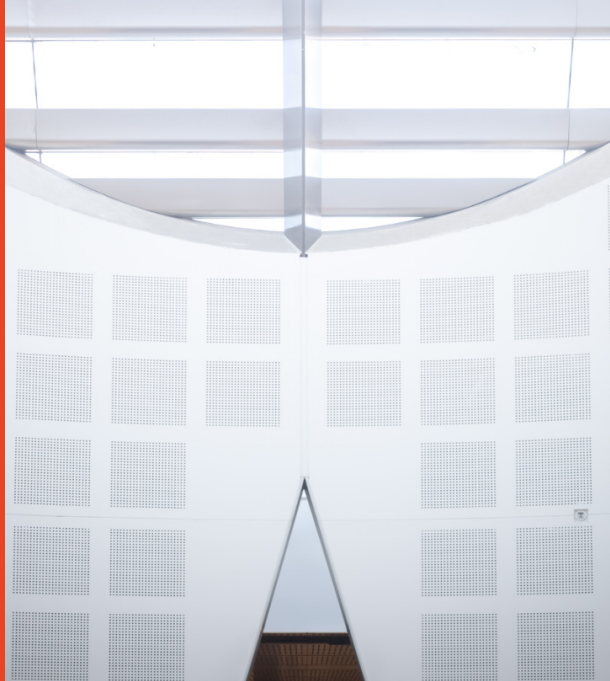
COMP5313/COMP4313 - Large Scale Networks

Week 3b: Networks in Their Surrounding Contexts

Lijun Chang

March 13, 2025

THE UNIVERSITY OF SYDNEY



Introduction

- So far, we have discussed the network as a standalone object
- Now, we will study how network changes/evolves, by focusing on the impact of the surrounding context in which a network is embedded
  - factors that exist outside the nodes and edges of a network, but which nonetheless affect how the network's structure evolves.
- We will talk about
  - Homophily
  - Link formation
  - Spatial Models

Outline

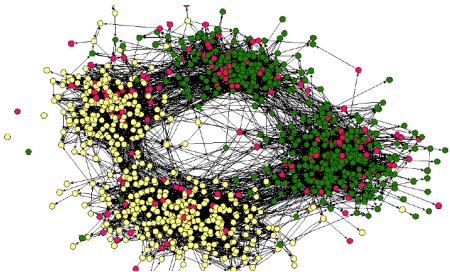
Homophily

Homophily is one of the most basic notions governing the structure of social networks

- Your friends do not look like a random sample of the population
- They are generally similar to you in terms of:
  - Ethnicity
  - Age
  - Mutable characteristics (place they live, occupations, interests, beliefs, opinions)
- Although most of us have specific friendships crossing these boundaries, in aggregate links in a social network tend to connect people who are similar to one another.

Homophily

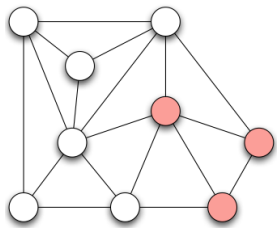
- In this social network from a town's middle school and high school which is divided by ethnicities (left to right) and by schools (top to bottom), can you conclude anything?



- Homophily can divide a social network into densely connected homogeneous parts that are weakly connected to each other.
- Is homophily genuinely present in the network, or is it an artifact of the way the network is drawn?

Homophily

- We need a way to measure homophily.
- Consider a friendship network of an elementary-school classroom
  - We suspect that it exhibits homophily by gender (boys are friends, girls are friends)
  - Consider this graph where pink nodes are girls, white nodes are boys:



  - If there were no inter-gender edges, it would be easy to identify homophily

Homophily

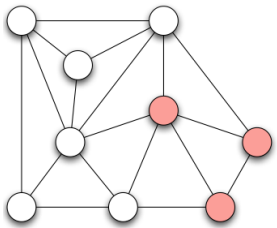
Let's try to define a precise metric of homophily

- Assume a node is a boy with probability p and a girl with probability q
- Consider a given edge of this network
  - Both ends would be boys with probability p^2
  - Both ends would be girls with probability q^2
  - An inter-gender edge (with one end a boy and one end a girl) has probability $2pq$
- We can summarize the test of homophily of gender as:

Homophily Test:

If the fraction of inter-gender edges is significantly less than 2pq, then there is evidence for homophily

Homophily



- Going back to the elementary-school graph
  - 5 of the 18 edges in the graph are inter-gender edges
  - $p = 6/9 = 2/3$ and $q = 3/9 = 1/3$
  - $2pq = 4/9 = 8/18$ (one should expect 8 inter-gender edges rather than 5)

  $\Longrightarrow$ This example seems to show evidence of homophily

Homophily

Remarks

- The percentage of inter-gender edges in a random assignment of genders will deviate somewhat from its expected value of $2pq$
- To perform the test in practice, one should quantify the significance of a deviation below a mean
  - Standard measures of statistical significance can be used here
- A network may have a fraction of inter-gender edges that is significantly more than $2pq$, hence exhibiting inverse homophily (e.g., more opposite sex partners than same-sex partners).
- We can generalize to characteristics taking more than two possible values
  - A heterogeneous edge connects two nodes that are different according to this characteristic
  - We compare the number of heterogeneous edges to their expected number

Homophily

Mechanisms underlying homophily

- **Selection:** the tendency of people to form friendship with people with similar characteristics
  - Individual characteristics drive the formation of links

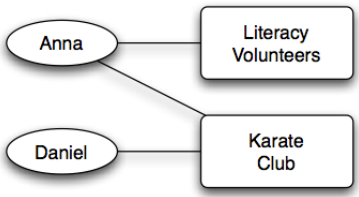
- **Social influence:** the tendency of people to change their behavior to be more closely into alignment with the behavior of their friends
  - The existing links shape people's characteristics

$\Longrightarrow$ Social influence can be considered as the reverse of selection

Outline

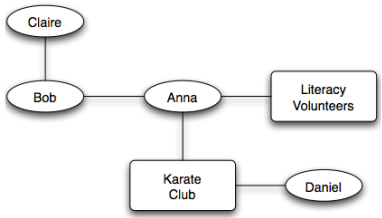
Affiliation Network

- To understand the impact of surrounding context on link formation, we put the context into the network
  - For simplicity and concreteness, focus on the activities in which a person takes part.
- Foci: focal points of social interactions constituting social, psychological, legal or physical entities around which joint activities are organized (e.g., workplaces, voluntary organizations, hangouts, etc.)
- We can represent the participation of people in a set of foci as an affiliation network or a bipartite graph whose nodes can be divided into two sets such that every edge connects a node in one set to a node in the other set.



Social-affiliation Network

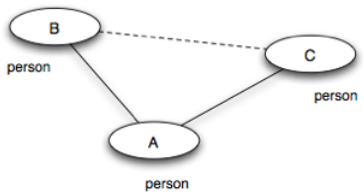
- Social networks and affiliation networks coevolve with one another
  - Participation in a shared focus gives opportunity for friendship
- Let us define a social-affiliation network with two kinds of edges:
  - A social one linking two people if they are friends
  - An affiliation one linking a person to a focus



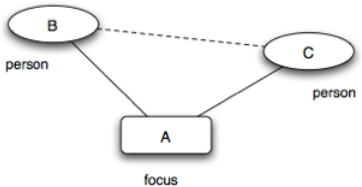
A social-affiliation network shows both the friendships between people and their affiliation with different social foci

Link Formation

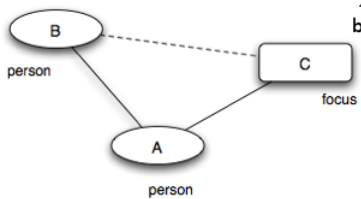
- Different edge creations in social-affiliation network are closure processes



1. Triadic closure: all nodes are persons



2. Focal closure: two people being linked under a common focus influence

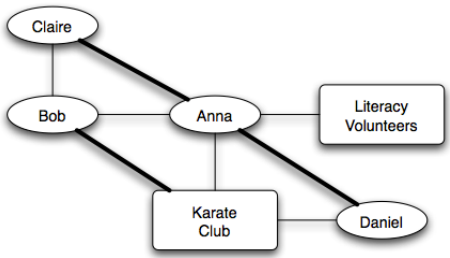


3. Membership closure: a new affiliation being created under the friendship influence

Link Formation

Examples

1. Bob introduces Anna to Claire (triadic closure)
2. Karate introduces Anna to Daniel (focal closure)
3. Anna introduces Bob to Karate (membership closure)



