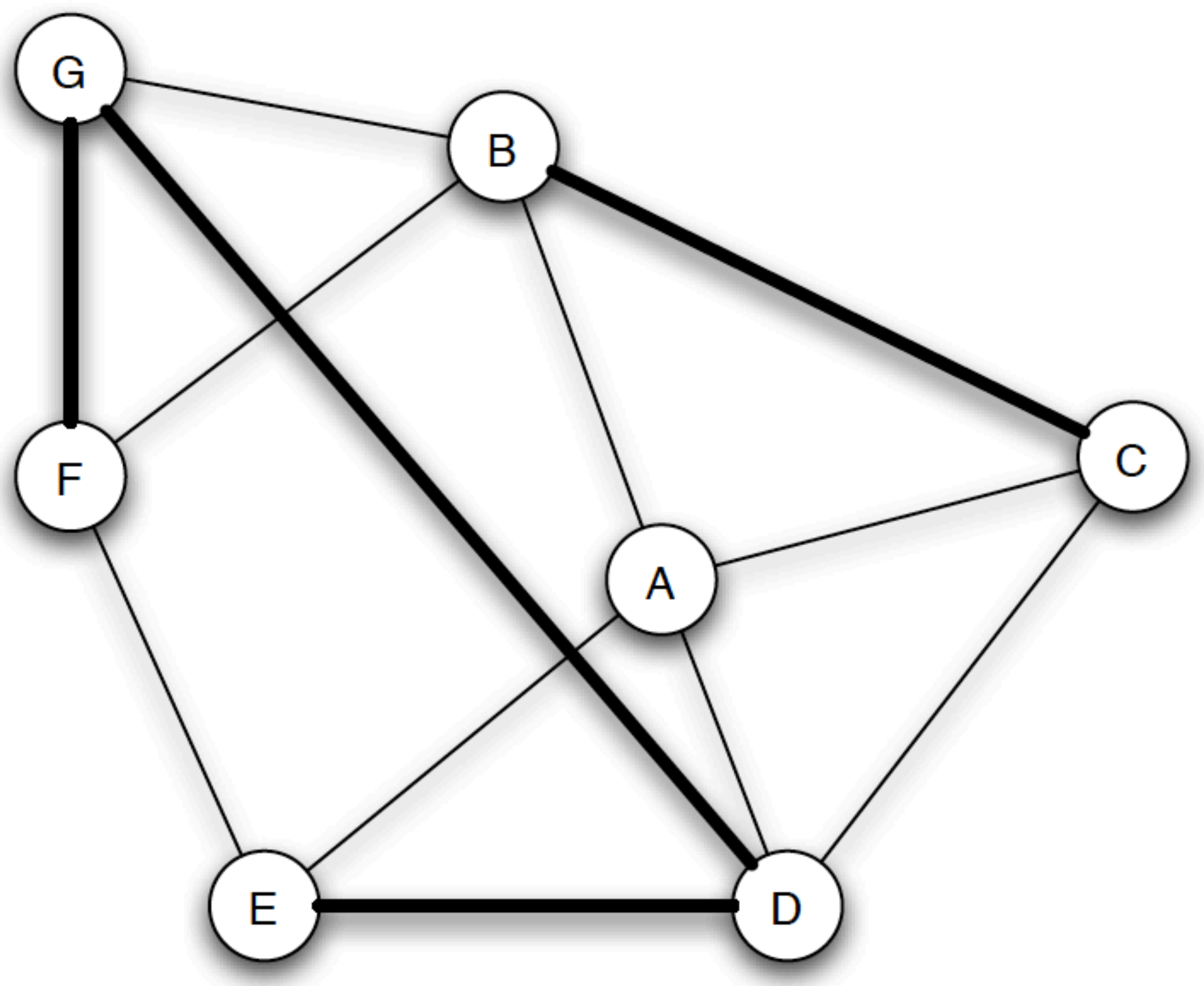
(b) *After B-C edge forms.*

Anatole Rapoport, *Spread of information through a population with socio-structural bias I: Assumption of transitivity*, Bulletin of Mathematical Biophysics 15 (6), 523-533 (1953)

Triadic closure



(a) *Before new edges form.*



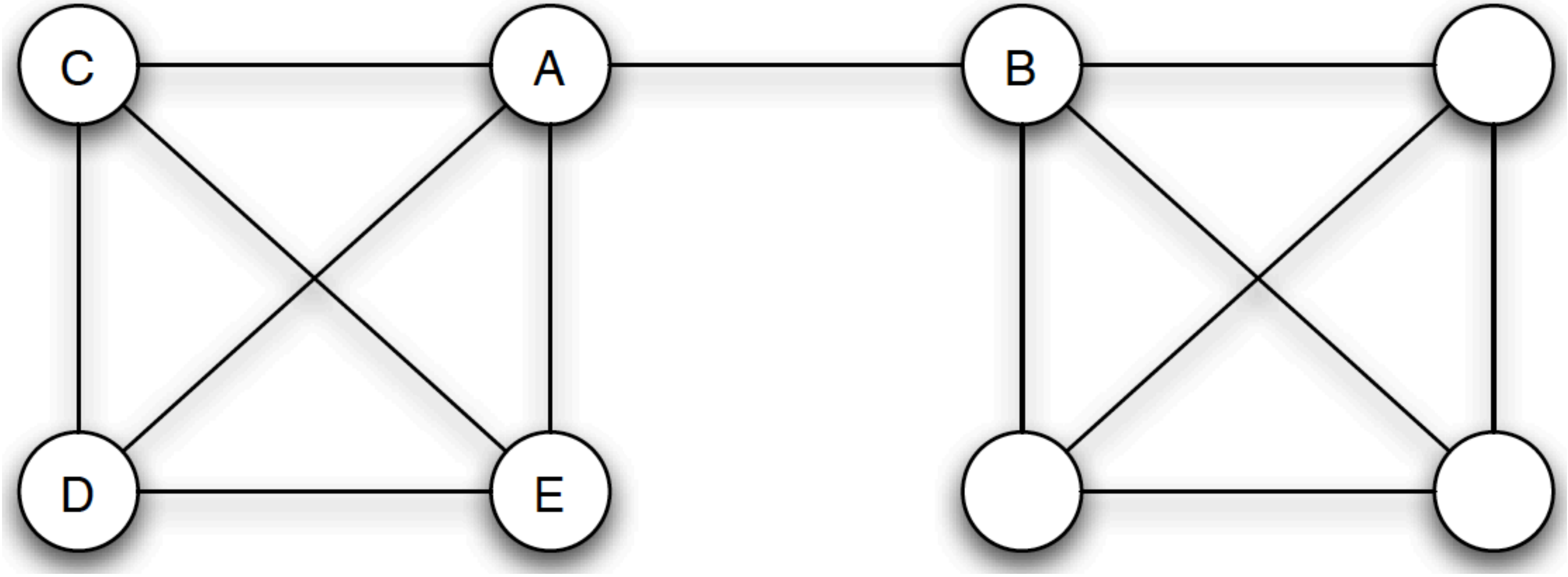
(b) *After new edges form.*

Triadic closure

Reasons:

1) Opportunity: if B and C are both friends of A they have higher chances to meet and get acquainted than if they had no common friends
2) Trust: if B and C are both friends of A they can become friends more easily than if they had no common acquaintances because they can trust each other through A
3) Incentive: if B and C are both friends of A but they do not become friends this may become a stress factor in the life of the group

Bridges



A link is a **bridge** if its removal leads to a breakup of the network in two parts

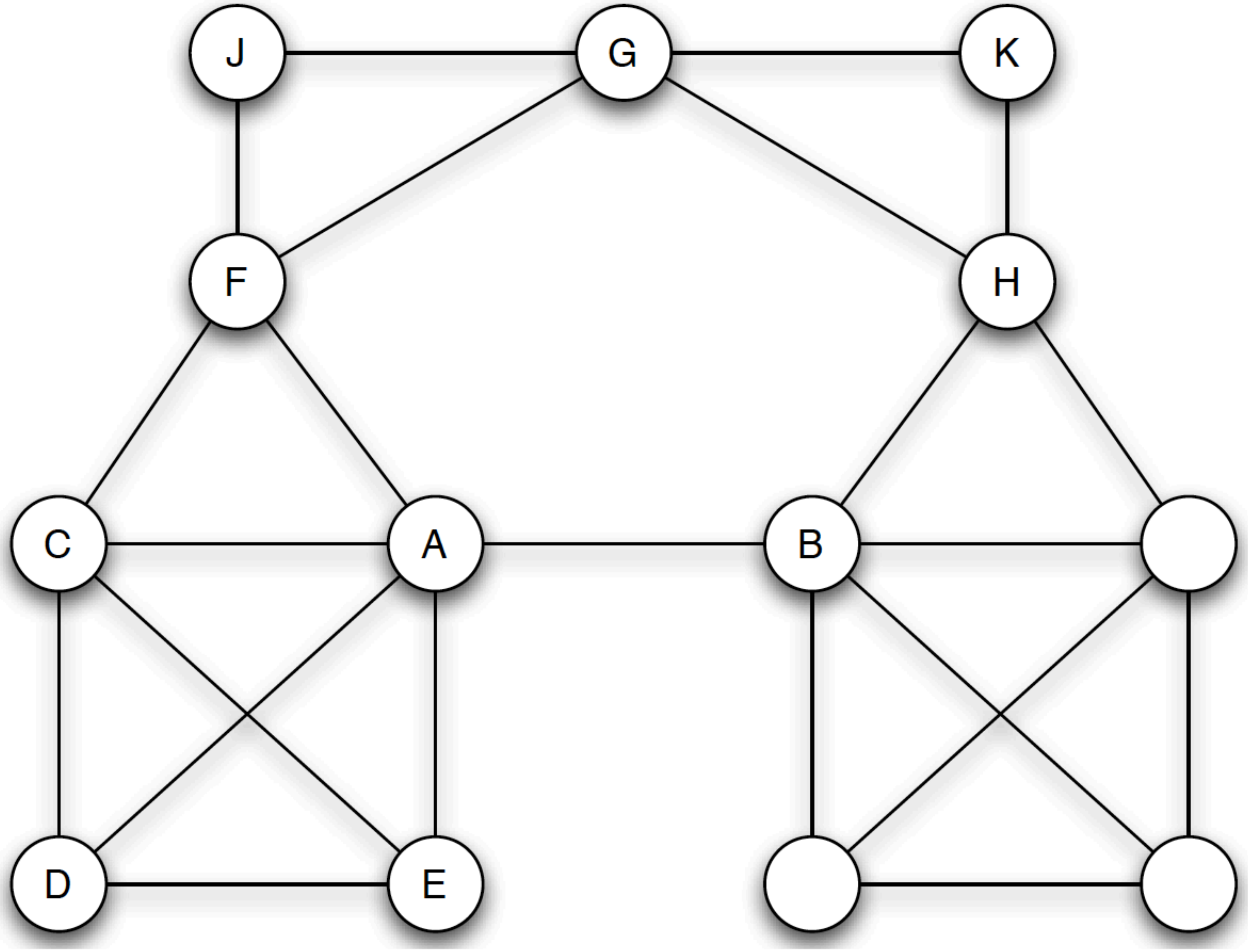
The link AB gives to A and B the opportunity to be in touch with people outside their closest circles

Local bridges

Issue: bridges are rare in social networks

Local bridge: link between two nodes having no friends in common

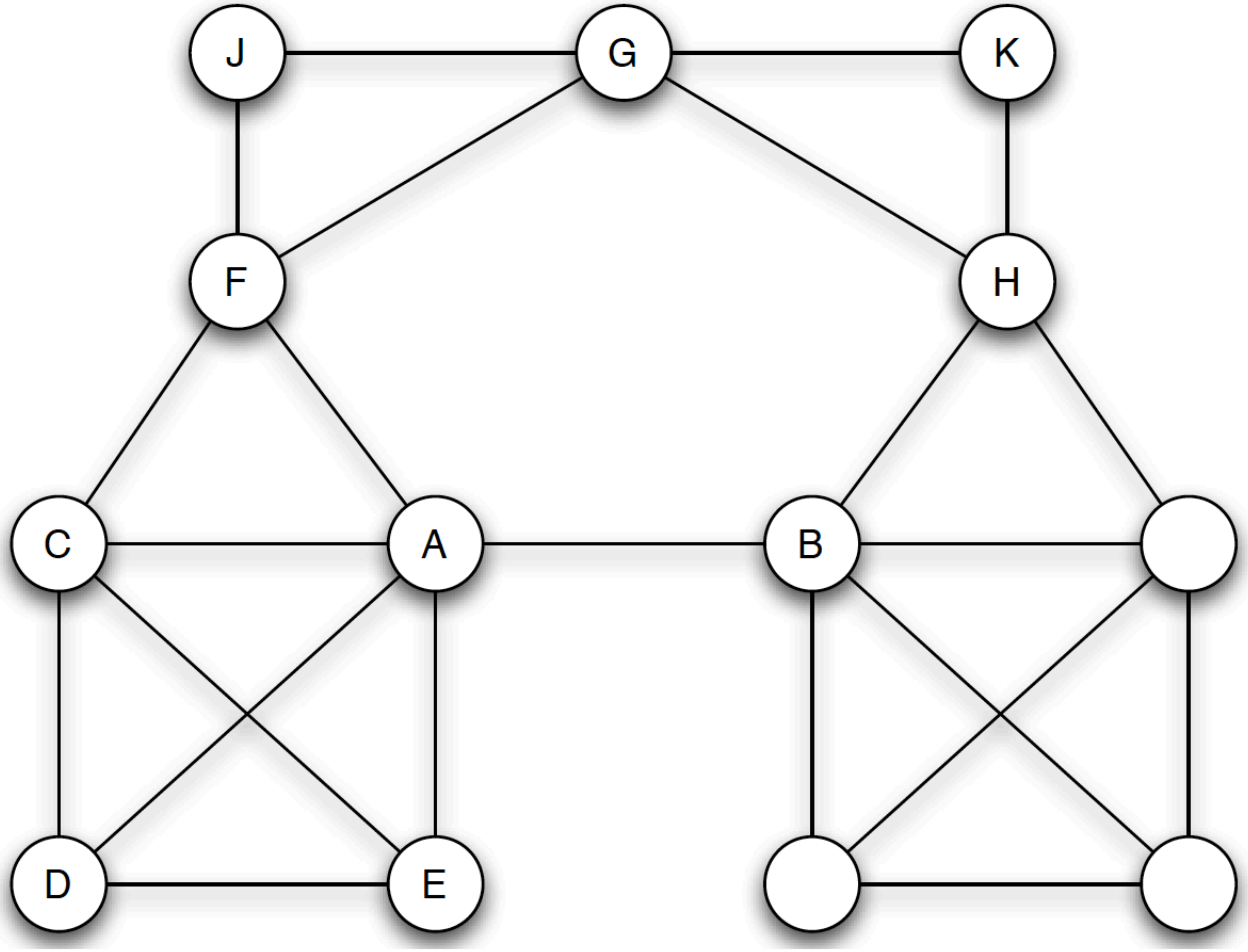
If A and B are joined by local bridge, its removal sets A and B more than two steps apart from each other



Local bridges

Local bridges are not sides of triangles, by definition

Local bridges connect areas of the network that would be otherwise distant



Strength of ties

Links (ties) usually are valued, to indicate the strength (importance) of the interaction

Two values:

1) Strong (S), for friendships
2) Weak (W), for acquaintances

Important: links can be created and destroyed and their strengths can vary in time!

