CSC200: Lecture 5

Allan Borodin

Announcements

- First quiz will take place Friday, October 2 in the tutorial rooms, SS 1069 and SS 1088. The quiz will cover material in chapters 2,3,4. I will be a little more specific in Wednesdays lecture.
- Tutorials are an integral part of this course. So whether you find all this initial material difficult or easy, it is important to attend tutorials. There will more things to discuss folloiwng the quiz, so best not to not leave right after the quiz.

- Last lecture:
  - Strong vs. weak ties and the strength of weak ties.
  - Correspondence between
    1. (quantitative) weakness of tie (a local property of an edge)
    2. closeness to being a local bridge measured by neighbourhood overlap (a somewhat more global network property)
  - Clustering coefficient of a node measuring the extent to which triadic closure has taken place (for whatever reason)
  - The strong triadic closure as a "working assumption".

- This lecture:
  - Quick review and some further discussion of Chapter 3.
  - Start of Chapter 4.

- This week's tutorial: Quiz followed by discussion of basic probability concepts which are central to chapter 4 of text.

Information spread in a passive network

- The maintained or passive relation network (as in the Facebook networks on slide 19 of Lecture 4) is said to occupy a middle ground between
  1. strong tie network (in which individuals actively communicate), and
  2. very weak tie networks (all "friends") with many old (and inactive) relations.

- "Moving to an environment where everyone is passively engaged with each other, some event, such as a new baby or engagement can propagate very quickly through this highly connect neighborhood."

- We can add that an event might be a political demonstration.

"Tightly knit communities" connected by weak ties

- The intuitive concept of tightly knit communities occurs several times in Chapter 3 but is deliberately left undefined.

- In a small network we can sometimes visualize the tightly knit communities but one cannot expect to do this is a large network. That is one needs algorithms and this is the topic of the advanced material in Section 3.6.

"Tightly knit communities" connected by weak ties

- The intuitive concept of tightly knit communities occurs several times in Chapter 3 but is deliberately left undefined.

- In a small network we can sometimes visualize the tightly knit communities but one cannot expect to do this is a large network. That is one needs algorithms and this is the topic of the advanced material in Section 3.6.

- Recalling the relation to weak ties, the text calls attention to how nodes at the end of one (or especially more) local bridges can play a pivotal role in a social network.

- These "gatekeeper nodes" between communities stand in contrast to nodes which sit at the center of a tightly knit community.

