Social Network Analysis

Networks in their surrounding context

Individuals

- Each individual in a social network has a distinctive set of personal characteristics.]
- Similarities and compatibilities among two people's characteristics can strongly influence whether a link forms between them.
- Each individual also engages in a set of behaviors and activities that can shape the formation of links within the network.

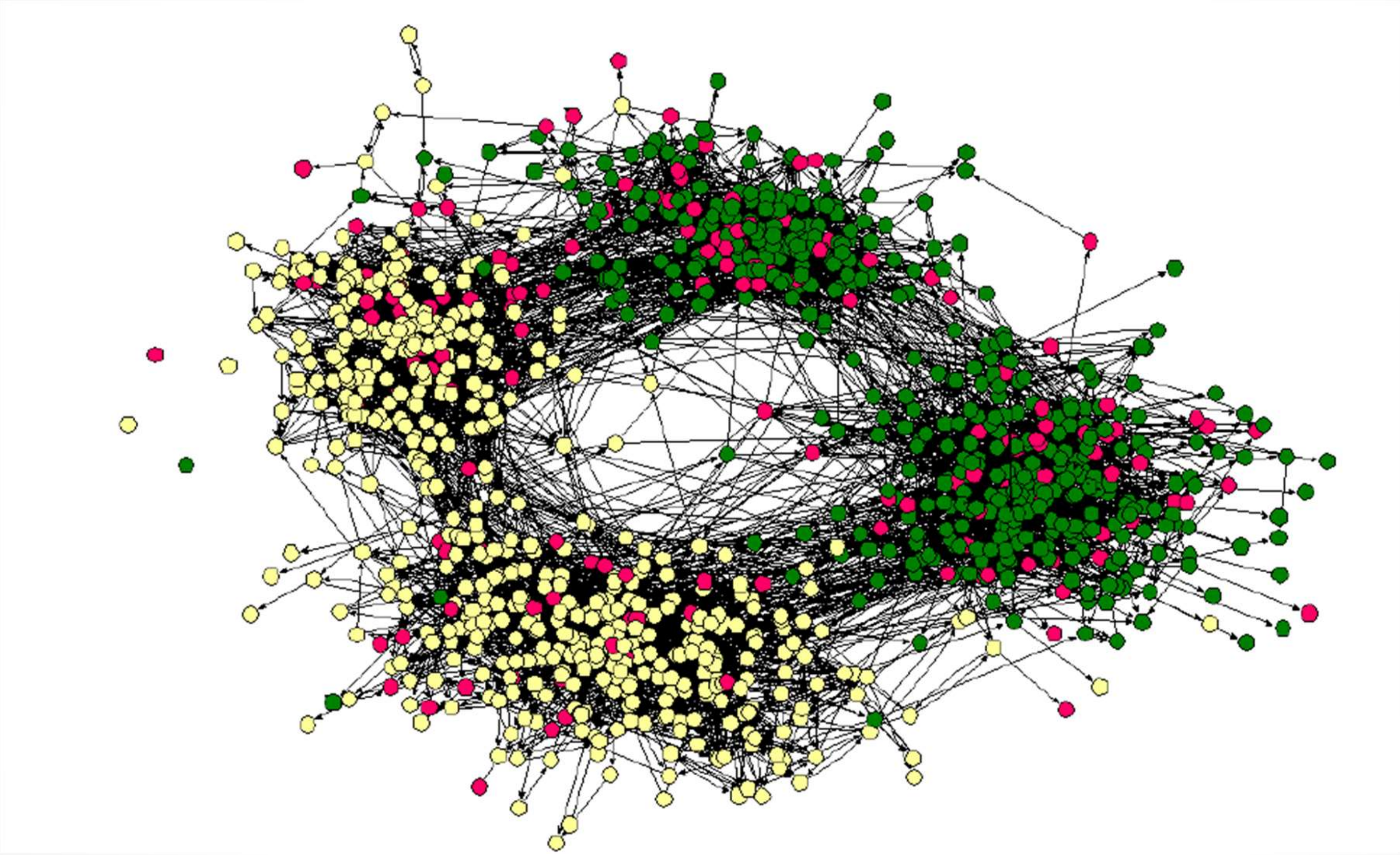
Network's Surrounding Context

- The contexts in which a social network is embedded will generally have significant effects on its structure,
- Network's surrounding contexts: Factors that exist outside the nodes and edges of a network, but which nonetheless affect how the network's structure evolves.
- The surrounding contexts affecting a network's formation can, to some extent, be viewed in network terms as well
- By expanding the network to represent the contexts together with the individuals - several different processes of network formation can be described in a common framework.

Homophily

- Homophily — the principle that we tend to be similar to our friends.
  - Typically, your friends don't look like a random sample of the underlying population
  - Similar in age, where-they-live, occupation, affluence, interests and opinions.
- Most of us have *specific* friendships that cross all these boundaries,
  - but in aggregate, the pervasive fact is that links in a social network tend to connect people who are similar to one another.
- Homophily provides us with fundamental illustration of how a network's surrounding contexts can drive the formation of its links.

Homophily - Example



- Homophily can produce a division of a social network into densely-connected, homogeneous parts that are weakly connected to each other.
- In this social network from a town's middle school and high school [1], two such divisions in the network are apparent:
  - One based on race (with students of different races drawn as differently colored circles
  - The other based on friendships in the middle and high schools respectively.