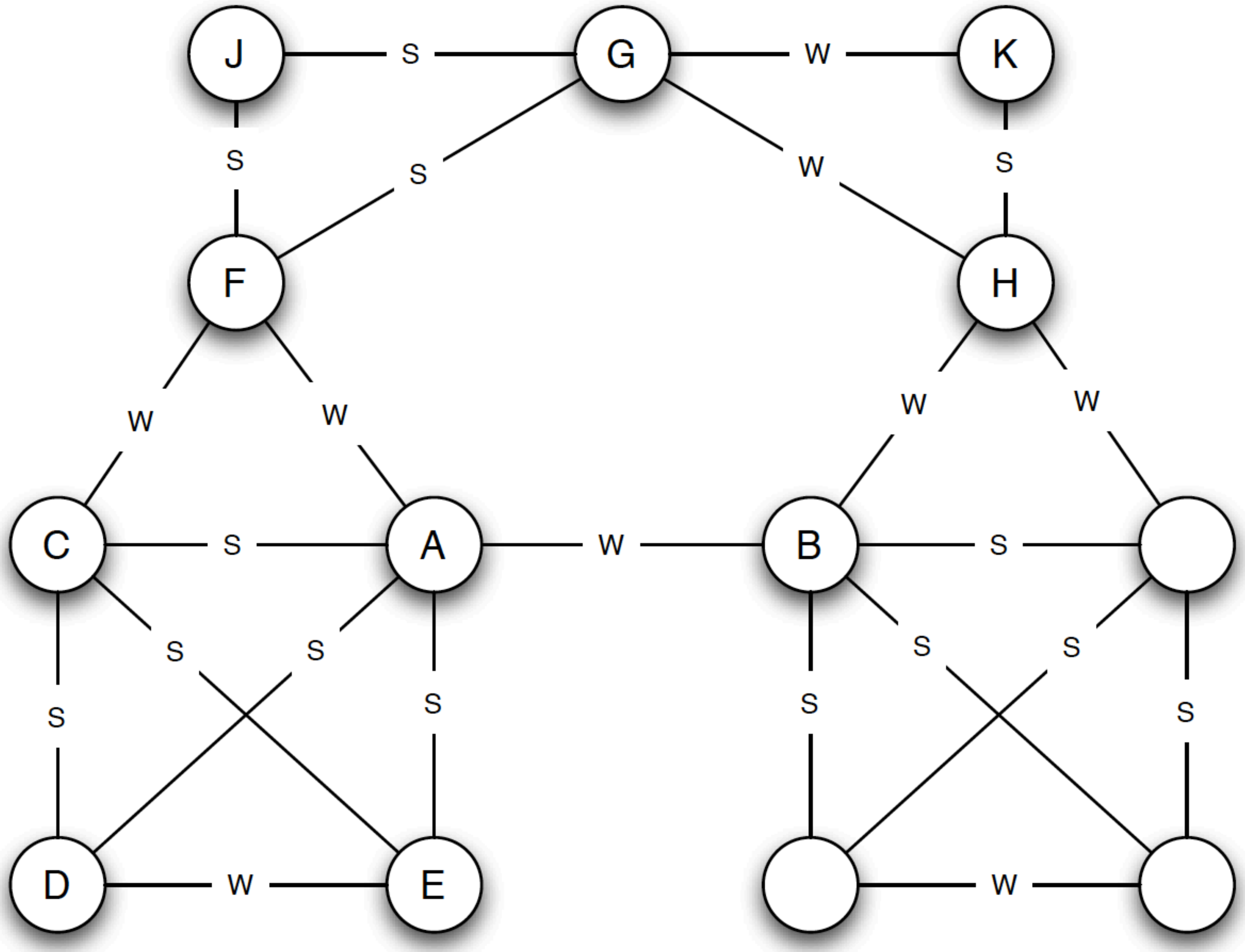
Expectation: links are often created to close triads

Strength of ties vs triadic closure

Assumption: *if a node A has links to nodes B and C, then the B-C link is especially likely to form if A's edges to B and C are both strong ties*

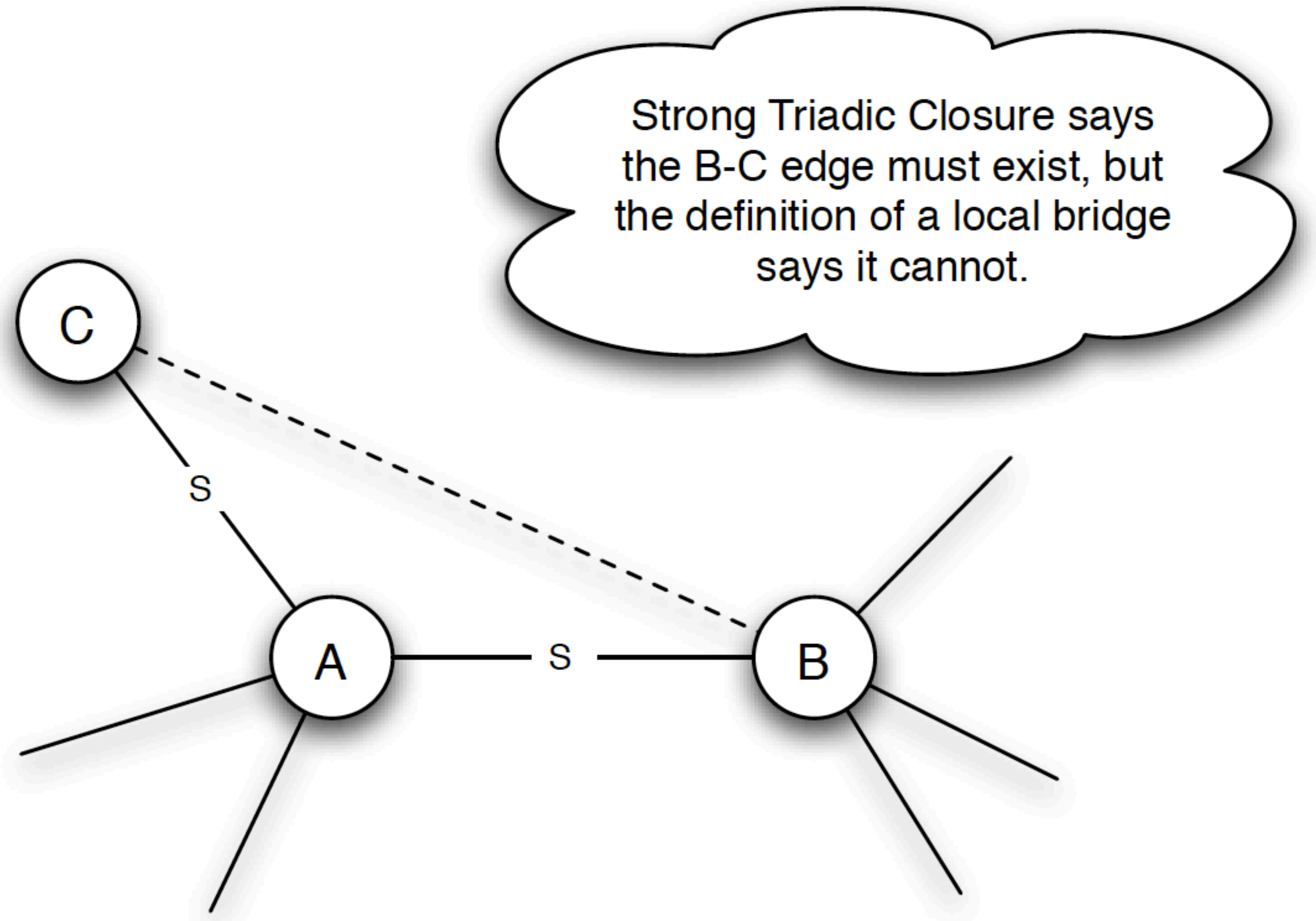
Granovetter's principle: *We say that a node A violates the Strong Triadic Closure Property if it has strong ties to two other nodes B and C, and there is no link at all between B and C. We say that a node A satisfies the Strong Triadic Closure Property if it does not violate it*

Strength of ties vs triadic closure



Local bridges vs weak ties

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*



Local bridges vs weak ties

Question: is it really so?

Now it is possible to validate these hypotheses on real data!

[Onnela 2007]: analysis of mobile phone communication network

Nodes: cell phone users

Links: reciprocal calls

J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, A.-L. Barabási, *Structure and tie strengths in mobile communication networks*, PNAS 104 (6), 7332-7336 (2007)

Weighing the ties

How can we measure the strength of ties?

Option: the weight of the link between callers A and B is the total duration (in minutes) of all calls between A and B

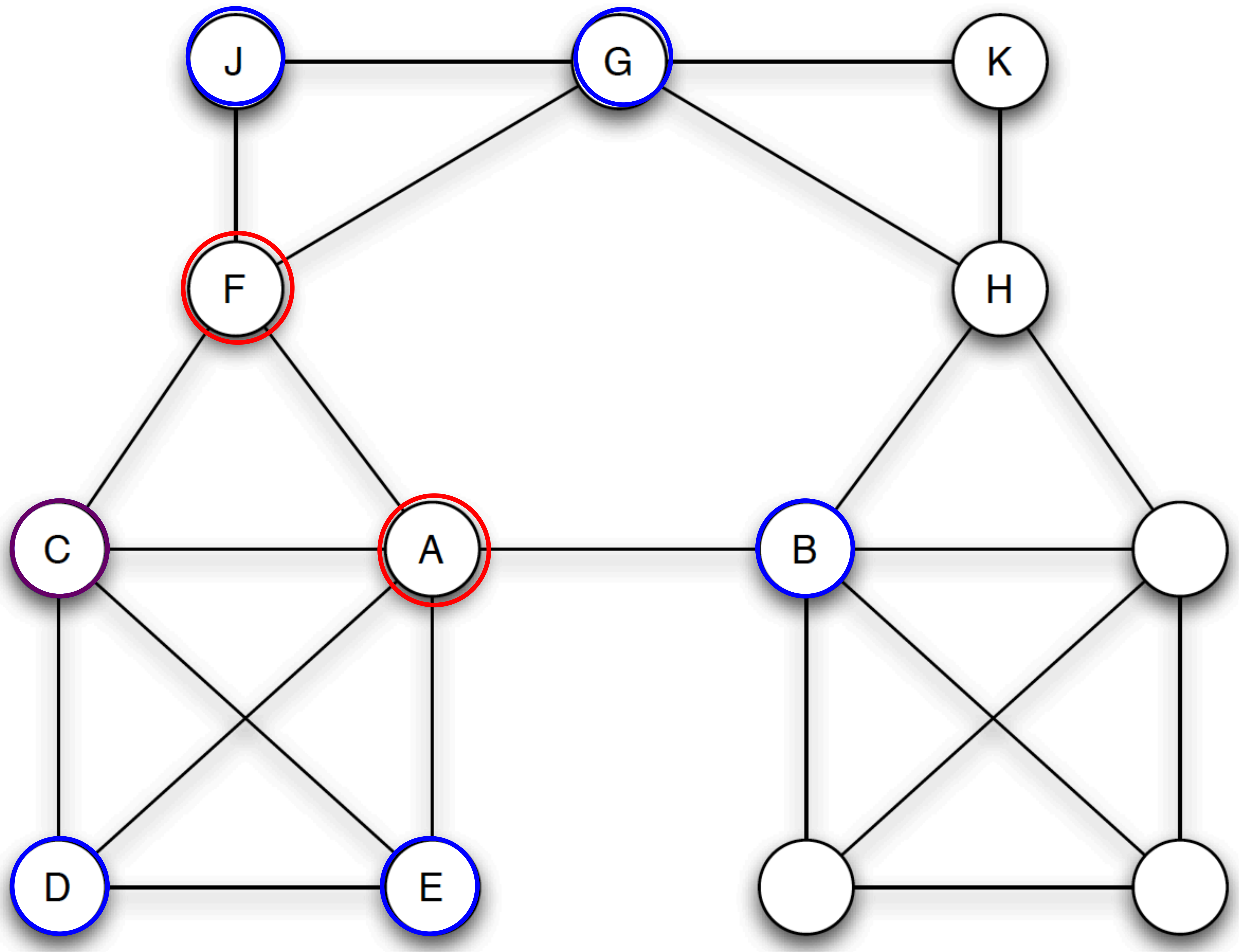
For a realistic definition of local bridges Onnela et al. defined the **neighborhood overlap**:

$$O(AB) = \frac{\#\ \text{shared neighbors of A and B}}{\#\ \text{neighbors of at least one of A or B}}$$

J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, A.-L. Barabási, *Structure and tie strengths in mobile communication networks*, PNAS 104 (6), 7332-7336 (2007)

Neighborhood overlap

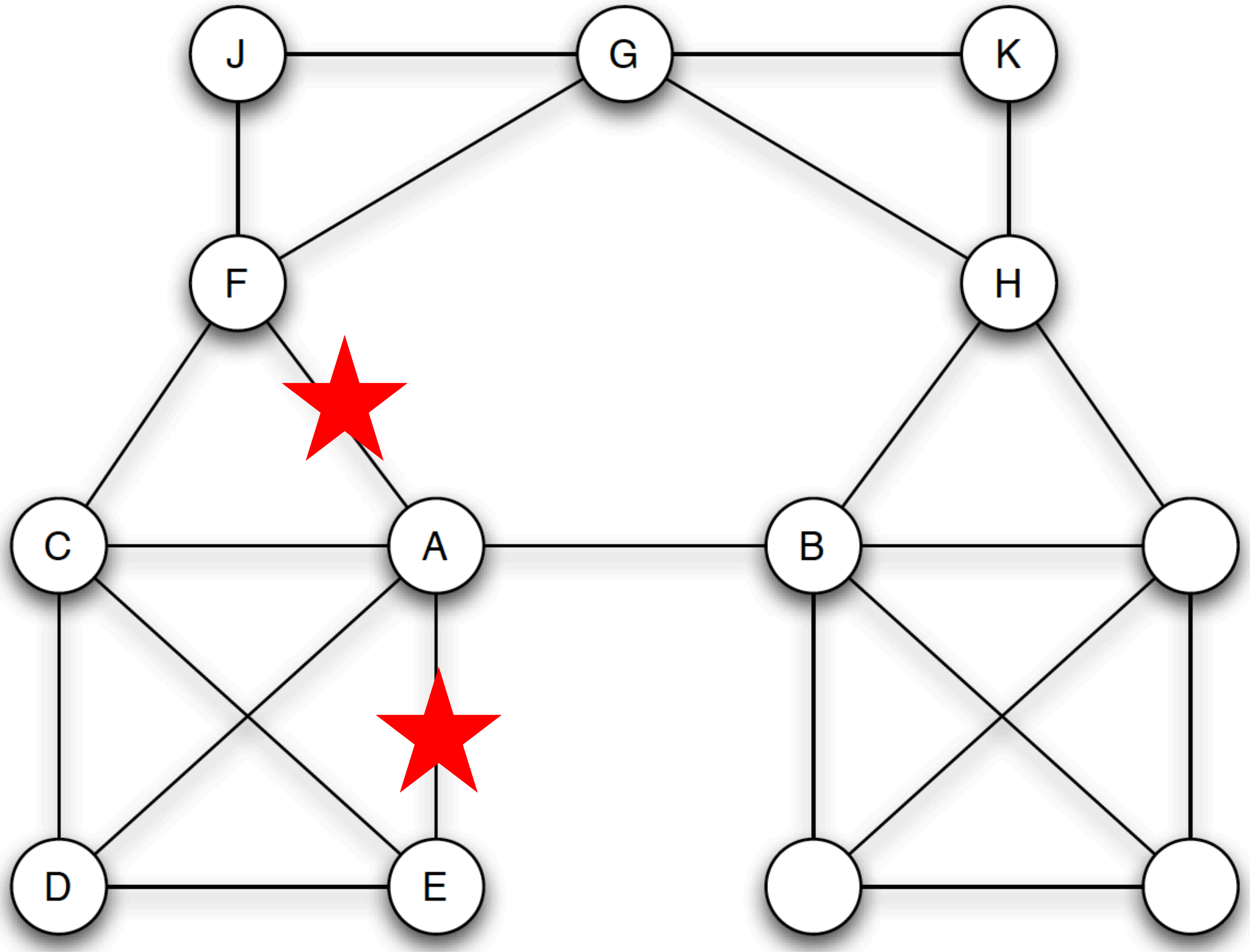
$$O(AB) = \frac{\#\ \text{shared neighbors of A and B}}{\#\ \text{neighbors of at least one of A or B}}$$



Neighborhood overlap

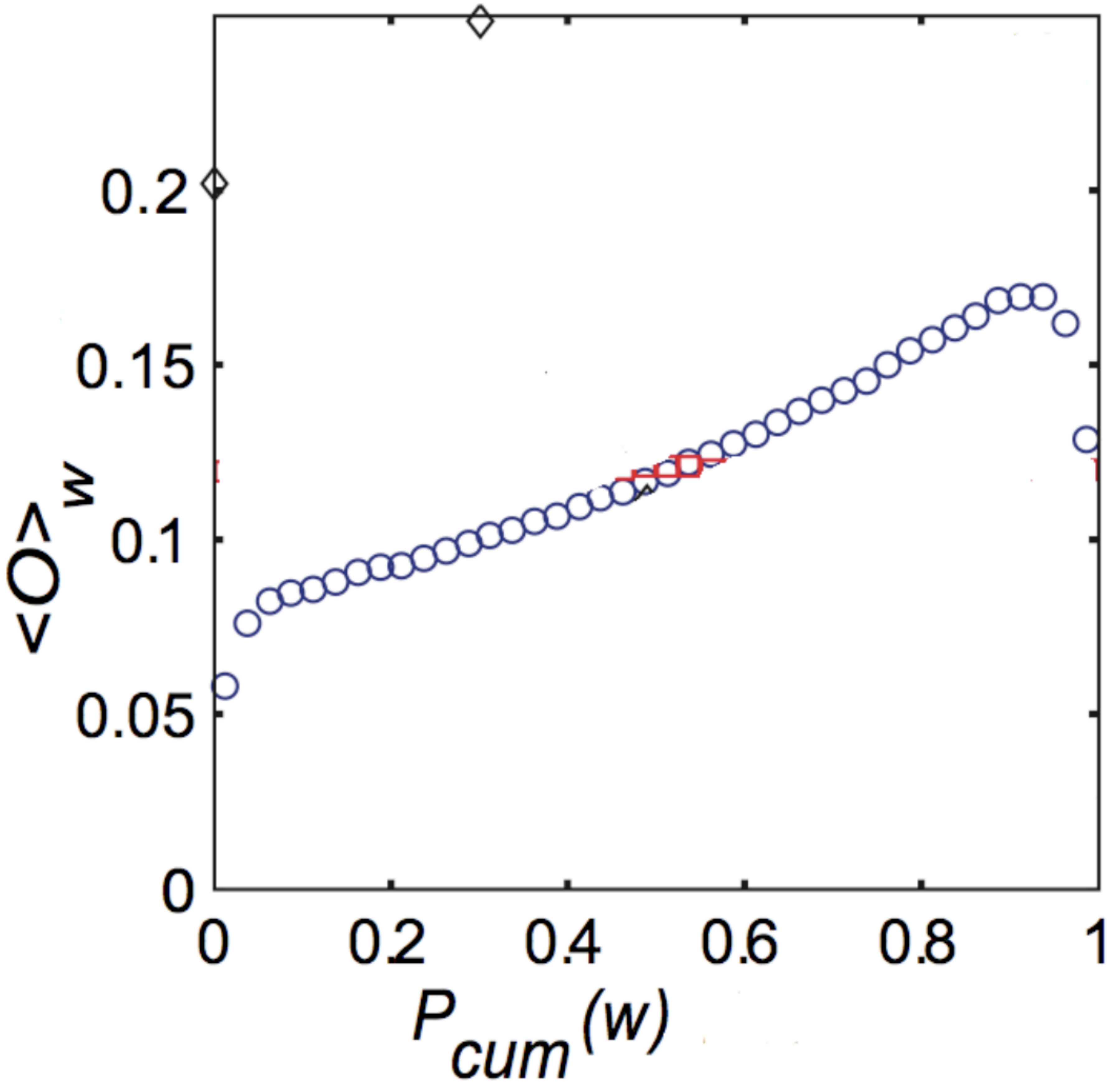
Special case: if O(AB)=0 -> AB is a local bridge

If O(AB) > 0 but low, AB is almost a local bridge



Tests of Granovetter’s hypotheses I

Question: do ``quasi’’ bridges also carry low weight?

