Networks in Their Surrounding Contexts

Objectives

- Examine additional processes (to triadic closure) that affect the formation of links in the network
- Surrounding contexts: factors that exist outside the nodes and edges of a network
- Represent the contexts together with the network in a common framework

Homophily

- Homophily principle: we tend have similar characteristics with our friends

“birds of a feather flock together”

- People of similar character, background, or taste tend to congregate or associate with one another (like likes like)
- expression appears in the 16th century, a literal translation of Plato's Republic

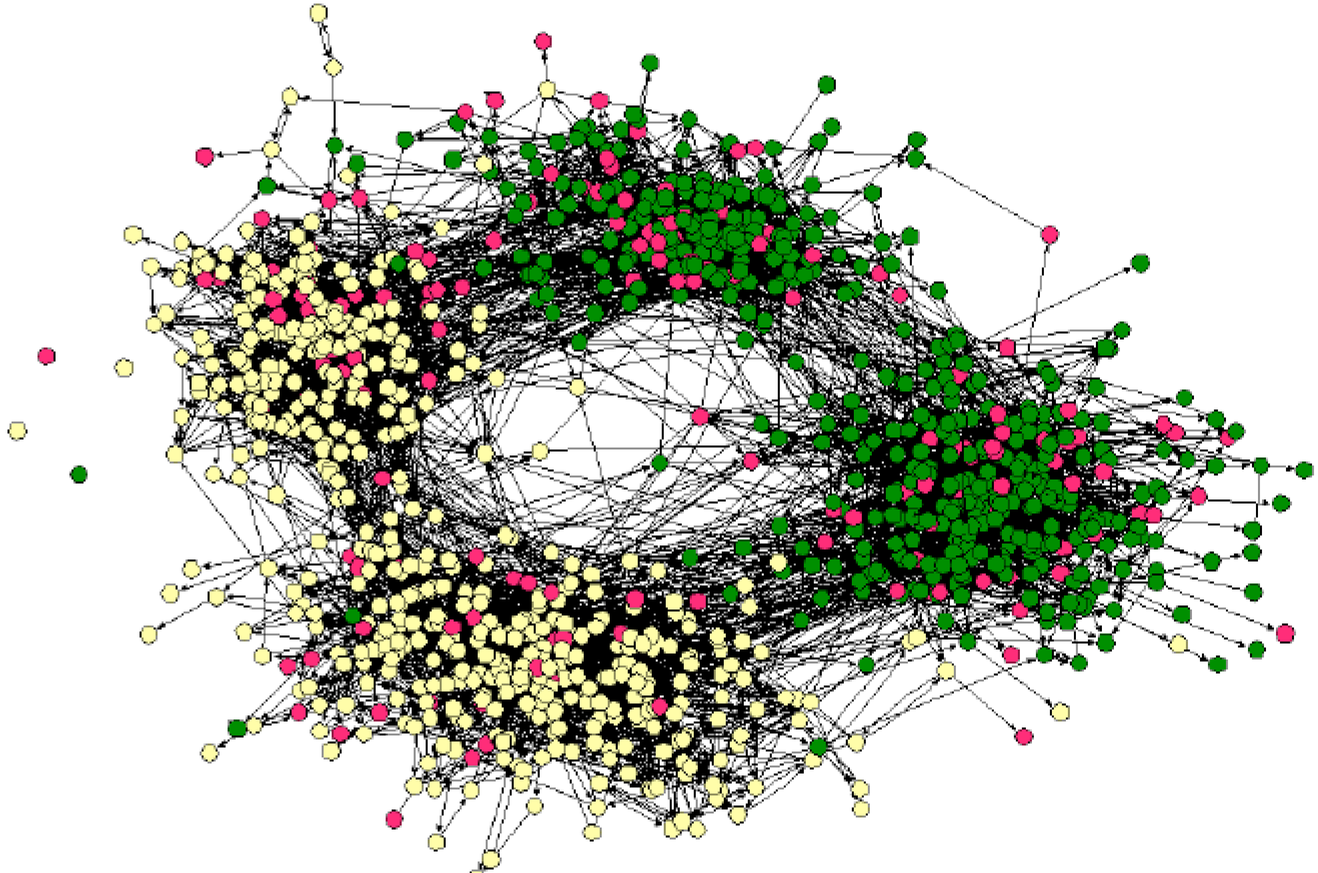
Homophily

- Links in a social network tend to connect people who are similar to one another
  - basic notions governing the structure of social networks
- Its role in modern sociology by influential work in the 1950s (Lazarsfeld and Merton)

Homophily vs. Triadic Closure for Link Formation

- With triadic closure:
  - a new link is added for reasons that are intrinsic to the network (need not look beyond the network)
  - Ex: a friendship that forms because two people are introduced through a common friend
- With homophily:
  - a new link is added for reasons that are beyond the network (at the contextual factors)
  - Ex: a friendship that forms because two people attend the same school or work for the same company

Example



Social network from a town's middle school and high school (students of different races drawn as differently colored circles)

2 divisions:

• one based on race and

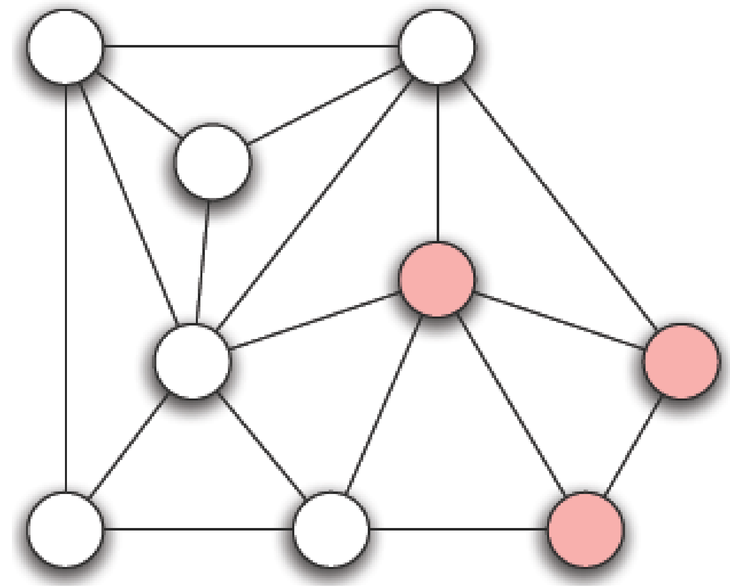
• the other based on friendships in the middle and high schools

Homophily vs. Triadic Closure for Link Formation

- Strong interactions between intrinsic and contextual effects
- Both operating concurrently
- Triadic closure (intrinsic mechanism):
  - B and C have a common friend A
  - B and C have increased opportunities to meet
- Homophily (contextual mechanism):
  - B and C are each likely to be similar to A in a number of dimensions
  - also possibly similar to each other as well
- Most links arise from a combination of several mechanisms
  - difficult to attribute any individual link to a single mechanism

Measuring Homophily

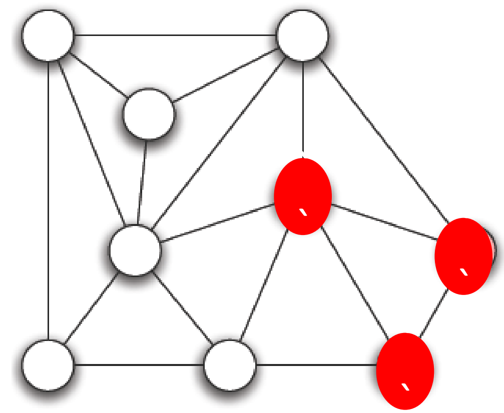
- Given a characteristic (like race, or age), how to test if a network exhibits homophily according to it?
- Ex friendship network:
  - Exhibits homophily by gender?
  - boys tend to be friends with boys, and girls tend to be friends with girls
  - cross-gender edges exist

