CSC200: Lecture 6

Allan Borodin

Announcements

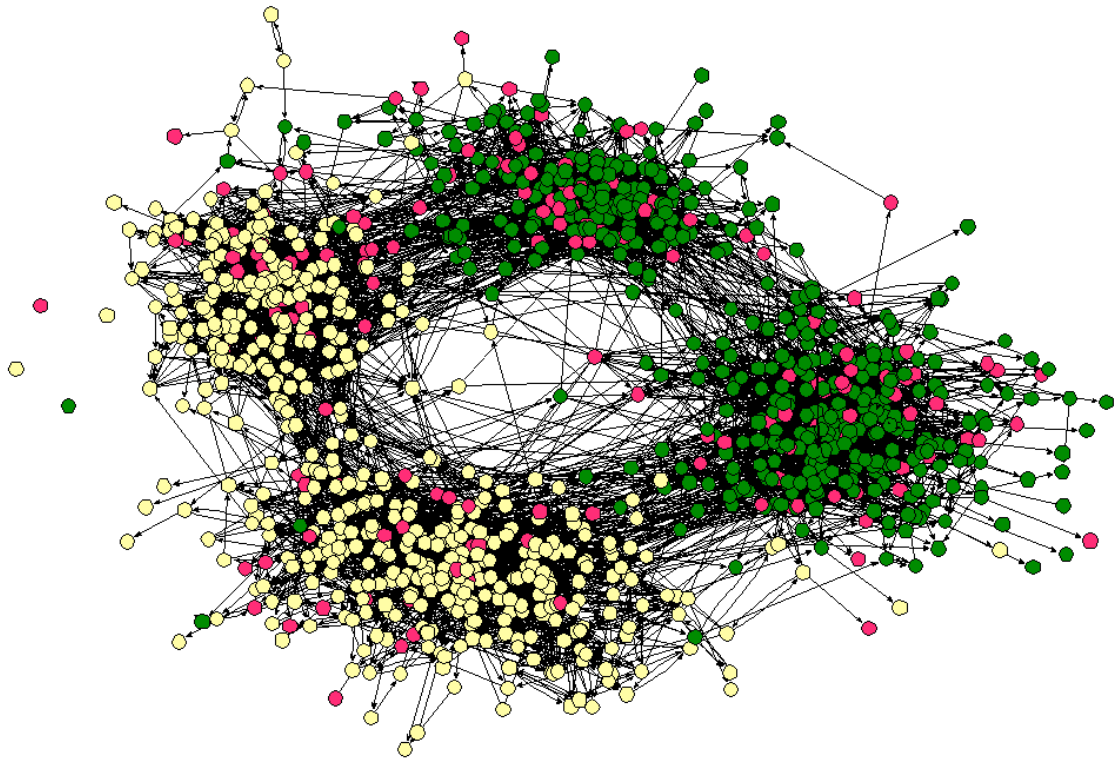
- First quiz in tutorial this Friday, October 2. Quiz will cover material in chapter 3.
  1. The quiz is worth 2 % of final grade.
  2. You are allowed to use one, two-sided 8.5" by 11" sheet of *handwritten* notes. No other materials or aids of any type are permitted.
- First assignment has now been posted and is due October 16.
- I am not available Tuesday, October 6 for office hours 2-3 but I am available 3-4 and as always by appointment or by luck (i.e. dropping if I am available).

- Last lecture:
  - We finished Chapter 3.
  - We then started the topic of homophily based on factors which are immutable (race, gender,age) or more mutable (interests, economic status, neighbourhoods).
  - For immutable factors, it seems that (to some extent) homophily comes about by "selection".

- This lecture: Continuing discussion of Chapter 4.

