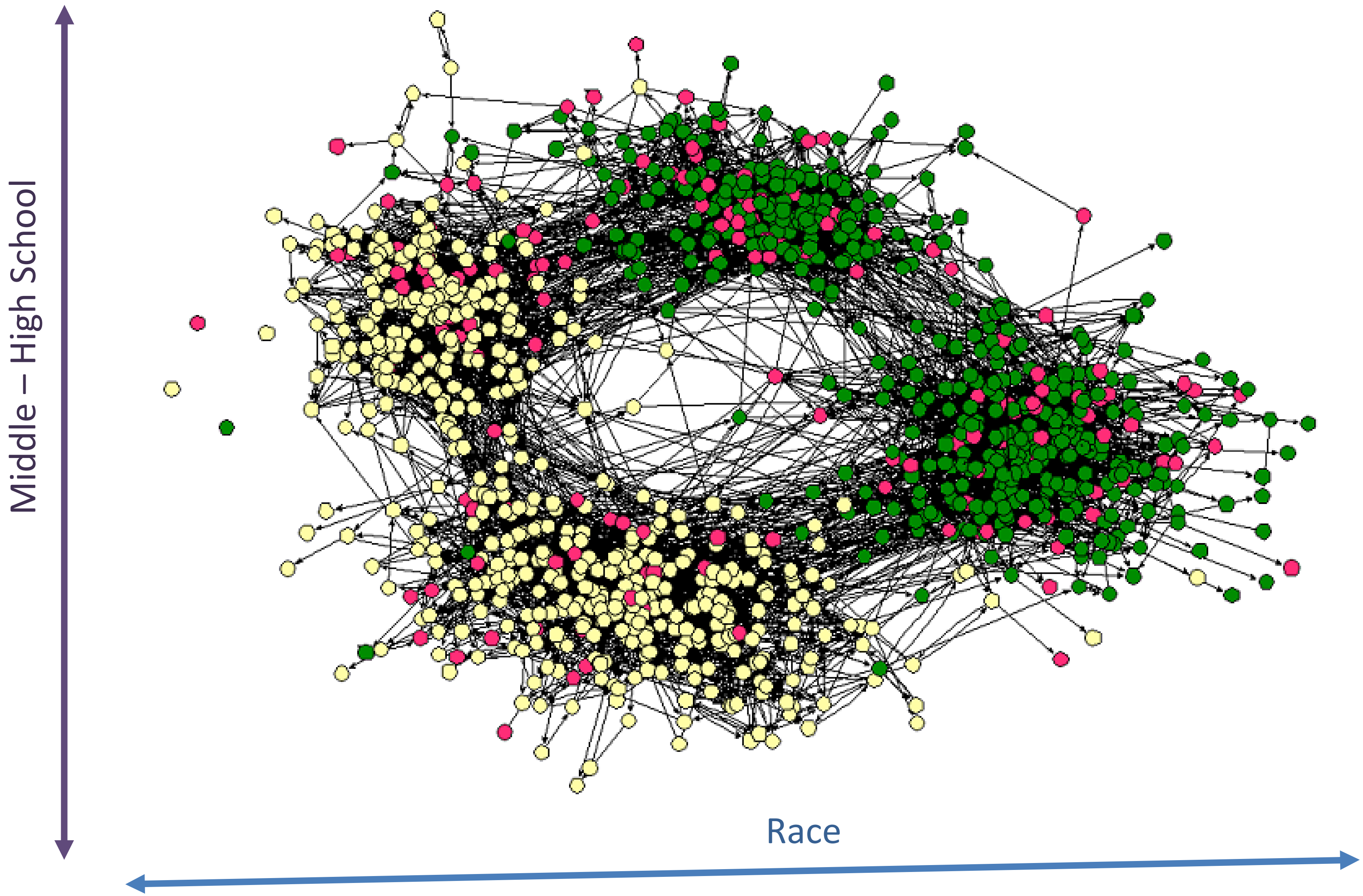
Online Social Networks and Media

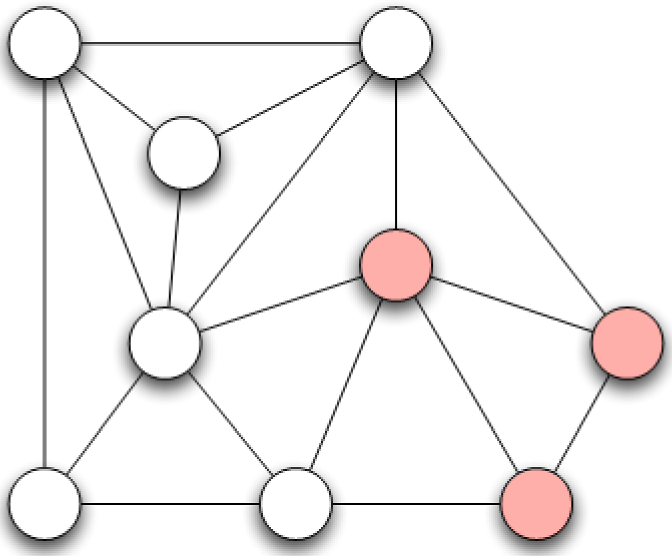
Networks and Surrounding Contexts

Chapter 4, from D. Easley and J. Kleinberg book

Homophily



Measuring Homophily



If the fraction of cross-gender edges is significantly less than expected, then there is evidence for homophily

gender male with probability p
gender female with probability q

Probability of cross-gender edge?

Measuring Homophily

- “significantly” less than
- Inverse homophily
- Characteristics with more than two values:
  - Number of heterogeneous edges (edge between two nodes that are different

Mechanisms Underlying Homophily:
Selection and Social Influence

Selection: tendency of people to form friendships with others who are like then

Socialization or Social Influence: the existing social connections in a network are influencing the individual characteristics of the individuals

Social Influence as the inverse of Selection

Mutable & immutable characteristics

The Interplay of Selection and Social Influence

Longitudinal studies in which the social connections and the behaviors within a group are tracked over a period of time

Why?

- Study teenagers, scholastic achievements/drug use (peer pressure and selection)
- Effect of possible interventions (example, drug use)

The Interplay of Selection and Social Influence

Christakis and Fowler on obesity, 12,000 people over a period of 32-years

Why?

(i) Because of selection effects, choose friends of similar obesity status,

(ii) Because of confounding effects of homophily according to other characteristics

(iii) Because changes in the obesity status of person's friends was exerting an influence that affected her