Central nodes vs. gatekeepers

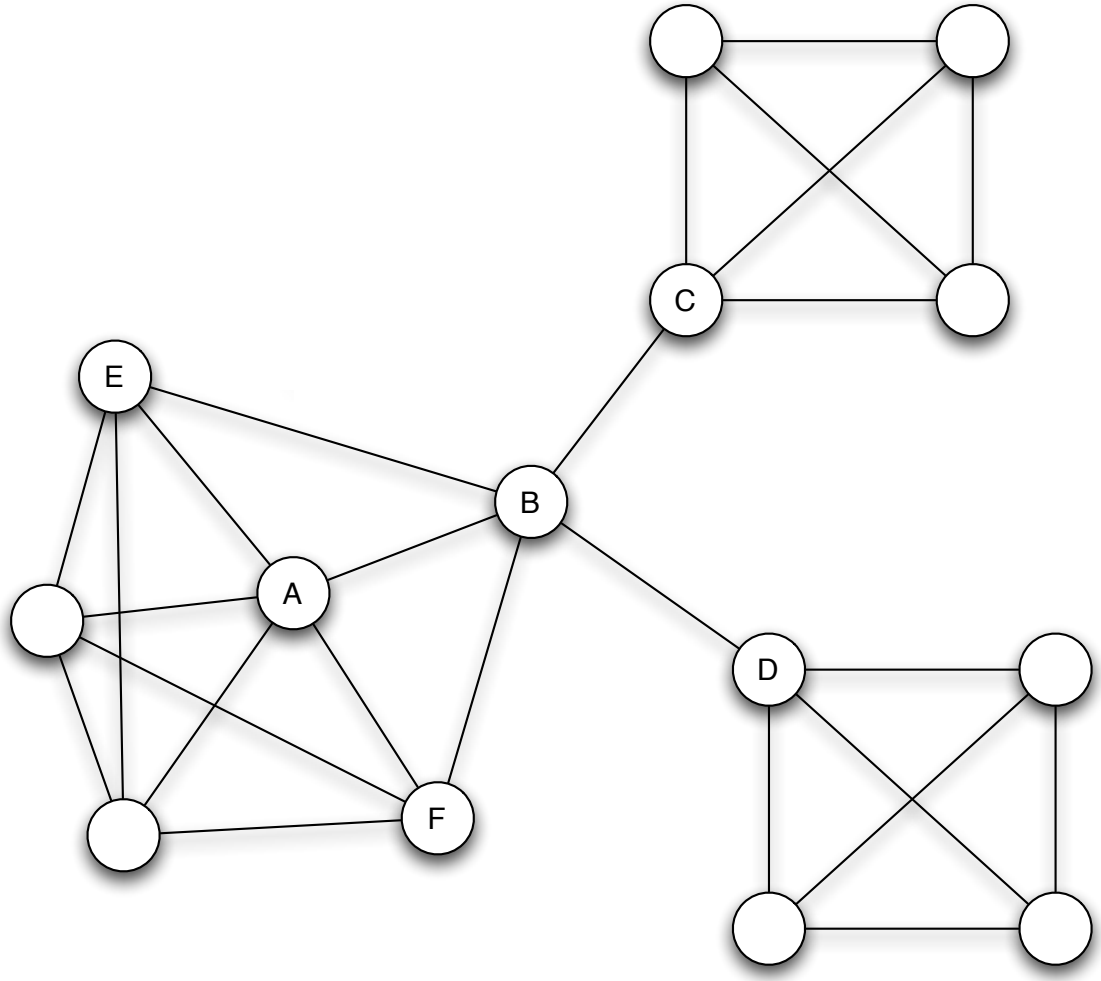


Figure : The contrast between densely-knit groups and boundary-spanning links is reflected in the different positions of central node A and gatekeeper node B in the underlying social network. [Fig 3.11, textbook]

Social capital of nodes A and B

- The edges adjacent to node A all have high embeddedness. Visually one sees node A as a central node in a tightly-knit cluster. As such, the social capital that A enjoys is its "bonding capital" in that the actions of A can (for example) induce norms of behaviour because of the trust in A.
- In contrast, node B is a bridge to other parts of the network. As such, its social capital is in the form of "brokerage" or "bridging capital" as B can play the role of a "gatekeeper" (of information and ideas) between different parts of the network. Furthermore, being such a gatekeeper can lead to creativity stemming from the synthesis of ideas.
- Some nodes can have both bonding capital and bridging capital.

Florentine marriages: Bridging capital of the Medici

- The Medici are connected to more families, but not by much.
- More importantly: Five of the seven edges adjacent to the Medici are bridges or local bridges and the Medici lie on the shortest paths between most pairs of families.