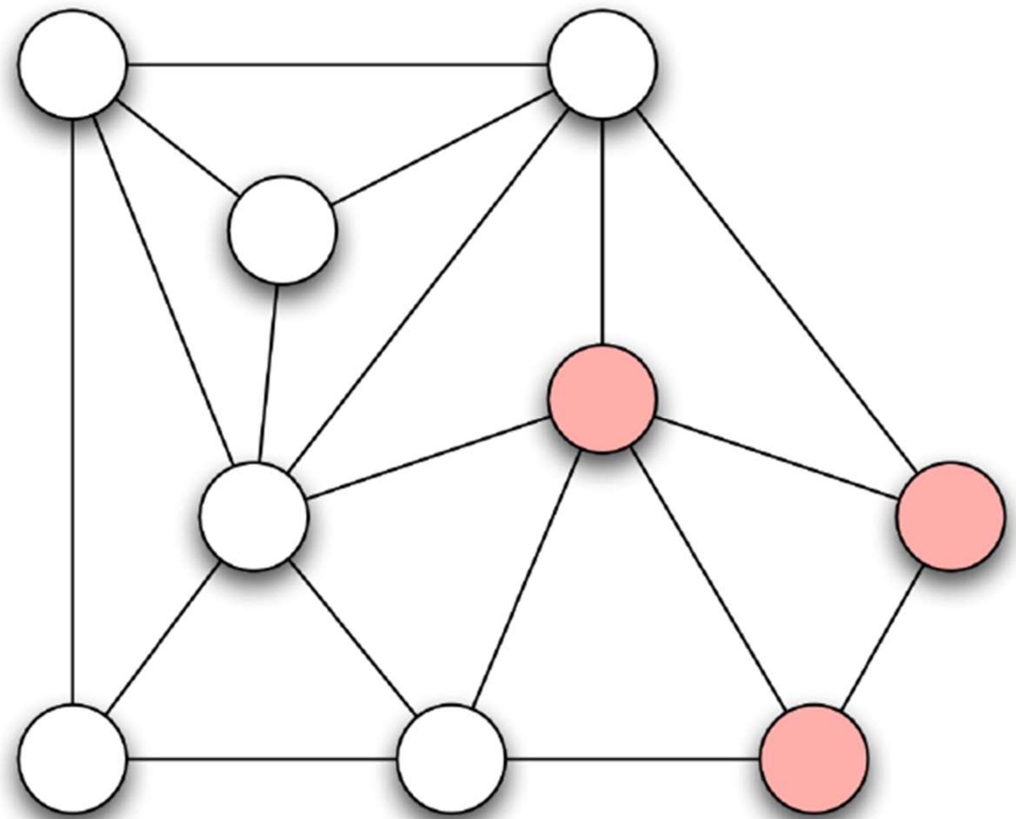
Intrinsic & Contextual Factors

- Consider the basic contrast between:
  - a friendship that forms because two people are introduced through a common friend
  - a friendship that forms because two people attend the same school, club or work for the same company.
- In the first case, a new link is added for reasons that are **intrinsic** to the network itself
  - we need not look beyond the network to understand where the link came from.
- In the second case, the new link arises for an equally natural reason, but one that makes sense only when we look at the **contextual** factors *beyond the network*
  - at some of the social environments (in this case schools, clubs and companies) to which the nodes belong.
- Of course, there are strong interactions between *intrinsic* and *contextual* effects on the formation of any single link; they are both operating *concurrently* in the same network.

Triadic Closure

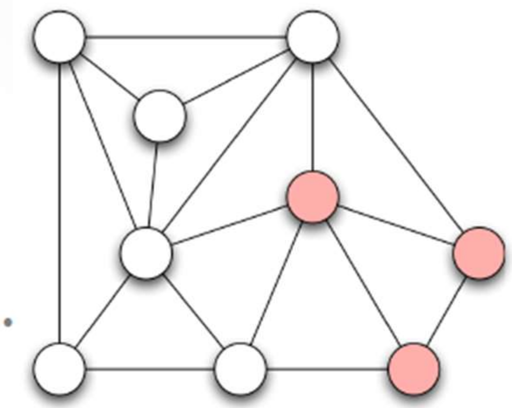
- Principle of triadic closure — is supported by a range of mechanisms that range from *the intrinsic to the contextual.*
- When individuals B and C have a common friend A, then there are increased opportunities and sources of trust on which to base their interactions, and A will also have incentives to facilitate their friendship.
- Since we know that A-B and A-C friendships already exist, the principle of homophily suggests that B and C are each likely to be similar to A in a number of dimensions, and hence quite possibly similar to each other as well.
  - As a result, based purely on this similarity, there is an elevated chance that a B-C friendship will form; and this is true even if *neither of them is aware that the other one knows A.*
- There is no single basis for triadic closure.
  - As we take into account more and more of the factors that drive the formation of links, it inevitably becomes difficult to attribute any individual link to a single factor.
  - One expects most links to in fact arise from a *combination* of several factors — partly due to the effect of other nodes in the network, and partly due to the surrounding contexts.

Is this homophily?



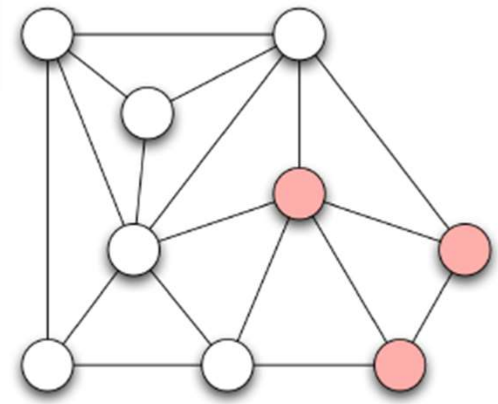
Hypothetical class: 3 shaded nodes are girls and the 6 unshaded are boys.

How to measure homophily?



- If there were no cross-gender edges at all, then the question of homophily would be easy to resolve: "homophily is present".
- Is the picture consistent with homophily?
- To motivate the measure, we can first ask:
    - What would it mean for a network not to exhibit homophily by gender?
        - It would mean that the proportion of male and female friends a person has looks like the background male/female distribution in the full population.

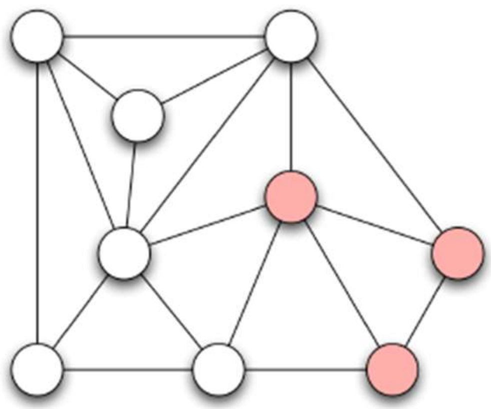
How to measure homophily?



- Suppose we have a network in which a p fraction of all individuals are male, and a q fraction of all individuals are female.

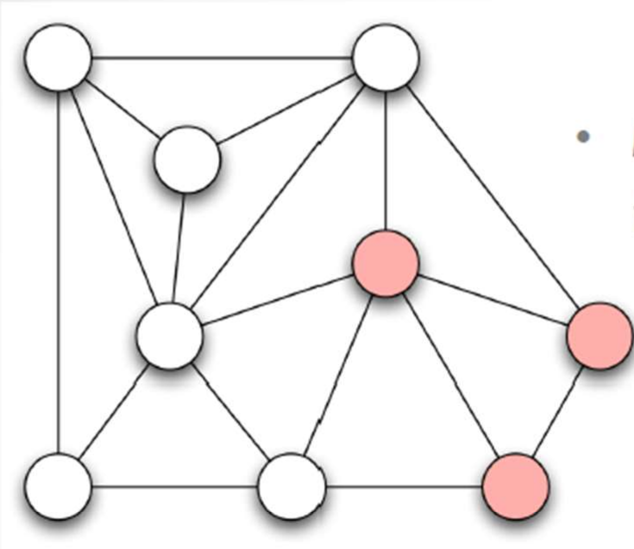
- Consider a given edge in this network. If we independently assign each node the gender male with probability p and the gender female with probability q, then
  - both ends of the edge will be male with probability p^2, and both ends will be female with probability q^2.
  - if the first end of the edge is male and the second end is female, or vice versa, then we have a cross-gender edge, so this happens with probability $2pq$.