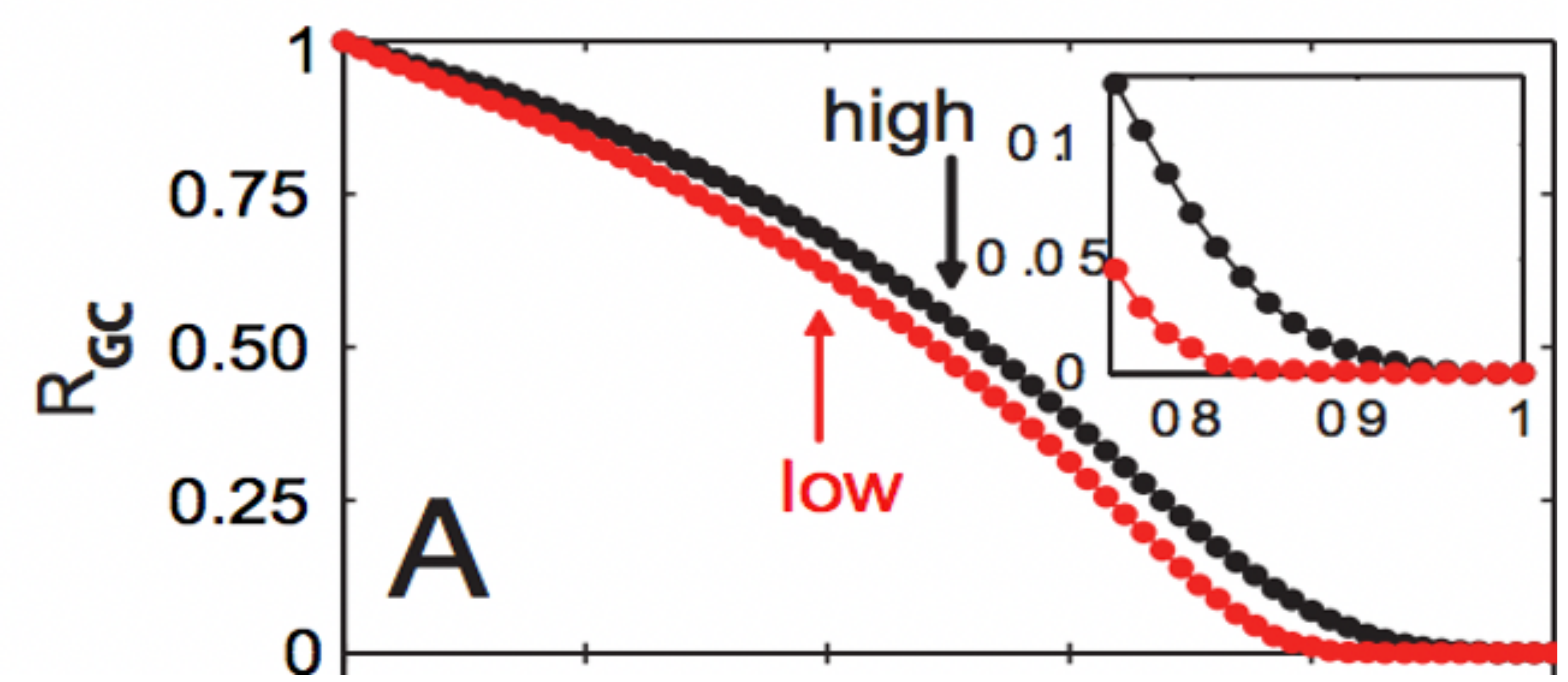


Looks good!

Tests of Granovetter’s hypotheses II

Question: do ``quasi’’ bridges connect different communities?

Tests by Onnela et al.: By removing the links from the weakest to the strongest the network is fragmented more easily -> weak ties act more as *connectors* than strong ties



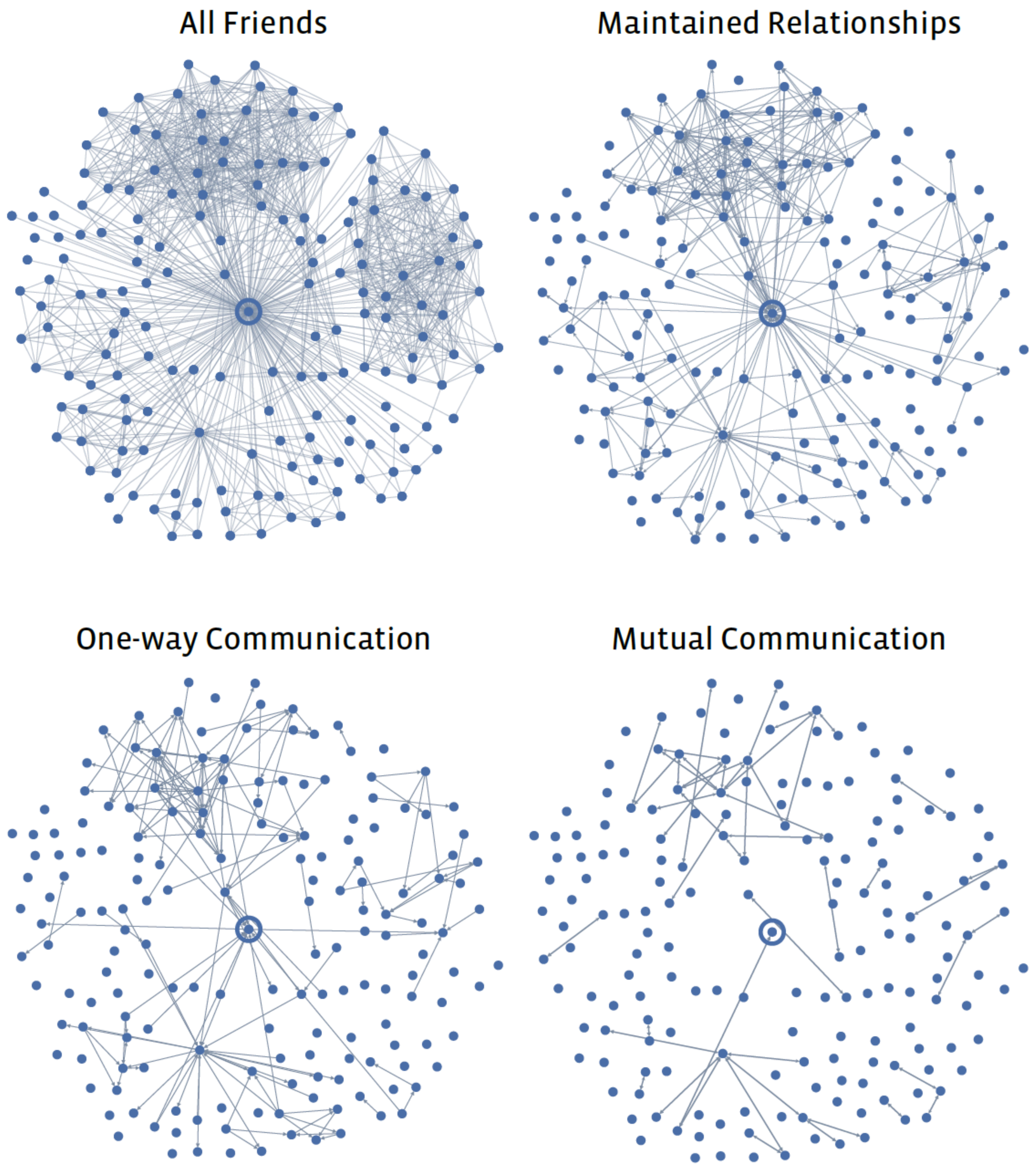
Tie strength on Facebook

Question: what are the strong ties among a user's friends?

Three categories of links [Marlow 2009]:
1) *Reciprocal (mutual) communication*: there were messages sent from a user to his friend and back
2) *One-way communication*: the user sent one or more messages to the friend (reciprocated or not)
3) *Maintained relationship*: the user follows what the friend does (e.g. by visiting her profile), with or without communication

C. Marlow, L. Byron, T. Lento, I. Rosenn, *Maintained relationships on Facebook*, (2009).
On-line at http://overstated.net/2009/03/09/maintainedrelationships-on-facebook

Tie strength on Facebook

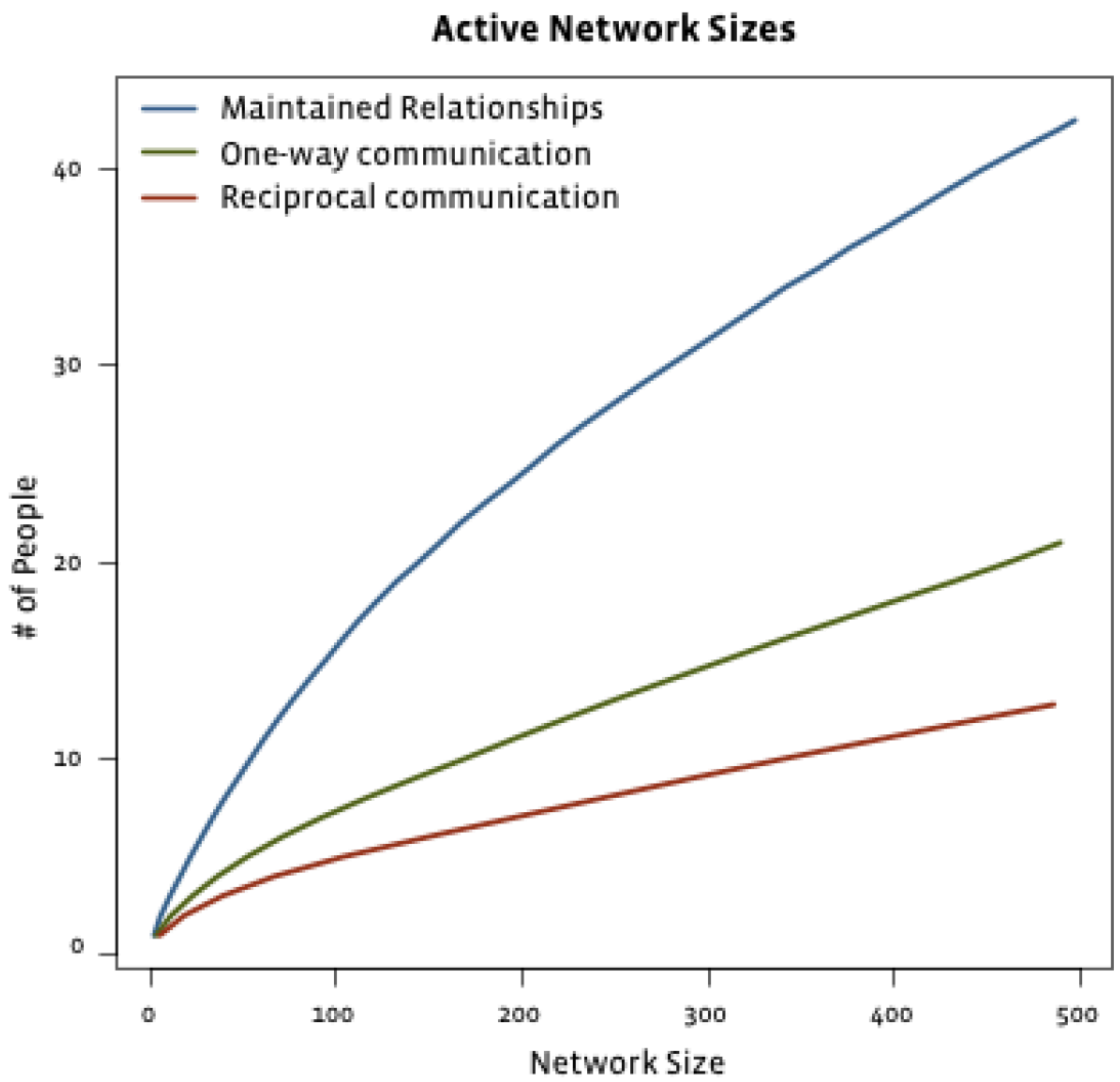


Remarks:

1) As ties get stronger, the network becomes sparser
2) As ties get stronger, certain parts of the network thin out much faster than others

C. Marlow, L. Byron, T. Lento, I. Rosenn, *Maintained relationships on Facebook*, (2009).
On-line at http://overstated.net/2009/03/09/maintainedrelationships-on-facebook

Tie strength on Facebook



Remarks:

1) Even users with many friends communicate with a small number of people
2) However, users can keep up with many more people (*passive engagement*)

C. Marlow, L. Byron, T. Lento, I. Rosenn, *Maintained relationships on Facebook*, (2009).
On-line at http://overstated.net/2009/03/09/maintainedrelationships-on-facebook

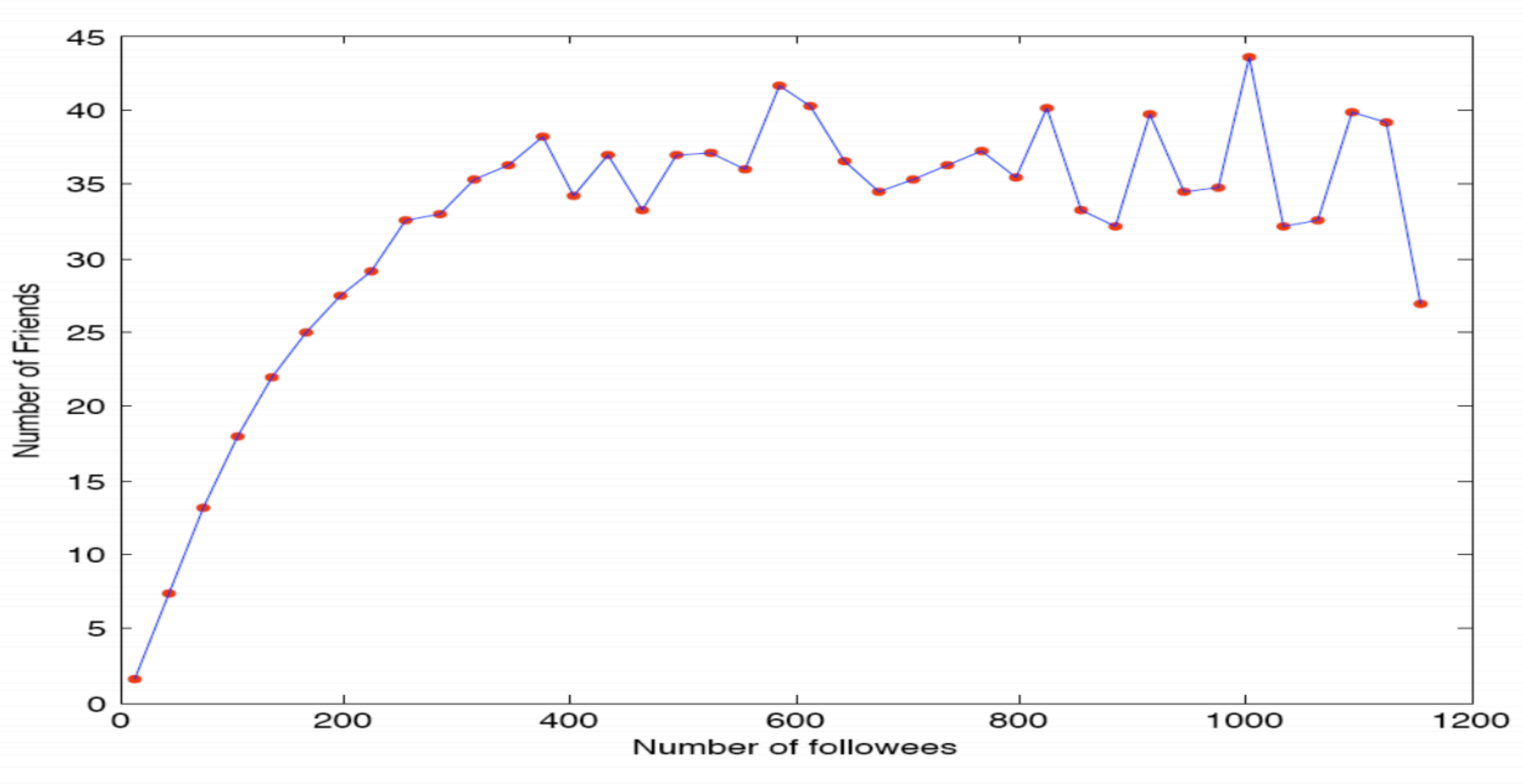
Tie strength on Twitter

Social network based on weak ties (as links are mostly not reciprocated)

For each user A, two categories of links [Huberman 2009]:
1) *Strong ties*: number of users followed by A and to whom A has sent at least two personal messages
2) *Weak ties*: number of users followed by A and to whom A has sent at most one personal message

B. A. Huberman, D. M. Romero, F. Wu, *Social networks that matter: Twitter under the microscope*, First Monday 14 (1) (2009)

Tie strength on Twitter



Remark: the number of strong ties quickly saturates!