(iii) As well -> "contagion" in a social sense

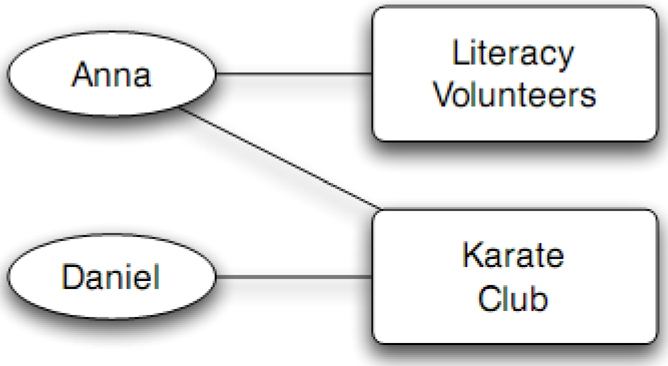
Affiliation

A larger network that contains both people and context as nodes

foci

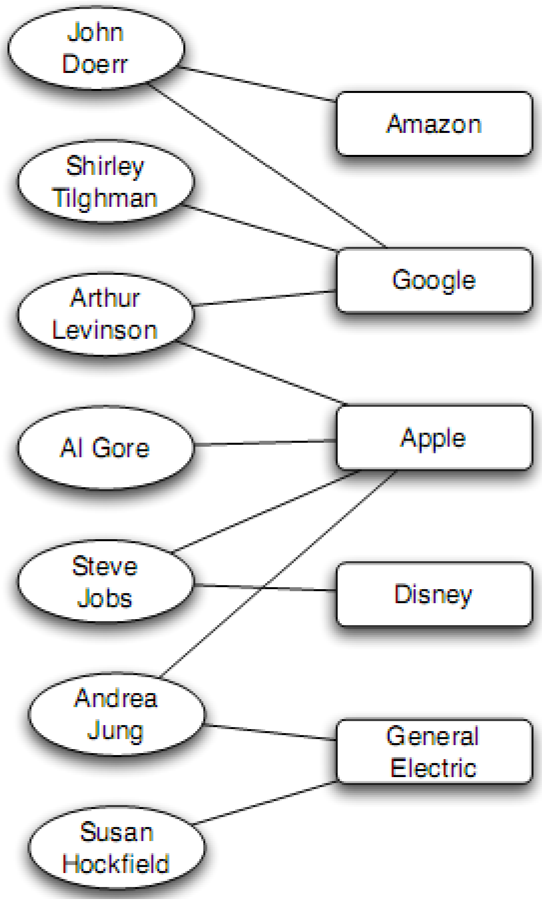
Affiliation network

Bipartite graph
A node for each person and a node for each focus
An edge between a person A and focus X, if A participates in X



Affiliation

Example:
Board of directors

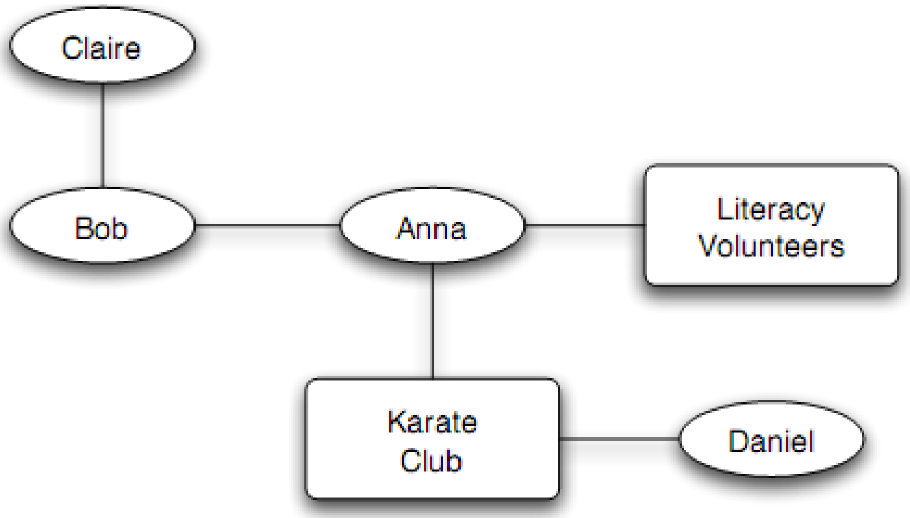


Co-evolution of Social and Affiliation Networks

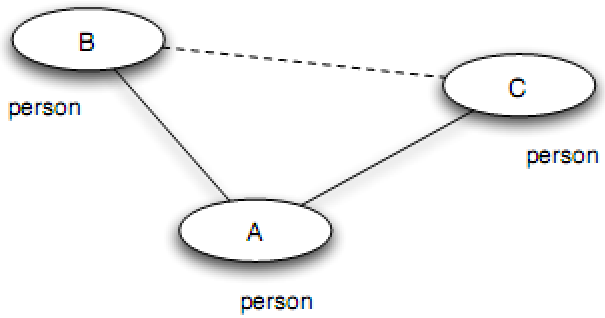
Social Affiliation Network

Two type of edges:

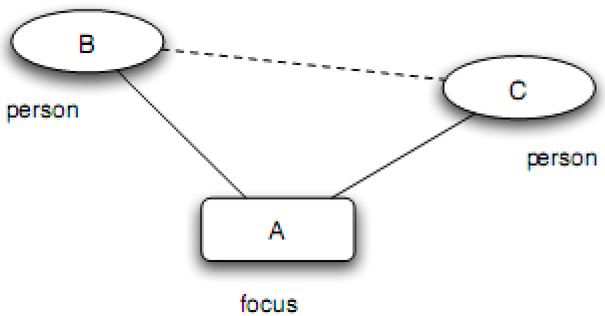
1. *Friendship:* between two people
2. *Participation:* between a person and a focus



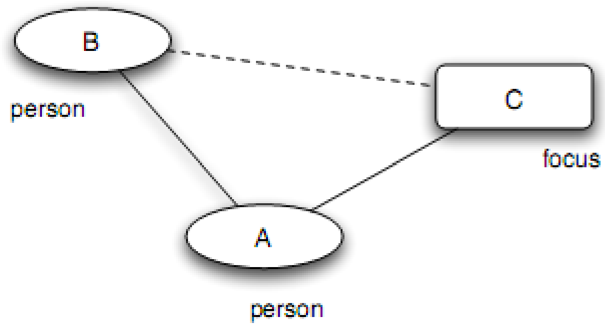
Co-evaluation of Social and Affiliation Networks: Closure process



Triadic closure: (two people with a friend in common - A introduces B to C)

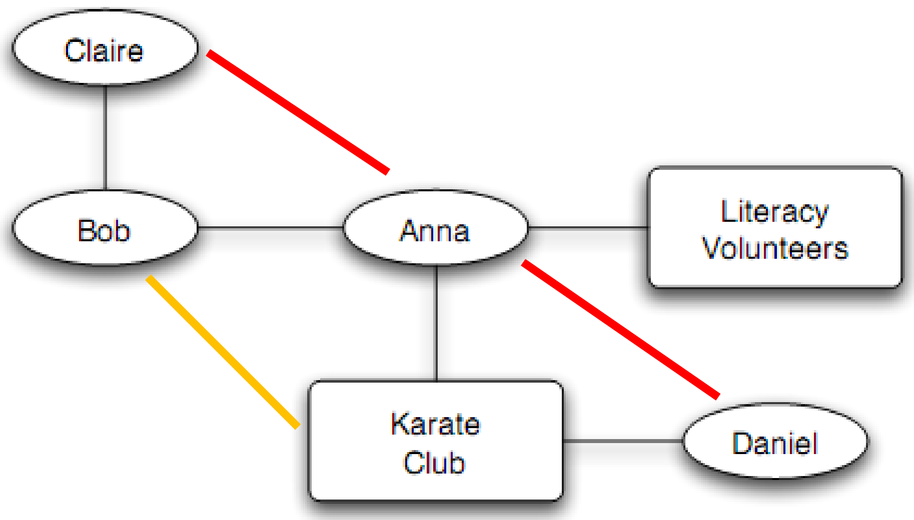


Focal closure: (two people with a focus in common - focus A introduces B to C)



Membership closure (social influence): (a person joining a focus that a friend is already involved in A introduces focus C to B)

