

Web Science: Selection & Social Influence (Part 1 - Homophily)

CS 432/532

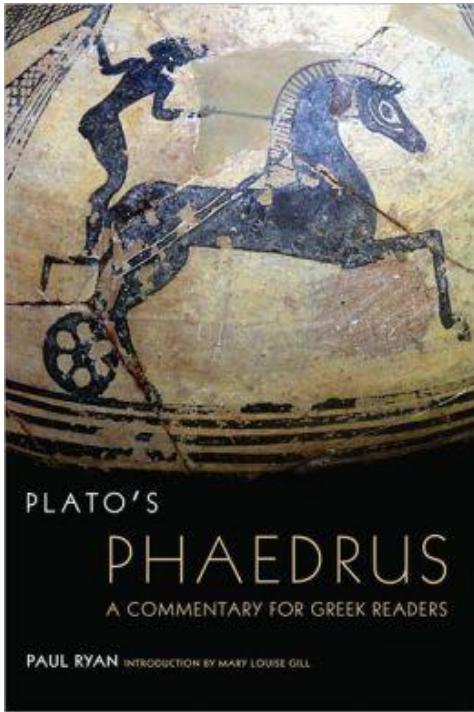
Old Dominion University

Permission has been granted to use these slides from Frank McCown, Michael L. Nelson, Alexander Nwala, Michele C. Weigle



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License](https://creativecommons.org/licenses/by-nc-sa/3.0/)

Most slides based on Ch 4 (“Networks in Their Surrounding Contexts”) of
Networks, Crowds and Markets by
Easley & Kleinberg (2010)
(abbreviated as EK10)



The old proverb says that
"birds of a feather flock together";
I suppose that equality of years inclines
them to the same pleasures, and
similarity begets friendship; ...



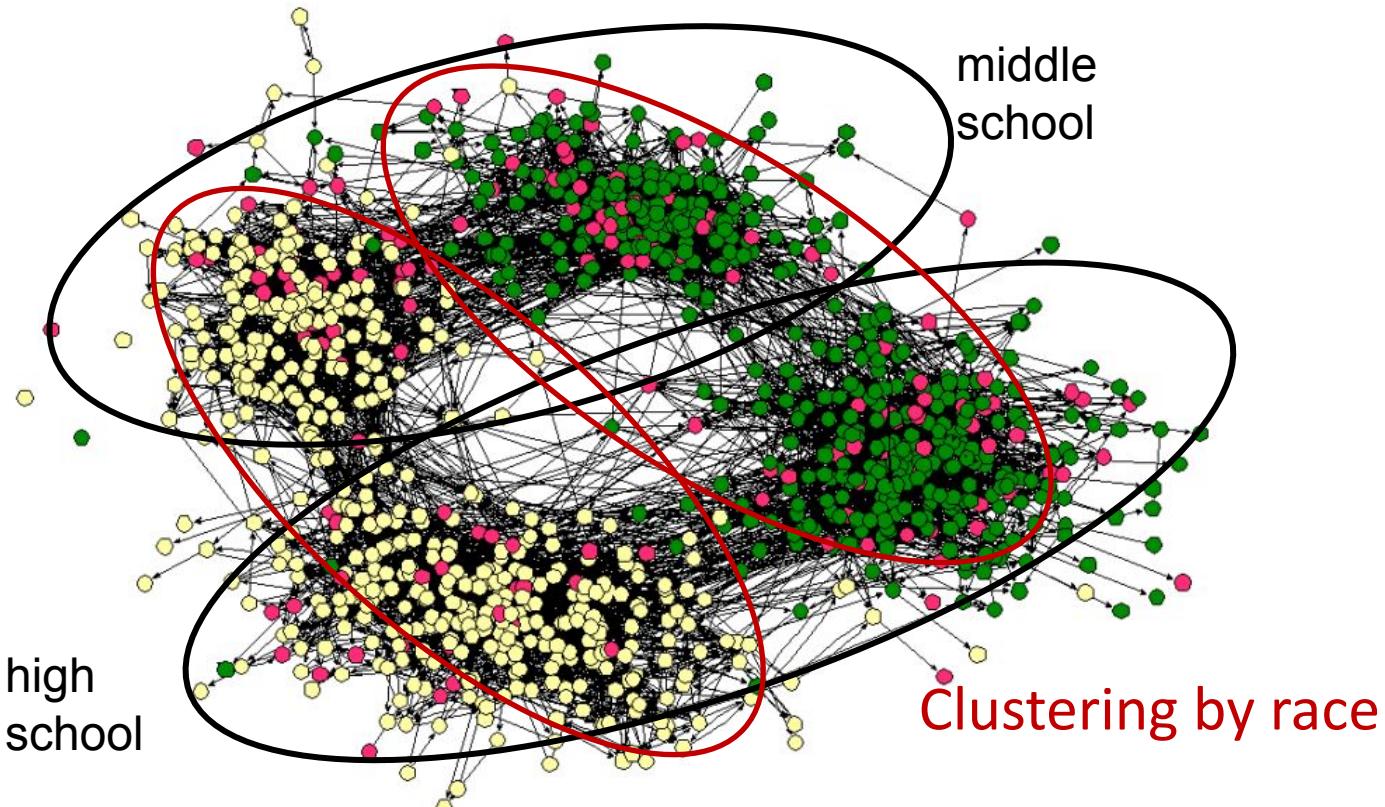
[Phaedrus by Plato](#)
[Phaedrus \(dialogue\)](#) (Wikipedia)
(ca. 370BC)

Homophily

- Your friends are more similar to you in age, race, interests, opinions, etc. than a random collection of individuals
- **Homophily:** principle that we tend to be similar to our friends



Social network from town's middle and high schools



Moody, [Race, School Integration, and Friendship Segregation in America](#) (2001)

Q: Why Do We Study This?

A: Recommending edges in the graph is life blood of social networks

The screenshot shows the Facebook homepage for a user named Michael. On the left, there's a sidebar with a profile picture and links for "Home", "Groups", "Messages", and "Settings". The main content area features a "SUGGESTED GROUPS" section with three cards: "Fullsize Flyers" (1,477 members), "Ford pre 68 Gassers and drag cars (World)" (457 members), and "WILLYS JEEP DUKC". At the bottom, there's a "WILLYS | Chat (7)" link.

Facebook recommending groups for Dr. Nelson

The screenshot shows a Twitter search results page for "Who to follow". It lists three users: "Strammer Max" (@strammer...), "nick ruest" (@ruebot), and "Larry Wayne Wilson" (@Larry...). Each user has a "Follow" button. Below the list is a "Find friends" section. At the bottom of the page, there's a footer with links to "About", "Help", "Terms", "Privacy", "Cookies", "Ads info", "Brand", "Blog", "Status", "Apps", "Jobs", "Advertise", "Businesses", "Media", and "Developers".

Twitter recommending people for Dr. Nelson to follow

The screenshot shows the Facebook homepage for Michael. On the left, there's a sidebar with a profile picture and links for "Home", "Groups", "Messages", and "Settings". The main content area features a "INVITE FRIENDS TO LIKE PAGES" section with several cards: "Denbigh High School - Class of 87" (School), "Delmarva Cougar Club" (Automotive), "Personal and Community Digital Archiving" (Community), "JCDL2012" (Community), "JCDL 2013" (Community), and "SitMyBaby" (Company). At the bottom, there's a footer with links to "English (US)", "Privacy", "Terms", "Cookies", "Advertising", "Ad Choices", "More", and "Facebook © 2016".

Facebook recommending to Dr. Nelson groups for him to recommend to his friends

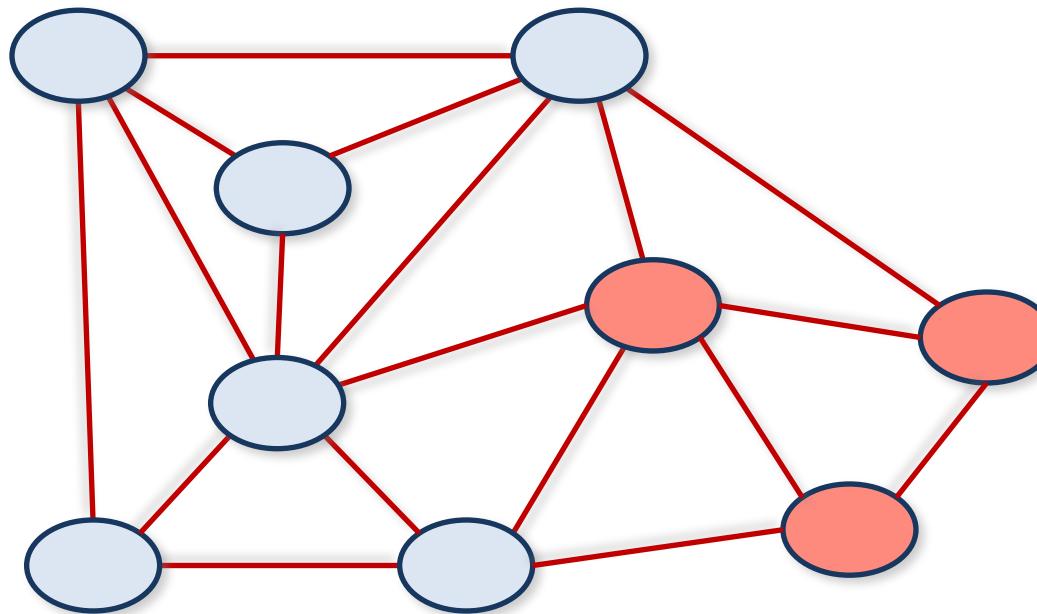
Homophily and Triadic Closure

- *Triadic closure*: When two individuals share a common friend, a friendship between the two is more likely to occur
- Homophily suggests two individuals are more alike because of common friend, so link may occur even if neither is aware of mutual friend!
- Difficult to attribute formation of link to any one factor

Can we develop a simple
test for the presence of
homophily?

Example Male/Female Network

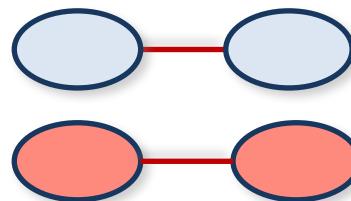
What would this network look like if it did *not* exhibit homophily?