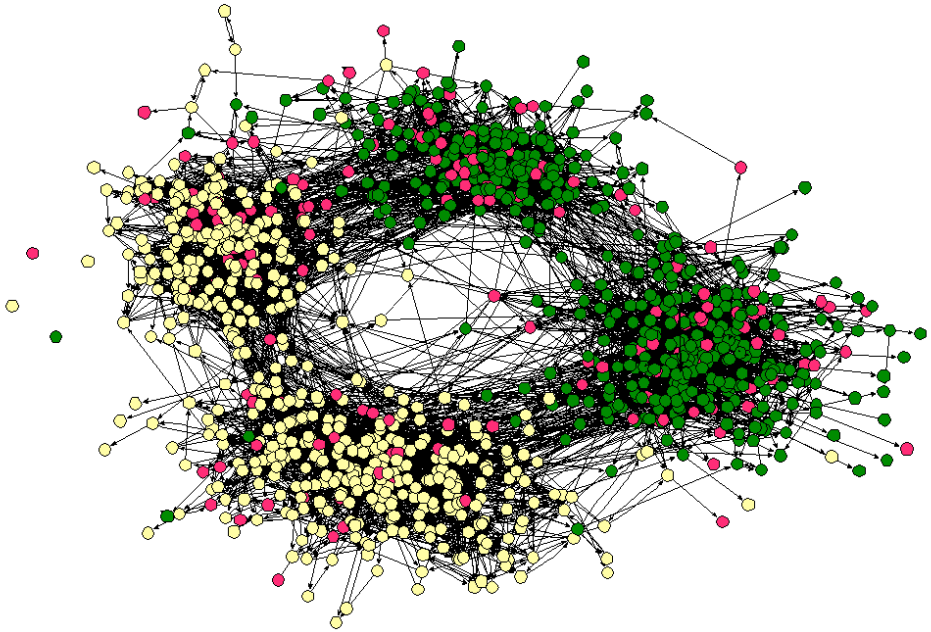
Homophily

- Homophily: we tend to be similar to our friends.
- Last lecture we discussed the network of friendships in a combined junior and senior high school visualized in Fig 4.1 of the text.



- There are questions about the meaning of such a perhaps controversial figure and the claim of homophily based on race.

Comments on figure 4.1



[Fig. 4.1, textbook]

- The visualization does suggest homophily based on race and the junior/senior high split.
- But how can we distinguish the impact of initial same race friendships (say based on residential neighbourhoods) from same race preferences?

Measuring homophily

- As mentioned before, when networks are large (and/or when homophily is less dramatic) it is difficult if not impossible to visualize various aspects of a network and so one needs a measure of homophily (whether the cause or the consequence of the network).

- Suppose we wish to study the likelihood of friendships according to some factor (with say two values) such as gender.

- Think Big!: Lets think in terms of large social networks where the presense or absense of a given individual will not have any noticeable impact on the probability of any phenomena.

Thought experiment

- What would it mean to say that a social network does or does not exhibit homophily according to some factor such as gender?

- Consider a given network where the fraction (i.e. probability) of males is p and the fraction of females is q.
  - Consider a given edge (u, v) in the network.
  - If gender has no correlation with relations, then the probability that the genders of u and v are different is $2pq$. Why?

  Precisely, what does it mean to say that friendship is "independent" of gender?

Thought experiment

- What would it mean to say that a social network does or does not exhibit homophily according to some factor such as gender?

- Consider a given network where the fraction (i.e. probability) of males is p and the fraction of females is q.
  - Consider a given edge (u, v) in the network.
  - If gender has no correlation with relations, then the probability that the genders of u and v are different is $2pq$. Why?

  Precisely, what does it mean to say that friendship is "independent" of gender?

- Random events E_1 and E_2 are independent if and only if (iff)
  $Prob[E_1 \text{ and } E_2] = Prob[E_1] \cdot Prob[E_2]$ iff
  $Prob[E_1] = Prob[E_1|E_2]$

A Test for Homophily

- This leads to a homophily test: If the actual fraction of cross-gender edges is "significantly less than" $2pq$ then there is evidence for homophily. What if the actual fraction of cross-gender edges is significantly higher?

A Test for Homophily

- This leads to a homophily test: If the actual fraction of cross-gender edges is "significantly less than" $2pq$ then there is evidence for homophily. What if the actual fraction of cross-gender edges is significantly higher?
- Clearly the meaning of an edge is an essential aspect of any study; e.g. consider the difference between an edge representing collaboration in a course project vs an edge meaning a romantic relationship.

Equivalent restatement of the test

- What would it mean to say that a network *does* exhibit homophily according some some factor such as gender?

Equivalent restatement of the test

- What would it mean to say that a network *does* exhibit homophily according some some factor such as gender?