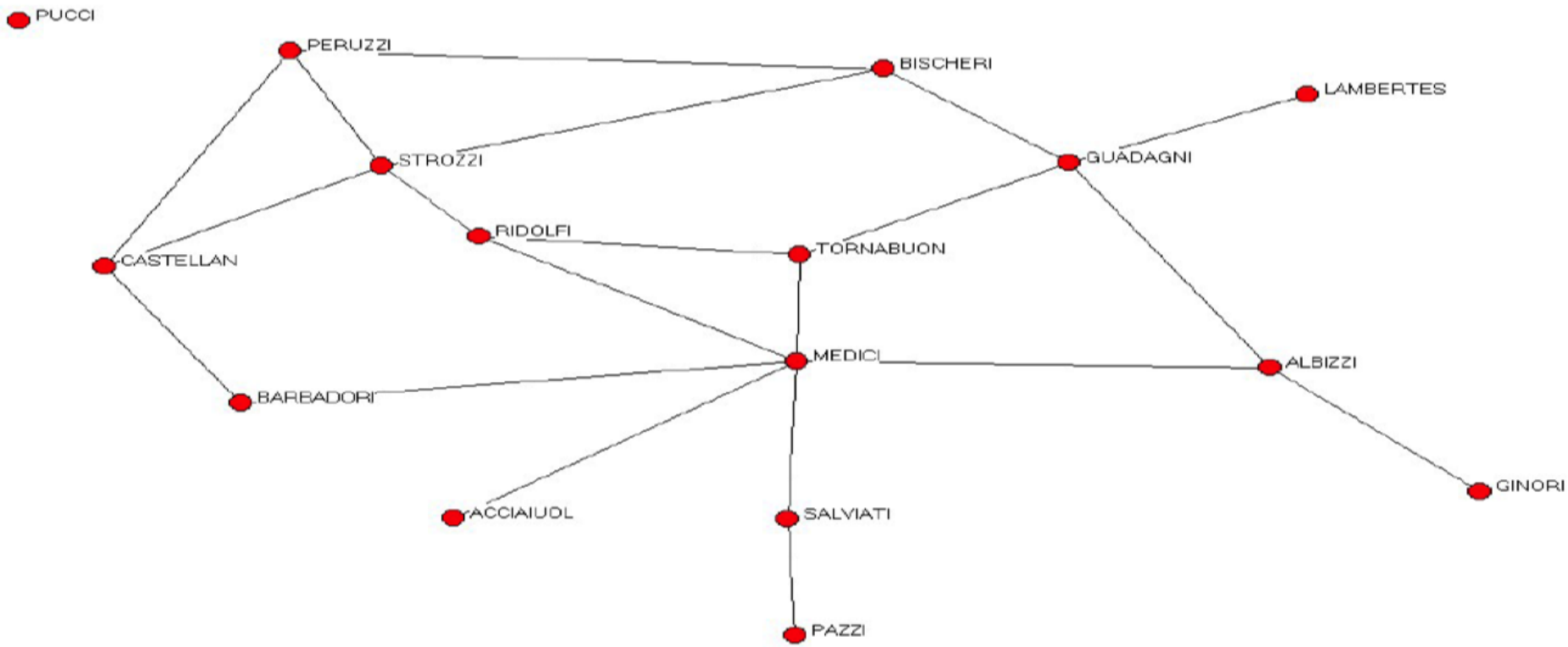
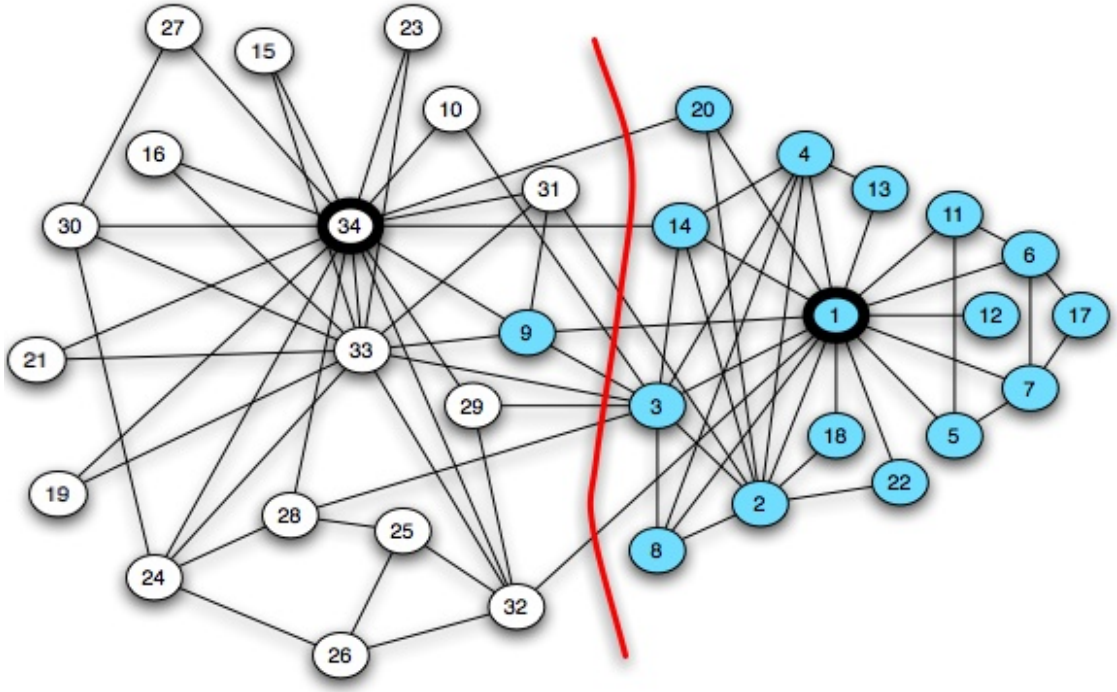