Randomly Generated Network

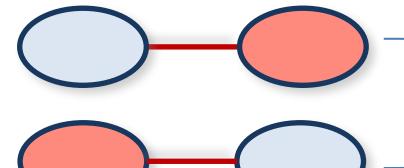
- Randomly assigned a gender according to gender balance of real network
- Number of cross-gender edges should not change significantly relative to real network

If fraction p are males and q are females:



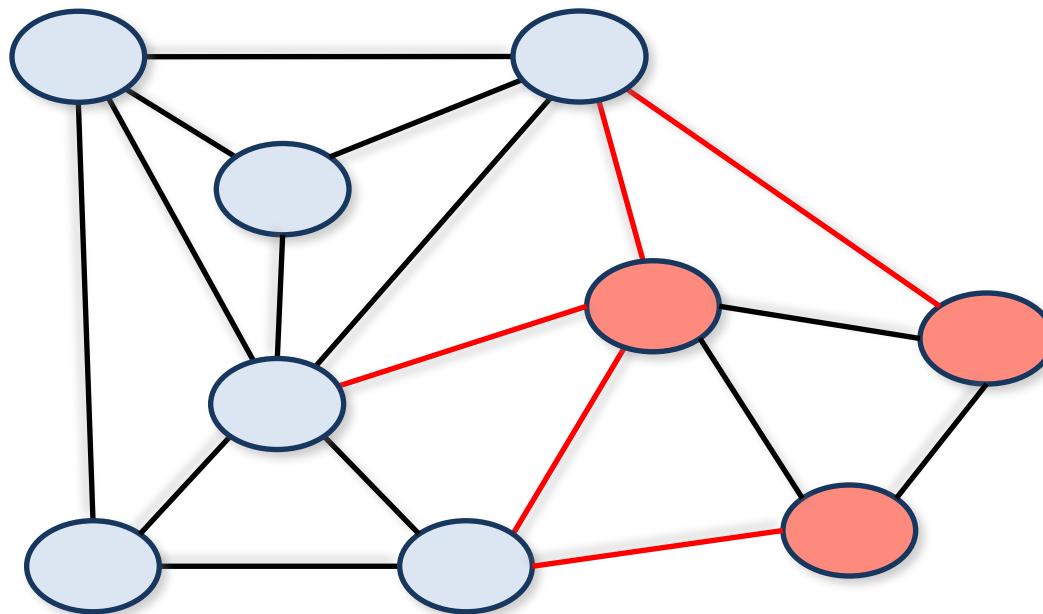
$$p^2 = 0.45$$

$$q^2 = 0.11$$



$$2pq = 0.44$$

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



Cross-gender edges:
5 of 18

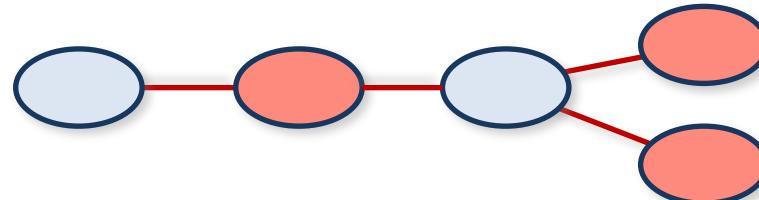
$$p = 6/9 = 2/3$$
$$q = 3/9 = 1/3$$

If no homophily, # of cross-gender edges should be $2pq = 4/9 = 8$ out of 18

∴ Evidence of homophily

Notes on Homophily Test

- "significantly less than" – How significant? A deviation below the mean is suitable
- What if network had significantly *more* than $2pq$ cross-gender edges?
 - Inverse homophily (i.e., "heterophily")
 - Example: Male-Female dating relationships



Notes on Homophily Test

- Homophily test can be used to test any characteristic like race, age, native language, preferences, etc.
- For characteristics that have more than 2 values, can perform a general version of the same type of calculation

Web Science: Selection & Social Influence

(Part 2 - Selection and Social Influence)

CS 432/532

Old Dominion University

Permission has been granted to use these slides from Frank McCown, Michael L. Nelson, Alexander Nwala, Michele C. Weigle



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 3.0
Unported License](https://creativecommons.org/licenses/by-nc-sa/3.0/)

Most slides based on Ch 4 (“Networks in Their Surrounding Contexts”) of
Networks, Crowds and Markets by
Easley & Kleinberg (2010)
(abbreviated as EK10)

Selection

- Why is homophily often present in a social network?
Selection
- **Selection:** People tend to *choose* friends who are like themselves



Selection

- Can operate at different scales and levels of intentionality
- You *actively choose* friends that are like yourself among a small group of people
- Your school's population is relatively homogeneous compared to overall population, so your *environment compels* you to choose friends like yourself



Mutable and Immutable Characteristics

- Selection operates differently based on type of characteristic
- **Immutable:** Characteristics that don't change (gender, race) or change consistently with the population (age, generation)
- **Mutable:** Characteristics that can change over time (behaviors, beliefs, interests, opinions)

Social Influence

- Research has shown that (surprise!) people are susceptible to **social influence**:
they may change their behaviors to more closely resemble the behaviors of their friends



"Follow the crowd! Do what the other sheep do! Gad, child, why can't you remember that?"

Cartoon credit: https://www.cartoonstock.com/directory/r/remember_to_follow_the_crowd.asp

Selection and Social Influence

- Social influence is reverse of selection
 - **Selection**: Individual *characteristics* drive the formation of *links*
 - **Social influence**: Existing *links* shape people's mutable *characteristics*



Longitudinal Studies

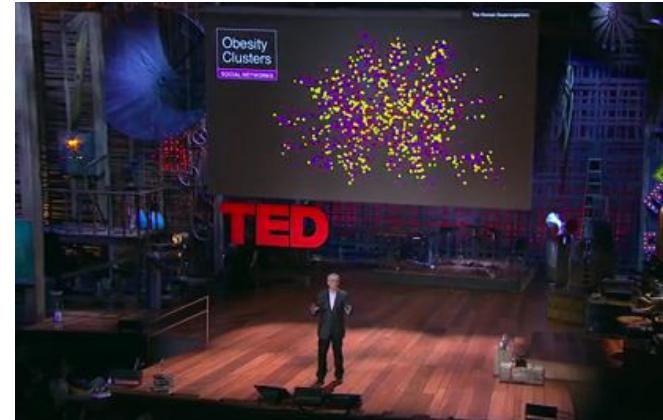
- Difficult to tell if selection or social influence at play with a single snapshot of a network
- Longitudinal studies tracking social connections and individual behaviors over time can help researchers uncover effect of social influence
- Does behavior change after changes to network, or does network change after changes in behavior?

Example: Teenage Drug Use

- Is drug use affected more by selection or social influence?
 - Understanding these effects can be helpful in developing interventions
- If drug use displays homophily in a network, targeting *social influence* (get friends to influence other friends to stop) might work best
- But if homophily due to *selection*, former drug users may choose new friends and drug-using behavior of others is not strongly affected

Social Influence of Obesity

- Christakis and Fowler tracked obesity status and social network of 12,000 people over 32 years
- Found homophily based on obesity status
- They wanted to know why



[Nicholas Christakis: The hidden influence of social networks | TED Talk](#)

Christakis & Fowler, [The Spread of Obesity in a Large Social Network over 32 Years](#) (2007)

Why Obesity Homophily?

1. *Selection effects* – people choose to befriend others of similar obesity status?
2. *Confounding effects of homophily* – other factors that correlate with obesity status?
3. *Social influence* – if friends changed their obesity status, did it influence person's future obesity status?

Discovered significant evidence for hypothesis 3 as well as 1 and 2: Obesity is a type of “contagion” that can spread through social influence!

Christakis & Fowler, [The Spread of Obesity in a Large Social Network over 32 Years \(2007\)](#)