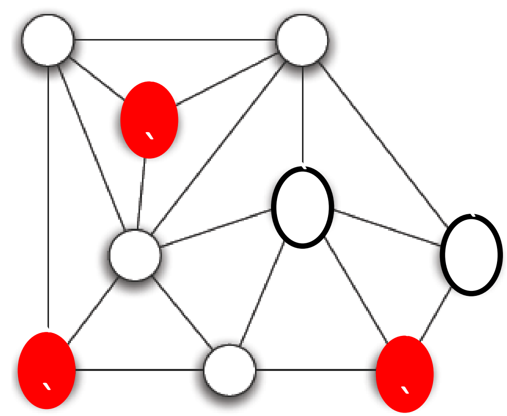


friendship network of a (hypothetical) classroom: shaded nodes are girls and the six unshaded nodes are boys

Measuring Homophily

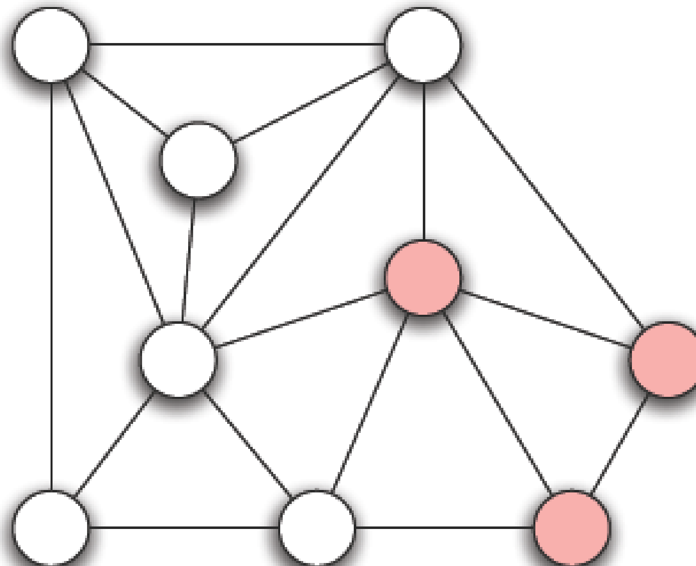
- Q: what would it mean for a network not to exhibit homophily by gender?
- A: number of cross-gender edges not **very different** from randomly assigning each node a gender
  - according to the gender balance in the original network





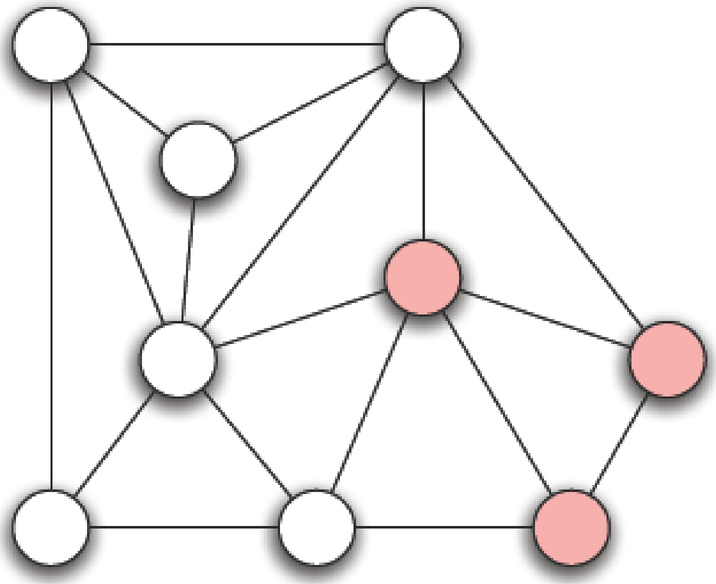
Measuring Homophily

- p the probability (fraction) of males
- q = 1-p the probability (fraction) of females
- For a given edge:
  - Homophily:
    - Prob(both ends male) = p*p
    - Prob(both ends female) = q*q
  - Cross gender:
    - Prob(ends male and female) = 2*p*q
- Homophily Test: If the fraction of cross-gender edges is **significantly** less than 2pq, then there is evidence for homophily



Measuring Homophily

- Ex:
  - p = 6/9 = 2/3
  - q = 1/3
  - 2pq = 4/9 = 8/18
  - 5/18 cross-gender edges
  - Test: 5/18 < 8/18 => some evidence of homophily
- Need definition of "significantly less than"
  - standard statistical significance
- What if cross-gender edges more than 2pq?
  - inverse homophily (Ex: network of romantic relationships)
- How to extend to characteristics with more than 2 states?



Mechanisms Underlying Homophily

- Homophily has 2 mechanisms for link creation
  - Selection: select friends with similar characteristics
    - individual characteristics drive the formation of links
    - involves immutable characteristics (determined at birth)
  - Social influence: modify behavior close to behaviors of friends
    - the reverse of selection
    - involves mutable characteristics (behaviors, activities, interests, beliefs, and opinions)

The Interplay of Selection and Social Influence

- Q: When homophily is observed, is it a result of selection or social influence?
  - Have people adapted their behaviors to become more like their friends, or have they selected friends who were already like them?
- A: Track the network and monitor the results of the two mechanisms (more details later)

The Interplay of Selection and Social Influence

- Most of the times, both mechanisms apply and interact with each other
- Studies show that teenage friends are similar to each other in their behaviors, and both selection and social influence apply:
  - teenagers seek social circles of people like them and peer pressure causes conform to behavioral patterns within these circles
- Q: how the two mechanisms interact and whether one is more strongly at work than the other? (more details later)

Affiliation

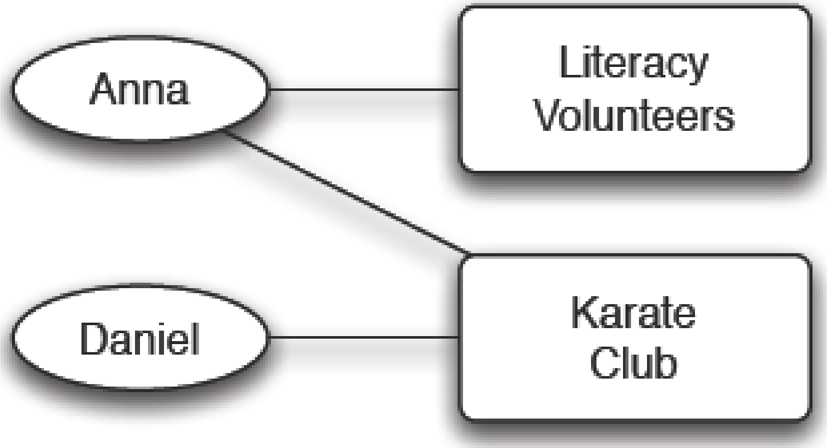
- Story so far:
  - Homophily groups together similar nodes
  - Selection and social influence determine the formation of links in a network
  - Similarity of nodes based on characteristics
- How to model these characteristics?
  - They represent surrounding contexts of networks
  - They exist "outside" the network
  - How to put these contexts into the network itself?