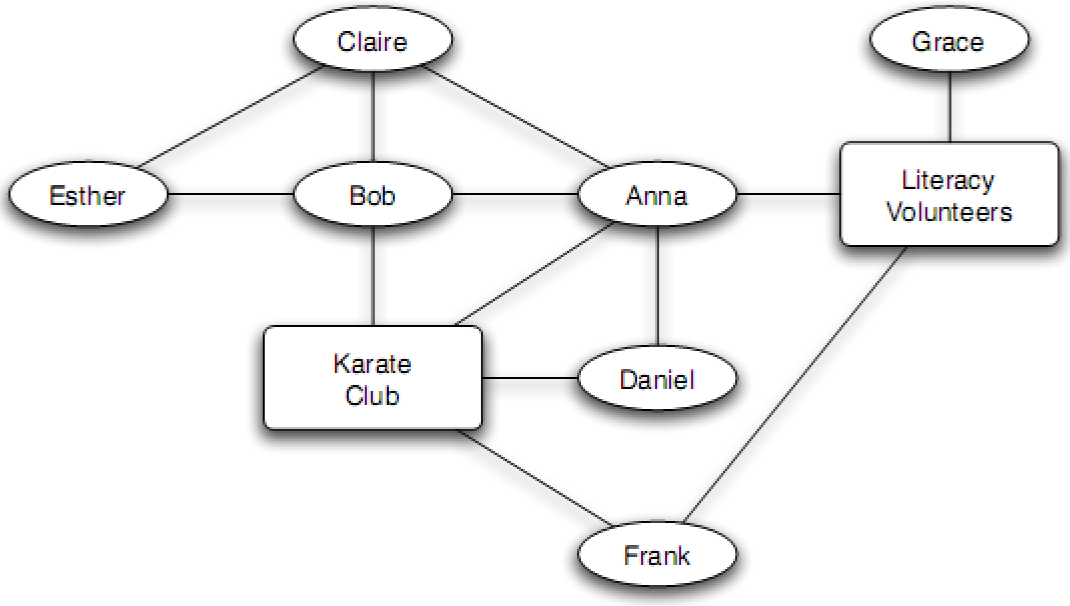
Co-evaluation of Social and Affiliation Networks



Tracking Link Formation in Online Data: triadic closure

Triadic closure:

- How much likely is a link to form between two people
- How much more likely a link to form between two people if they multiple friends in common?



Tracking Link Formation in Online Data: triadic closure

We take two snapshots of the network at different times

I. For each k, identify all pairs of nodes that have exactly k friends in common in the first snapshot, but who are not directly connected
II. Define T(k) to be the fraction of these pairs that have formed an edge by the time of the second snapshot
III. Plot(k) as a function of k

T(0): does not close any triangle
T(k): the rate at which link formation happens when it does close a triangle

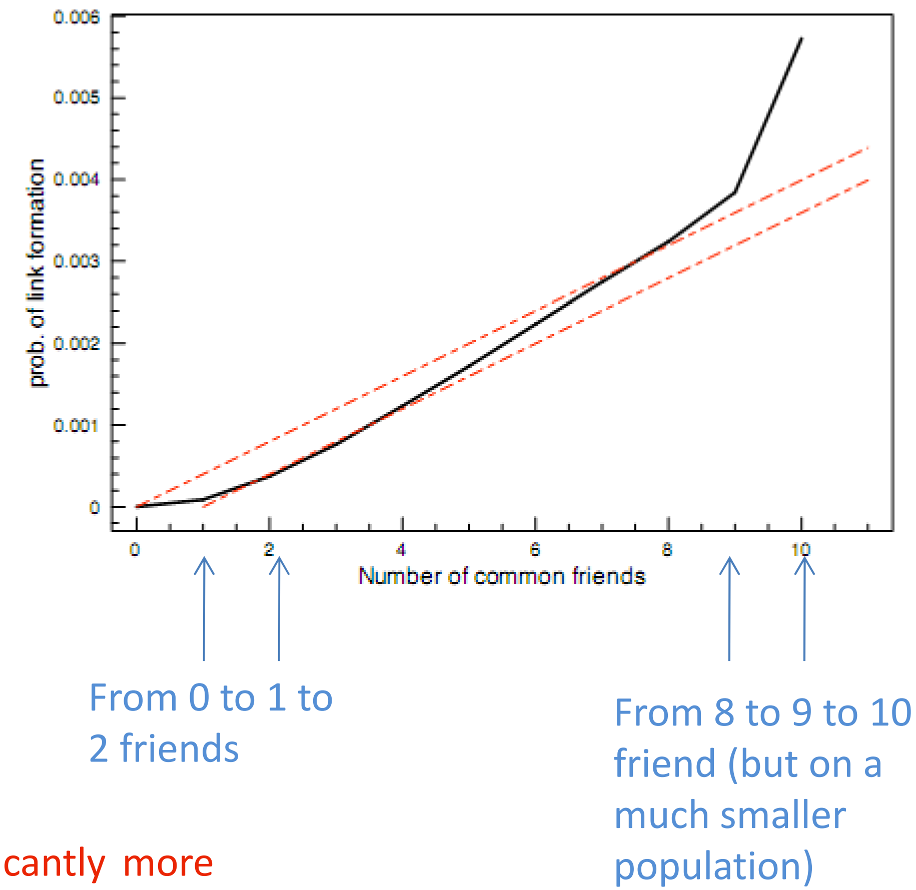
Tracking Link Formation in Online Data: triadic closure