Figure : see [Jackson, Ch 1]

A Balanced Min Cut in Graph: Bonding capital of nodes 1 and 34



- Note that node 34 also seems to have bridging capital.
- Wayne Zachary's Ph.D. work (1970-72): observed social ties and rivalries in a university karate club.
- During his observation, conflicts intensified and group split.
- Could the club boundaries be predicted from the network structure?
- Split could almost be explained by minimum cut in social network.

Strong and weak ties in the karate club network

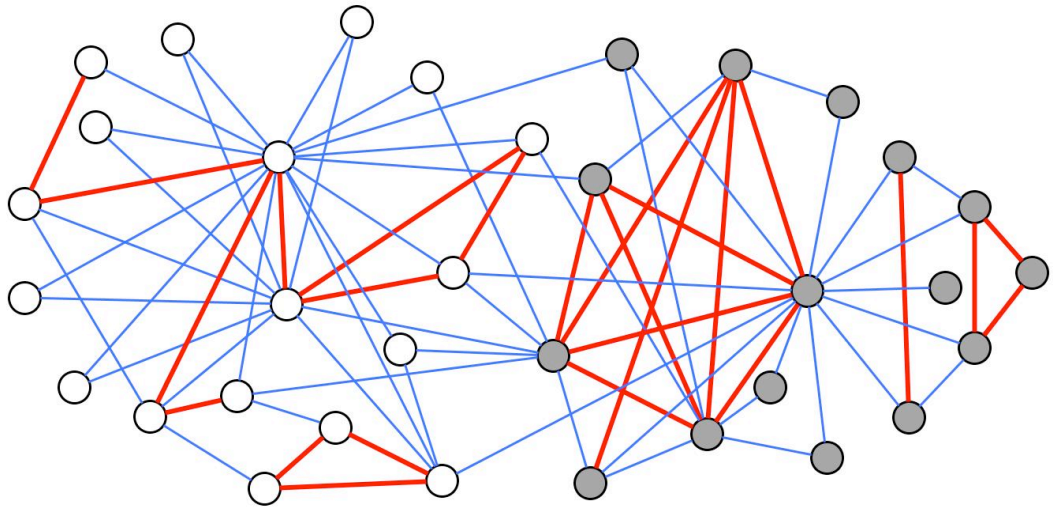


Figure 1: Karate Club graph. Blue light edges represent the weak edges, while red thick edges represent the strong edges.

- Sintos and Tsaparas apply their algorithm to the karate club network.
- Note that all the strong links are within one of the two "computed communities"; that is, links between the communities are all weak links.

Chapter 4: The context of network formation

- In this chapter, we study social networks within their context, considering factors outside of the nodes and edges of the network that impact how the network structure evolves.

- The chapter introduces a very important (and often controversial) issue, namely the relative roles of selection (similarity) vs influence in social relations.

- As we have already noted, Easley and Kleinberg have already indicated that there is a limit to what one can understand just in terms of the network structure.