Affiliation

- Represent the set of activities a person takes part in (a general view of "activity")
  - Ex: part of a particular company, organization, frequenting a particular place, hobby
- Refer to activities as foci: "focal points" of social interaction

Affiliation Networks

- Affiliation network:
  - bipartite graph
    - nodes divided into 2 sets
    - no edges joining a pair of nodes that belong to the same set
  - people affiliated with foci
- Ex:
  - Anna participates in both of the social foci on the right
  - Daniel participates in only one

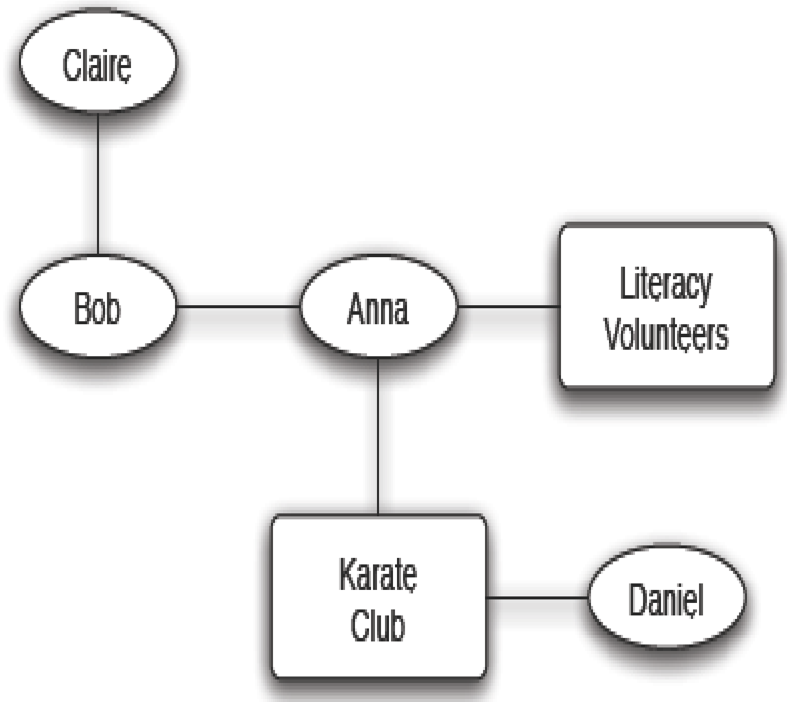


Co-Evolution of Social and Affiliation Networks

- Social networks change over time
  - new friendship links are formed
- Affiliation networks change over time
  - people become associated with new foci
- Co-evolution reflects interplay between selection and social influence
  - 2 people participate in a shared focus can become friends
  - if 2 people are friends, they can share their foci
- How to represent co-evolution with a single network?

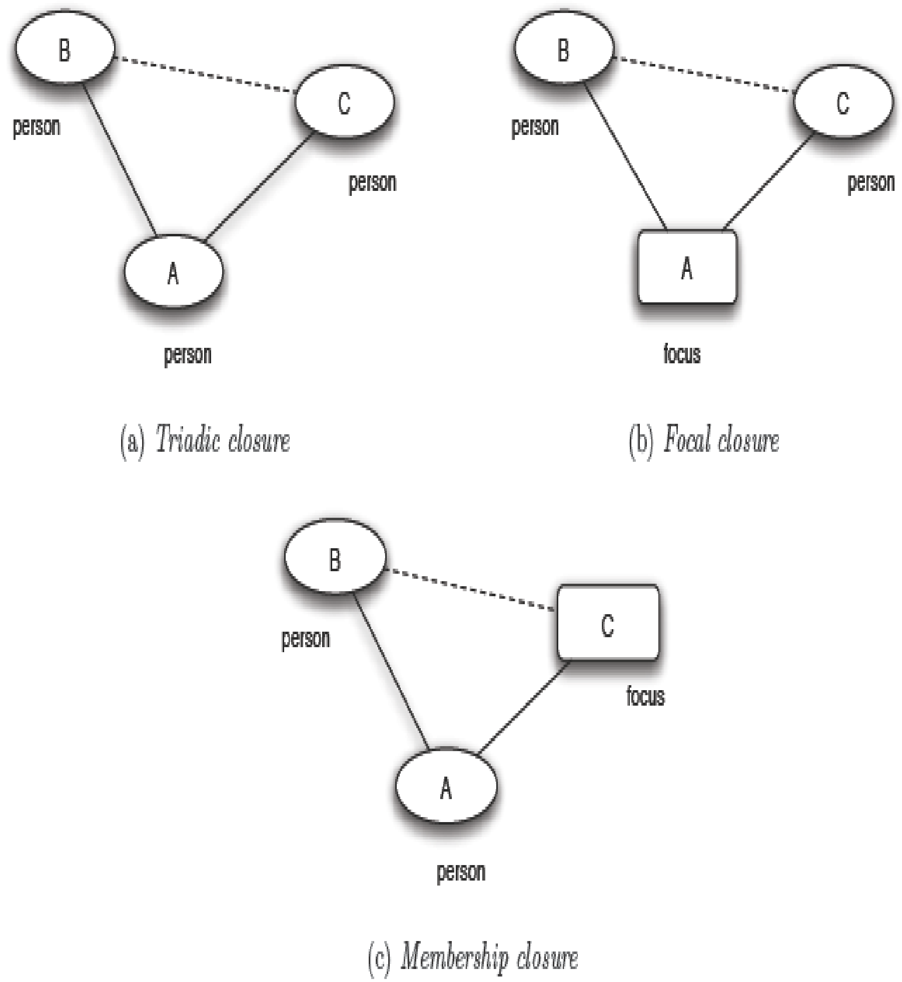
Social-affiliation networks

- Social-affiliation network contains:
  - a social network on the people and
  - an affiliation network on the people and foci



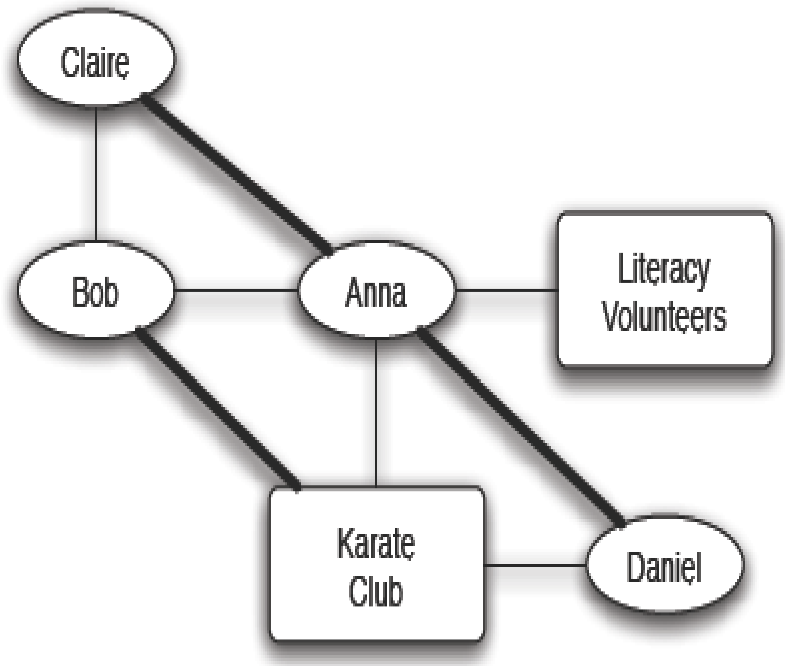
Social-affiliation networks

- In social-affiliation networks link formation as a closure process
- Several options for “closing” B-C
  - triadic closure: A, B, and C represent a person (already examined)
  - focal closure: B and C people, A focus
    - **selection**: B links to similar C (common focus)
  - membership closure: A and B people, C focus
    - **social influence**: B links to C influenced by A



Example

- Bob introduces Anna to Claire
- Karate "introduces" Anna to Daniel
- Anna introduces Bob to Karate



Edges with **bold** are the newly formed

Tracking Link Formation in On-Line Data

- Story so far: a set of mechanisms that lead to the formation of links
  - triadic closure
  - focal closure
  - membership closure
- Tracking these mechanisms in large populations
  - their accumulation observable in the aggregate

Tracking triadic closure

- Likelihood of link as a function of common friends?