Triadic Closure

- The triadic closure principle: [1]
  - If two people in a social network have a friend in common, then there is an increased probability that they will become friends at some point in the future

- Triadic closure is natural
  - The reason why B and C are more likely to become friends if they are friends of A is simply based on the opportunity for B and C to meet.
    1. If A spends time with both B and C, then there is an increased chance that they will end up knowing each other and potentially becoming friends
    2. The fact that each of B and C is friends with A (provided they are aware) gives them a basis for trusting each other that may be lacking in an arbitrary pair of unconnected people
    3. If A is friends with B and C, then it becomes a source of latent stress in these relationships if B and C are not friends with each other

[1] A. Rapoport, "Spread of information through a population with socio-structural bias: I. assumption of transitivity," *The bulletin of mathematical biophysics*, vol. 15, no. 4, pp. 523–533, 1953.

Triadic Closure

How to evaluate empirically triadic closure in a dynamic network?

- Get an estimate of the probability of friendship formation $T(k)$ under the effect of k common friends

Empirical estimation of the probability of link formation under the effect of common friends

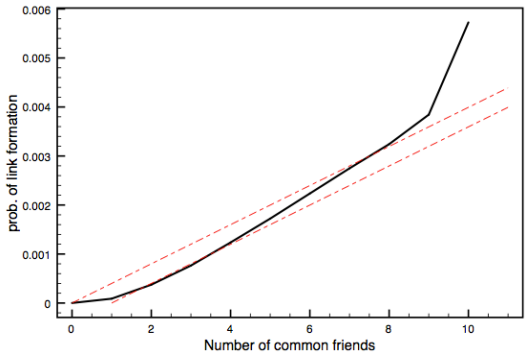
1. Take 2 snapshots of the network at different times
2. For each k, identify all pairs of nodes who have exactly k friends in common in the first snapshot, but who are not directly connected by an edge.
3. $T(k) =$ the fraction of these pairs that have formed an edge by the time of the second snapshot

- Plot $T(k)$ as a function of k to illustrate the effect of common friends on the formation of links

Triadic Closure

Kossinets and Watts computed $T(k)$ on a dataset of email communication among $22,000$ undergraduate and graduate students over a one-year period at a large US university.[2]

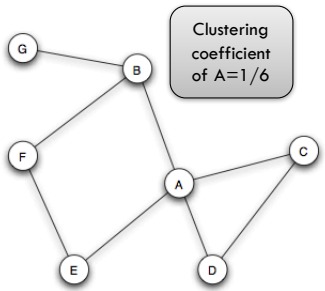
- They linked two people at some instant if they had exchanged an email in the last 60 days, and computed $T(k)$ on each pair of one-day apart snapshots.
- Observations:
  - $T(0)$ is close to 0.
  - Link formation increases overall linearly with the number of common friends



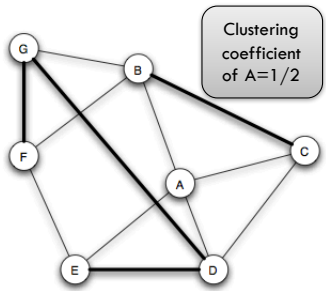
[2] G. Kossinets and D. J. Watts, "Empirical analysis of an evolving social network," *science*, vol. 311, no. 5757, pp. 88–90, 2006.