Word of caution repeated

Easley and Kleinberg (end of Section 3.3):

> *Given the size and complexity of the (who call whom) network, we cannot simply look at the structure. . . Indirect measures must generally be used and, because 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. . .*

We should also add that we may know very little about the reasons for the formation (or disappeaance) of an edge.

Word of caution repeated

Easley and Kleinberg (end of Section 3.3):

> *Given the size and complexity of the (who call whom) network, we cannot simply look at the structure. . . Indirect measures must generally be used and, because 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. . .*

We should also add that we may know very little about the reasons for the formation (or disappeaance) of an edge.

Yogi Berra(1925-2015):

> *In theory there is no difference between theory and practice. In practice there is.*

Homophily

- Homophily: we tend to be similar to our friends.

Homophily

- Homophily: we tend to be similar to our friends.
- This observation is captured in various writings and proverbs perhaps most notably by "Birds of a feather flock together" suggesting that friendships (and membership in groups) are selectively formed due to similar interests.
- In contrast we also have "opposites attract" but the quote might better be "opposites attract but the like-minded last".

Homophily

- Homophily: we tend to be similar to our friends.
- This observation is captured in various writings and proverbs perhaps most notably by "Birds of a feather flock together" suggesting that friendships (and membership in groups) are selectively formed due to similar interests.
- In contrast we also have "opposites attract" but the quote might better be "opposites attract but the like-minded last".
- Why triadic closure? In Chapter 3: some network "intrinsic" reasons (opportunity, trust, incentive) for forming a freindship and now we consider "contextual" reasons for homophily.

Characteristic factors

- Factors which help determine our friendships and relations can be immutable or more transient.

- Some (essentially) immutable factors: race, birth date, gender; religion, height. What other such (mainly permanent) factors exist?

Characteristic factors

- Factors which help determine our friendships and relations can be immutable or more transient.
- Some (essentially) immutable factors: race, birth date, gender; religion, height. What other such (mainly permanent) factors exist?
- Some more mutable (often related) factors: membership in clubs or courses, educational level, recreational interests, professional interests and career, income level, residential neighbourhood

Characteristic factors

- Factors which help determine our friendships and relations can be immutable or more transient.
- Some (essentially) immutable factors: race, birth date, gender; religion, height. What other such (mainly permanent) factors exist?
- Some more mutable (often related) factors: membership in clubs or courses, educational level, recreational interests, professional interests and career, income level, residential neighbourhood
- Of course, immutable factors can and do influence mutable factors. Furthermore, one's friendships can and do influence mutable factors such as say recreational interests.

The influence vs selection issue

- So the selection vs influence issue can be seen as the relative extent to which our friendships are formed selectively due to similarity vs friendships influencing our interests and other similarity traits.