1. Two snapshots of the network
2. For each k, find all pairs of nodes with k common friends in the first snapshot, but not directly connected
3. T(k): fraction of these pairs connected in the second snapshot
   - empirical estimate of probability that a link will form between two people with k common friends
4. Plot T(k) as a function of k
   - T(0) is the rate of link formation when it does not close a triangle

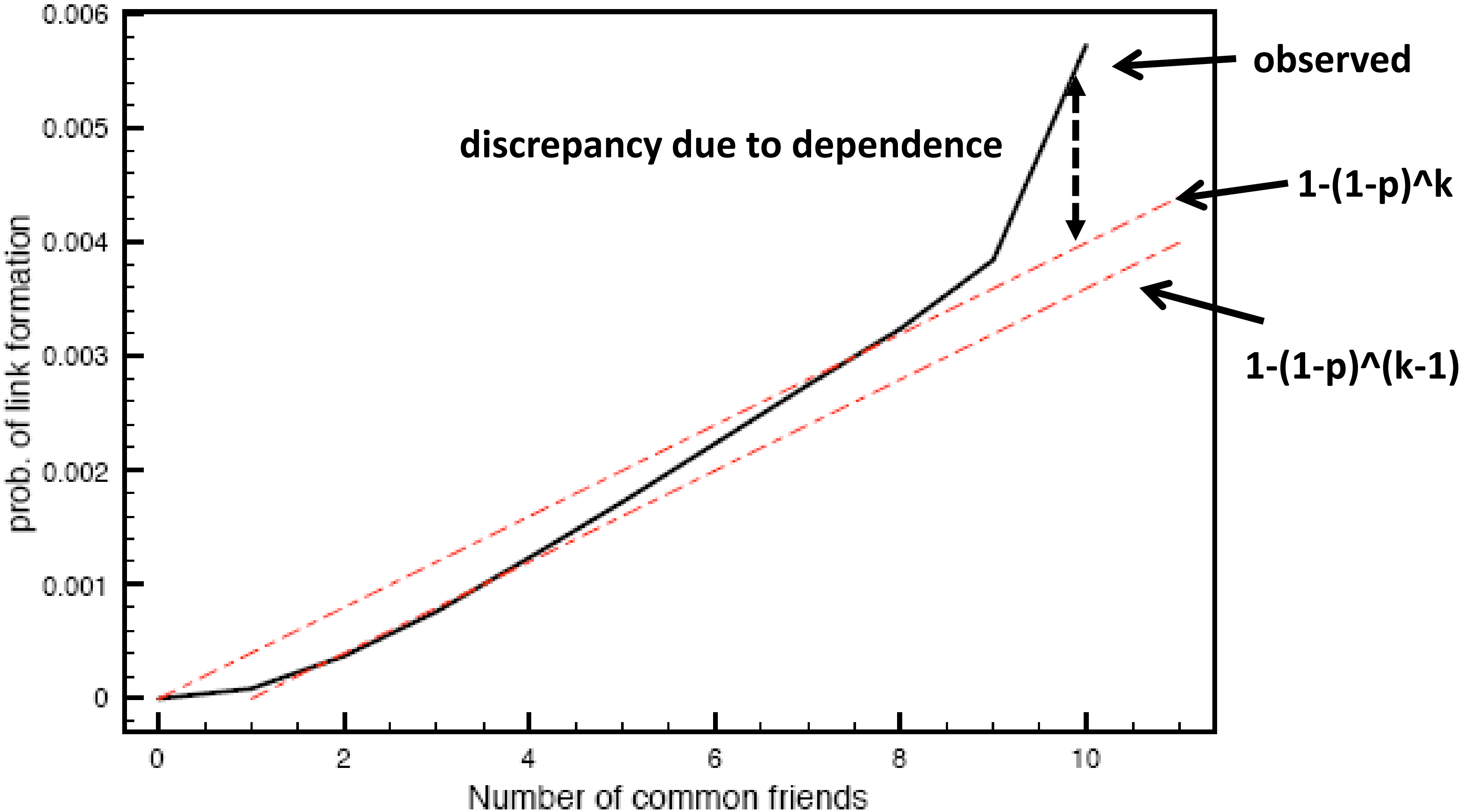
Tracking triadic closure

- Kossinets and Watts computed T(k)
  - full history of e-mail communication (“who-talks-to-whom”)
  - a one-year period
  - 22,000 students at a large U.S. university
  - observations in each snapshot were one day apart (average over multiple snapshots)

Tracking triadic closure

- Interpret the result compared to a baseline
- Assume that each common friend that 2 people have, gives them an independent probability **p** of forming a link
  - 2 people have k friends in common => the probability they fail to form a link is **(1-p)^k**
  - 2 people have k friends in common => probability that they form a link is **1-(1-p)^k**