Triadic Closure

- A consequence of triadic closure is that the clustering coefficient tends to increase
  - The clustering coefficient of a node A is the probability that two randomly selected friends of A are friends with each other



(a) *Before new edges form.*



(b) *After new edges form.*

Membership Closure

Wikipedia is an updatable online encyclopedia

- Wikipedia consists of a set of pages, each providing information on a particular topic
- Editors are people who edit Wikipedia pages
- Editors have a user account and a user talk page where someone else can leave a message for the editor
- Every action (edit, message) on Wikipedia is recorded and timestamped

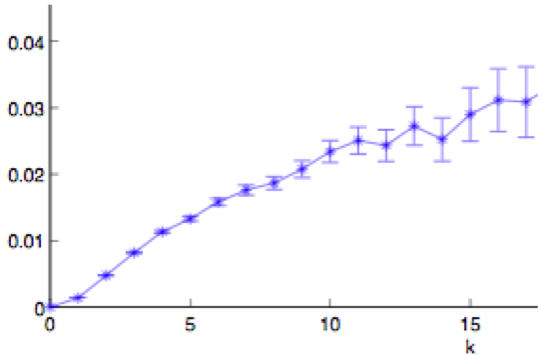
Membership Closure

- Consider the Wikipedia social-affiliation network [3]
  - Each node is a Wikipedia editor with a maintained user account and user talk page on the system
  - An edge joins two editors if one has written on the user talk page of the other
  - The foci are Wikipedia articles
  - There is an affiliation between an editor and a focus if it has edited the corresponding article

[3] D. Crandall, D. Cosley, D. Huttenlocher, J. Kleinberg, and S. Suri, "Feedback effects between similarity and social influence in online communities," in *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 160–168, ACM, 2008.

Membership Closure

- The probability that a person edits a Wikipedia article as a function of the number of prior editors of that article with whom he or she has communicated



- The probability increases with the number k of common neighbors, representing friends associated with the foci

Effect of Selection and Social Influence

- How selection and social influence work together to produce homophily?
  - Let's consider the Wikipedia affiliation network
  - How do similarities in behavior between two Wikipedia editors relate to their pattern of social interaction over time?

- Let's define behavior similarity
  - An editor's behavior corresponds to the set of articles she has edited
  - A simple behavior similarity definition is the neighborhood overlap (we have seen in previous weeks) of the bipartite affiliation network

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

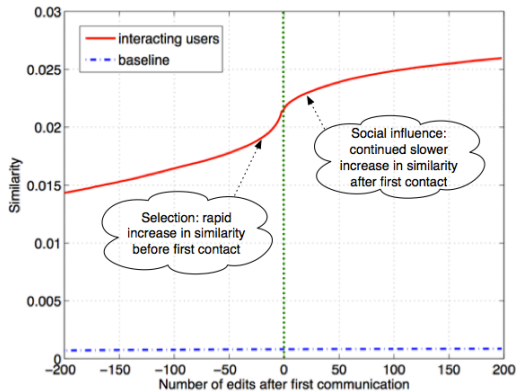
Effect of Selection and Social Influence

- Pairs of Wikipedia editors who have communicated are significantly more similar in their behavior than pairs who have not communicated.
  - So we have a case where homophily is clearly present.

- Does the homophily arise because editors are forming connections with those who have edited the same articles (selection), or is it because editors are led to the articles of those they talk to (social influence)?
  - For each pair of editors A and B who have ever communicated, record their similarity over time, where "time" in this case moves in discrete time units, advancing by one "tick" whenever either A or B edits articles or sends messages
  - Declare time 0 for the pair (A,B) when A and B first communicated
  - We obtain many curves showing similarities as a function of time
    - One for each pair of editors who ever communicated
  - All these curves are averaged into a single plot, showing the average level of similarity as a function of time relative to the first interaction

Effect of Selection and Social Influence

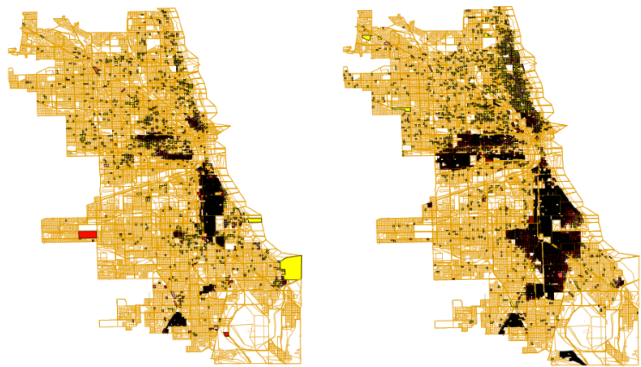
- Similarity increases both before and after the first interaction
  $\Longrightarrow$ Both selection and social influence are at work
- Non-symmetric around time 0: a rapid rise in similarity before first interaction



Outline

Spatial Model of Segregation

Percentage of African Americans per city block of Chicago for the years 1940 and 1960 (lower is lighter): concentration intensifies with time [4]



[4] M. M. Möbius and T. S. Rosenblat, "The process of ghetto formation: evidence from chicago," *manuscript, Harvard University*, 2001.