B. A. Huberman, D. M. Romero, F. Wu, *Social networks that matter: Twitter under the microscope*, First Monday 14 (1) (2009)

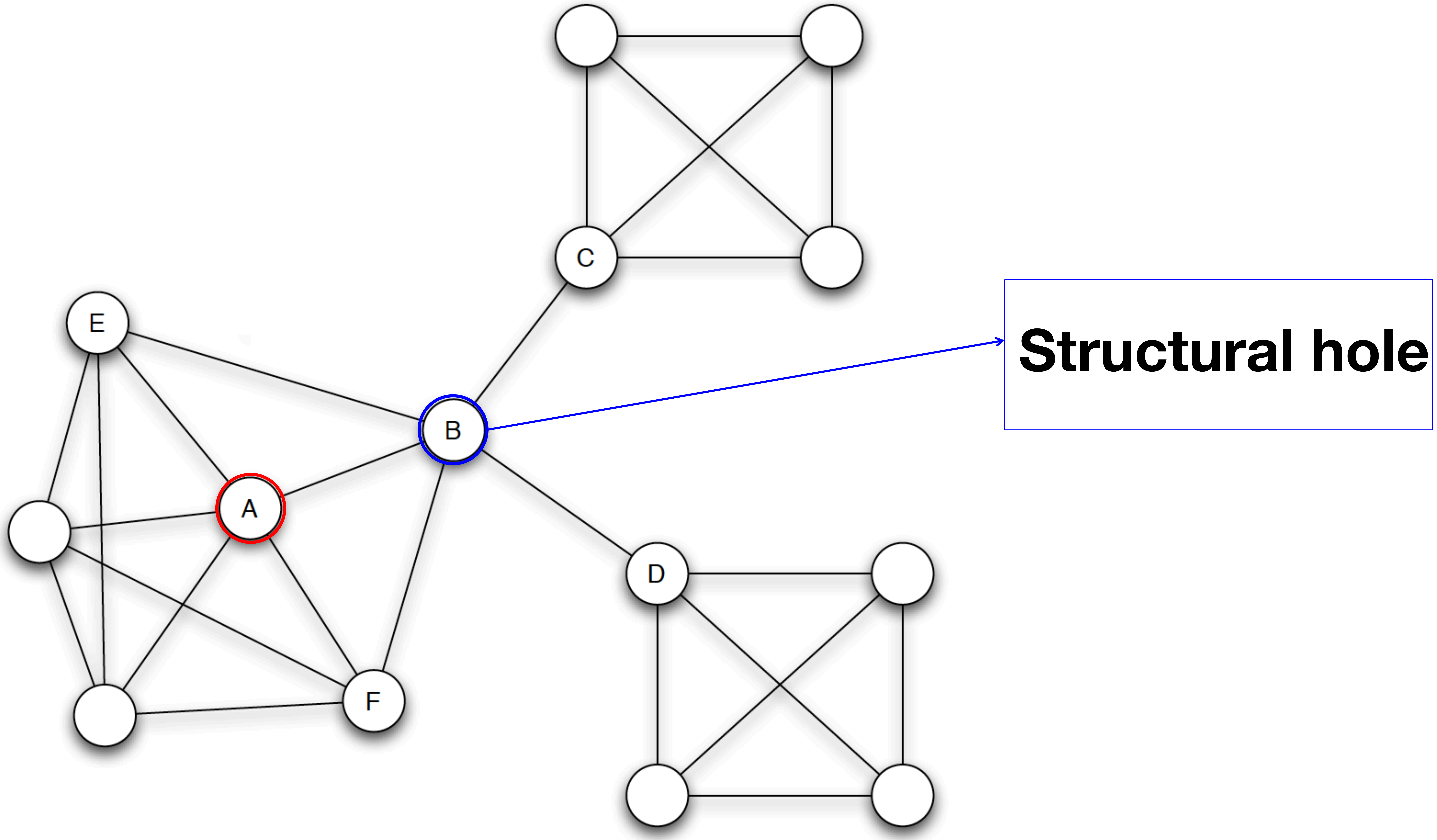
Roles of nodes

Link embeddedness: number of common neighbors of its endpoints

Two categories of nodes:
1) *Embedded nodes*: nodes whose links have high embeddedness
2) *Brokers*: nodes whose links have low to zero embeddedness

Nodes connected by embedded links are more likely to trust each other, via their mutual friendships

Roles of nodes



Roles of nodes

Features:

1) Embedded nodes are in a trusted environment, relationships are reliable and are more difficult to be broken
2) Brokers have risky interactions but easier access to information originating in distinct portions of the network. They can pool from diverse environments and expertise and have more chances to develop novel ideas. They regulate the access of embedded nodes to information coming from different groups

Homophily

Contexts: intrinsic features of the nodes affecting network structure and dynamics

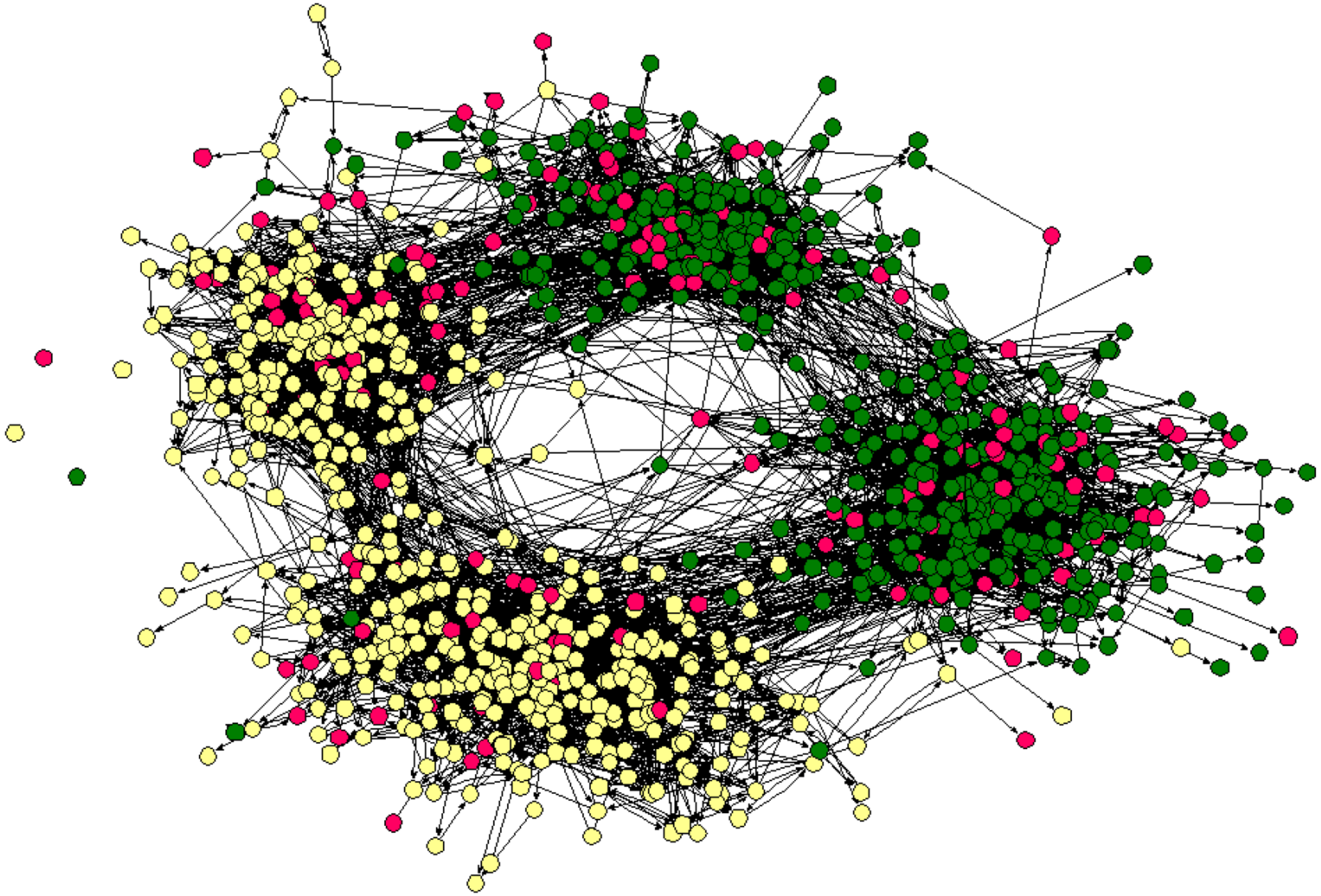
Examples: attributes of nodes (e.g. age), environments (e.g. work)

Homophily: we are similar to our friends

Homophily affects network dynamics: edges can eventually be formed between similar people

Homophily

Social network from a town's middle and high school



Homophily & triadic closure

Premise: A is friend with both B and C

Conclusion: A is similar to B and C

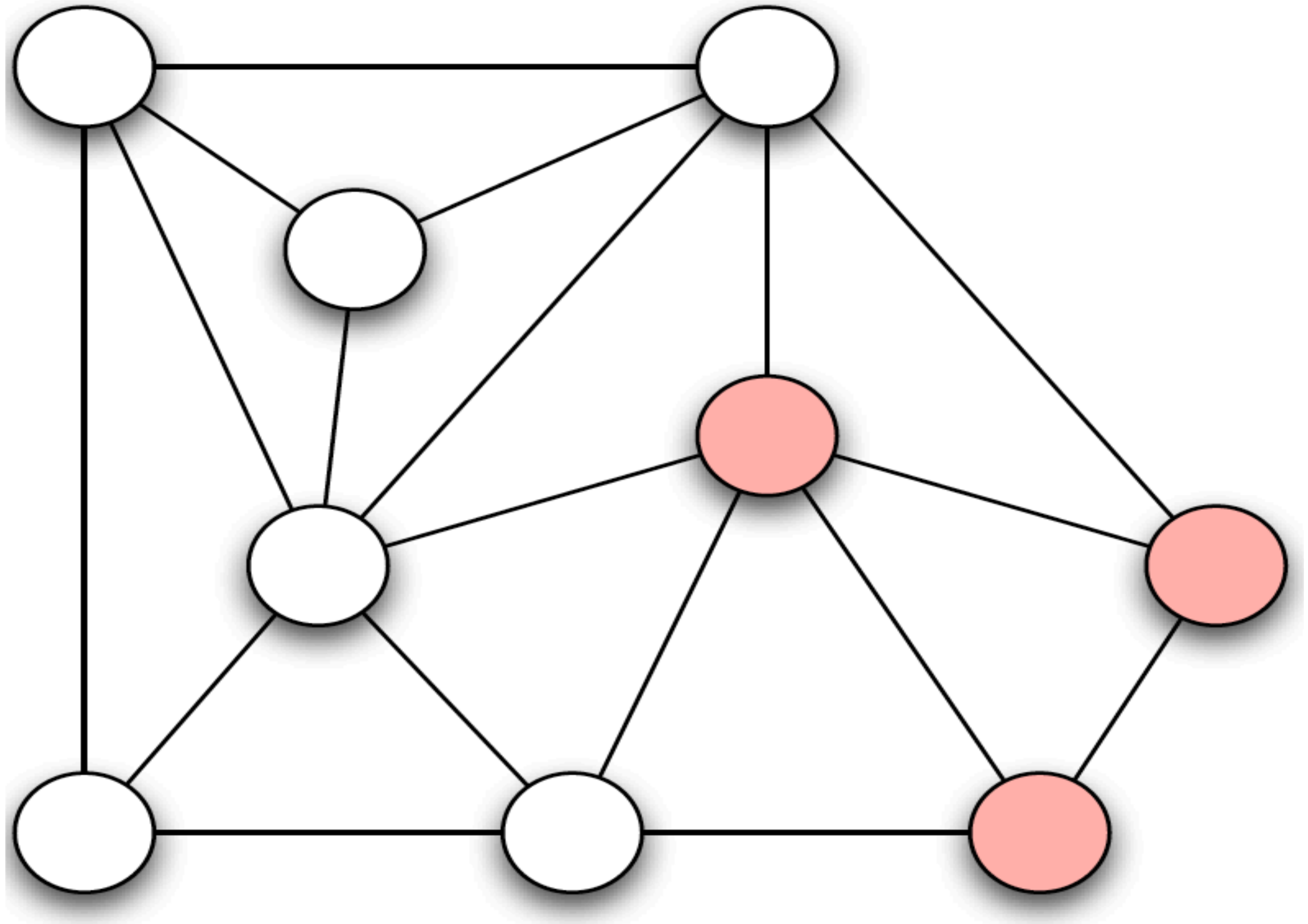
Consequence: B is similar C

Result: B and C may become friends, even if they ignore that they both know A

Measuring homophily

Question: given a particular characteristic of interest (race, age, etc.), is there a simple test we can apply to a network in order to estimate whether it exhibits homophily according to this characteristic?

Example: hypothetical school class (white circles= boys, red circles = girls)



Measuring homophily

Hint: what happens if there were no homophily?

Answer: edges would be present regardless of nodes' attributes

Fraction p of boys and q of girls:

1) The fraction of edges connecting pairs of boys is p^2
2) The fraction of edges connecting pairs of girls is q^2
3) The fraction of edges connecting a boy and a girl is $2pq$

Homophily Test: If the fraction of cross-gender edges is significantly less than $2pq$, then there is evidence for homophily

Measuring homophily

Measuring homophily

Measuring homophily

Measuring homophily

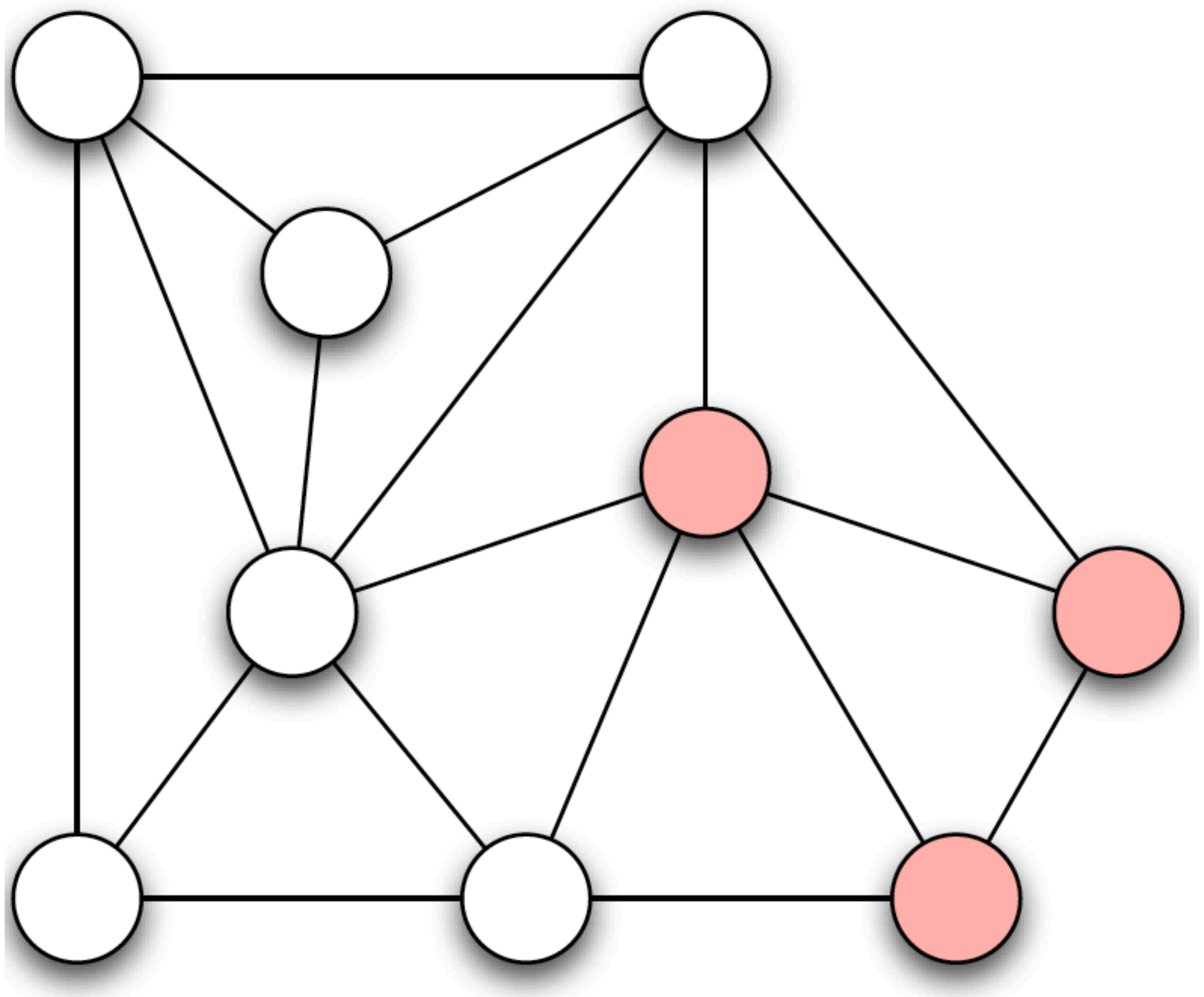
Measuring homophily

Number of edges = 18
p=6/9=2/3
q=3/9=1/3

2pq = 4/9

Actual fraction of edges between boys and girls = 5/18 < 4/9

Evidence of homophily!