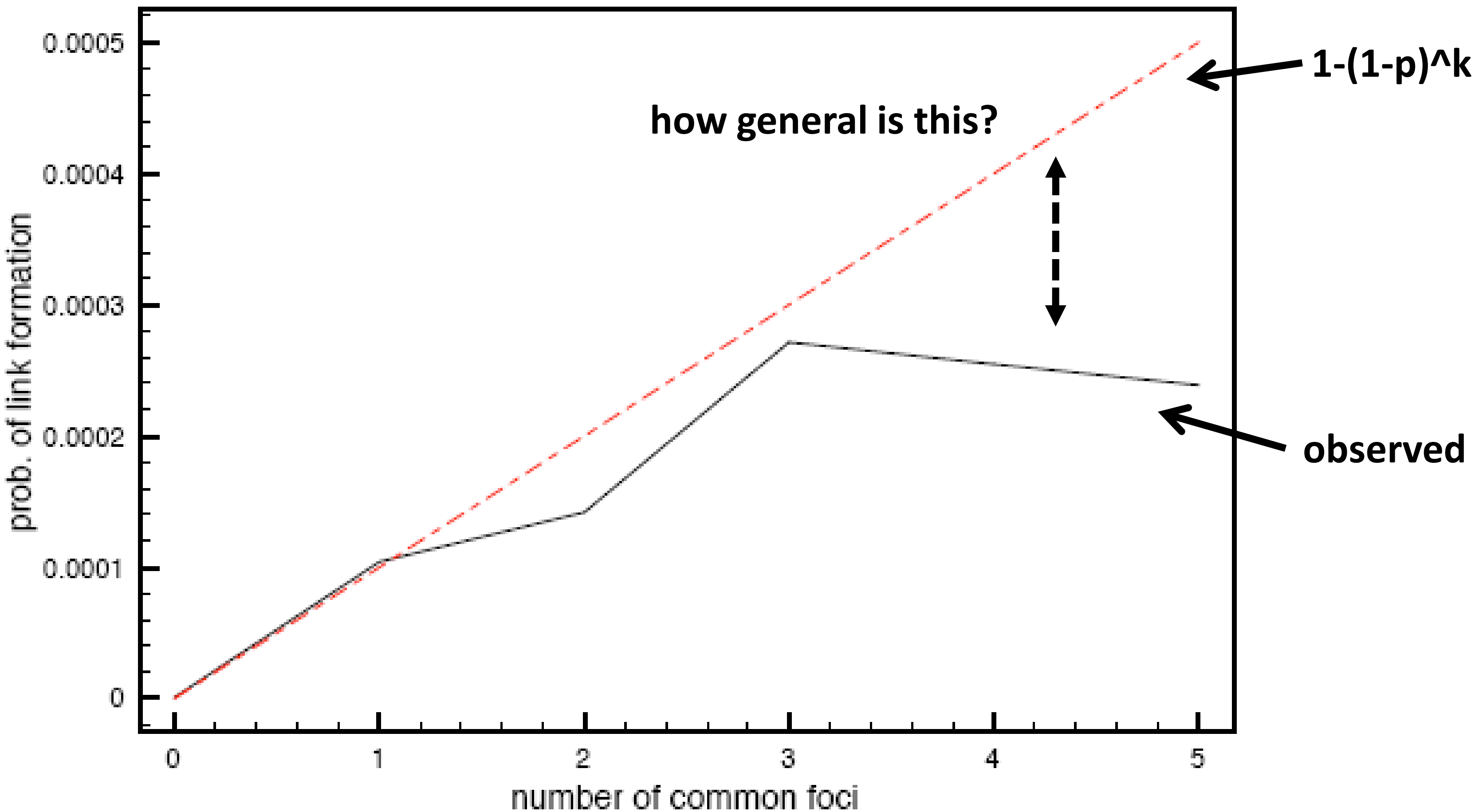
Tracking triadic closure



Tracking focal closure

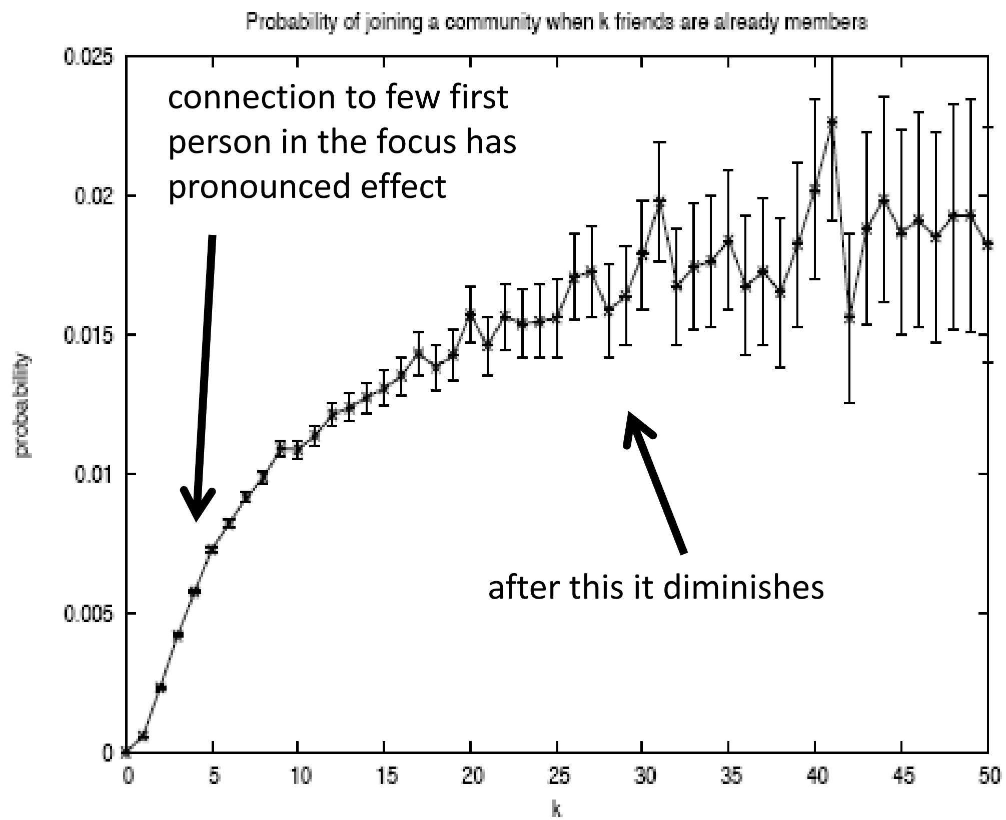
- Likelihood of link formation as a function of the number of common foci?
- Kossinets and Watts supplemented their university e-mail dataset with information about the class schedules
  - each class became a focus
  - students shared a focus if they had taken a class together

Tracking focal closure



Tracking membership closure

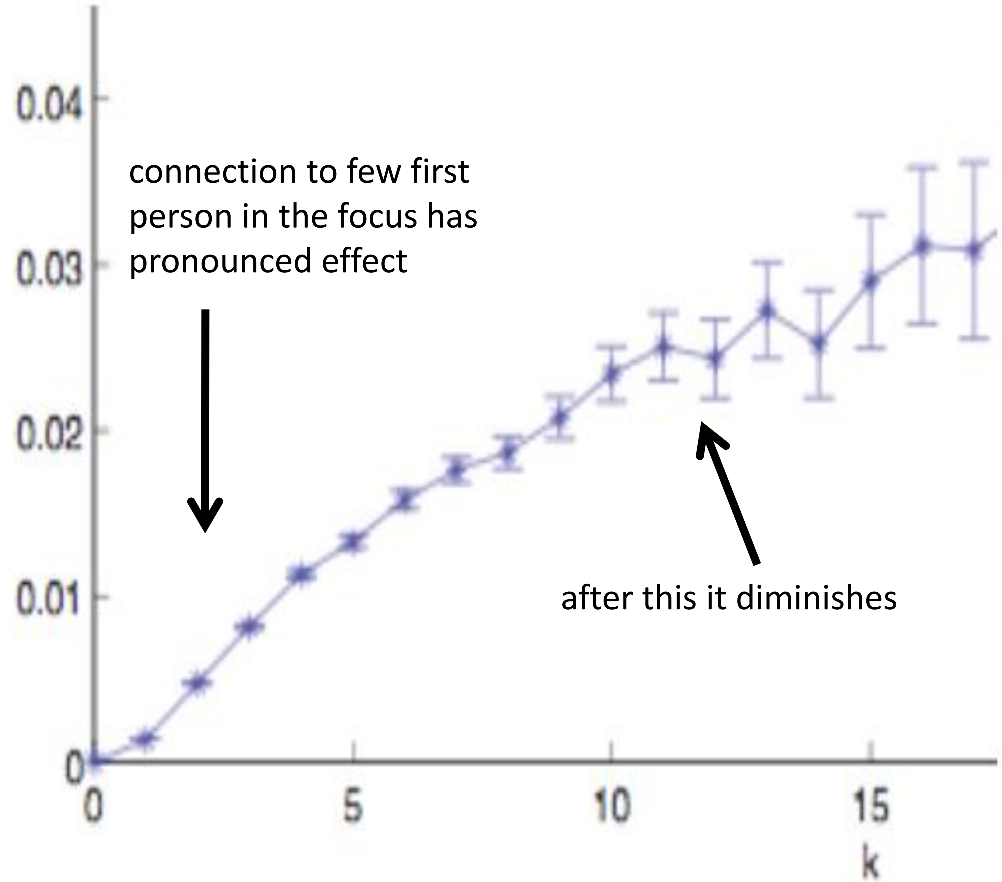
- Blogging site LiveJournal
  - social network (friendship links)
  - foci correspond to membership in user-defined communities



probability of joining a LiveJournal community as a function of the number of friends who are already members