Homophily Test



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

Homophily Test

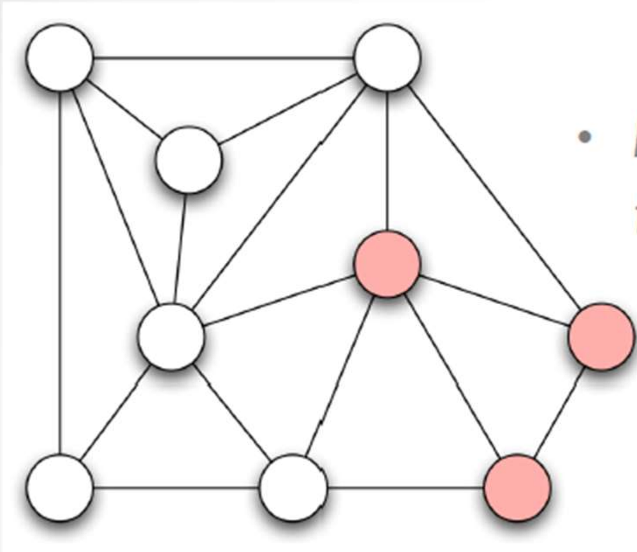


- *If the fraction of cross-gender edges is significantly less than $2pq$, then there is evidence for homophily.*

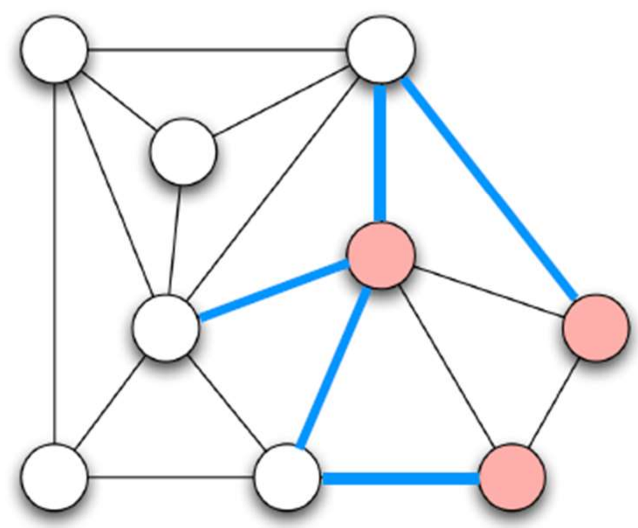
- *N = ? E = ?*
- *N = 9 ; |E| = 18*
- *p,q = ?*

- $p = 6/9 = \frac{2}{3}$
- $q = 3/9 = \frac{1}{3}$

Homophily Test



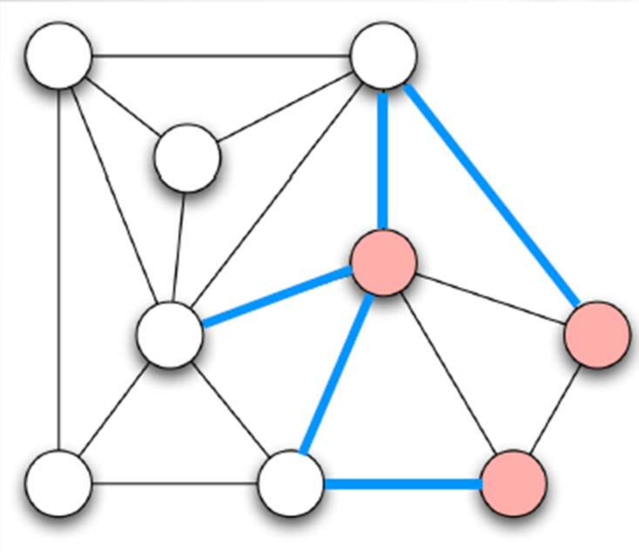
- *If the fraction of cross-gender edges is significantly less than $2pq$, then there is evidence for homophily.*



- $N = 9, |E| = 18$
- $p = 6/9 = \frac{2}{3}$
- $q = 3/9 = \frac{1}{3}$

- $2pq = 2 \times 2/3 \times 1/3 = \frac{4}{9}$
- fraction of cross-edges = $\frac{5}{18}$
- $\frac{5}{18} < \frac{4}{9} = \frac{8}{18}$

Homophily Test



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

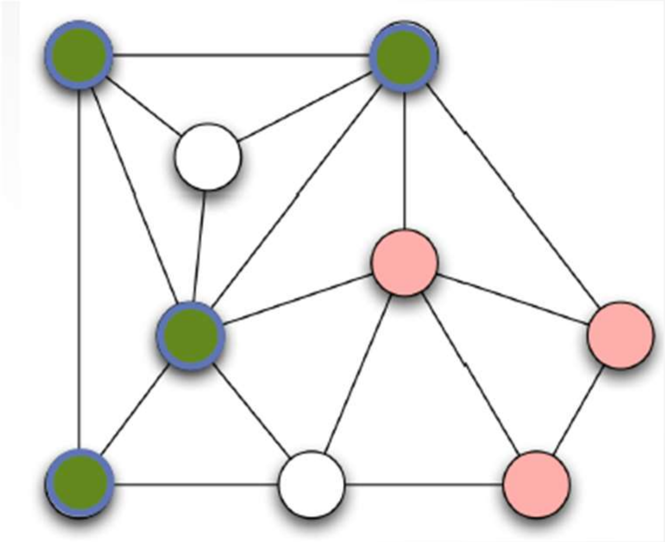
- The number of cross-gender edges in a random assignment of genders will deviate some amount from its expected value of 2pq.
  - To perform the test in practice one needs a working definition of "significantly less than."
  - Standard measures of statistical significance (quantifying the significance of a deviation below a mean) can be used for this purpose.

Inverse Homophily

- It's also easily possible for a network to have a fraction of cross-gender edges that is *significantly more than 2pq!*
- Network exhibits "inverse homophily" (for example, buyer-seller relationships)

More than 2 classes

- Homophily may extend to test any characteristic (race, ethnicity, age, etc.)
- If there are more than two classes, we can still do a similar calculation
- We say that an edge is *heterogeneous* if it connects two nodes that are different according to the characteristic in question.
- We then ask how the number of heterogeneous edges compares to what we'd see if we were to randomly assign values for the characteristic to all nodes in the network — using the proportions from the real data as probabilities.
- In this way, even a network in which the nodes are classified into many groups can be tested for homophily using the same underlying comparison to a baseline of random mixing.