- Consider again the network where the fraction (i.e. probability) of males is p and the fraction of females is q with (of course) $p + q = 1$.
  - Consider a given edge (u, v) in the network.
  - If gender has no effect on relations being formed, then the probability that the genders of u and v are the same is $p^2 + q^2$.

Equivalent restatement of the test

- What would it mean to say that a network *does* exhibit homophily according some some factor such as gender?

- Consider again the network where the fraction (i.e. probability) of males is p and the fraction of females is q with (of course) $p + q = 1$.
  - Consider a given edge (u, v) in the network.
  - If gender has no effect on relations being formed, then the probability that the genders of u and v are the same is $p^2 + q^2$.

- This leads to the equivalent homophily test: If the actual fraction of same-gender edges is "significantly more than" $p^2 + q^2$ then there is evidence for homophily.

- Sanity check: $1 - (p^2 + q^2) = 2pq$. What is this equation saying and why is it true?

Selection vs social influence continued

- With immutable factors (such as race and gender), when we observe evidence of homophily, we often attribute increased friendships to selection. (Note again that race often correlates with other factors.)
  - Selection: tendency to form friendships with others who are like them.

Selection vs social influence continued

- With immutable factors (such as race and gender), when we observe evidence of homophily, we often attribute increased friendships to selection. (Note again that race often correlates with other factors.)
  - Selection: tendency to form friendships with others who are like them.

- But when considering more mutable factors, there is a feedback between similar characteristics and social links.
  - To what extent does behaviour get modified by our social network? That is, to what extent is social influence determining behaviour?

- Of course, both selection and social influence can be interacting in the same social network. How does one understand the relative interplay?

Selection vs social influence continued

- With immutable factors (such as race and gender), when we observe evidence of homophily, we often attribute increased friendships to selection. (Note again that race often correlates with other factors.)
  - Selection: tendency to form friendships with others who are like them.

- But when considering more mutable factors, there is a feedback between similar characteristics and social links.
  - To what extent does behaviour get modified by our social network? That is, to what extent is social influence determining behaviour?

- Of course, both selection and social influence can be interacting in the same social network. How does one understand the relative interplay?

Longitudinal studies make it possible to see behavioral changes that occur after changes in an individuals network connections, as well as changes to the network that occur after an individual changes his or her behavior.

Two interesting longitudinal studies

- In academic success (or drug usage) in teenage friendship networks, Cohen and Kandel claim that peer pressure (i.e. social influence) is less a factor here than previously believed.

- Christakis and Fowler (2007) report on obesity patterns of 32,000 people observed over a 32 year period: (in contrast to the above example) they claim that obesity (or, less controversially, keeping fit) is to some extent a contagious disease spread within a social network.

Why the obsesity homophily?

- Three possibilities identified by Christakis and Fowler:
  1. selection
  2. homophily being driven by other factors that correlate with obsesity (e.g. poverty)
  3. the social influence of peer pressure.

- Christakis and Fowler conclude that even accounting for 1 and 2, social influence is a significant factor.

Why the obsesity homophily?

- Three possibilities identified by Christakis and Fowler:
  1. selection
  2. homophily being driven by other factors that correlate with obsesity (e.g. poverty)
  3. the social influence of peer pressure.

- Christakis and Fowler conclude that even accounting for 1 and 2, social influence is a significant factor.

- Upon further thought, is it so surprising that obesity is to some extent a contagious disease? Perhaps friendships are leading to increased obesity but the possible implication that this is caused by "peer pressure" (or even "social acceptance") should not necessarily be infered. That is, social influence (i.e. that friendships lead to a change in behaviour) should be interpreted more broadly. How else can friendships lead to an increased chance of obesity?

- Clearly, observing homophily is only a starting point.

Selection vs social influence: Why do we care?

- How do we study the relative interplay between selection and social influence: why do we want to consider this chicken vs. egg question?
- If indeed social influence is a significant factor, then targeting key individuals and trying to modify undesirable behaviour can be effective since to some extent we are then viewing such behaviour as a process of influence spread.
- If not, focusing on a few individuals will at best change the behaviour of a few individuals.

Social-affiliation networks: incorporating context into the network

- Up to now we have viewed contextual (mutable and immutable) factors that affect the formation of links to be outside of the social network being considered.
- Section 4.3 discusses how to include context in the network so as to have a common framework for studying the interplay between the extent of (social) triadic closure (common friendships induce new friendships), homophily determined by selection, and mutual activity determined by social influence.