Measuring homophily

Remarks:

1) It is not sufficient that the fraction of edges connecting classes of nodes is lower than 2pq, we need to show that it is *significantly* lower than 2pq
2) The fraction of edges connecting classes of nodes could also *exceed* 2pq (inverse homophily)
3) The argument can be extended to more than 2 classes. Ex. n classes with proportions p_1, p_2 , …, p_n: there is homophily between class i and j if the proportion of edges connecting nodes of i to nodes of j is significantly lower than $2p_ip_j$

Selection and social influence

Selection: tendency to form links with people with similar characteristics

Social influence: tendency to acquire similar features as the people one is connected to

Selection typically has to do with *immutable features* of individuals (age, gender, race)

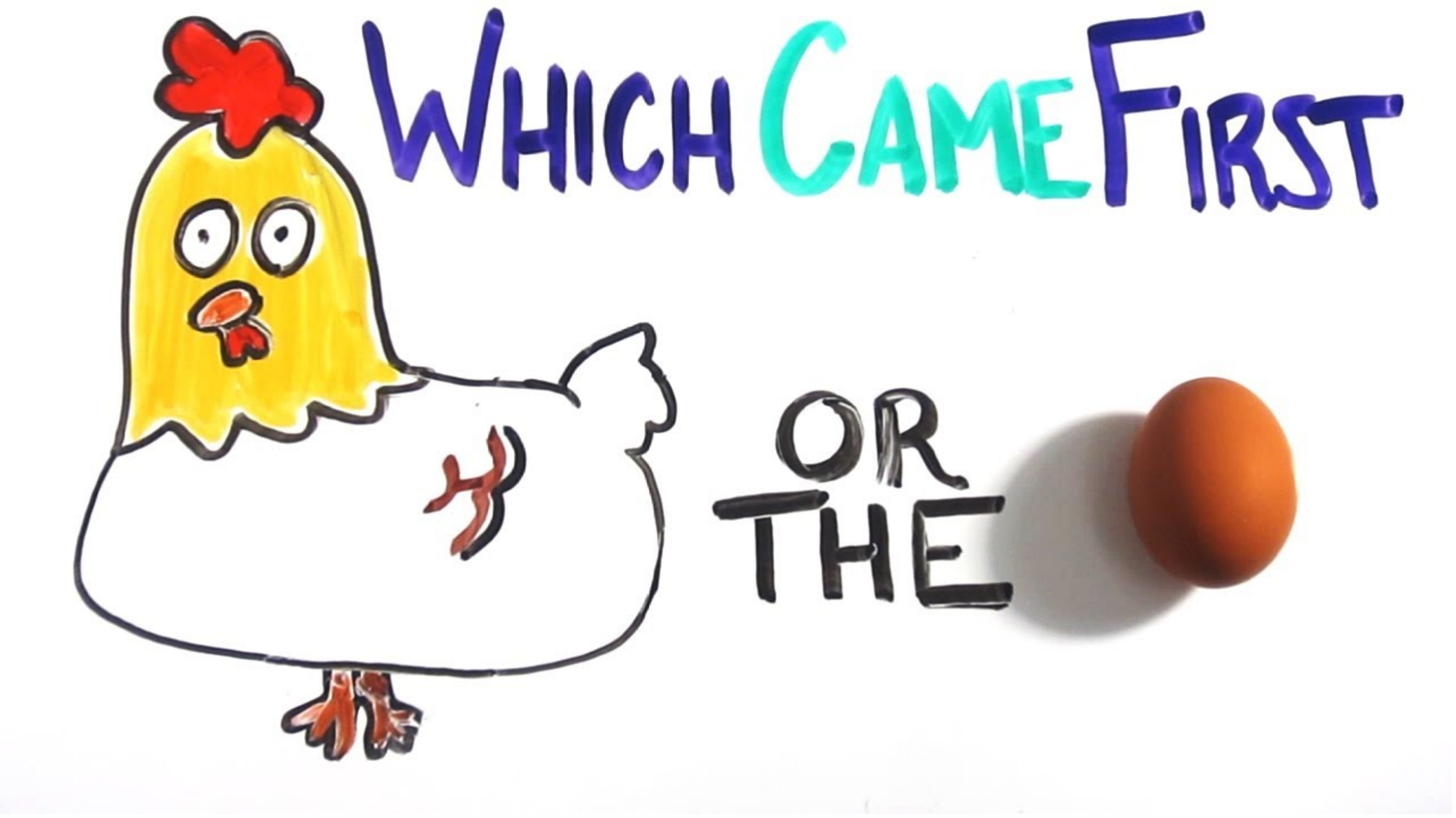
Social influence typically has to do with *mutable features* of individuals (beliefs, opinions)

Selection drives the formation of links, social influence determines how links affect people’s (mutable) features

Selection and social influence

Dilemma: how can we discriminate between selection and social influence in real networks?

Question: given a link between two people, is that link there because people are similar (selection) or are people similar because of that link (social influence)?



Selection and social influence

Solution: longitudinal study of networks

Discriminating the two effects allows us to find good policies to fight social problems

Example: if social influence dominates in drug abuse, one could target influential drug users; if selection dominates this would hardly change anything

Problem: what is the effect of social networks on health-related problems?

Selection and social influence

Christakis & Fowler: how does obesity spread?

Result I: obese and non-obese people cluster according to homophily!

Question: selection or social influence?

Answer: social influence plays an important role!

Caveat: statistical analysis not reliable!

N. A. Christakis, J. H. Fowler, *The spread of obesity in a large social network over 32 years*, New England Journal of Medicine 357, 370-379 (2007)

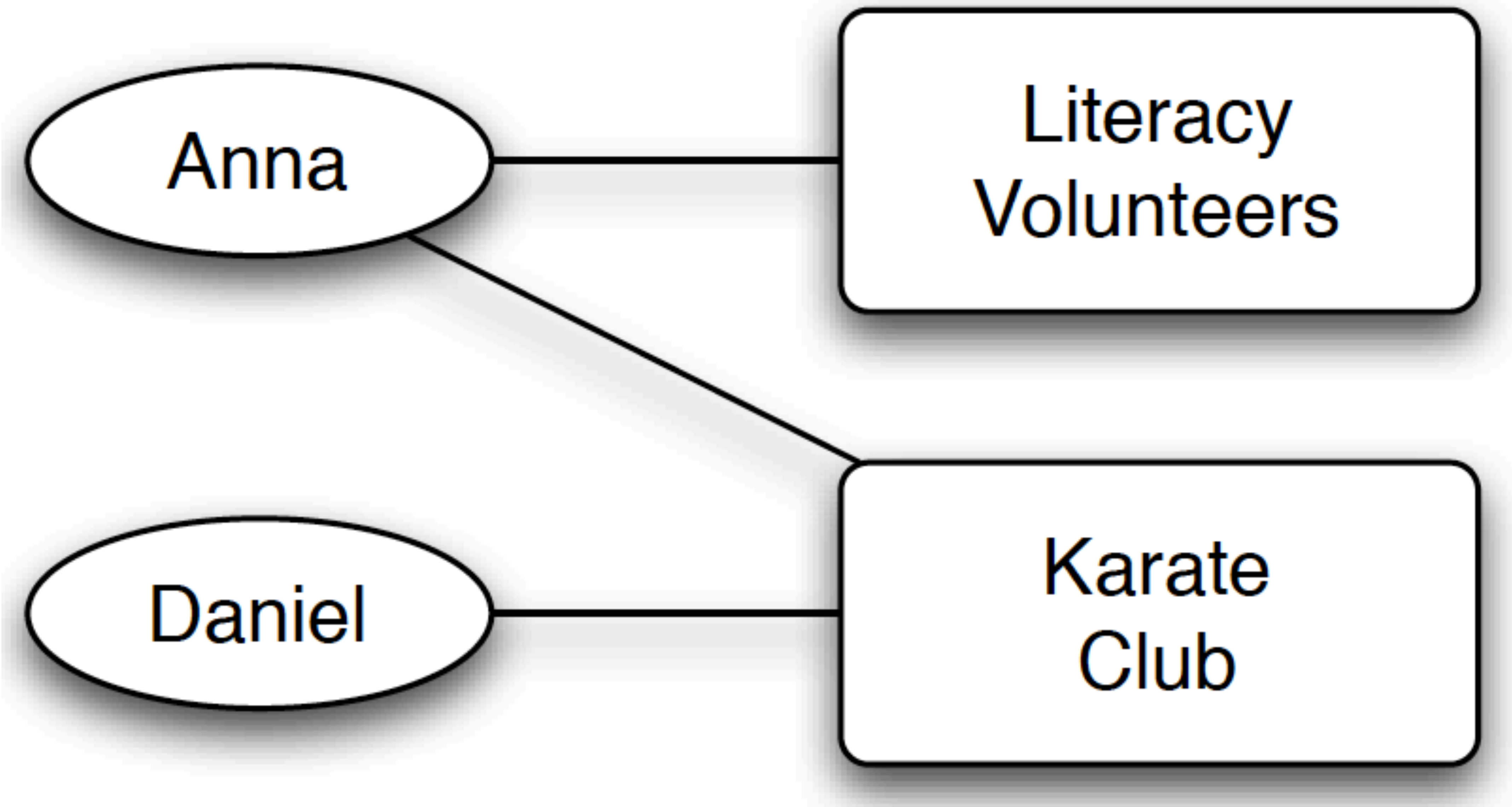
C. R. Shalizi, A. C. Thomas, *Homophily and contagion are generically confounded in observational social network studies*, Sociological Methods and Research 40 (2), 211-239 (2011)

Affiliation networks

Embedding the contexts in a network

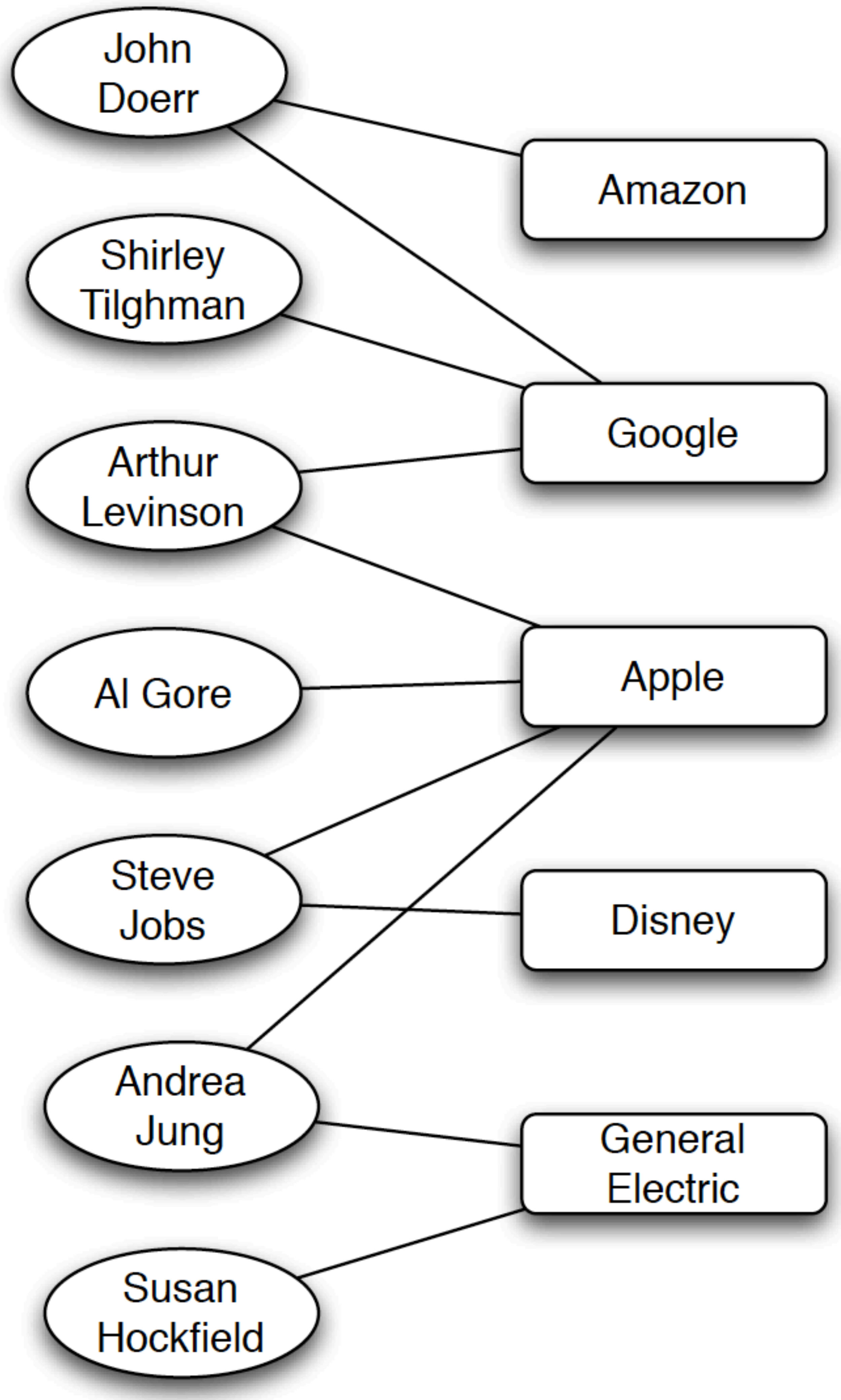
Affiliation network: bipartite network of people and activities (*foci*)

A bipartite graph consists of two classes of nodes, with edges joining *only* nodes of different classes