Homicide is Also Contagious

SOCIAL NETWORKS AND THE RISK OF GUNSHOT INJURY

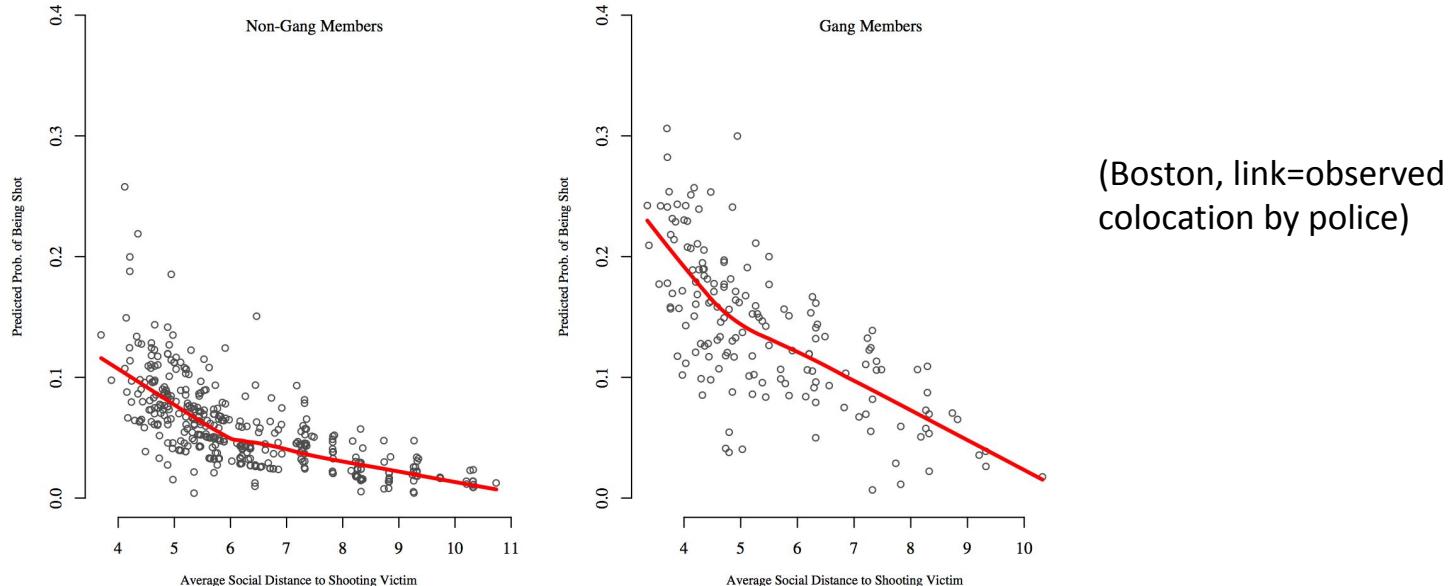
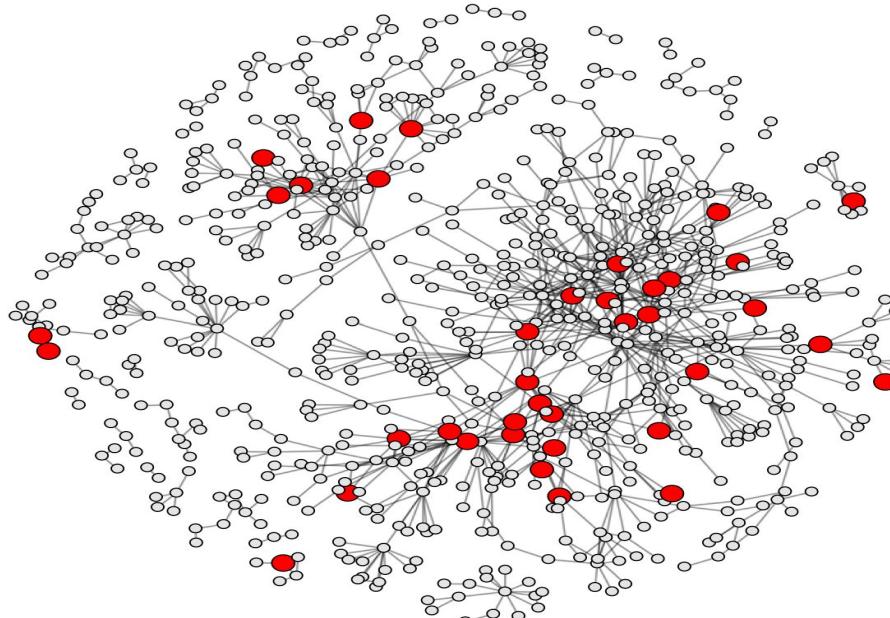


FIGURE 3. The relationship between the predicted probability of being a shooting/homicide victim and distance to another shooting/homicide victim.

Papachristos, Braga, Hureau, “[Social Networks and the Risk of Gunshot Injury](#)”, 2012,

85% of Victims in Giant Component



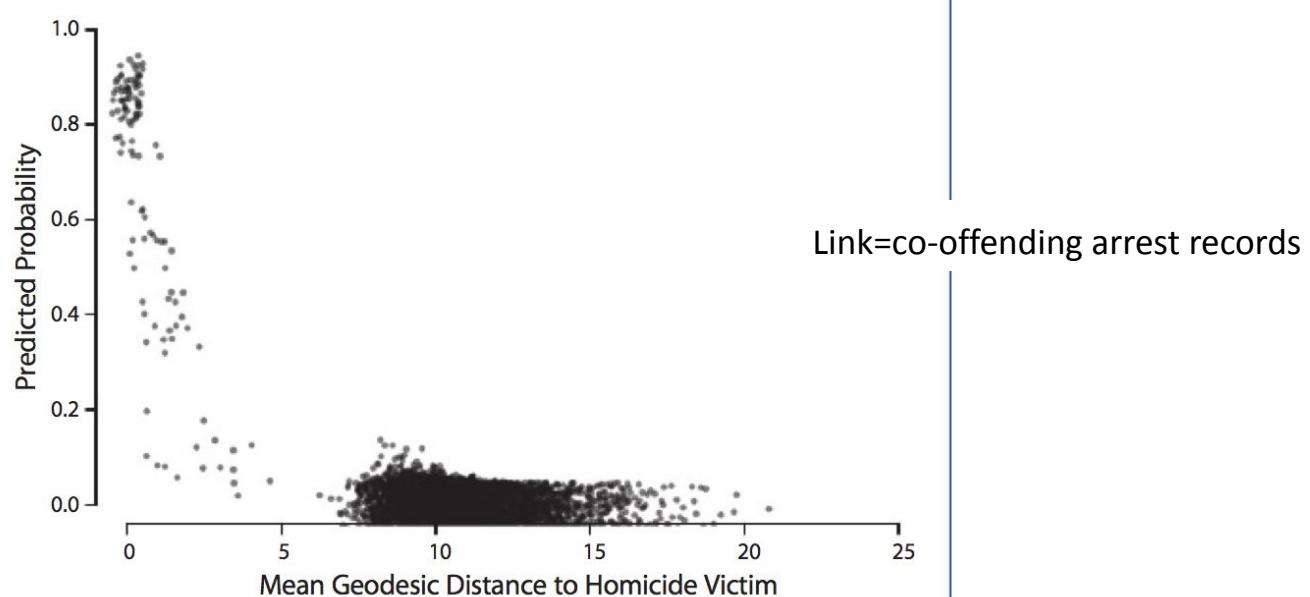
(Boston, link=observed
colocation by police)

FIGURE 2. The social network of high-risk individuals in Cape Verdean community in Boston, 2008.

Probability of gunshot victimization is related to one's network distance to other gunshot victims

Papachristos, Braga, Hureau, ["Social Networks and the Risk of Gunshot Injury"](#), 2012,

41% of Homicides in 4% of Population



Note. The geodesic distance refers to the shortest path between 2 nodes, n_i and n_j , where the distance is simply $d(i, j)$.³⁷

FIGURE 2—Predicted Probability of Homicide Victimization in an African American Community and Mean Geodesic Distance to a Homicide Victim: Chicago, IL, 2006–2011.

[Study: Odds Of Being Murdered Closely Tied To Social Networks](#), NPR, 2013

Papachristos and Wildeman, "[Network Exposure and Homicide Victimization in an African American Community](#)", 2013

Web Science: Selection & Social Influence (Part 3 - Affiliation and Closure)

CS 432/532

Old Dominion University

Permission has been granted to use these slides from Frank McCown, Michael L. Nelson, Alexander Nwala, Michele C. Weigle

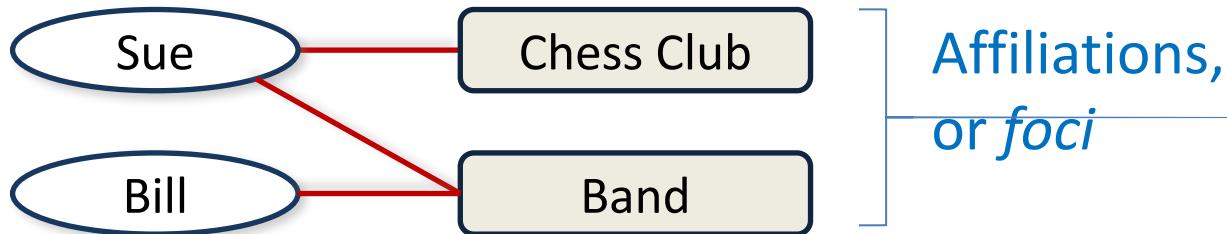


This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License](https://creativecommons.org/licenses/by-nc-sa/3.0/)

Most slides based on Ch 4 (“Networks in Their Surrounding Contexts”) of
Networks, Crowds and Markets by
Easley & Kleinberg (2010)
(abbreviated as EK10)

Social-Affiliation Network

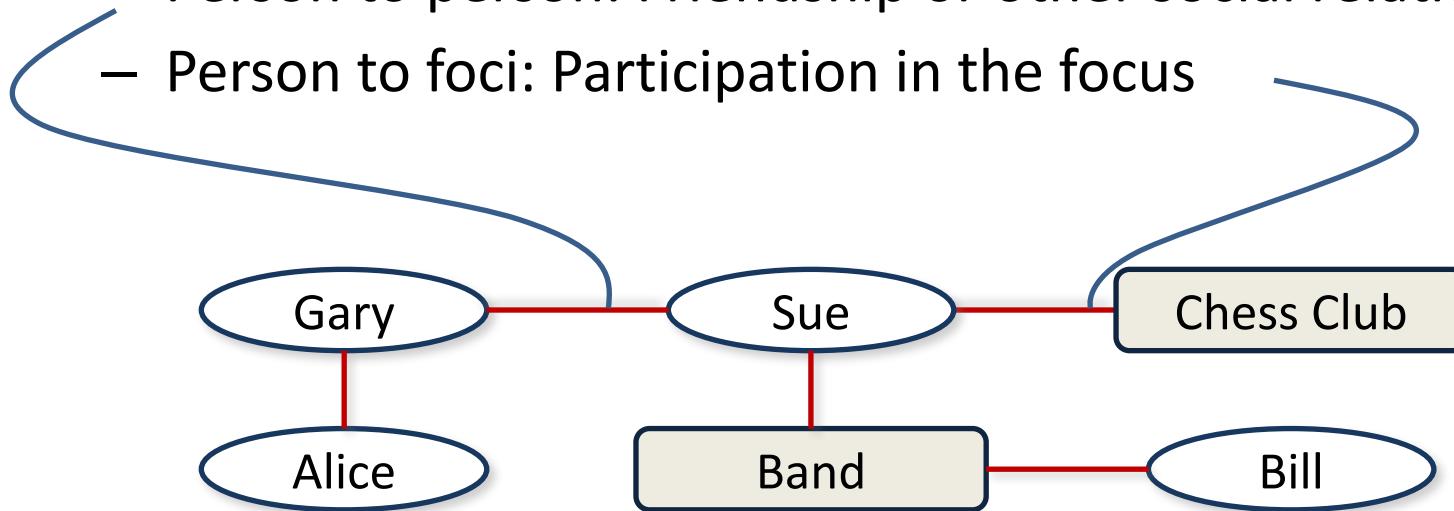
- Putting context into the network by showing connections to activities, companies, organizations, neighborhoods, etc.