Homophily Mechanism: Selection

- Homophily: "people tend to have links to others who are similar to them"
- But how are the links formed?
- **Immutable characteristics**:
  - A person's attributes that are determined at birth
  - Race, Ethnicity
  - Play a role in how this person's connections are formed over the course of his or her life.
- The tendency of people to form friendships with others who are *like them* is often termed ***selection.***
  - people are selecting friends with similar characteristics.
  - neighborhoods, schools,…

Homophily Mechanism: Social Influence

- Homophily: "people tend to have links to others who are similar to them"
- But how are the links formed?
- **Mutable characteristics**:
  - a person's behaviours, activities, interests, beliefs, and opinions.
  - people may modify their behaviours to bring them more closely into alignment with the behaviours of their friends.
- **Social influence** can be viewed as the reverse of selection:
  - with selection, the individual characteristics drive the formation of links
  - while with social influence, the existing links in the network serve to shape people's (mutable) characteristics.

Selection and Social Influence: Interplay

- When we look at a single snapshot of a network and see that people tend to share mutable characteristics with their friends,
  - It can be very hard to sort out the distinct effects and relative contributions of selection and social influence.
  - Have the people in the network adapted their behaviors to become more like their friends, or
  - Have they sought out people who were already like them?
- How do we answer this question?
- By doing longitudinal studies of a social network
  - Social connections and the behaviors within a group are both tracked over a period of time.
  - Behavioral changes that occur after changes in an individual's network connections,
  - vs. changes to the network that occur after an individual changes his or her behavior.

Selection and Social Influence: Interplay

- Example
  - Study the processes that lead pairs of adolescent friends to have similar outcomes in terms of scholastic achievement and delinquent behavior such as drug use.
- Empirical evidence confirms the intuitive fact that:
  - Teenage friends are similar to each other in their behaviors,
  - Both selection and social influence have a natural resonance in this setting.
    - Teenagers seek out social circles composed of people like them
    - Peer pressure causes them to conform to behavioural patterns within their social circles.
- One line of work has suggested that while both effects are present in the data [2]
  - The outsized role that earlier informal arguments had accorded to peer pressure (i.e. social influence) is actually more moderate
  - The effect of selection here is in fact comparable to (and sometimes greater than) the effect of social influence.

Selection and Social Influence: Interplay

- Understanding the tension between these different forces can be important not just for identifying underlying causes,
- Also for reasoning about the effect of *possible interventions* one might attempt in the system.
- Suppose we find that illicit drug use displays homophily across a social network — with students showing a greater likelihood to use drugs when their friends do:
  - We can ask about the effects of a program that targets certain high-school students and influences them to stop using drugs.
  - If the observed homophily is based on some amount of social influence, such a program could have a broad impact across the social network,
  - If the observed homophily is arising instead almost entirely from selection effects, then
    - the program may not reduce drug use beyond the students it directly targets:
    - as these students stop using drugs, they change their social circles and form new friendships.

Homophily Reasons

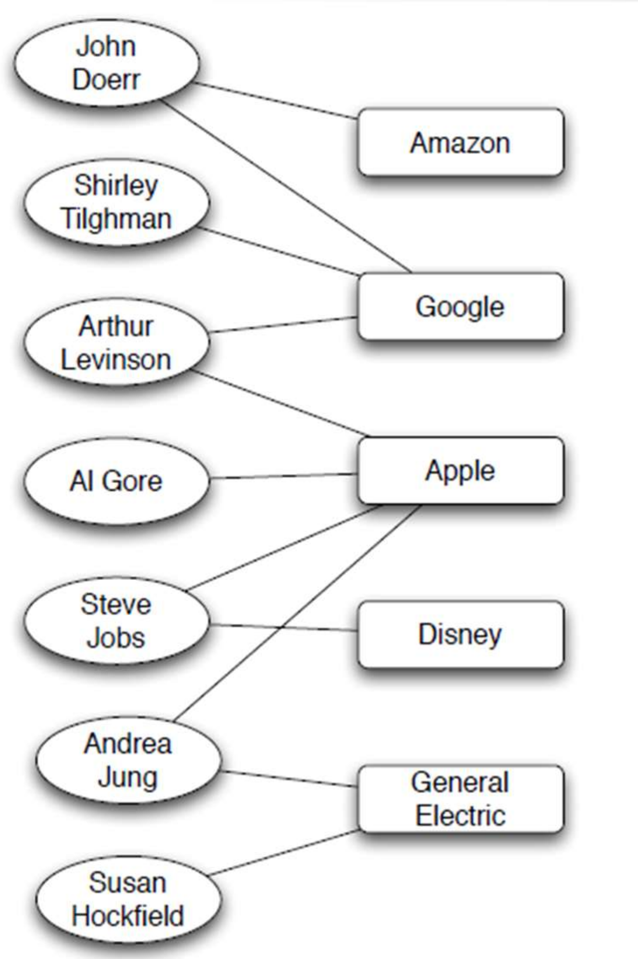
- An observation of homophily is often not an endpoint in itself.
- Why the homophily is present?
- How its underlying mechanisms will affect the further evolution of the network?
- How these mechanisms interact with possible external factors to influence the behavior of people in the network?

Affiliation

- Link formation in a network –
  - based on similarities in characteristics of the nodes, and
  - based on behaviors and activities that the nodes engage in.
- Surrounding contexts have been viewed as existing “outside” the network.
- We can place contexts into the network itself, by working with a larger network that contains both people and contexts as nodes.
- In principle we could represent any context this way
- Focus on how to represent the set of activities a person takes part in, and how these affect the formation of links.
- A very general view of the notion of an “activity" here.
  - Being part of a particular company, organization, or neigborhood
  - frequenting a particular place;
  - pursuing a particular hobby or interest
  - These are all activities that, when shared between two people, tend to increase the likelihood that they will interact and hence form a link in the social network
- We'll refer to such activities as **foci**
  - focal points of social interaction

Affiliation Network

- Bipartite Graph
- Two companies are implicitly linked by having the same person sit on both their boards;
  - possible conduits for information and influence to flow between different companies.
- Two people, on the other hand, are implicitly linked by serving together on a board
  - particular patterns of social interaction among some of the most powerful members of society