Social-affiliation networks: incorporating context into the network

- Up to now we have viewed contextual (mutable and immutable) factors that affect the formation of links to be outside of the social network being considered.
- Section 4.3 discusses how to include context in the network so as to have a common framework for studying the interplay between the extent of (social) triadic closure (common friendships induce new friendships), homophily determined by selection, and mutual activity determined by social influence.
- Let's consider the (mutable) context of affiliation in a group/participation in an activity. Such an activity is referred to as a foci, a focal point for social interaction.

Social-affiliation networks: incorporating context into the network

- Up to now we have viewed contextual (mutable and immutable) factors that affect the formation of links to be outside of the social network being considered.
- Section 4.3 discusses how to include context in the network so as to have a common framework for studying the interplay between the extent of (social) triadic closure (common friendships induce new friendships), homophily determined by selection, and mutual activity determined by social influence.
- Let's consider the (mutable) context of affiliation in a group/participation in an activity. Such an activity is referred to as a foci, a focal point for social interaction.
- We incorporate such foci into social networks by considering a focus to be a different type of node, distinct from a node representing an individual. We first consider a pure affiliation network, an example being of which we have already seen in a bipartite graph with individuals and corporate boards.

Example of a pure affiliation network

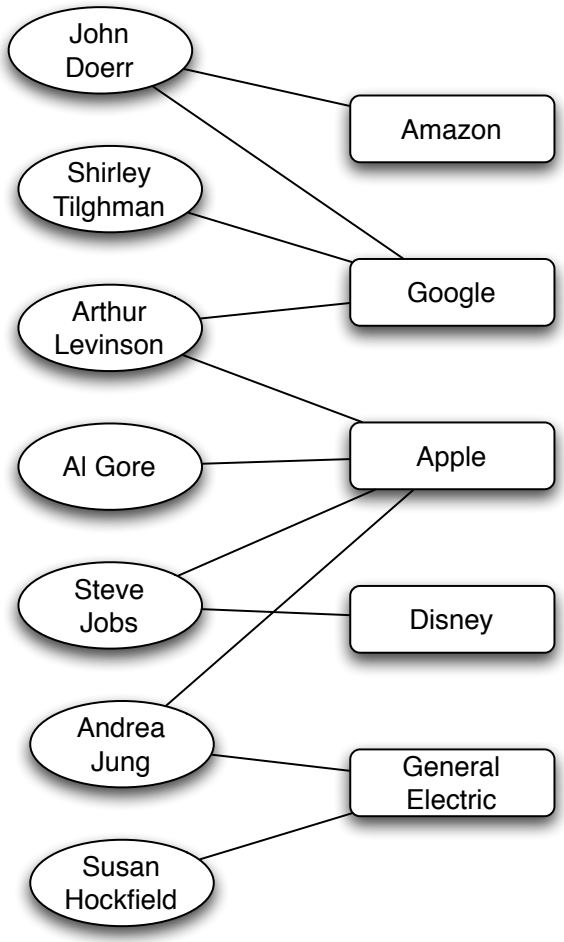


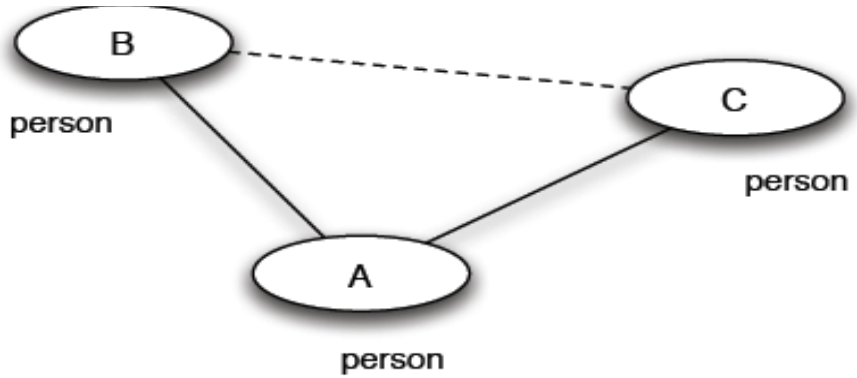
Figure : [E&K, Fig 4.4] One type of affiliation network that has been widely studied is the memberships of people on corporate boards of directors. A very small portion of this network (as of mid-2009) is shown here.