Affiliation networks

Board of directors

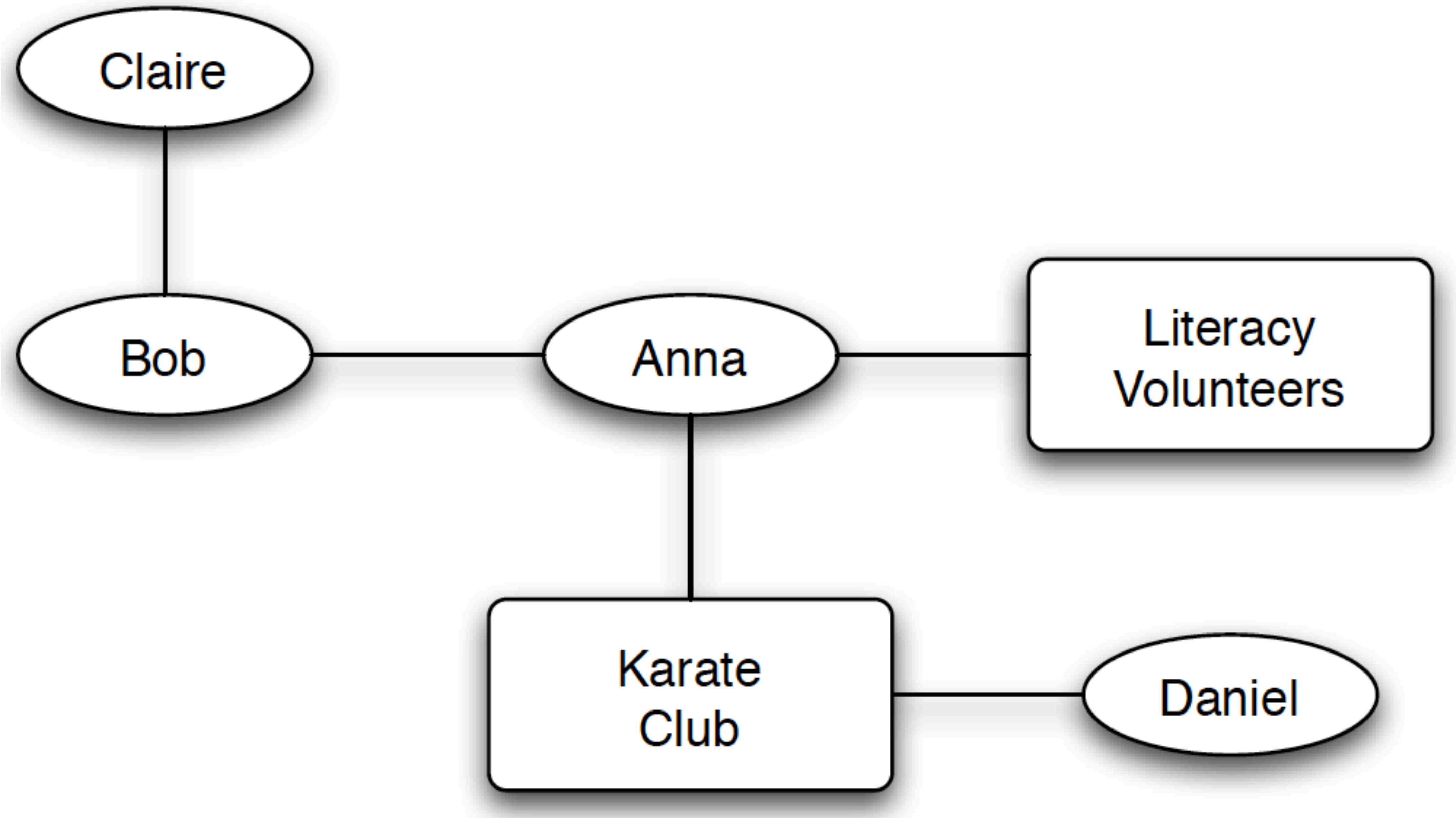


Coevolution of social and affiliation networks

Shared focus: higher probability to become friends

Friendship: higher probability to join the same activities

Social-affiliation network

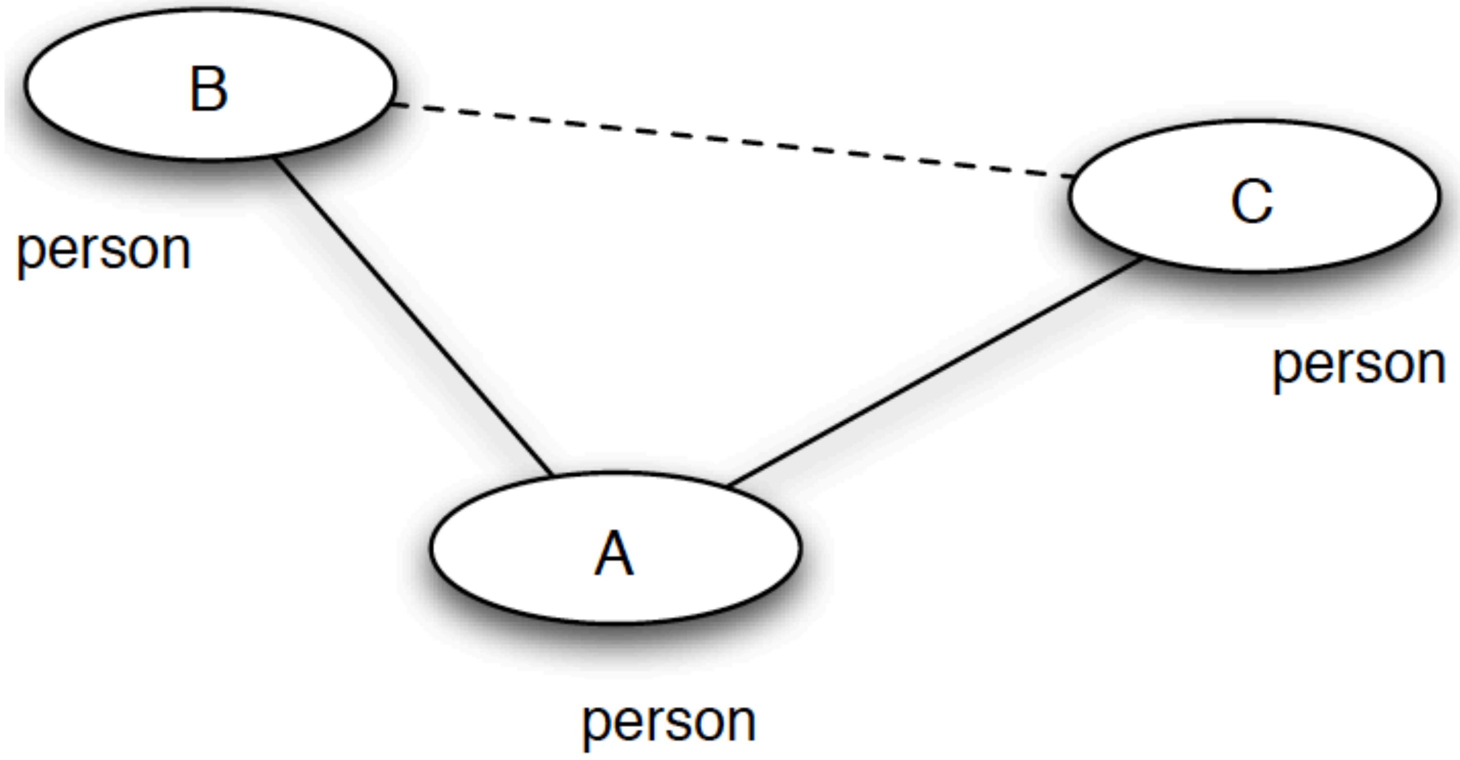


Coevolution of social and affiliation networks

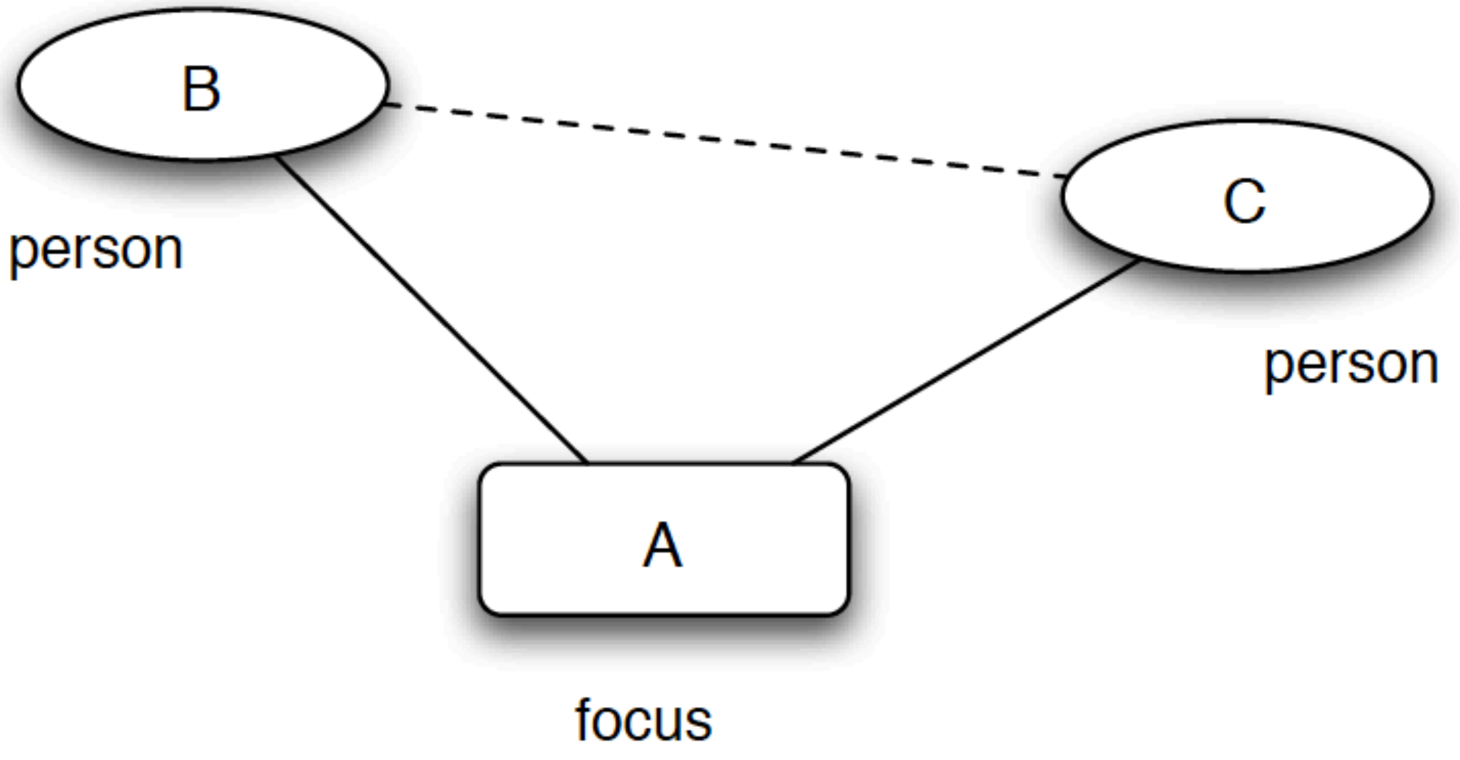
Different mechanisms for link formation:

1) **Triadic closure:** (A introduces B to C: A, B, C are persons)
2) **Focal closure:** (A introduces B to C: A is an activity, B and C persons)
3) **Membership closure:** (A introduces B to C: A and C persons, B activity)

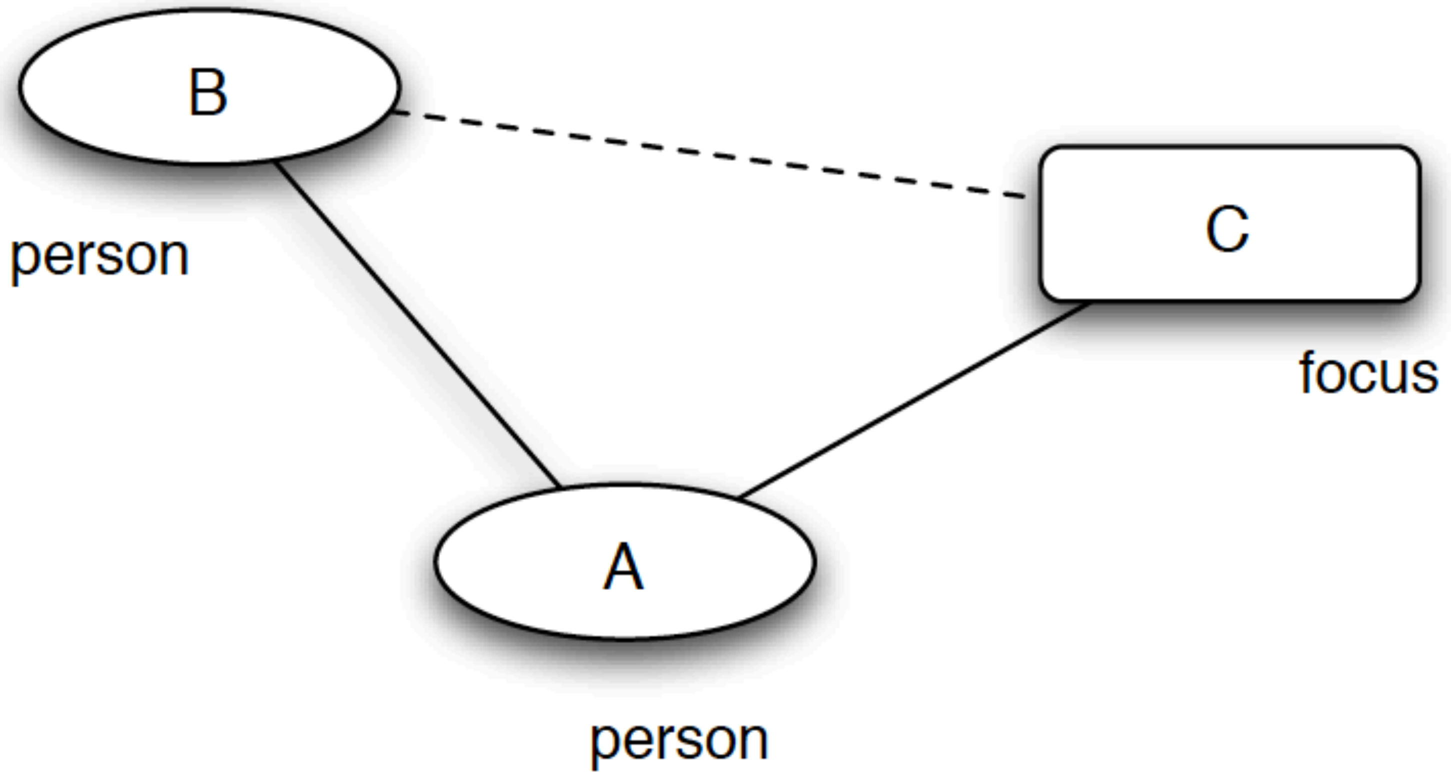
Coevolution of social and affiliation networks



(a) *Triadic closure*

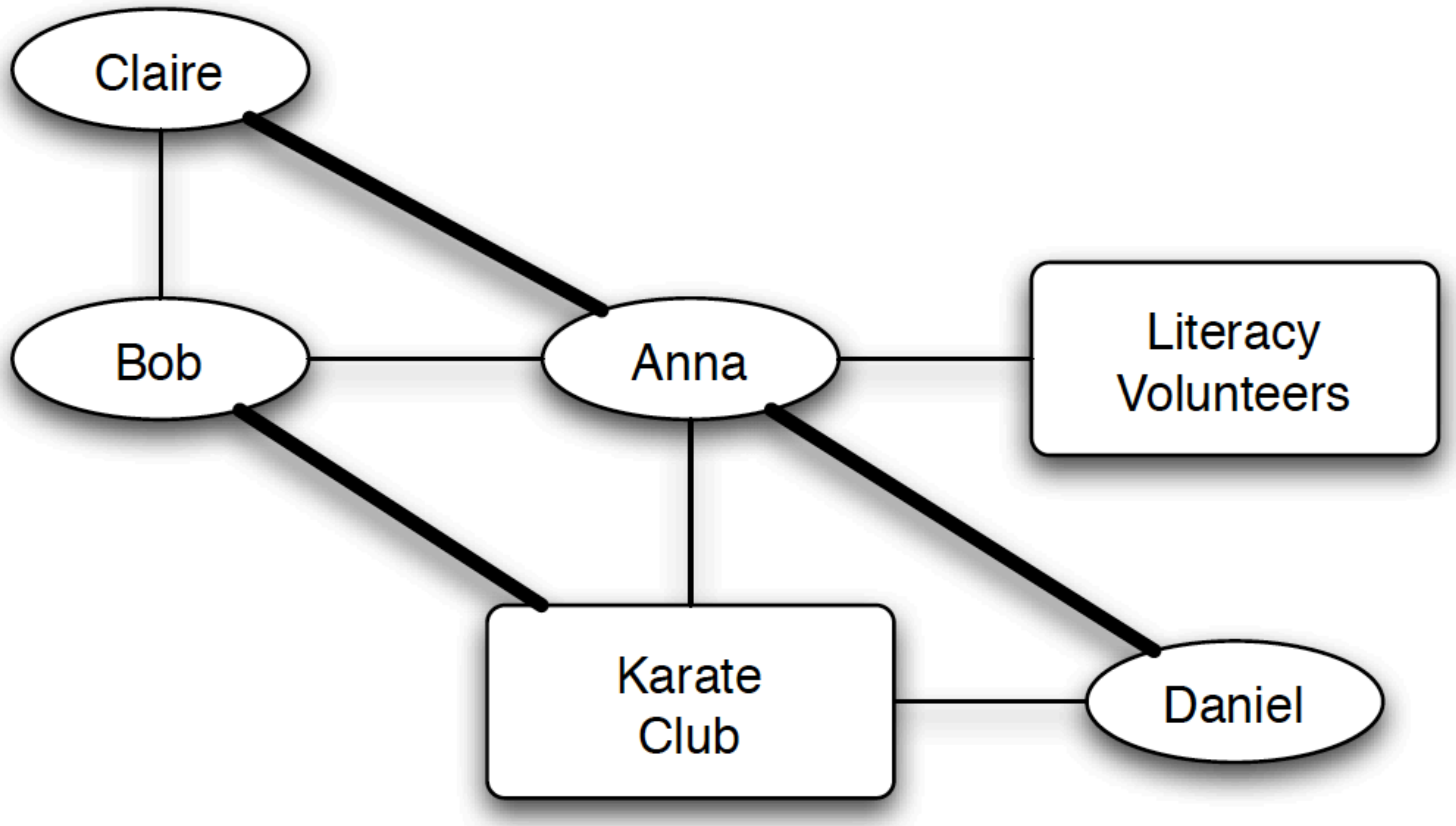


(b) *Focal closure*



(c) *Membership closure*

Coevolution of social and affiliation networks



1) Bob introduces Anna to Claire
2) Karate introduces Anna to Daniel
3) Anna introduces Bob to Karate

Tracking link formation in online data

Question: do we have quantitative evidence for the link formation mechanisms above?

Approach: longitudinal studies of large online social networks

Question I: how much more likely is it that a link is formed between people sharing friends than between people with no common friends?

Question II: how much does the link probability depend on the number of common friends?