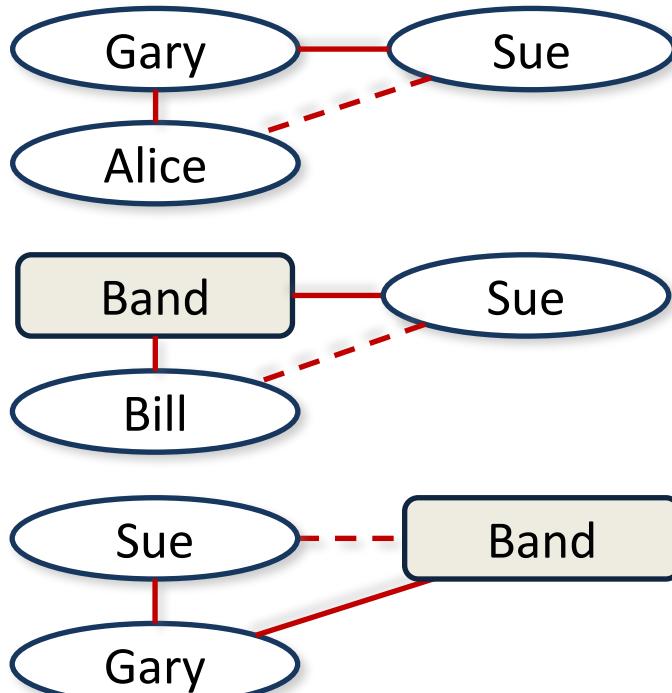
- Bipartite graph:** every edge joins two nodes belonging to different sets (and no edges join nodes that belong to the same set)

Social-Affiliation Network

- Two types of edges:
 - Person to person: Friendship or other social relationship
 - Person to foci: Participation in the focus



Closure Processes



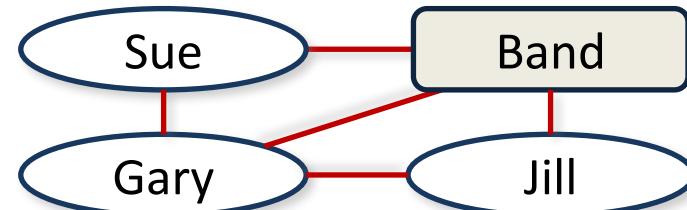
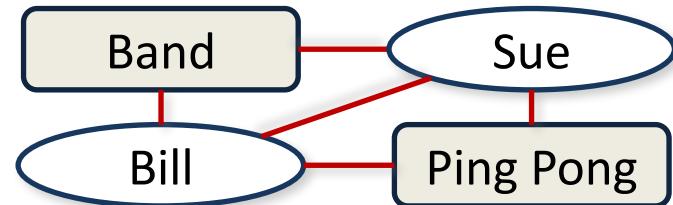
Triadic closure

Focal closure: closure due to selection

Membership closure: closure due to social influence

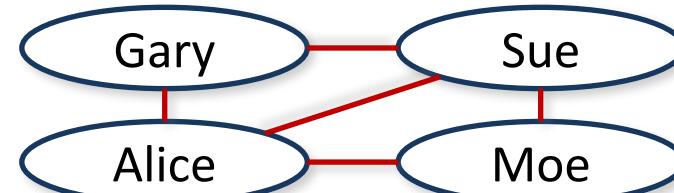
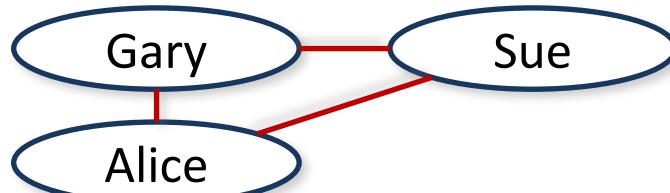
Research Questions About Closure

- *Focal closure*: What is the probability that two people form a link as a function of the number of foci they are jointly affiliated with?
- *Membership closure*: What is the probability that a person becomes involved in a particular focus as a function of the number of friends who are already involved in it?



Research Questions About Closure

- Would Sue be *more likely* to become friends with Alice if they shared more than one friend?
- In other words, is triadic closure dependent on the number of shared friends?
- More formally: What is the *probability* that two people form a link as a function of the number of mutual friends they share?



Research Methodology: Measuring Triadic Closure

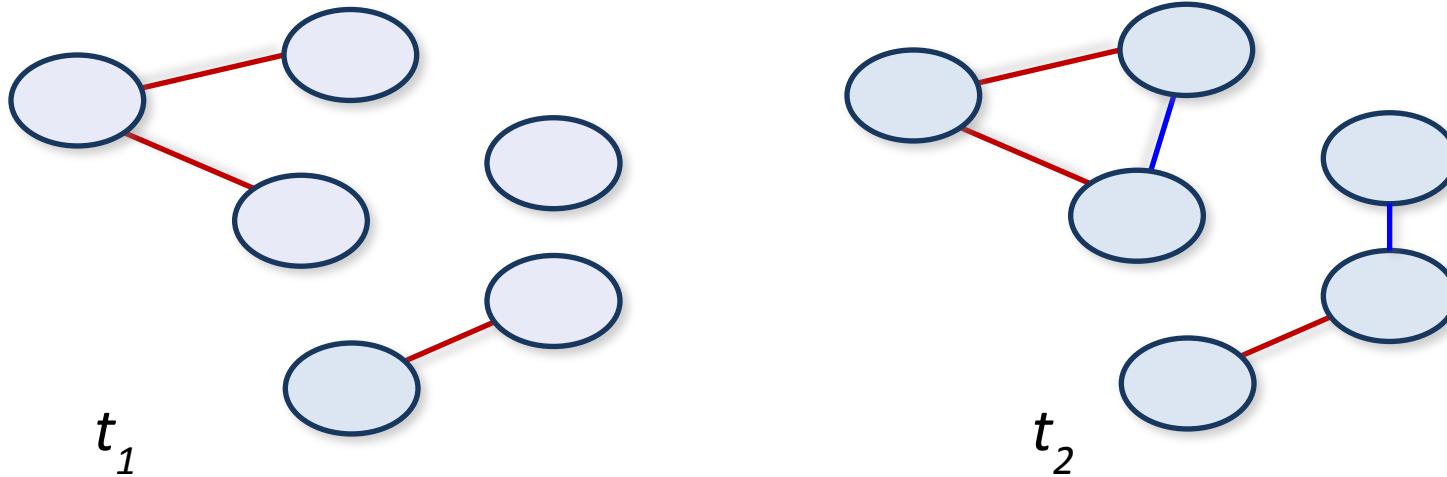
1. Take two snapshots of network at times t_1 and t_2
2. For each k , identify all pairs of nodes who have exactly k friends in common at t_1 but who are not directly connected by an edge
3. Define $T(k)$ to be fraction of these pairs that form an edge by t_2 . This is the probability that a link will form between two people with k friends in common
4. Plot $T(k)$ as function of k to illustrate effect of common friends on link formation

main idea: $k=0$, we might become friends, but chances ($T(0)$) are low

$k=10$, we have a lot of friends in common, so chances ($T(10)$) are high

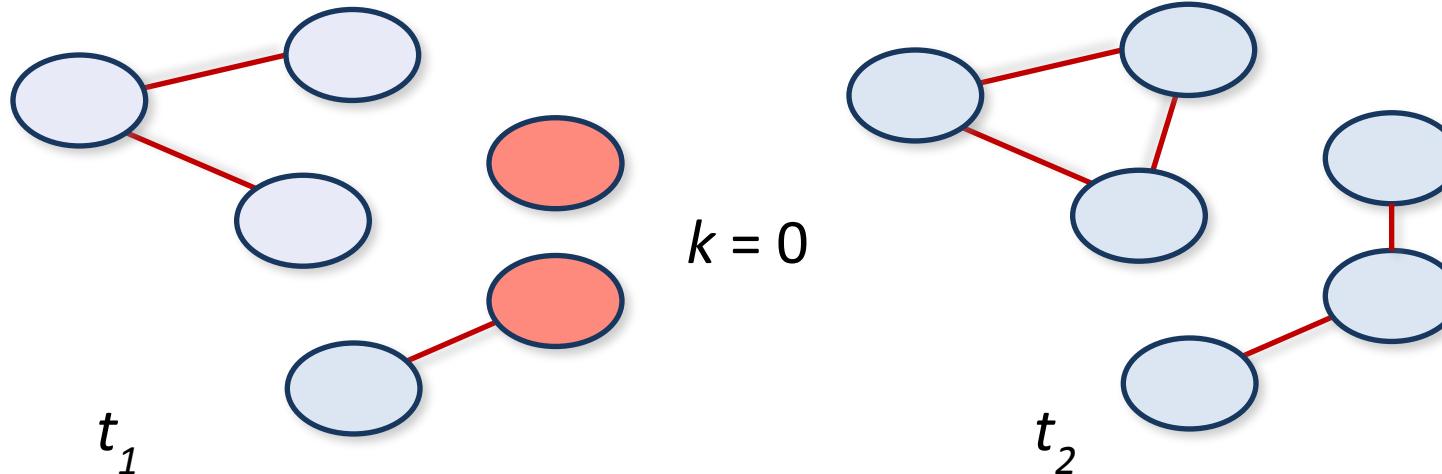
Research Methodology: Measuring Triadic Closure

1. Take two snapshots of network at times t_1 and t_2



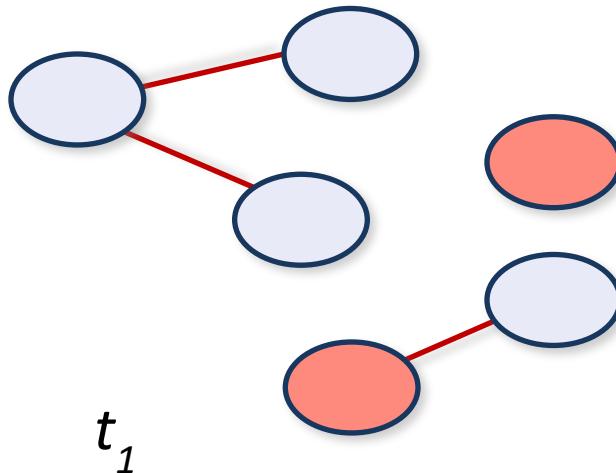
Research Methodology: Measuring Triadic Closure