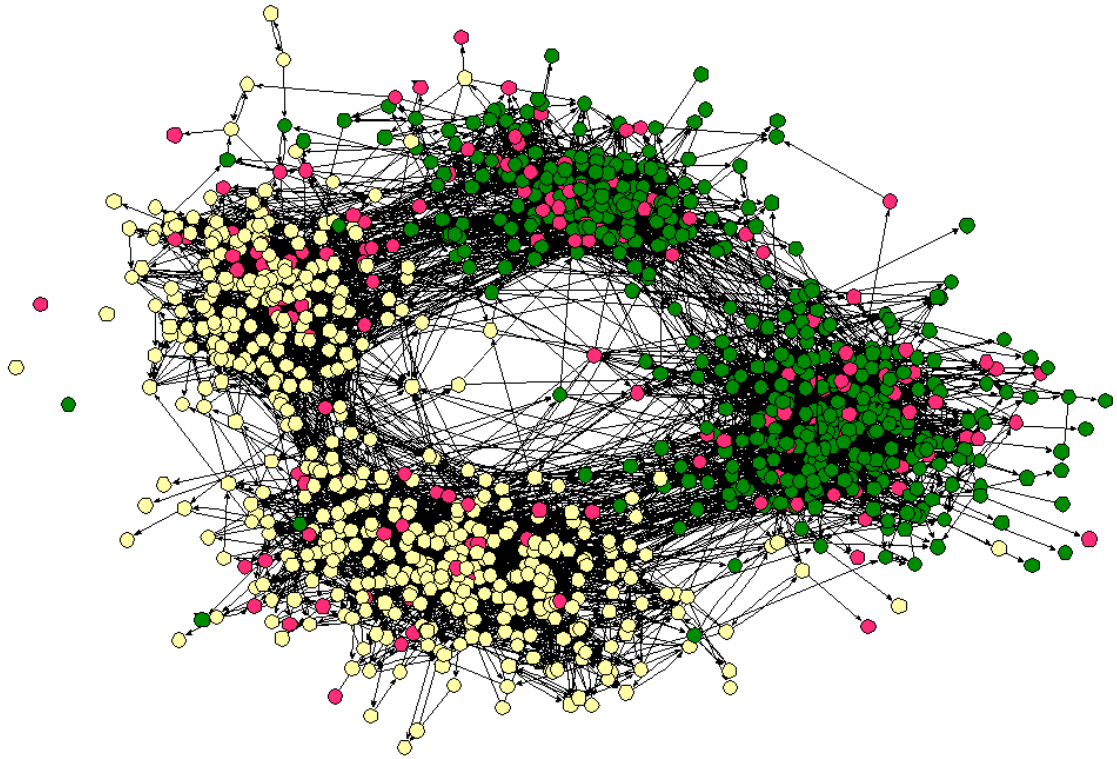
The influence vs selection issue

- So the selection vs influence issue can be seen as the relative extent to which our friendships are formed selectively due to similarity vs friendships influencing our interests and other similarity traits.
- Homophily (which we will use just to note the correlation between friendships and similarity) can be more easily attributed (directly or indirectly) to similarity leading to friendships when similarity factors are immutable or not easily changeable. The issue becomes much less clear and sometimes quite controversial when the similarity factors are mutable.

The influence vs selection issue

- So the selection vs influence issue can be seen as the relative extent to which our friendships are formed selectively due to similarity vs friendships influencing our interests and other similarity traits.
- Homophily (which we will use just to note the correlation between friendships and similarity) can be more easily attributed (directly or indirectly) to similarity leading to friendships when similarity factors are immutable or not easily changeable. The issue becomes much less clear and sometimes quite controversial when the similarity factors are mutable.
- And to further complicate matters, the "environment" of various (perhaps unobserved) external events or hidden influences can also impact one's friendships and/or interests and affiliations.
- For example, Alice and Bob are not friends nor have any interest in political issues. Then a popular entertainer is performing in a rally for a political candidate. Alice and Bob meet at the event and become friends as well as becoming more politically involved.