Social-affiliation networks continued

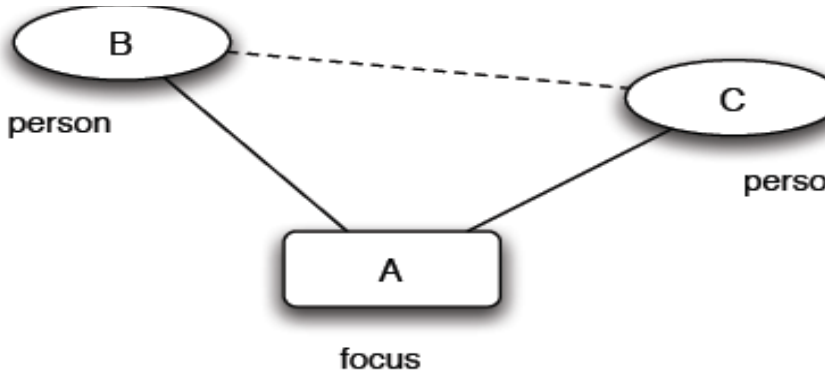
We can then combine the people-people edges of a social network with the people-focus edges of an affiliation network to form a social-affiliation network. Within such a combined network, we can discuss three types of graph triangle closures:

- triadic closure as introduced in chapter 3 where common friends of one or more individuals become friends
- focal closure where individuals become friends based on their common interest(s)
- membership closure where an individual joins an activity because a friend (or a group of friends) is (are) already in that activity

Three types of closure

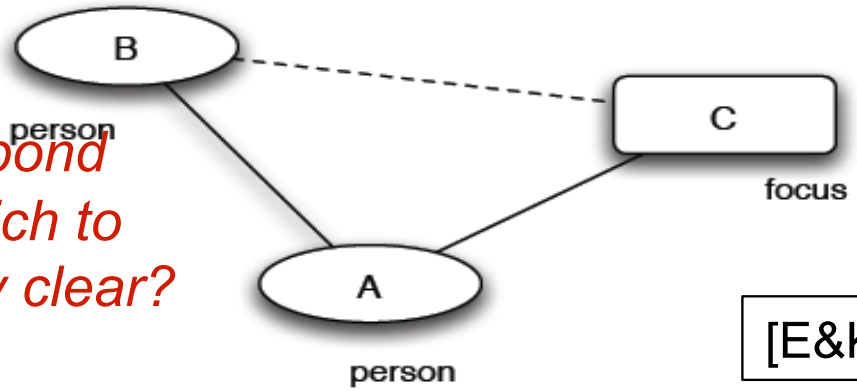


(a) *Triadic closure*



(b) *Focal closure*

Which of these correspond to social influence, which to selection? Is it still fully clear?



(c) *Membership closure*

[E&K, Ch.4, Fig. 4.6]