2. For each k , identify all pairs of nodes who have exactly k friends in common at t_1 but who are not directly connected by an edge



Research Methodology: Measuring Triadic Closure

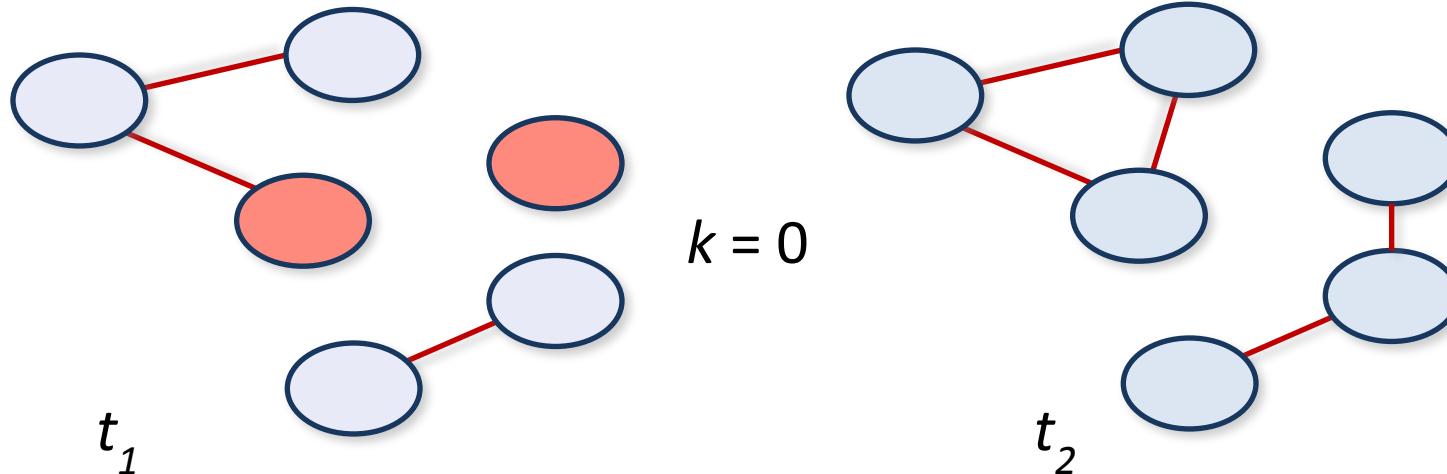
2. For each k , identify all pairs of nodes who have exactly k friends in common at t_1 but who are not directly connected by an edge



$$k = 0$$

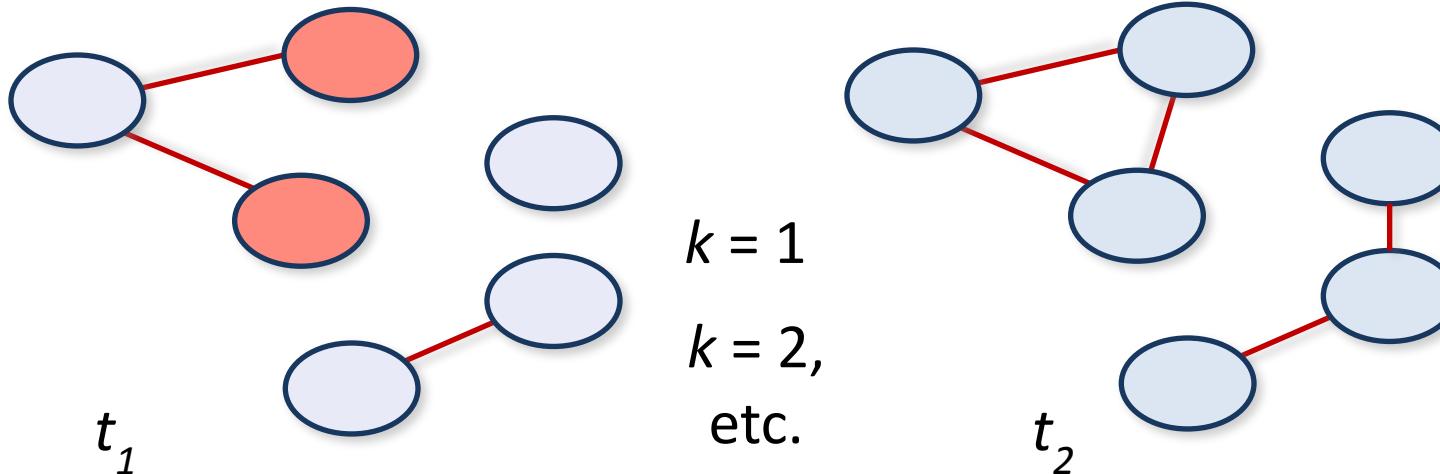
Research Methodology: Measuring Triadic Closure

2. For each k , identify all pairs of nodes who have exactly k friends in common at t_1 but who are not directly connected by an edge



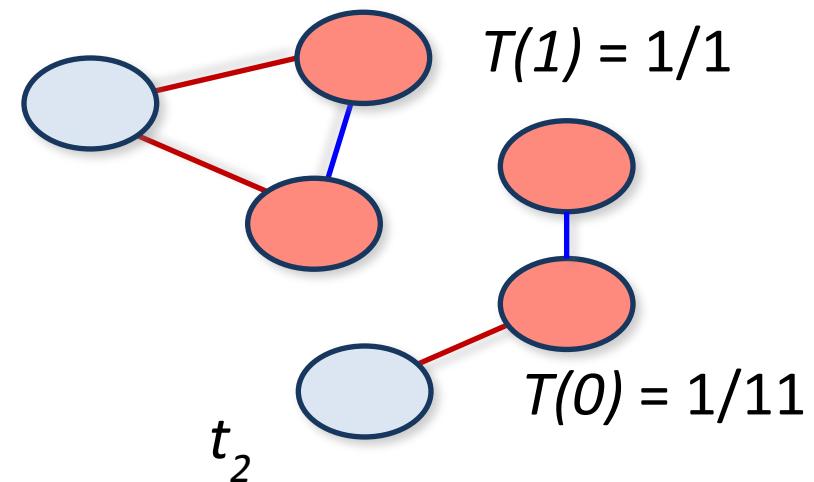
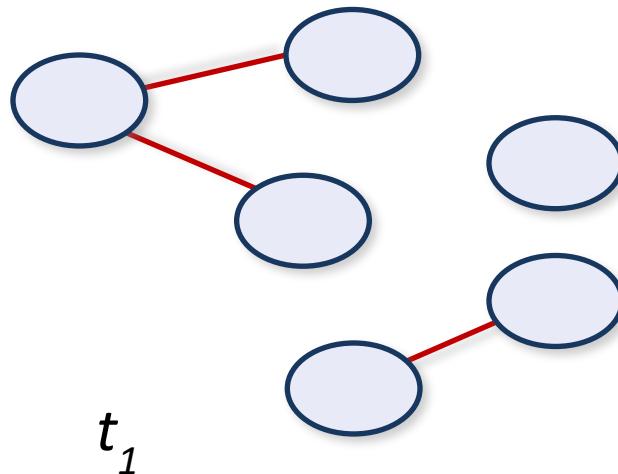
Research Methodology: Measuring Triadic Closure

2. For each k , identify all pairs of nodes who have exactly k friends in common at t_1 but who are not directly connected by an edge



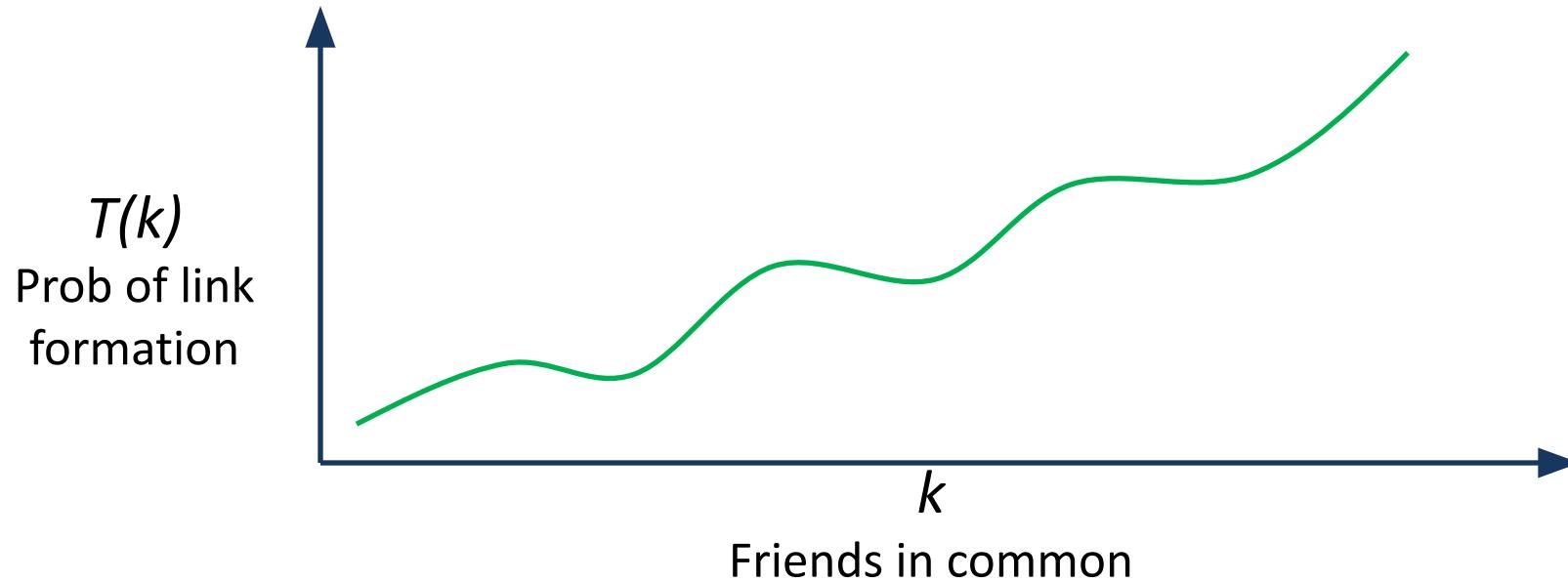
Research Methodology: Measuring Triadic Closure