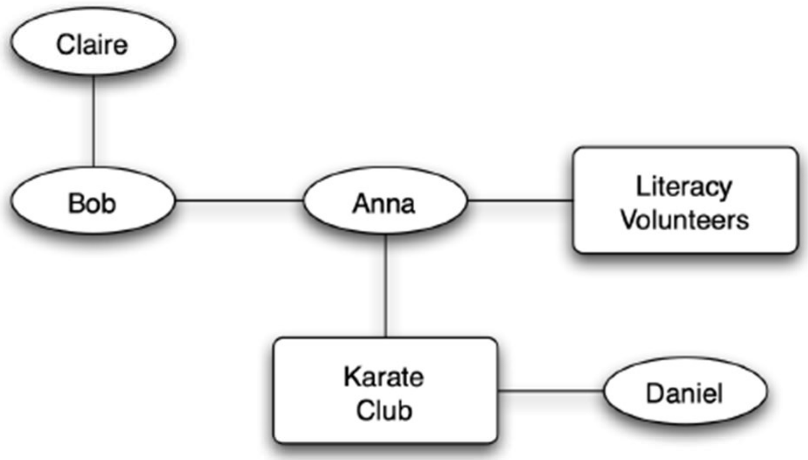
Co-Evolution of Social & Affiliation Networks

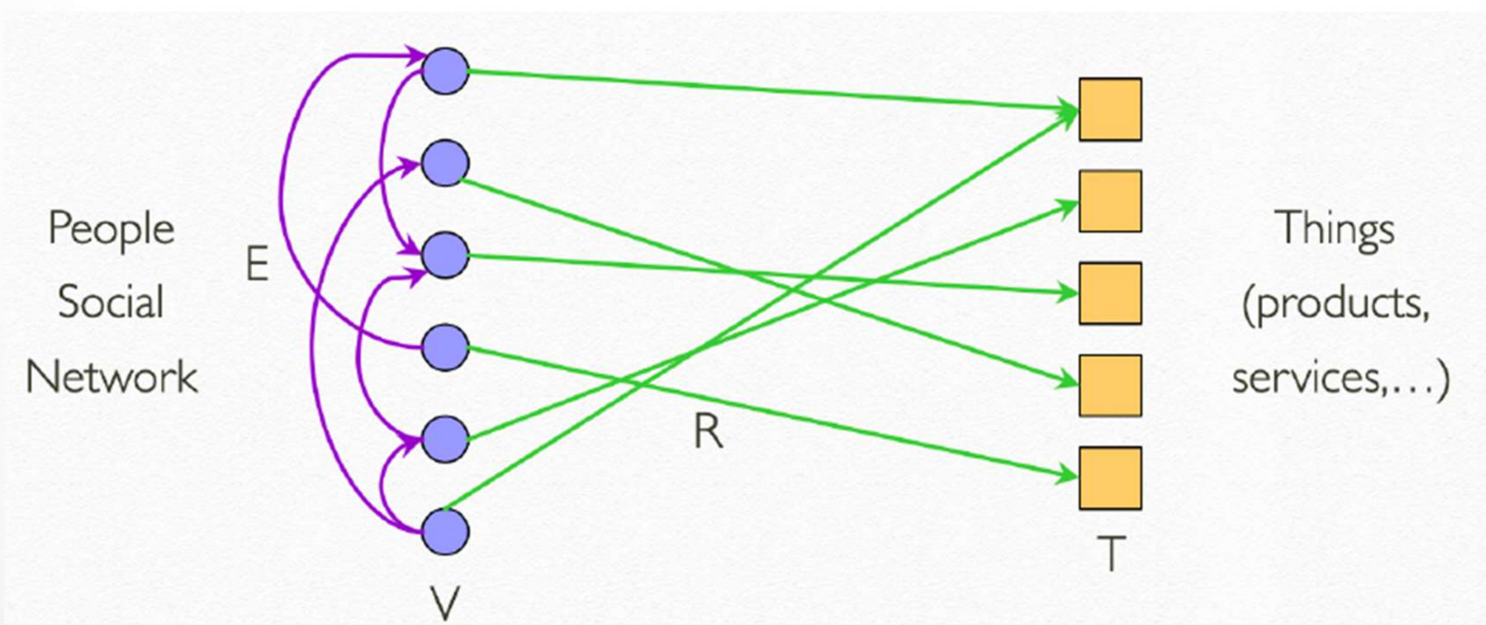
- Social networks and affiliation networks change over time.
  - New friendship links are formed
  - People become associated with new foci.
- Represent a kind of co-evolution that reflects the interplay between selection and social influence:
  - Selection: If two people participate in a shared focus, this provides them with an opportunity to become friends
  - Social Influence: If two people are friends, they can influence each other's choice of foci.

Social-Affiliation Networks

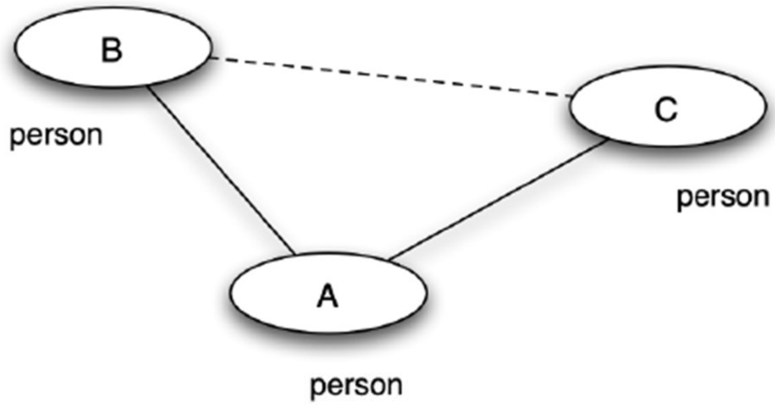
- Contains:
  - A social network on the people
  - An affiliation network on the people and foci
  - Two type of edges:
    1. An edge in a social network: it connects two people.
    2. An edge in an affiliation network: it connects a person to a focus.