Figure : [E&K, Fig 4.6] Three types of closure

Toy example of a social-affiliation network

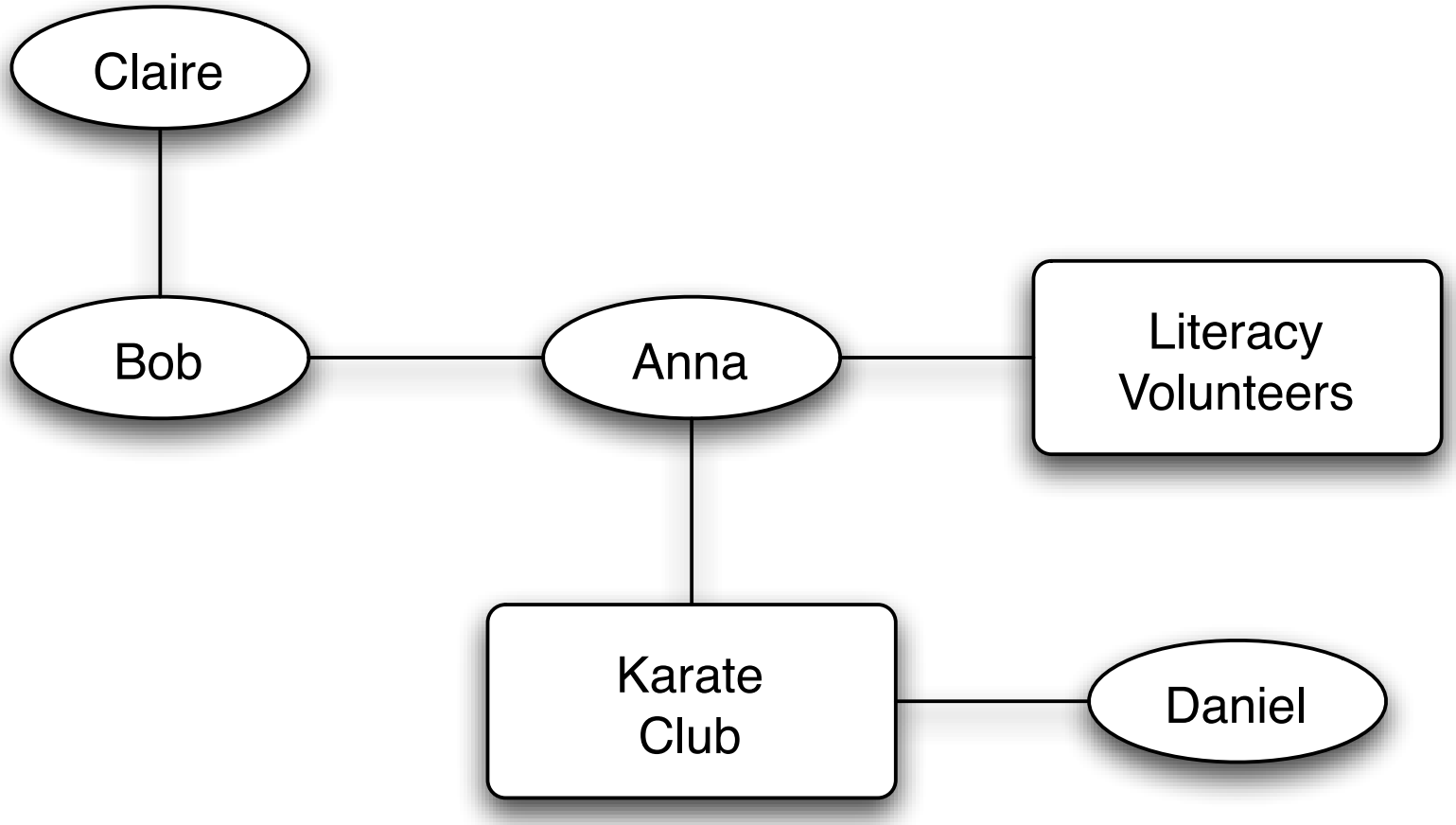


Figure : [E&K, Fig 4.5] In this social-affiliation network, the oval nodes are people and the rectangular nodes are activities. What kinds of triagular closures can occur?