Tracking link formation in online data

Procedure:

1) Take two snapshots of the network at different times
2) For each k we identify all pairs of nodes with exactly k friends in common, but not connected to each other
3) We define T(k) to be the fraction of these pairs that are connected by an edge in the second snapshot
4) We plot T(k) as a function of k

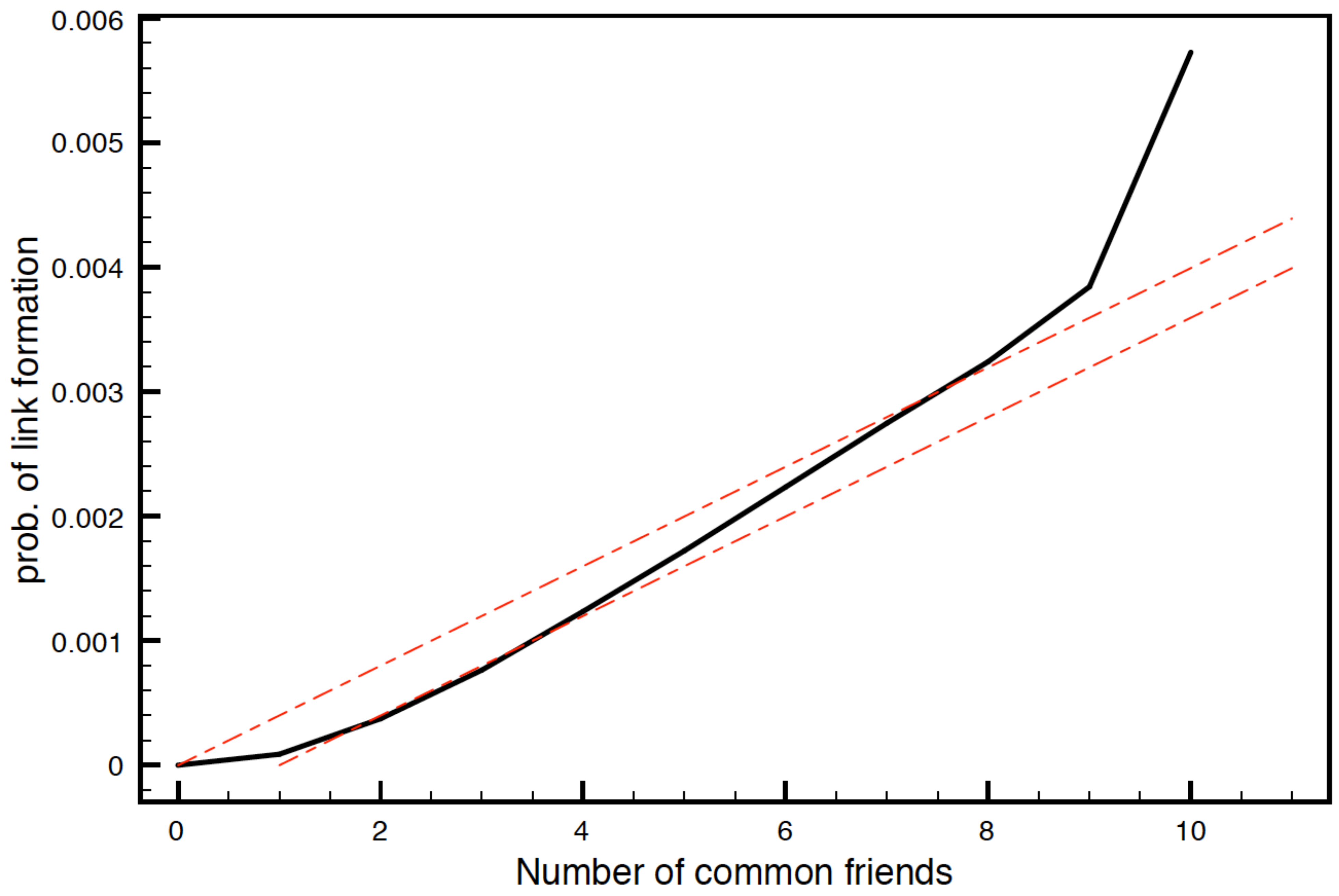
Expectation: T(k) is an increasing function of k

G. Kossinets, D. J. Watts, *Empirical analysis of an evolving social network*, Science 311, 88-90 (2006)

Tracking link formation in online data

Remarks:

1) T(0) is close to zero
2) T(k) increases monotonically with k
3) The increase is roughly linear
4) The curve turns upwards for k=1 and k=2



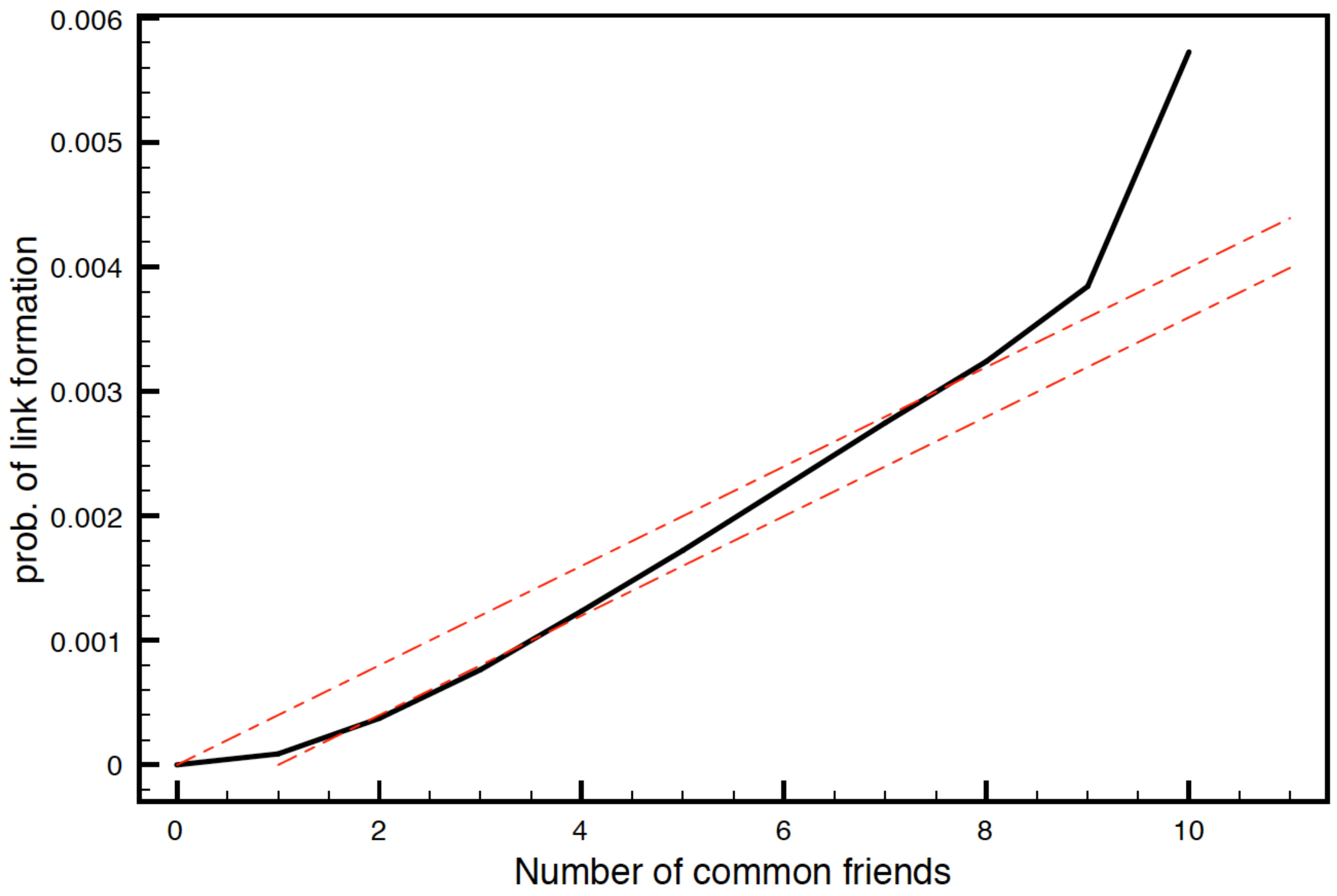
G. Kossinets, D. J. Watts, *Empirical analysis of an evolving social network*, Science 311, 88-90 (2006)

Tracking link formation in online data

Baseline model:

1) Assumption: each common friend gives a link probability p
2) For k common friends the probability of having no link is $(1\text{-}p)^k$

$$T_{baseline} = 1 - (1-p)^k$$



G. Kossinets, D. J. Watts, *Empirical analysis of an evolving social network*, Science 311, 88-90 (2006)

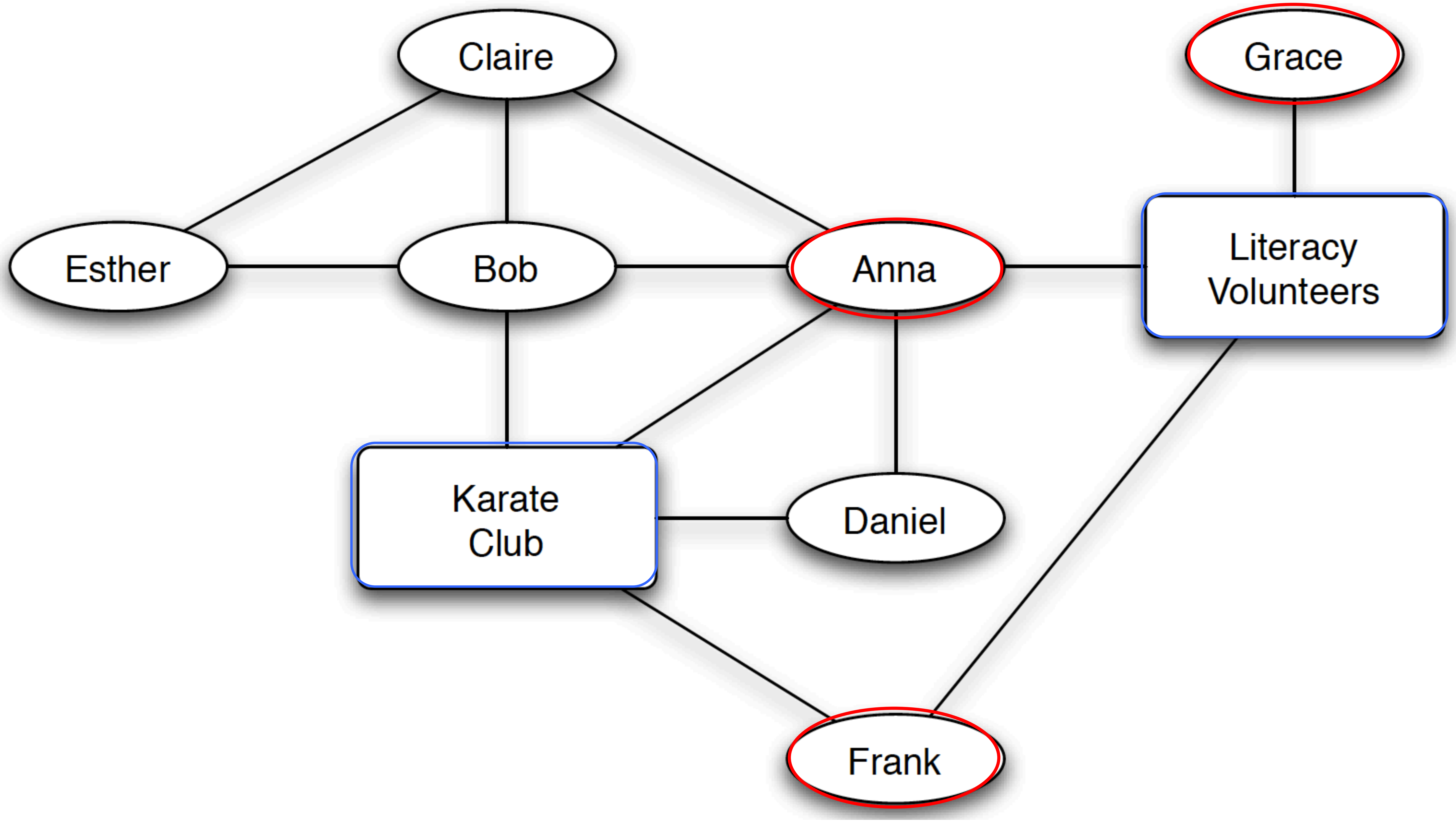
Tracking link formation in online data

Focal closure

Question: What is the probability that two people form a link as a function of the number of foci they are jointly affiliated with?