Social-Affiliation Networks



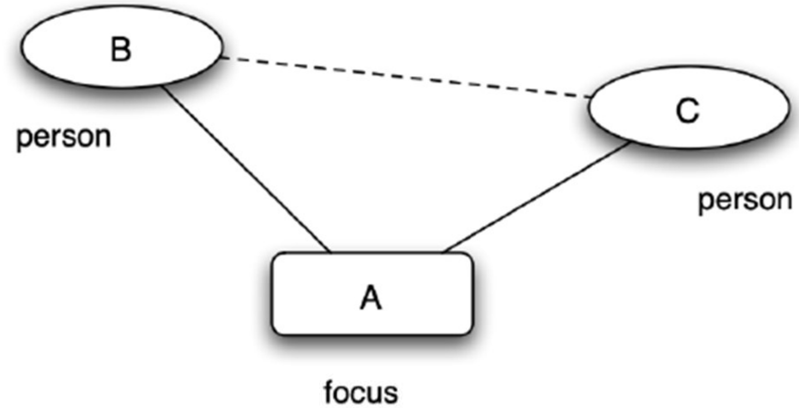


Social-Affiliation Networks: Closures



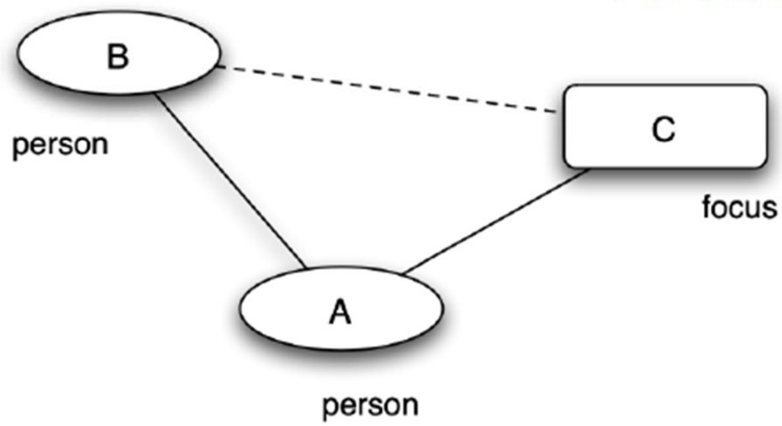
(a) *Triadic closure*

Bob introduces Anna to Claire



(b) *Focal closure*

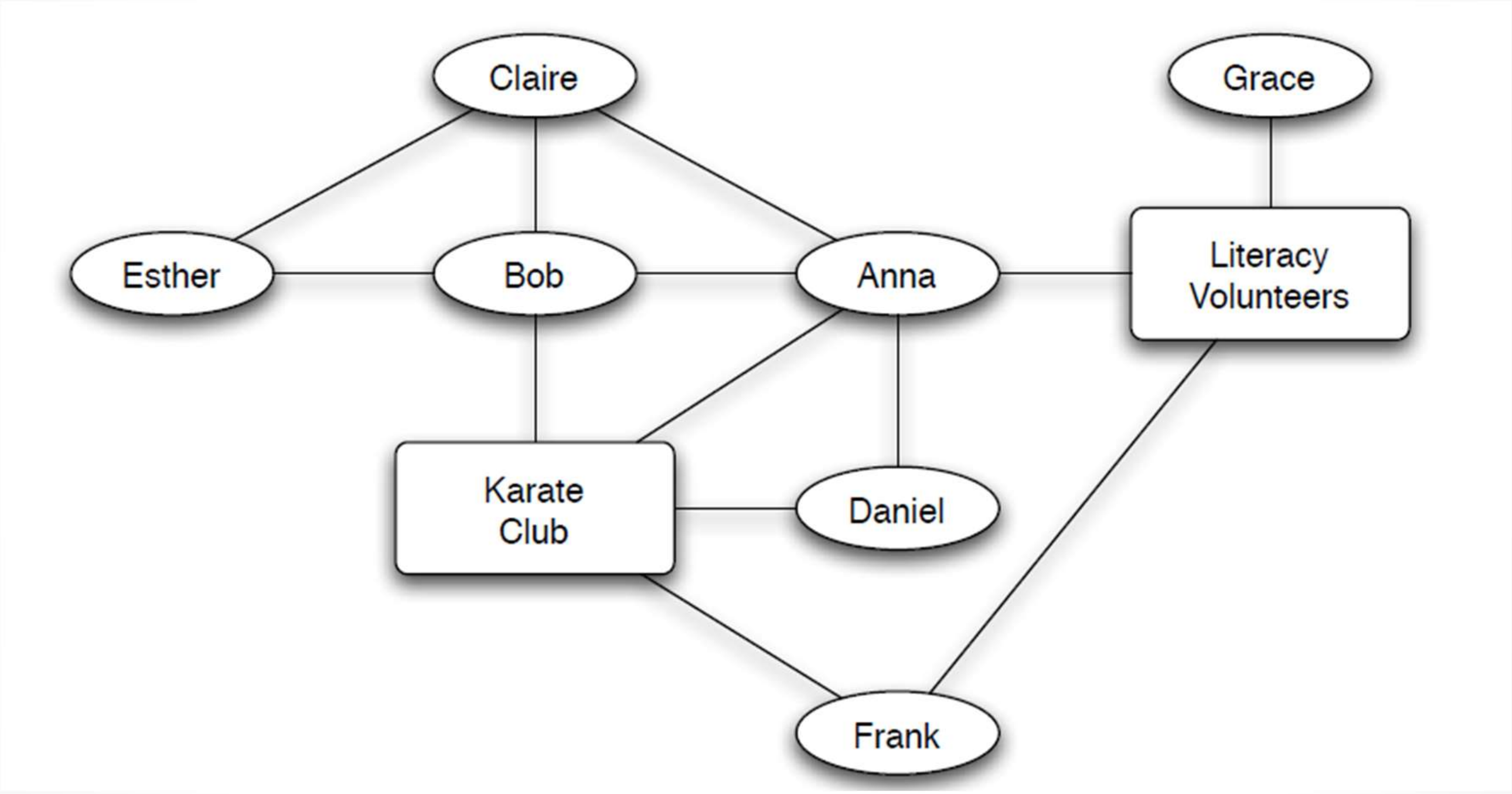
Art introduces Bob to Claire



(c) *Membership closure*

Anna introduces Bob to Chess

Social-Affiliation Networks: Link Formation



In Social-Affiliation Network pairs of people can have more than one friend (or more than one focus) in common;
how does this increase the likelihood that an edge will form between them?

Tracking Link Formation

- Given that most of the forces responsible for link formation go largely unrecorded in everyday life,
  - it is a challenge to select a large, clearly delineated group of *people* and social *foci*,
  - accurately quantify the relative contributions that these different mechanisms make to the formation of real network links.