Toy example showing three types of closure

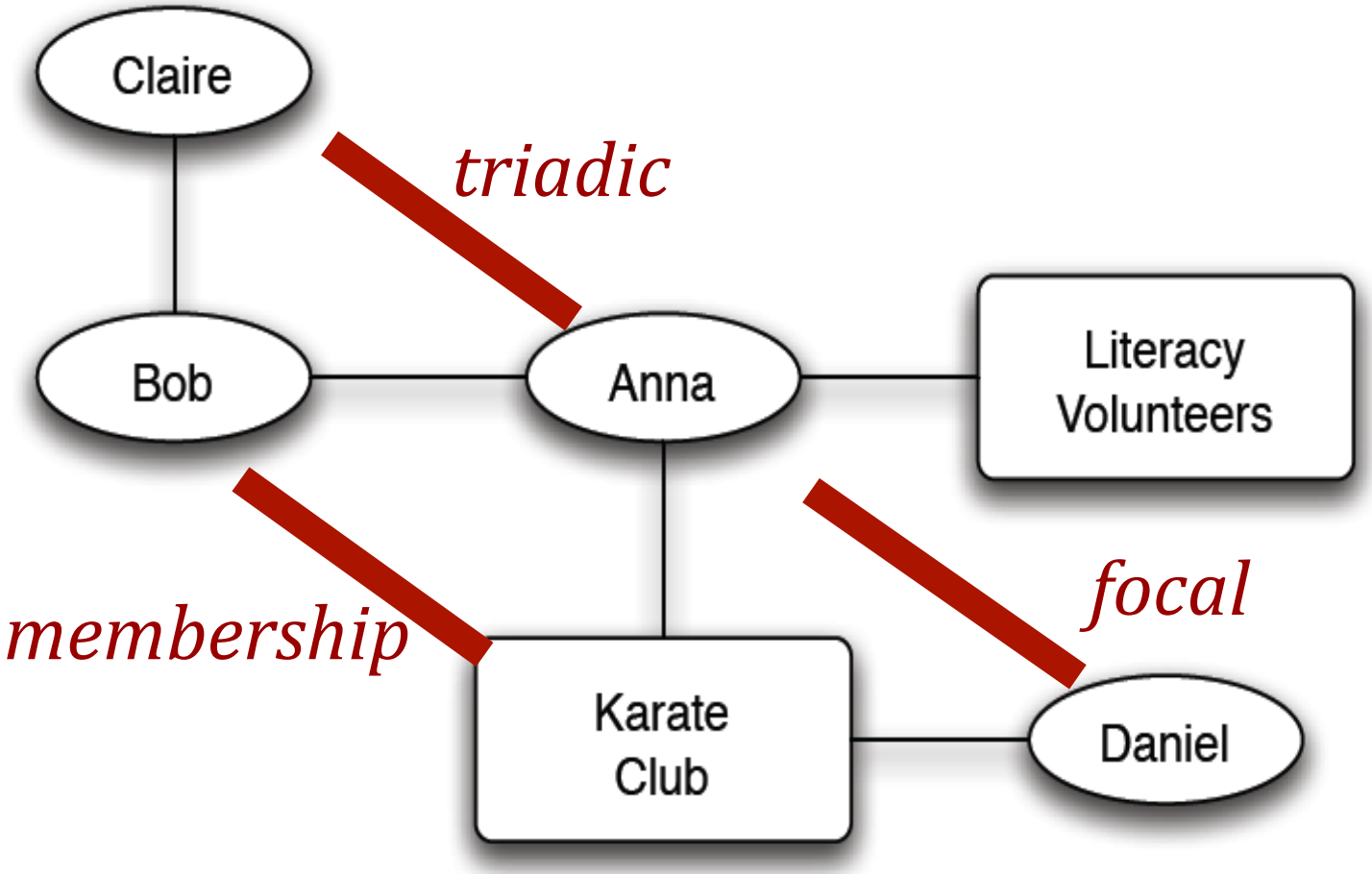


Figure : [E&K, Fig 4.7] We can observe the three types of triangular closures that have occured in some time period.