3. Define $T(k)$ = fraction of pairs that form an edge by t_2 . This is the probability that a link will form between two people with k friends in common



Research Methodology: Measuring Triadic Closure

4. Plot $T(k)$ as function of k to illustrate effect of common friends on link formation



Web Science: Selection & Social Influence (Part 4 - Closure in Real-World Networks)

CS 432/532

Old Dominion University

Permission has been granted to use these slides from Frank McCown, Michael L. Nelson, Alexander Nwala, Michele C. Weigle



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License](https://creativecommons.org/licenses/by-nc-sa/3.0/)

Most slides based on Ch 4 (“Networks in Their Surrounding Contexts”) of
Networks, Crowds and Markets by
Easley & Kleinberg (2010)
(abbreviated as EK10)

Email Social Network

- Kossinets and Watts (2006) examined email communication of 22,000 students over one year at a large US university
- Made link between two people if an email was sent between the two in the last 60 days
- Each snapshot is one day apart
- $T(k)$ averaged over multiple pairs of snapshots

Triadic Closure in Email Data Set

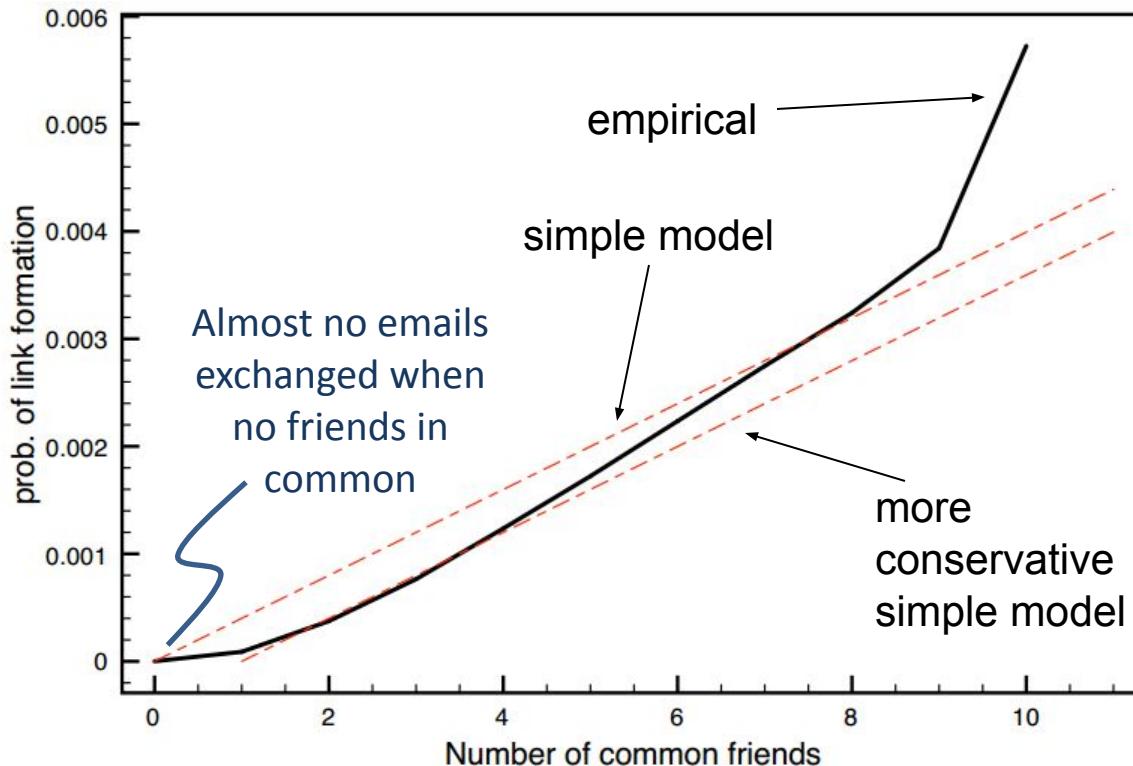


Figure 4.9 from [EK10]

Triadic Closure in Email Data Set

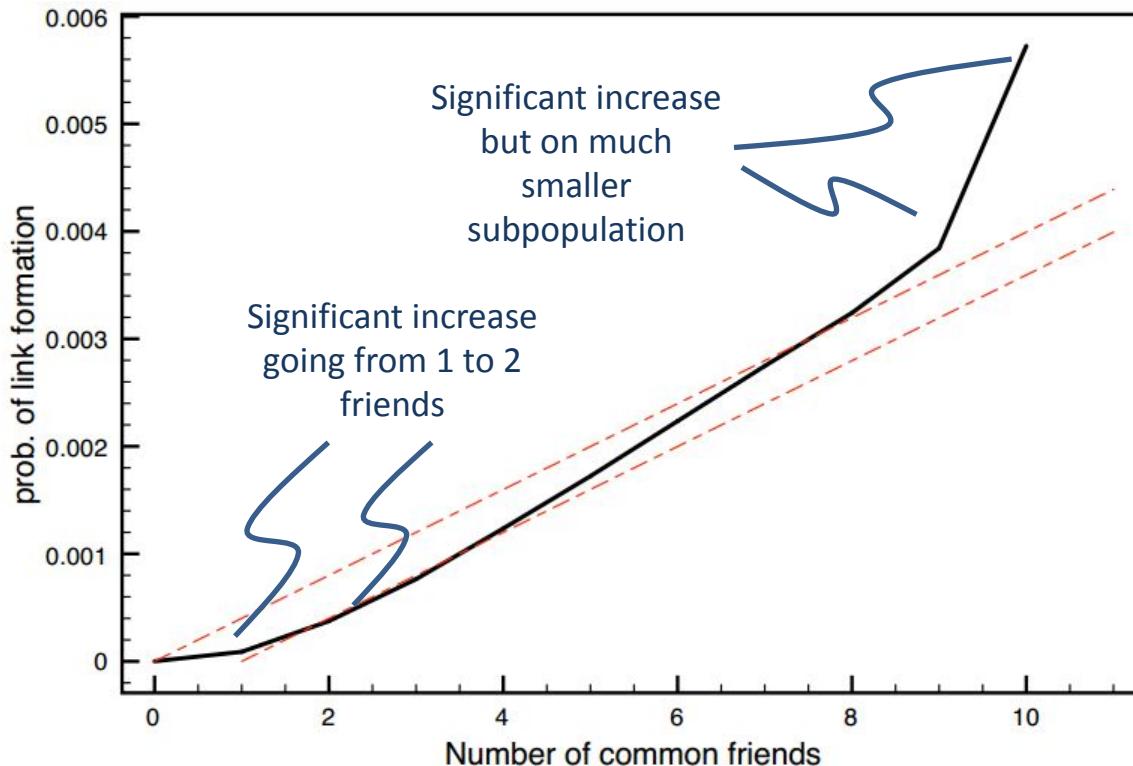


Figure 4.9 from [EK10]

Triadic Closure in Email Data Set

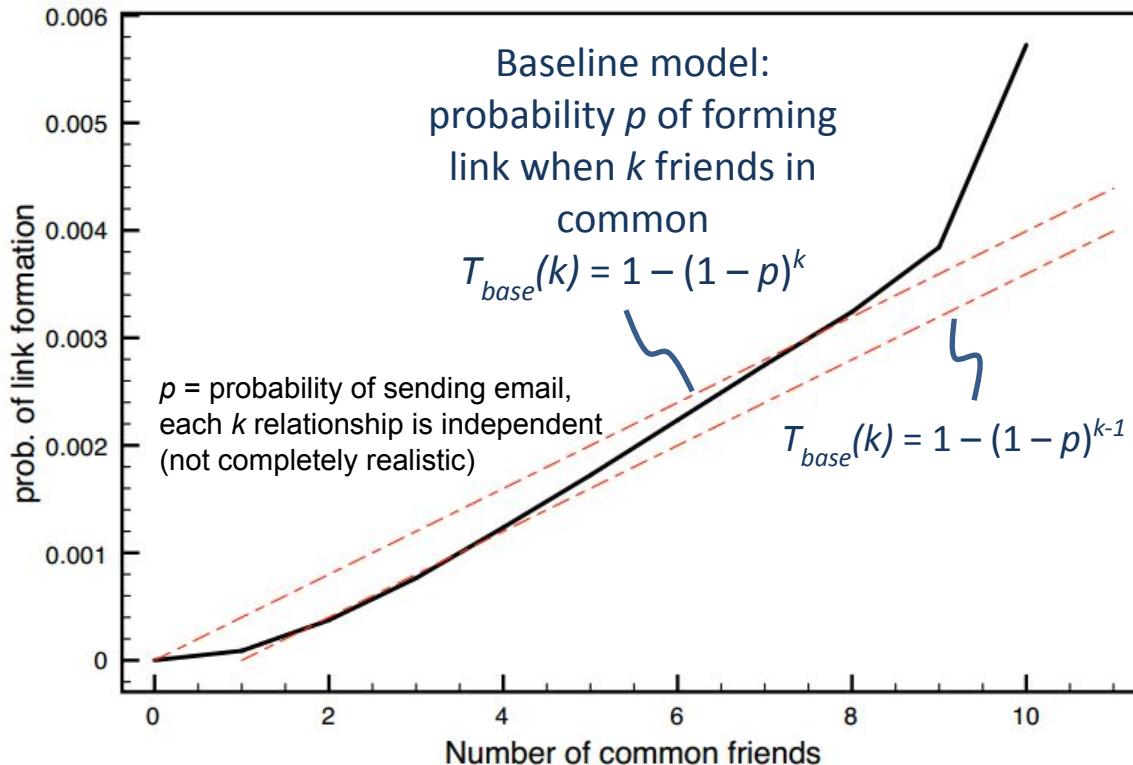
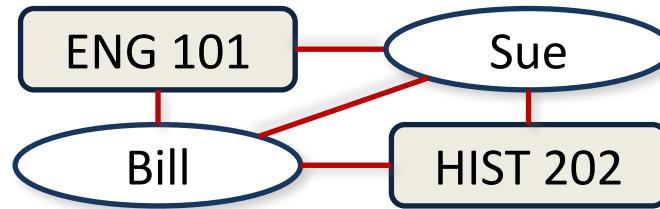