G. Kossinets, D. J. Watts, *Empirical analysis of an evolving social network*, Science 311, 88-90 (2006)

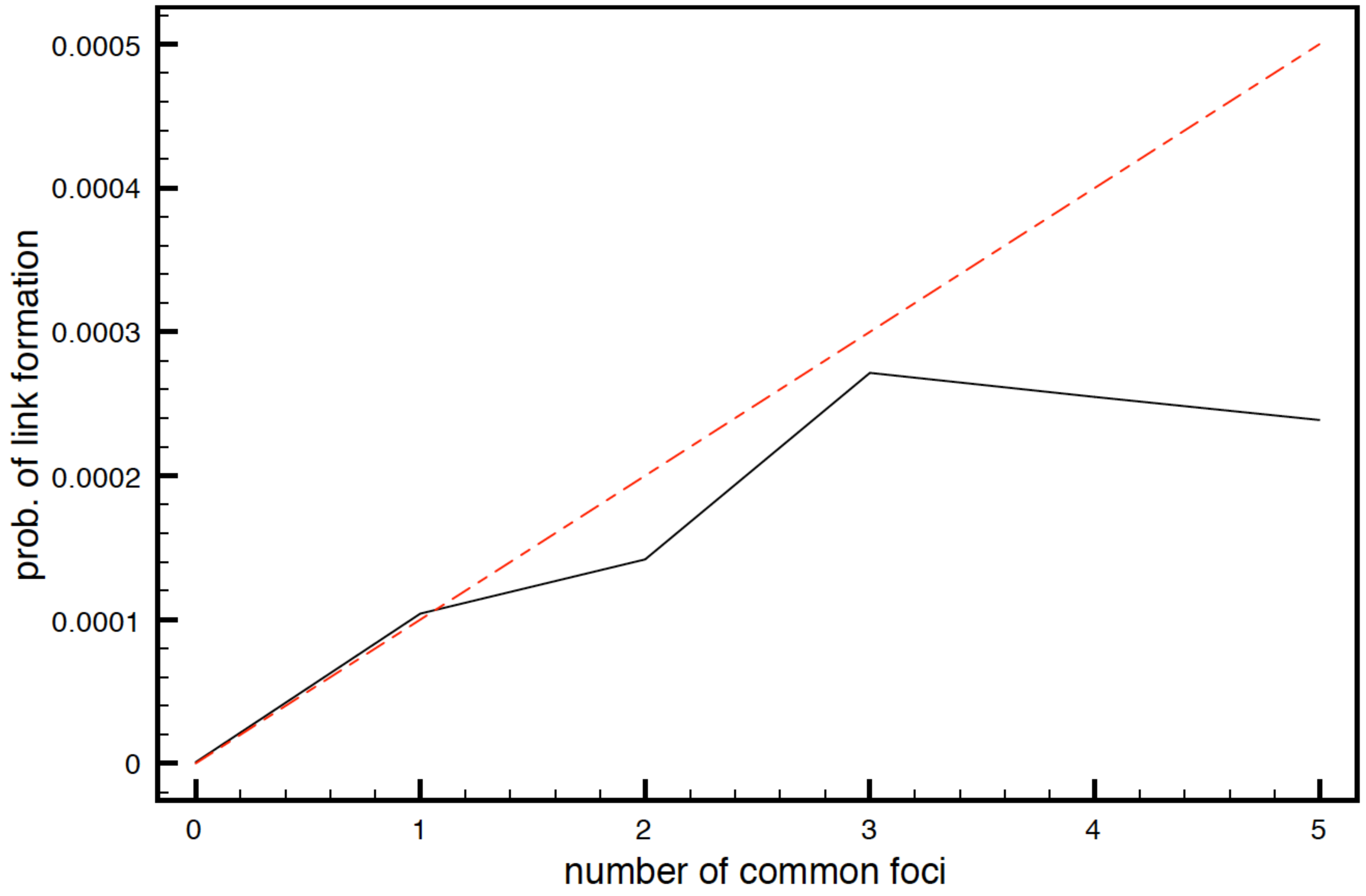
Tracking link formation in online data



Tracking link formation in online data

Focus: class(es) attended by students

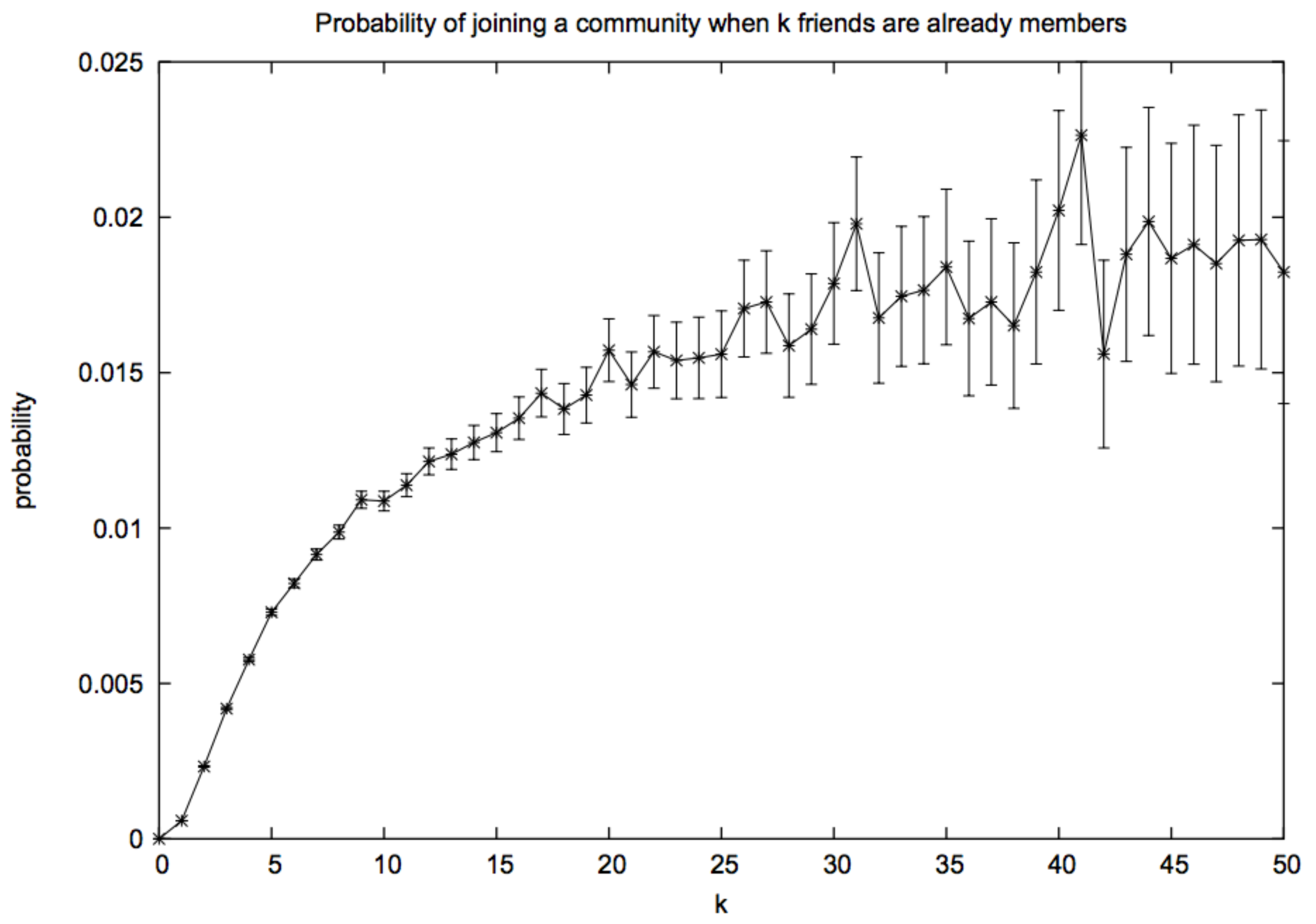
Remark: curve bends down a bit from 3 foci onwards



Tracking link formation in online data

Membership closure

Question: What is the probability that a person becomes involved with a particular focus as a function of the number of friends who are already involved in it?



Network: friendships of LiveJournal users and their community memberships

L. Backstrom, D. Huttenlocher, J. Kleinberg, X. Lan, *Group formation in large social networks: Membership, growth, and evolution*, Proceedings 12ft ACM SIGKDD (2006)

Interplay of selection and social influence

System: Wikipedia

Network nodes: Wikipedia editors

Network edges: pairs of Wikipedia editors who communicate via user talk pages

Similarity of editors A and B:

$$\frac{\text{number of articles edited by both A and B}}{\text{number of articles edited by at least one of A and B}}$$

Example: A edited the pages “USA” and “Earthquake”, B the pages “USA” and “Galaxy”. The similarity is 1/3

Interplay of selection and social influence

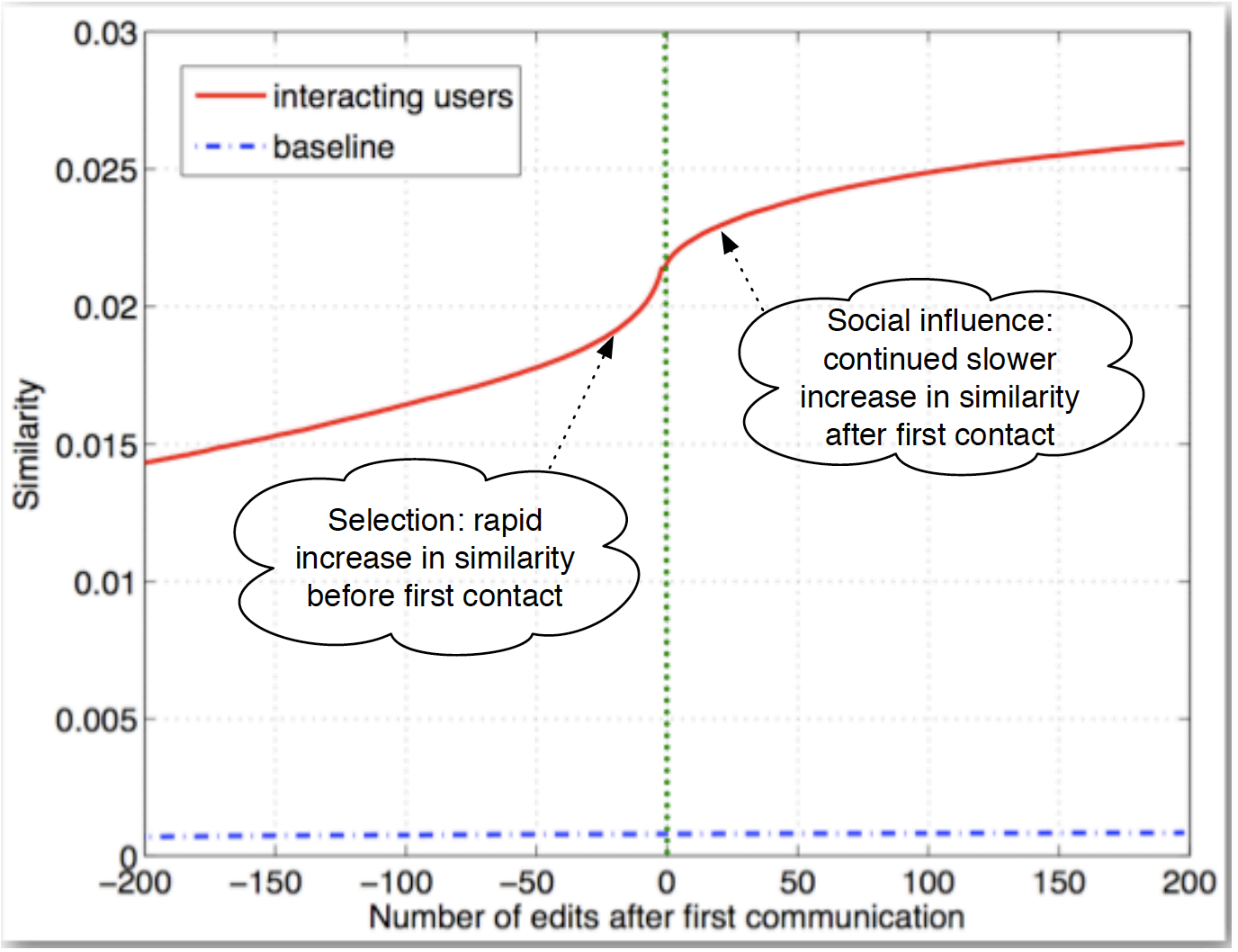
Premise: pairs of editors that communicate are significantly more similar than pairs who do not communicate

Question: is the homophily arising because editors talk to the peers that are similar to them (selection), or do editors become similar because they talk to each other (social influence)?

Interplay of selection and social influence

Remarks:

1) Similarity grows in time, both before and after the first communication: *selection + social influence*
2) Fastest increase just before first communication: *selection plays more important role*
3) Similarity way higher than for pairs of non-interacting editors



Network assortativity

- Two possible mechanisms by which assortativity emerges naturally:

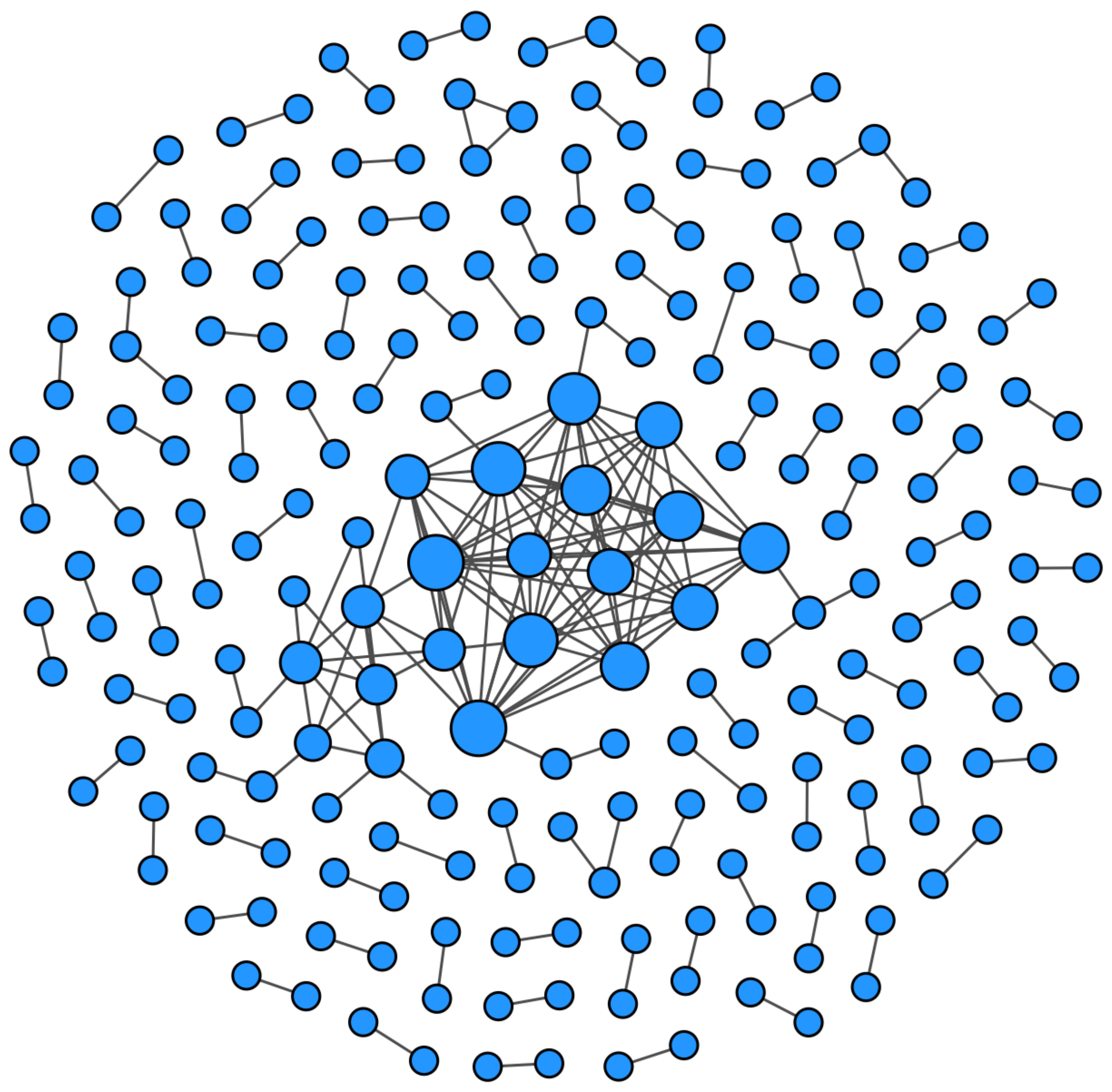
  1. **Selection** or **homophily**: similar nodes become connected

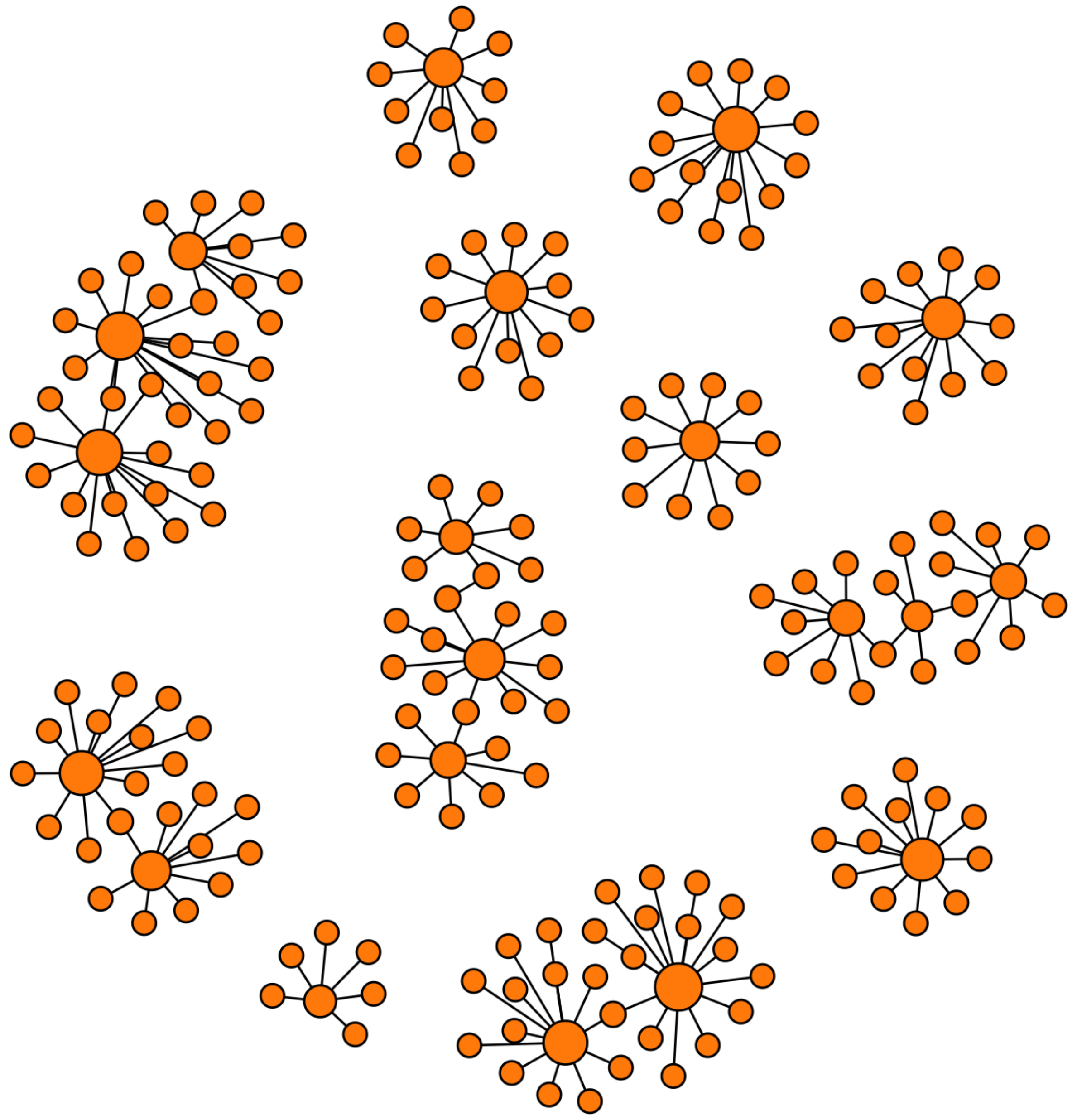
  2. (Social) **influence**: connected nodes become more similar

- It can also be a bad thing. For example "**echo chambers**" and "**groupthink**" are situations where your friends are like you, diversity is killed, and you are only exposed to opinions that reinforce your pre-existing beliefs…

Degree assortativity

- A.k.a. **degree correlation**
- Assortative networks have a **core-periphery** structure with hubs in the core
  - Ex: social networks
- Disassortative networks have **hub-and-spoke** (or **star**) structure
  - Ex: Web, Internet, food webs, bio networks





Assortativity in NetworkX

```
# based on an a categorical attribute, such as gender
assort_a = nx.attribute_assortativity_coefficient(G, category)

# based on a numerical attribute, such as age
assort_n = nx.numeric_assortativity_coefficient(G, quantity)

# based on degree (Pearson correlation
# between degree of adjacent nodes)
r = nx.degree_assortativity_coefficient(G)
```

Assortativity in NetworkX

Another way to compute the degree assortativity is by measuring the correlation between the degree and the average degree of the neighbors of nodes with that degree:

$$k_{nn}(i) = \frac{1}{k_i} \sum_j a_{ij} k_j$$

$$\langle k_{nn}(k) \rangle = \langle k_{nn}(i) \rangle_{i:k(i)=k}$$

```
import  scipy.stats
knn_dict = nx.k_nearest_neighbors(G)
k, knn = list(knn_dict.keys()), list(knn_dict.values ())
r, p_value = scipy.stats.pearsonr(k, knn)
```