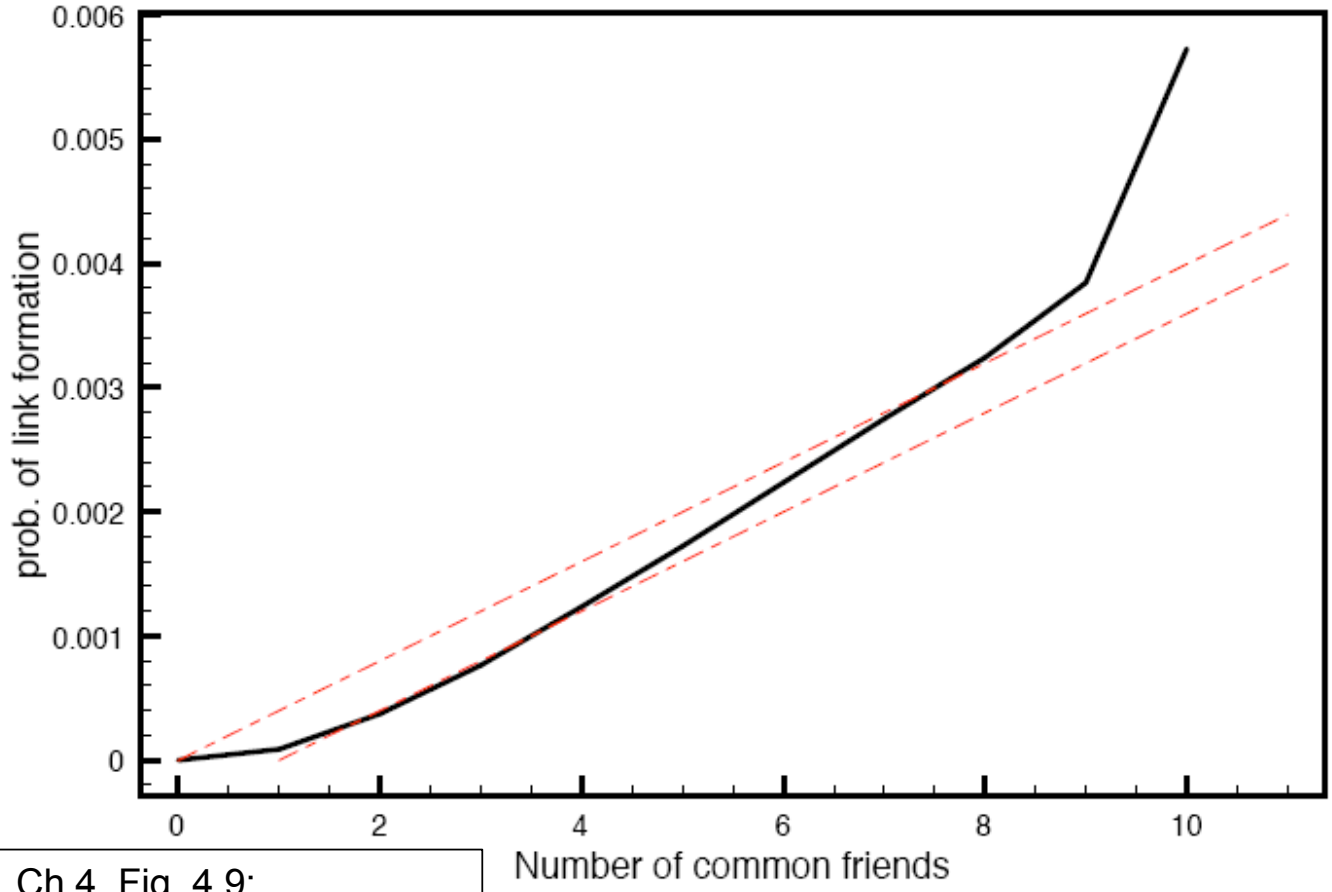
How do we measure extent of these processes?

- Closure is inherently dynamic
  - So we need to take snapshots of the network at different times to see how the relationships evolve and to what extent each form of closure occurs
  - If common friends or common interests are causing new links (i.e., closures) then the more friends or interests in common, the more we should see this effect.
- We briefly look at a couple studies stemming from online interactions, but realize the usual warning about limitations of such studies
  - As in all modeling we may be missing many factors
  - The timing of the snapshots may influence results
  - These particular studies look at link formation, but not link dissolution. What would the network look like if links formed but never dissolved?

Triadic closure: dependence on number mutual friends

- Student email use, large US university [Kossinets, Watts 2006]
- "Friends" defined as two-way email communication (prev. 60 days)
- Measure probability $T(k)$ of a new friendship emerging between a pair of students as a function of the number k of mutual friends
- that is, the probability of it happening in any given day (averaging over many such pairs)
- Compare data (black) with baseline theoretical model (red) baseline: assume any single mutual friend will generate a new friendship with probability p and that this will happen independently for each common friend. Thus $T(k) = 1 - (1 - p)^k$ Why?
- For small p, $(1 - p)^k \approx 1 - pk$ so that $T(k) \approx pk$.

Probability (per-day) of triadic closure as a function of the number of common friends



[E&K, Ch.4, Fig. 4.9;
from Kossinets and Watts, 2006]

Figure : [E&K, Fig 4.9]

Observations

- Data does not show much more propensity for friendship when going from zero to one mutual friend.
  - The second dashed red line shifts the curve over by one friend so as to better compare the actual data and baseline model.
  - Why no major impact with one common friend?
- Increasing from 1 to 9 friends shows linear curve (greater slope than baseline)
- A sharp difference going beyond 9 friends
  - The theoretical model (and its assumption of independence) no longer supported
  - Is there some threshold of mutual friends which escalates the pressure for triadic closure?

Exercise: translate per-day probability into per-month or per-year probability