Figure 4.9 from [EK10]

Email Social Network

- To evaluate focal closure, Kossinets and Watts obtained class schedules for 22,000 students
- Created social-affiliation network where classes are foci



- Determined probability of link formation as function of number of shared foci

Focal Closure in Email Data Set

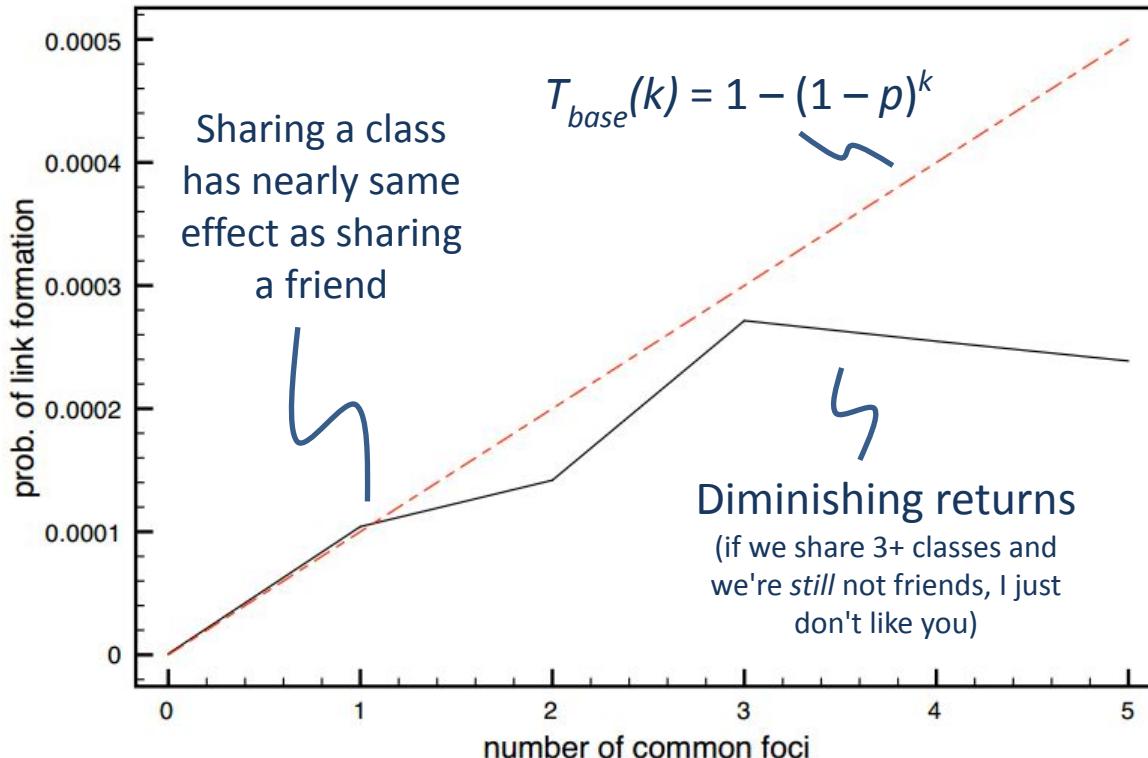
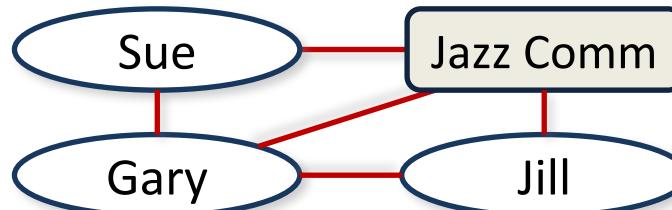


Figure 4.10 from [EK10]

Measuring Membership Closure

- Backstrom et al. (2006) created social-affiliation network for LiveJournal
 - Friendships designated by users in their profile
 - Foci are membership in user-defined communities



LiveJournal User Profile Example

The screenshot shows a LiveJournal user profile for the user [jazzfish](#). The top navigation bar includes links for [Profile](#), [Journal](#), [Comments](#), [Photos](#), [Links](#), [Friends](#), [Communities](#), and [About](#). Below the navigation is a search bar and a date selector set to 10/15/2014.

FRIENDS:

FRIENDS 195 FEEDS 12

baranoouji, berac Jade, blaise Pascal, bradhicks, chelona, cithra, cthulhia, daghain, dimmaus, elf, elisem, elvenyukiryu, fuzzyam, idoru, jadelennox, jadine, jazzfish, jude, karawynn, kiarrith, kittenchan, marc magus, mariness, markush, mikailborg, mneme, nihilistic_kid, planetalyx, platypusgirl, playing_tragic, prog, psilan, rachel_swirsky, rebelsheart, rivka, robbat2, scathach, shadesong, shadowsong, silmaril, skreidle, solestria, stakebait, subbes, timotab, voltbang, vond, xiphias, zaph, zyxwvut [MORE](#)

COMMUNITIES:

WATCHING 35

4th_st_fantasy, backupproject, chez_turtle, dot_poly_snark, farthingparty, firbolg, getfuzzyfeed, helptheproject, labcats, lj_biz, lj_maintenance, lj_releases, lolscience, m15m, magick4terri, mono_poly, motivationofmen, nemesis_draco, news, paidmembers, pearlswine, peril_book_club, poly_misanthrop, polyamorous, polyamory, potlatchcon, racism_101, readercon, slow_poke_poly, tjhsst_alumni, tjpeople, vancouver, vanpoly, viable_paradise, wiscon

[jazzfish - LiveJournal](#)

Membership Closure in LiveJournal Data Set

Probability of joining a community when k friends are already members

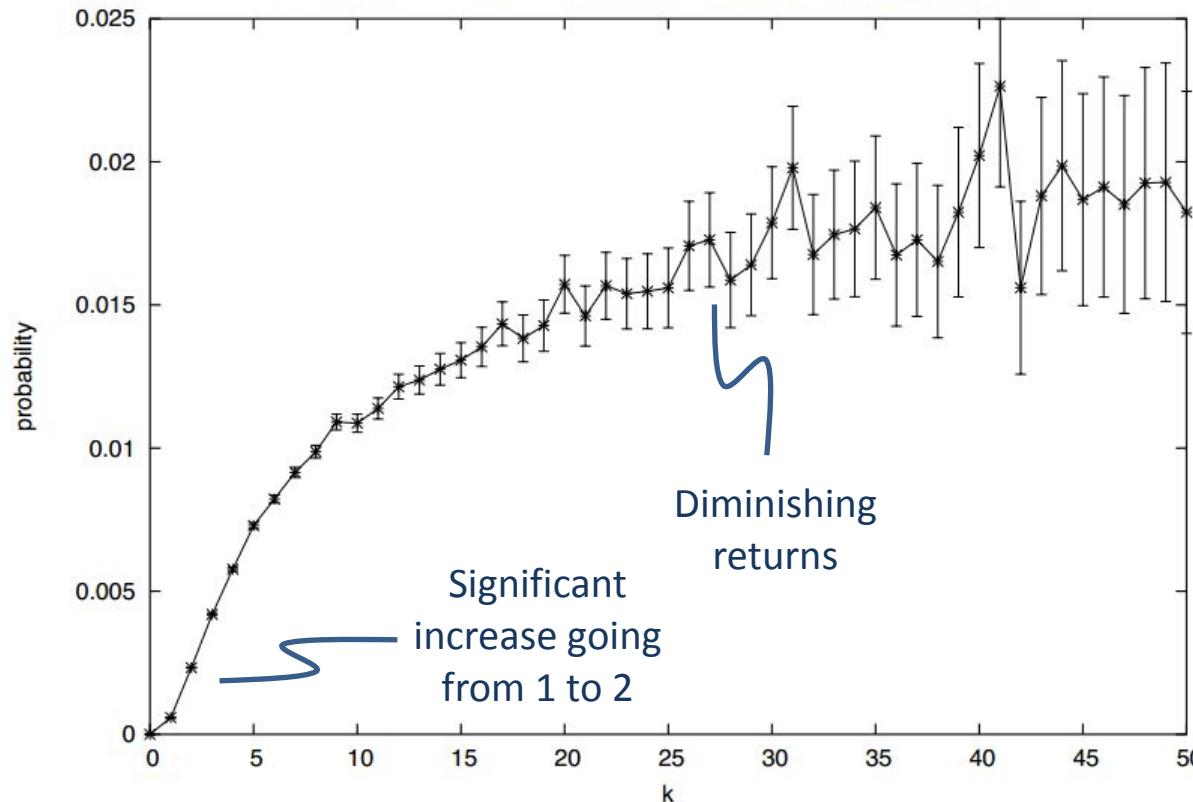
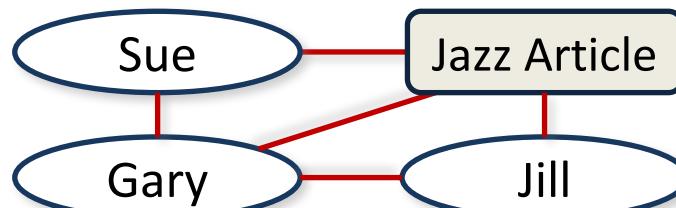


Figure 4.11 from [EK10]

Measuring Membership Closure

- Crandall et al. (2008) created social-affiliation network for Wikipedia
 - Node for editors maintaining a user talk page
 - Link if two editors have communicated using talk page(s)
 - Foci are articles edited by editors



Wikipedia User Talk Pages

W: User talk:GoingBatty – Wikipedia Secure https://en.wikipedia.org/wiki/User_talk:GoingBatty

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
User contributions
Logs
View user groups
Upload file
Special pages
Permanent link
Page information

Print/export
Create a book
Download as PDF
Printable version

Languages

This is **GoingBatty's talk page**, where you can send messages and comments to GoingBatty.