Tracking membership closure

- Wikipedia editors
  - link editors when they communicated (user talk page)
  - each Wikipedia article defines a focus (editor associated with the articles he/she edited)



probability of editing a Wikipedia articles as a function of the number of friends who have already done so

Quantifying the Interplay Between Selection and Social Influence

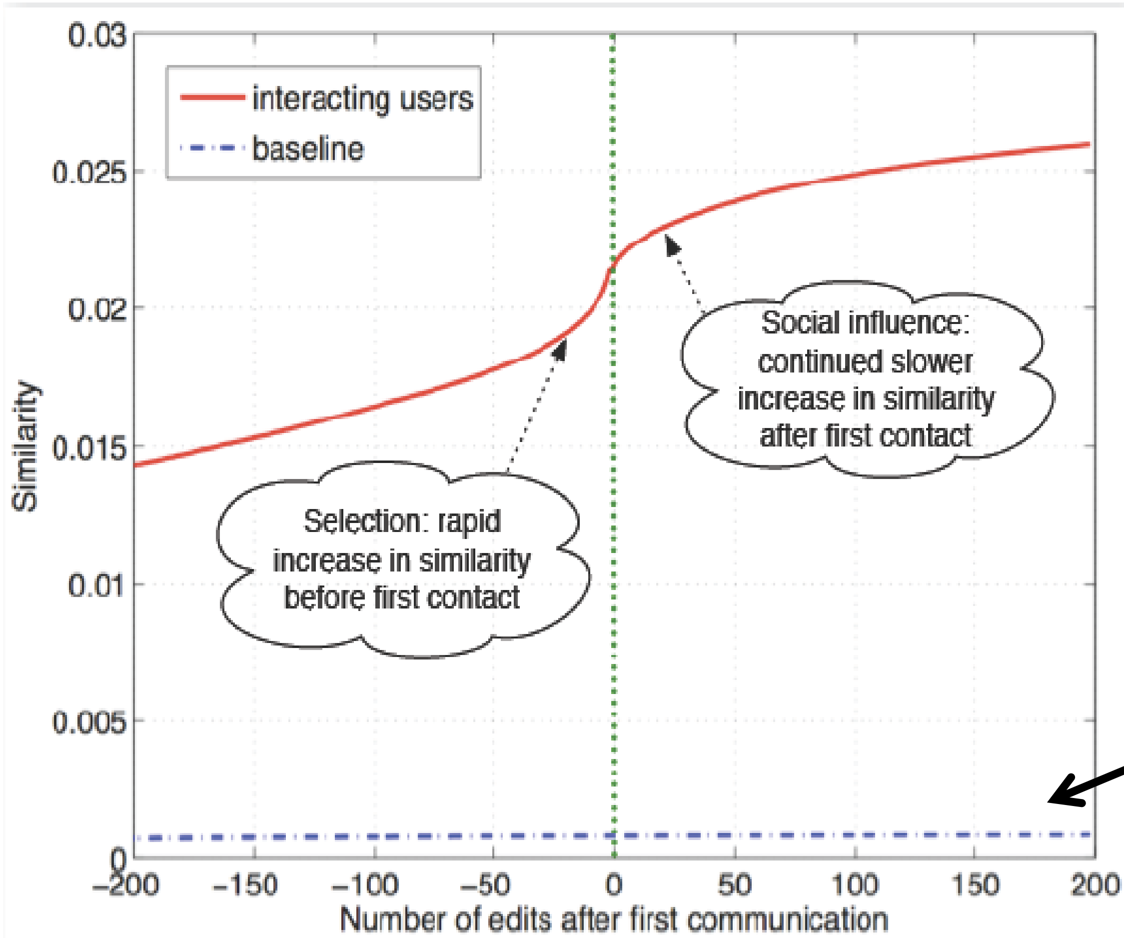
- How selection and social influence work together to produce homophily?
  - How do similarities in behavior between two Wikipedia editors relate to their pattern of social interaction over time?
  - Similarity between 2 Wikipedia editors A, B:

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

- Is homophily (similarity) due to editors connected (talk) with those edited the same articles (selection), or because editors are led to edit articles by those they talk to (social influence)?

Quantifying the Interplay Between Selection and Social Influence

Record similarity over time for each pair of editors A and B who have ever talked



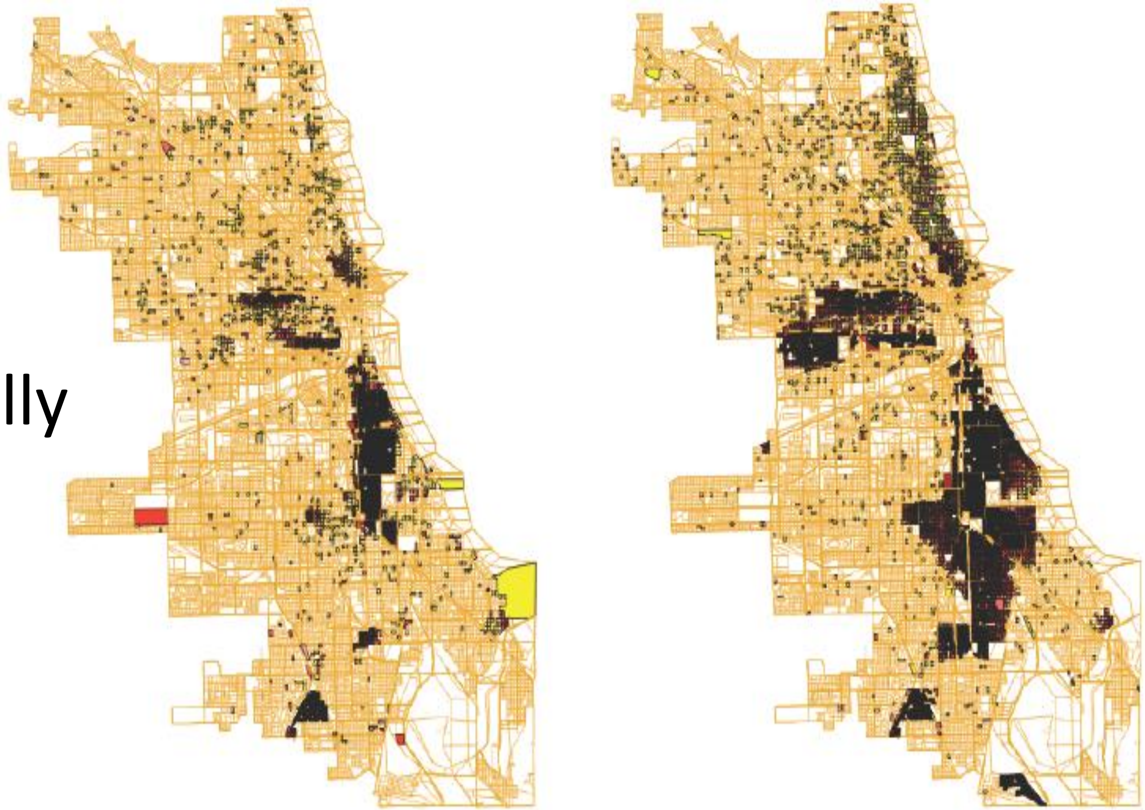
similarity of non-interacting pairs

“tick” in time whenever either A or B performs an action (editing or talking). Time 0 is the point at which they first talked

A SPATIAL MODEL OF SEGREGATION

Spatial patterns of segregation

- One of the most strong effects of homophily is in the formation of ethnically and racially homogeneous neighborhoods in cities
  - a process with a dynamic aspect
  - what mechanisms?