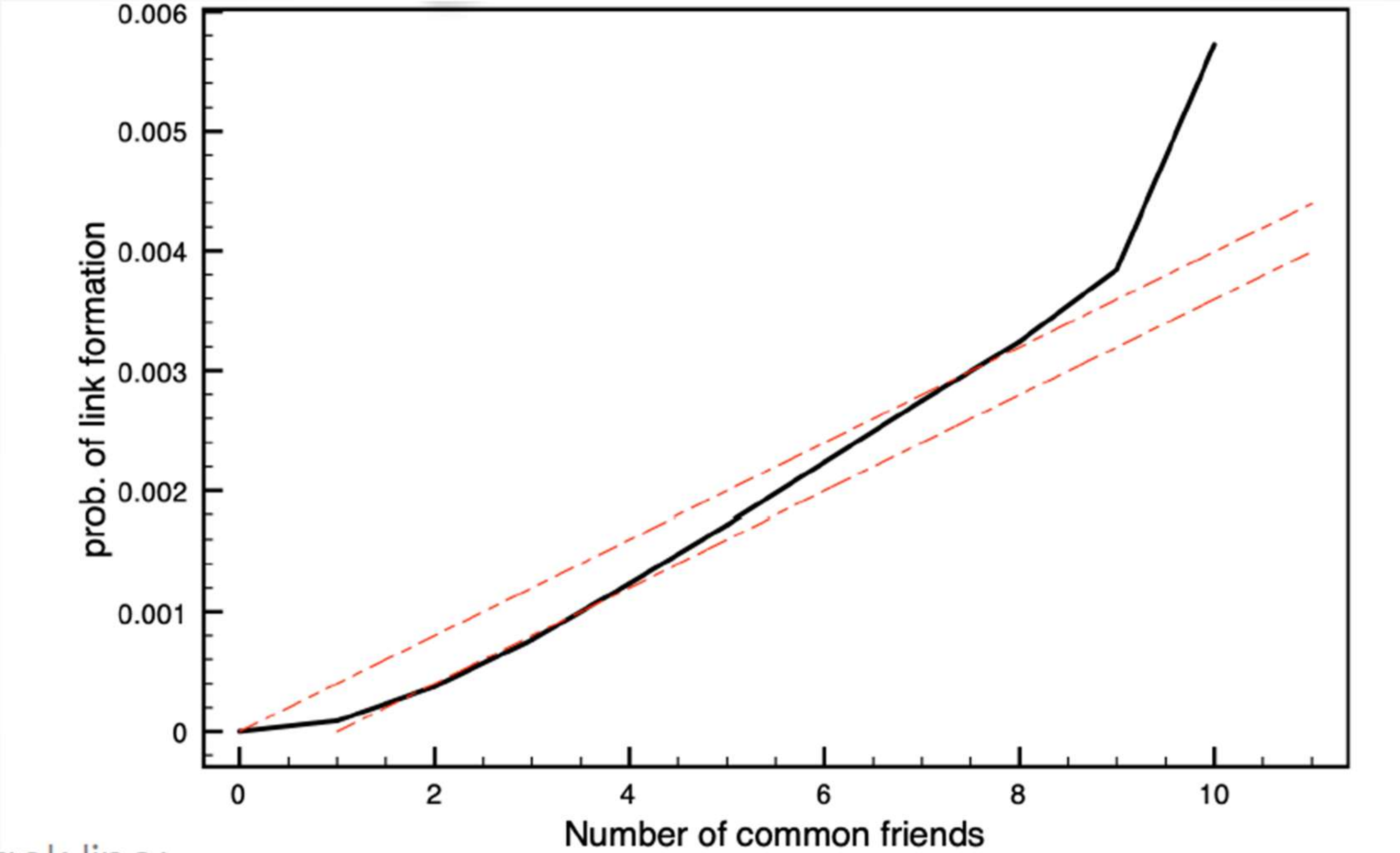
Tracking Link Formation in online data

- Given that most of the forces responsible for link formation go largely unrecorded in everyday life
- We take two snapshots of the network at different times.
- For each *k*, we identify all pairs of nodes who have exactly friends in common in the first snapshot, but who are not directly connected by an edge.
- We define *T(k)* to be the fraction of these pairs that have formed an edge by the time of the second snapshot.
  - (This is our empirical estimate for the probability that a link will form between two people with friends in common.)
- We plot *T(k)* as a function of *k* to illustrate the effect of common friends on the formation of links.
- *T(0)* is the rate at which link formation happens when it does not close a triangle
- The values of *T(k)* for larger *k* determine the rate at which link formation happens when it does close a triangle.

Tracking Link Formation in e-mail data

- Kossinets and Watts [3] computed this function *T(k)* using a dataset encoding the full history of e-mail ("who-talks-to-whom") communication among roughly 22,000 undergraduate and graduate students over a one-year period at a large U.S. university.
- Two people are connected by a link at a given instant if they had exchanged e-mail in each direction at some point in the past 60 days.
- They then determined an "average" version of *T(k)* by taking multiple pairs of snapshots:
  - A curve for *T(k)* on each pair of snapshots, and then
  - Averaged all the curves
  - The observations in each snapshot were one day apart
  - So their computation gives the average probability that two people form a link per day, as a function of the number of common friends they have.

Tracking Link Formation in e-mail data



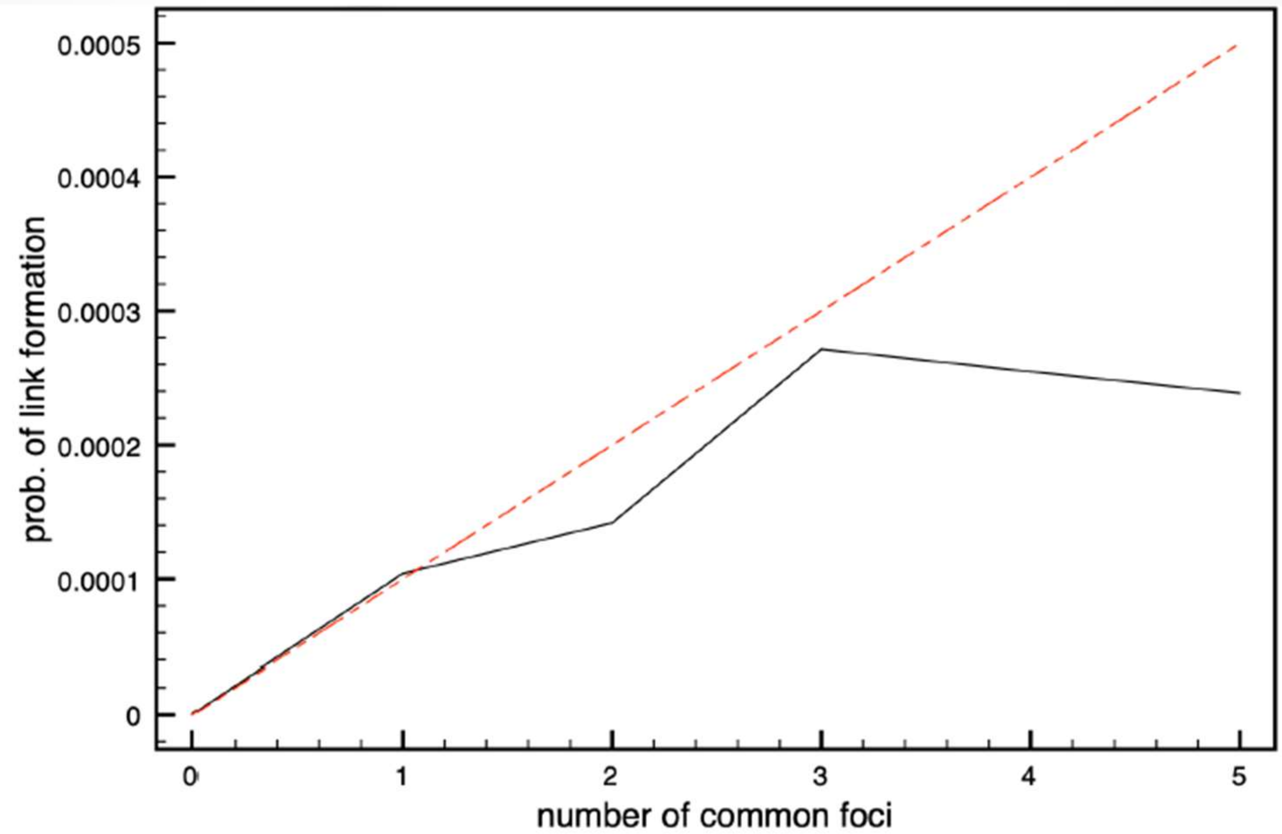
Solid-black line:

- T(0) is very close to 0.
- Probability of link formation increases steadily as the number of common friends increases.
- Particularly pronounced upward bend from 0 to 1 to 2 friends.

Tracking Link Formation in e-mail data - Baseline

- Baseline model - describing what one might have expected the data to look like in the presence of triadic closure.
- Suppose that for some small probability p, each common friend that two people have gives them an independent probability p of forming a link each day.
- So if two people have k friends in common, the probability they fail to form a link on any given day is $(1-p)^k$
  - this is because each common friend fails to cause the link to form with probability $1-p$, and these k trials are independent.
- Since $(1-p)^k$ is the probability the link fails to form on a given day, the probability that it does form, according to our simple baseline model, is $T_{baseline}(k) = 1-(1-p)^k$
- We plot this curve in as the upper dotted line.
- Given the small absolute effect of the first common friend in the data, we also show a comparison to the curve $1-(1-p)^{k-1}$, which just shifts the simple baseline curve one unit to the right.

Tracking Link Formation – Focal Closure



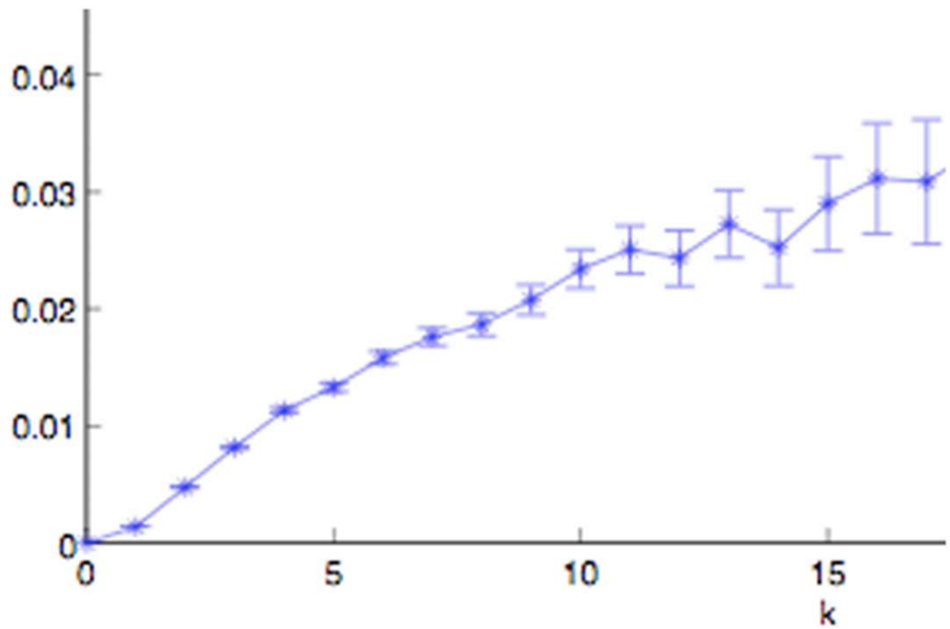
For focal closure, Kossinets and Watts supplemented their university e-mail dataset with information about the class schedules for each student.

- This turns downward and appears to approximately level off, rather than turning slightly upward. Thus, subsequent shared classes after the first produce a "diminishing returns" effect.

Tracking Link Formation – Membership Closure

- The social-affiliation network contains a node for each Wikipedia editor who maintains a user account and user talk page on the system
  - There is an edge joining two such editors if they have communicated, with one editor writing on the user talk page of the other.
- Each Wikipedia article defines a focus
  - An editor is associated with a focus corresponding to a particular article if he or she has edited the article.

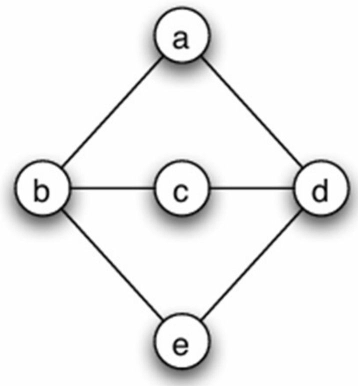
Tracking Link Formation – Membership Closure



Quantifying the effects of membership closure in a large online dataset: The plot shows the probability of editing a Wikipedia articles as a function of the number of friends who have already done so [4]

Exercise

- Consider the social network represented in the figure Suppose that this social net work was obtained by observing a group of people at a particular point in time and recording all their friendship relations. Now suppose that we come back at some point in the future and observe it again. According to the theorie based on empirical studies of triadic closure in networks, which new edge is most likely to be present? (I.e. which pair of nodes, who do not currently have an edge connecting them, are most likely to be linked by an edge when we return to take the second observation?) Also, give a brief explanation for your answer



References

1. James Moody. Race, school integration, and friendship segregation in America. American Journal of Sociology, 107(3):679{716, November 2001.
2. Jere M. Cohen. Sources of peer group homogeneity. Sociology in Education, 50:227-241,
3. Gueorgi Kossinets and Duncan Watts. Empirical analysis of an evolving social network. Science, 311:88-90, 2006.
4. David Crandall, Dan Cosley, Dan Huttenlocher, Jon Kleinberg, and Siddharth Suri.
5. Feedback effects between similarity and social inuence in online communities. In Proc. 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2008.

Readings

- Networks, Crowds, and Markets: Reasoning About a Highly Connected World
  https://www.cs.cornell.edu/home/kleinber/networks-book/
  - Chapter 4.1-4.4