- Put new text under old text. [Click here to start a new topic.](#)
- Please sign and date your posts by typing four tildes (~~~~).
- New to Wikipedia? [Welcome!](#) Ask questions, get answers.

- If I left you a message: please answer on [your](#) talk page, as I am watching it.
- If you leave me a message: I will answer on [my](#) talk page, so please add it to your watchlist.
- [Please click here to leave me a new message.](#)

This talk page is automatically archived by lowercase sigmabot III. Any sections older than 14 days will be automatically archived to the 2017 archive. Archives prior to 2011 were compiled manually and can be found at the right hand side of this page. Sections without timestamps are not archived.

Discussion at Wikipedia:Articles for deletion/Tungsten Network

[edit]

Hello! I see you made edits and improvements to this page as well. Once it was marked for deletion, I put some time into improving on the article to meet its criteria. I'm anxious to get it up to standard and to preserve the contribution, but in a form that meets guidelines. I'm also feel very reluctant and cautious to remove the "Flagged for Deletion" warning and move it to the talk page. Can you help me understand the best way to make this happen without violating any guidelines or breaking any unwritten community codes? I really want to do this right. [MushuNeak](#) (talk) 21:35, 23 February 2017 (UTC)

Archives
2010 · 2011 · 2012 · 2013
2014 · 2015 · 2016 · 2017

[edit]

Hi, I hope that you're well. Could you please help me with the article about [Mark White](#), and also his band [Vice Versa](#) please? They've recently reformed, and are currently working on some new tracks which I'm very happy about! And thank you if so, [Joe Vitale 5](#) (talk) 22:57, 27 February 2017 (UTC)

@[Joe Vitale 5](#): ✓ Done! GoingBatty (talk) 03:01, 28 February 2017 (UTC)

Thank you, have a nice day! Joe Vitale 5 (talk) 22:29, 28 February 2017 (UTC)

Membership Closure in Wikipedia Data Set

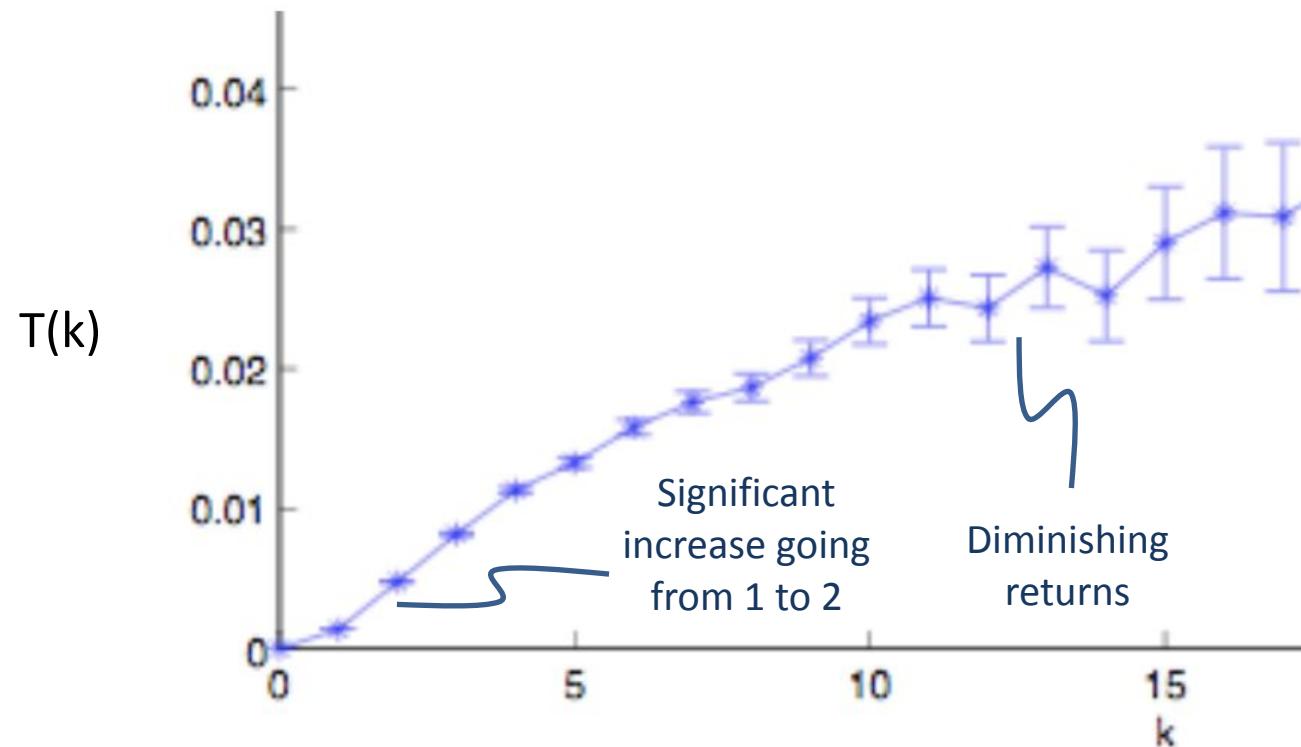


Figure 4.12 from [EK10]

Selection and Social Influence

- Further examination of Wikipedia data set to see evidence of homophily produced by selection and social influence

- How do we measure similarity of two editors?

$$\frac{\text{number of articles edited by both editors}}{\text{number of articles edited by at least one editor}}$$

- Similar to *neighborhood overlap* of two editors in an affiliation network of editors and articles

Selection and Social Influence

- Pairs of Wikipedia editors who have communicated are significantly more similar in behavior than editors who have never communicated
- Does homophily arise because...
 - editors are talking with editors who have edited the same article? (*selection*)
 - editors are led to the articles by those they talk to? (*social influence*)

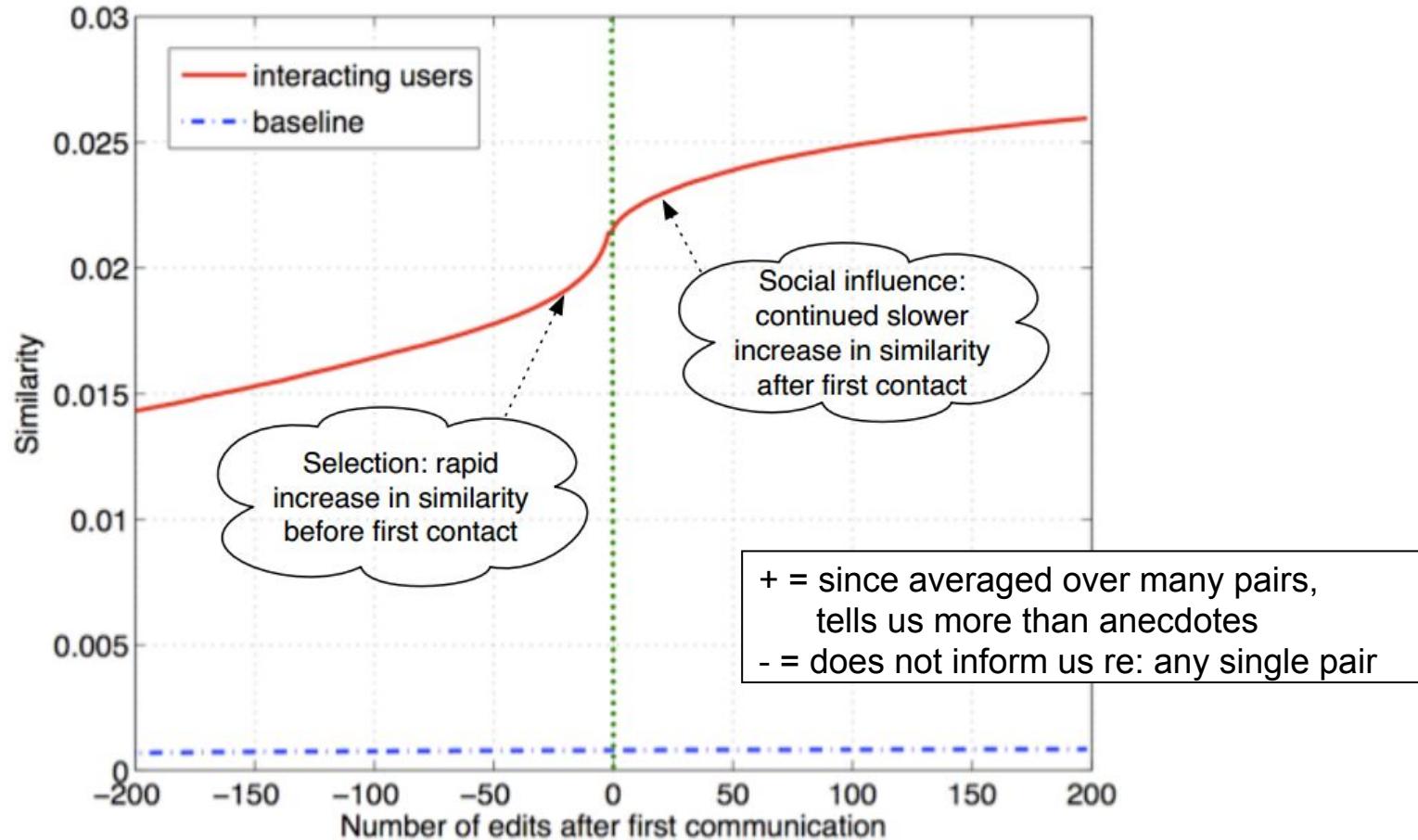


Figure 4.13 from [EK10]

Objectives

- Define homophily.
- Explain how selection affects homophily.
- Explain how social influence affects homophily.
- Explain the difference between focal closure and membership closure in a social-affiliation network.
- Explain the process/steps for measuring how triadic closure is dependent upon the number of shared friends.