Spatial Model of Segregation

The Schelling Model [5]

- Assume a population of individuals of immutable type O or X, these are *agents*
- A grid is a 2-dimensional representation of a city
- Each agent resides in a cell of this grid, each having 8 neighbor cells
- Below a threshold t of similar neighbors, an agent moves to another cell

X	X				
X	O		O		
X	X	O	O	O	
X	O			X	X
	O	O	X	X	X
		O	O	O	

X1*	X2*				
X3	O1*		O2		
X4	X5	O3	O4	O5*	
X6*	O6			X7	X8
	O7	O8	X9*	X10	X11
		O9	O10	O11*	

- **Example:** right fig. shows with * unsatisfied people from the left fig for $t = 3$

[5] T. C. Schelling, "Dynamic models of segregation," *Journal of mathematical sociology*, vol. 1, no. 2, pp. 143–186, 1971.

Spatial Model of Segregation

Agent movements

- Agents move in a sequence of rounds
- In each round we move the unsatisfied agents in turn to a satisfying cell

X1*	X2*				
X3	O1*		O2		
X4	X5	O3	O4	O5*	
X6*	O6			X7	X8
	O7	O8	X9*	X10	X11
		O9	O10	O11*	

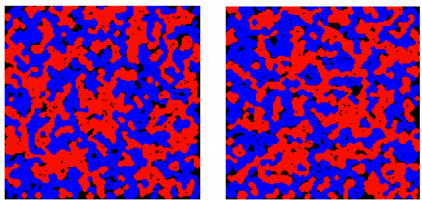
X3	X6	O1	O2		
X4	X5	O3	O4		
	O6	X2	X1	X7	X8
O11	O7	O8	X9	X10	X11
	O5	O9	O10*		

- **Example:** agents from left figure are moved, one at time from top-left to bottom-right, to the nearest satisfying cell, resulting in the right figure

Spatial Model of Segregation

Model simulated on $10,000$ agents of each type

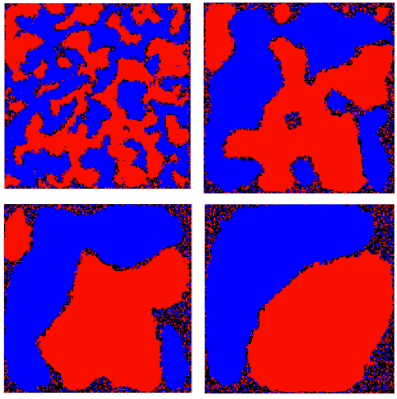
- Grid of 150×150 cells
- Threshold $t = 3$
- Unsatisfied agents move to a random cell



- **Example:** satisfaction states reached after $\sim$50 rounds from two different starting states
- The model produces large homogeneous regions, interlocking with each other as they stretch across the grid

Homophily

- With threshold $t = 4$, the segregation is more intense
- In the same settings as before, we report the state after:
  - 20 rounds
  - 150 rounds
  - 350 rounds
  - 800 rounds
- We obtain a single significant region of each type



Conclusions

- Homophily is one of the most basic notions governing the structure of social networks
  - Your friends are generally similar to you
- There are two mechanism underlying homophily
  - Selection
  - Social influence
- The surrounding context has impact on link formation
- Spatial model of segregation

Reading

- Reading for this week
  - Chapters 4 and 5 of the textbook
- Reading for next week
  - Advanced material of Chapter 3 of the textbook
  - Chapters 9.1–9.4 of Network Science (http://networksciencebook.com/chapter/9)