(a) *Chicago, 1940* (b) *Chicago, 1960*

In blocks colored yellow and orange the percentage of African-Americans is below 25, while in blocks colored brown and black the percentage is above 75

The Schelling Model

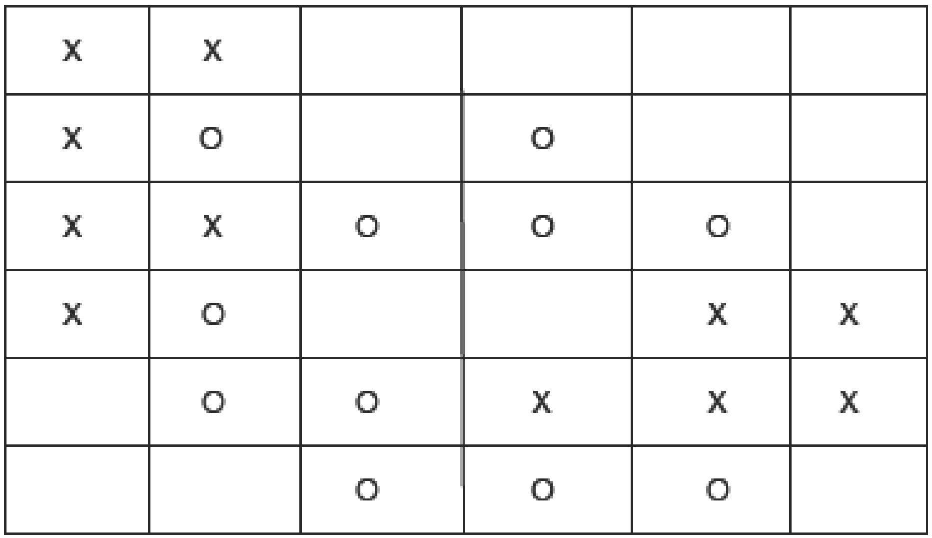
- How global patterns of spatial segregation can arise from the effect of homophily operating at a local level (Thomas Schelling)
  - an intentionally simplified mechanism
  - works even when no one individual explicitly wants a segregated outcome

The Schelling Model

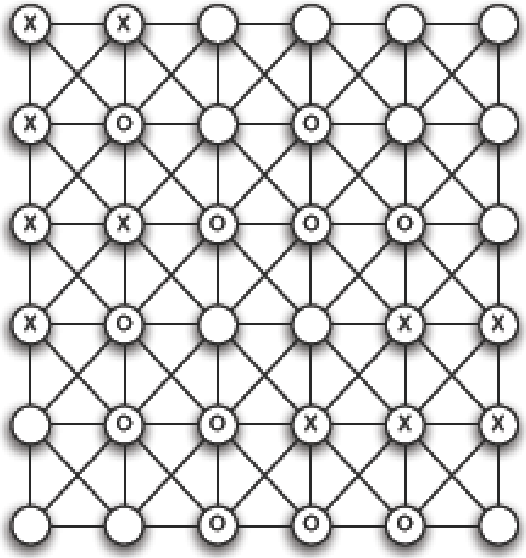
- Model assumptions:
  - Population of individuals called agents
  - Each agent of type X or type O
  - The two types represent some characteristic as basis for homophily (race, ethnicity, country of origin, or native language)
  - Agents reside in cells of a grid (simple model of a 2-D city map)
  - Some cells contain agents while others are unpopulated
  - Cell’s neighbors: cells that touch it (including diagonal contact)

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

The Schelling Model



(a) *Agents occupying cells on a grid.*



(b) *Neighbor relations as a graph.*

Cells are the nodes and edges connect neighboring cells.
We will continue with the geometric grid rather than the graph.

The Schelling Model

- Local mechanism:
  - each agent wants to have at least some t other agents of its own type as neighbors (t the same for all)
  - unsatisfied agents have fewer than t neighbors of the same type as itself and move to a new cell
- Ex (figure):
  - agents with ID
  - t = 3

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

(a) *An initial configuration.*

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

(b) *After one round of movement.*

The Dynamics of Movement

- Unsatisfied agents move in rounds
  - consider unsatisfied agents in some order
    - random or row-sweep
  - unsatisfied agents move to an unoccupied cell where will be satisfied
    - random or to nearest cell that satisfies them
  - may cause other agents to be unsatisfied
  - deadlocks may appear (no cell that satisfies)
    - stay or move randomly
- All variations have similar results
- Ex (figure):
  - t=3, one round, row-sweep, move to nearest cell, stay when deadlocks

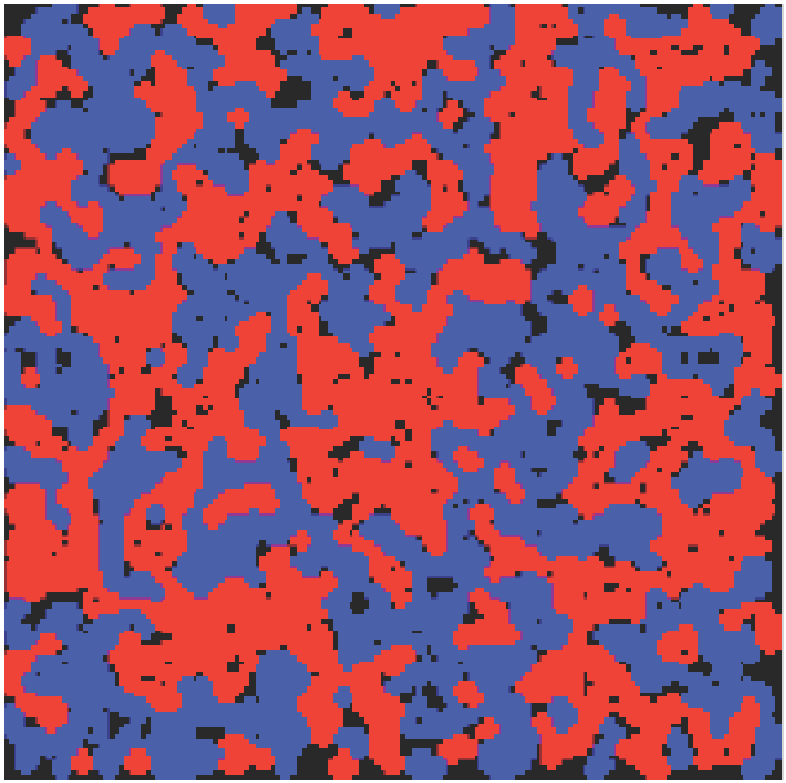
X1*	X2*				
X3	O1*		O2		
X4	X5	O3	O4	O5*	
X6*	O6			X7	X8
	O7	O8	X9*	X10	X11
		O9	O10	O11*	