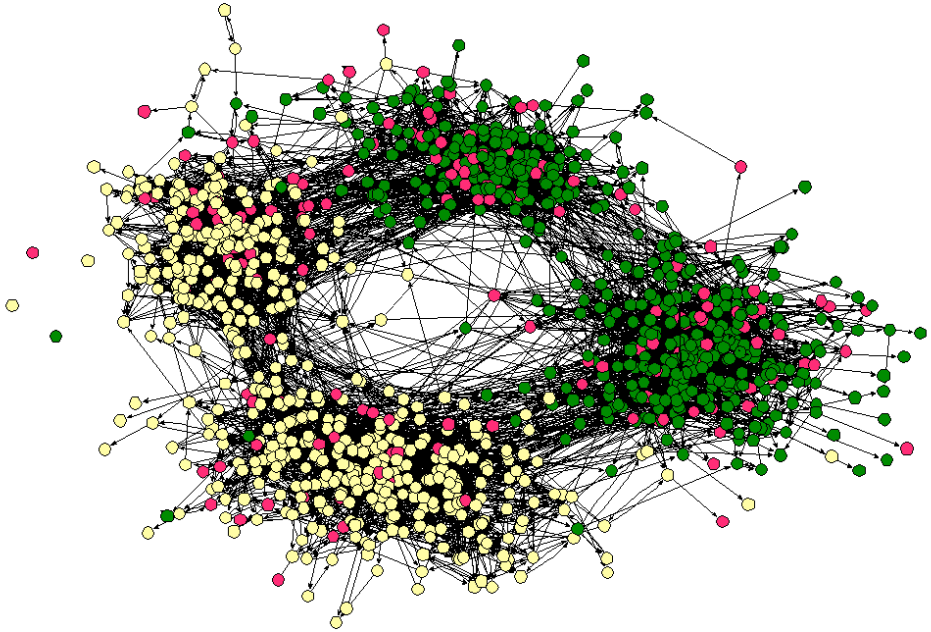
Graphic visualization of homophily



[Fig. 4.1, textbook]

- Homophily can divide 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, two divisions are apparent: one based on race (students of different races drawn as differently-colored circles), and the other based on friendships in the middle and high schools.

Comments on figure 4.1



[Fig. 4.1, textbook]

- Such a visualization is not at a scale that one can see most of the individual relations. The visualization clearly shows homophily based on race and the junior/senior high split (both immutable factors).
- We can measure the extent of homophily (as we will next see) but observing any such phenomena (even for immutable factors) is just the starting point in truly understanding the phenomena.
- The figure does show some detailed information; i.e. individuals without any friends (isolated nodes) or with few friends (low degree).

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 (whatever 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?

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?

- 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.

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?

- 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 would this say about same gender (male-male) or (female-female) edges?

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?

- 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 would this say about same gender (male-male) or (female-female) edges?

- 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.

Reviewing selection vs social influence

- With immutable factors (such as race and gender), when we observe evidence of homophily, we often attribute increased friendships to selection, which is the tendency to form friendships with others who are like you in some way(s). (But note that race often correlates with neighbourhoods or academic programs.)

Reviewing selection vs social influence

- With immutable factors (such as race and gender), when we observe evidence of homophily, we often attribute increased friendships to selection, which is the tendency to form friendships with others who are like you in some way(s). (But note that race often correlates with neighbourhoods or academic programs.)

- 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 interests and behaviour?

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

Reviewing selection vs social influence

- With immutable factors (such as race and gender), when we observe evidence of homophily, we often attribute increased friendships to selection, which is the tendency to form friendships with others who are like you in some way(s). (But note that race often correlates with neighbourhoods or academic programs.)

- 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 interests and 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 may make it possible to see the behavioral changes that occur after changes in an individuals network connections, as opposed to the 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 (1977) and Kandel (1978) claim that peer pressure (i.e. social influence) is less a factor here than previously believed. We can speculate (without having read the studies) that similar (for example) family environments is a significant determining factor for such behaviour amongst friends.

- In contrast to the above example, in a controversial report on obesity patterns of 32,000 people observed over a 32 year period, Christakis and Fowler (2007) claim: obesity or keeping fit is (perhaps surprisingly) to some extent a contagious disease spread within a social network. "You dont necessarily catch it from your friends the way you catch the flu, but it nonetheless can spread through the underlying social network via the mechanism of social influence." (Later in the course we will discuss models for the spread of influence in a network.)

Why the obsesity homophily?

- Three possibilities identified by Christakis and Fowler:
  1. selection
  2. homophily being driven by other factors that correlate with obesity (e.g. poverty)
  3. the social influence of peer pressure say as in the case of drug use or academic performance or fitness.

- Christakis and Fowler conclude that even accounting for 1 and 2, social influence is a significant factor.
  Aside: I am not sure as to the extent that they consider the relative role of genetics vs diet.

- Once again, we caution that observing homophily is clearly only a starting point.

Why do we care?

- How do we study the relative interplay (selection vs. social influence) and why do we want to answer this chicken vs. egg type question?
- If indeed social influence is a significant factor, then targeting key individuals and trying to modify undesirable behaviour (or promote positive behaviour) can be effective since 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.