(a) *An initial configuration.*

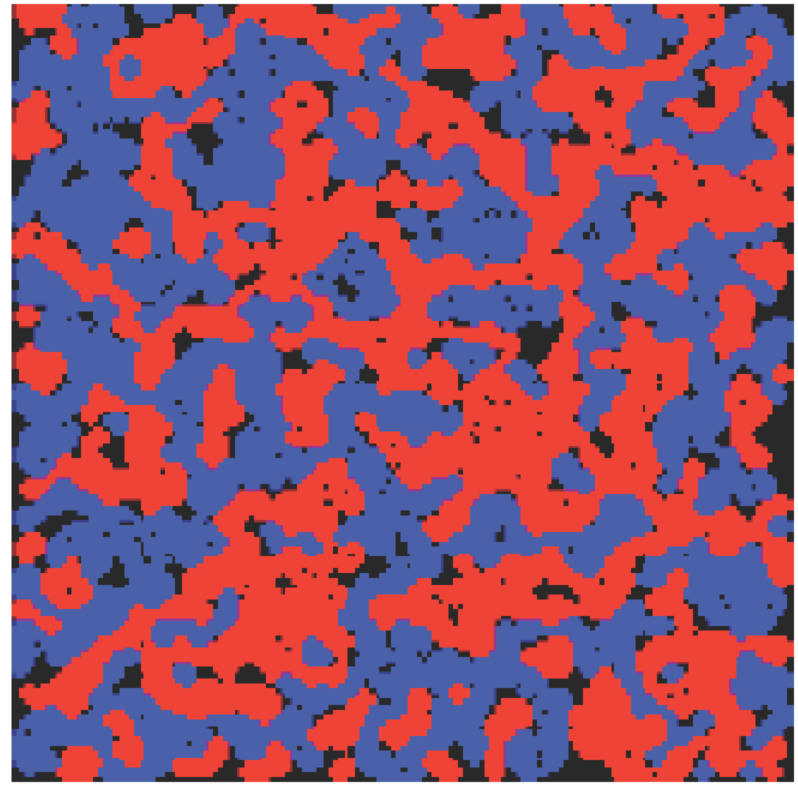
X3	X6	O1	O2		
X4	X5	O3	O4		
	O6	X2	X1	X7	X8
O11	O7	O8	X9	X10	X11
	O5	O9	O10*		

(b) *After one round of movement.*

Larger examples



(a) *A simulation with threshold 3.*



(b) *Another simulation with threshold 3.*

Two runs (50 rounds) of the Schelling model with unsatisfied agents moving to a random location. Threshold t=3, 150-by-150 grid with 10, 000 agents. Each cell of first type is red, of second type blue, or black if unoccupied.

Interpretations of the Model

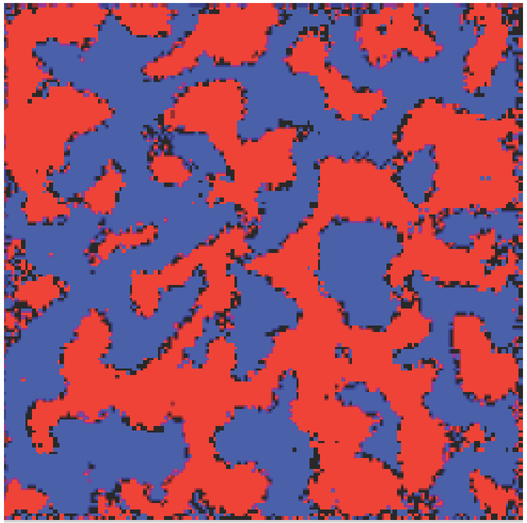
- Spatial segregation is taking place even though no individual agent is seeking it
  - agents just want to be near **t** others like them
  - when t=3, agents are satisfied being minority among its neighbors (5 neighbors of the opposite type)
- Ex (figure):
  - a checkerboard 4x4 pattern can make all agent satisfied (even for large grids)
  - we don't see this result in simulations

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

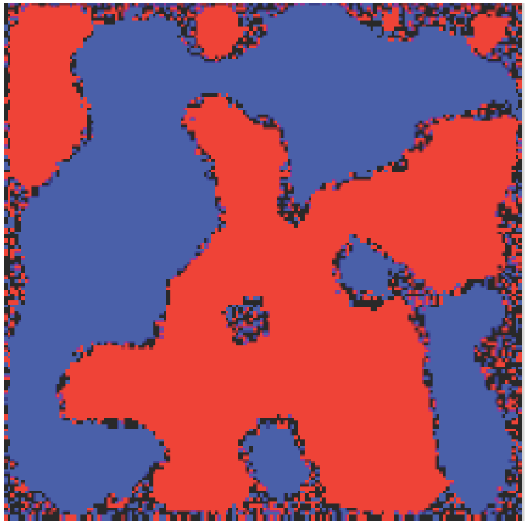
Interpretations of the Model

- More typically, agents form larger clusters
  - agents become unsatisfied and attach to larger clusters (where higher probability to be satisfied)
- The overall effect:
  - local preferences of individual agents have produced a global pattern that none of them necessarily intended

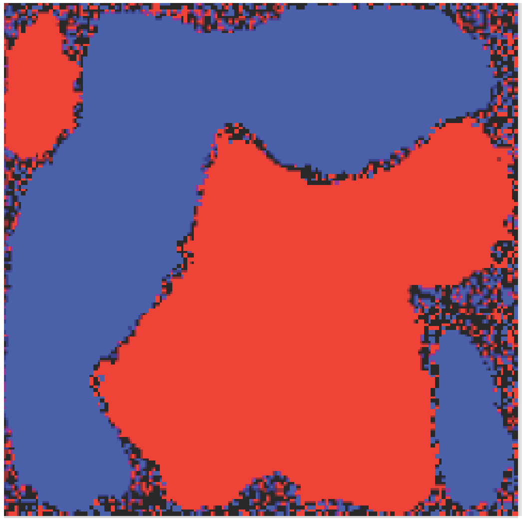
Interpretations of the Model



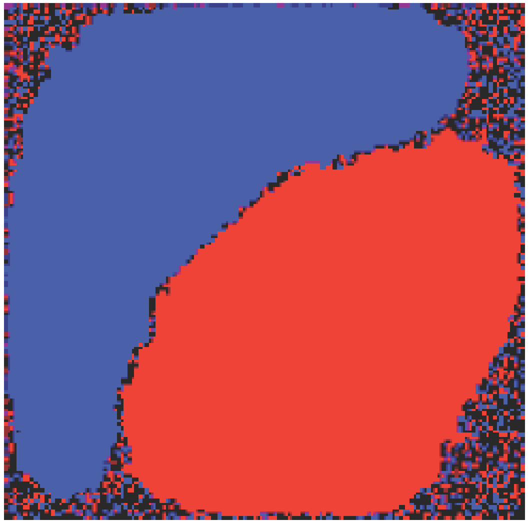
(a) *After 20 steps*



(b) *After 150 steps*



(c) *After 350 steps*



(d) *After 800 steps*

t=4, 150-by-150 grid, 10, 000 agents, varying number of rounds (steps), not shown until the end

Schelling model and Homophily

- The Schelling model is an example that, as homophily draws people together along immutable characteristics (race or ethnicity), it creates a natural tendency for mutable characteristics (decision about where to live) to change in accordance with the network structure