E-mail ("who-talks-to-whom" dataset type)
Among 22,000 undergraduates and graduate students (large US universty)
For 1-year

Network evolving over time

- At each instance (snapshot), join two people if they have exchanged e-mail in each direction at some point in the past 60 days
- Multiple pairs of snapshots ->
- Built a curve for T(k) on each pair, then average all the curves

Snapshots – one day apart



Having two common friends produces significantly more than twice the effect compared to a single common friend

Tracking Link Formation in Online Data: triadic closure

Assume triadic closure:
Each common friend two people have gives them an independent probability p of forming a link each day

For two people with k friend in common,
probability not forming a link on any given day

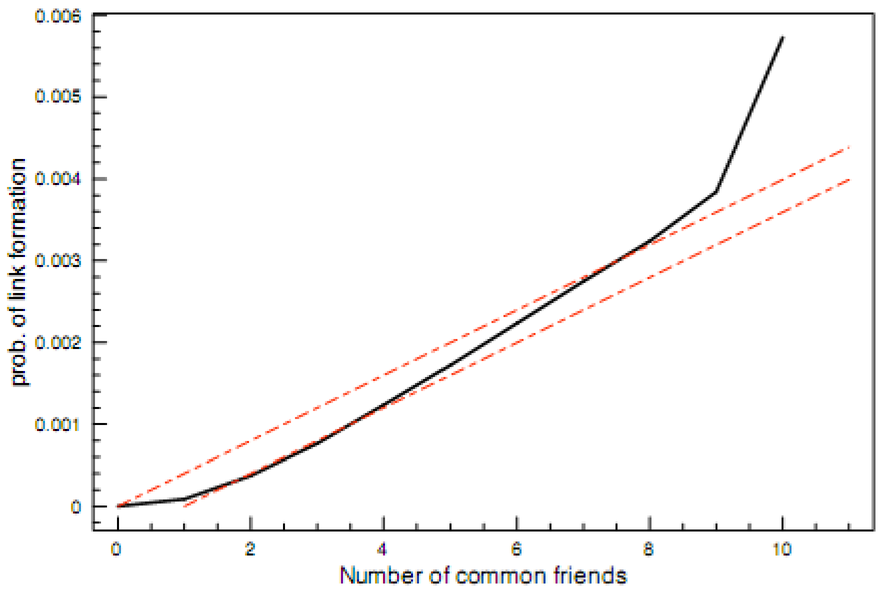
$$(1-p)^k$$

probability forming a link on any given day

$$T_{baseline}(k) = 1 - (1-p)^k$$

Given the absolute effect of the first common friend in the data

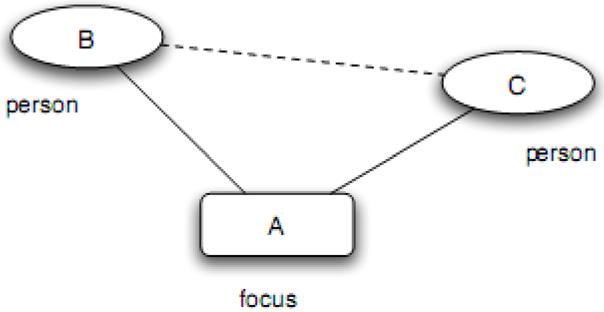
$$T_{baseline}(k) = 1 - (1-p)^{k-1}$$

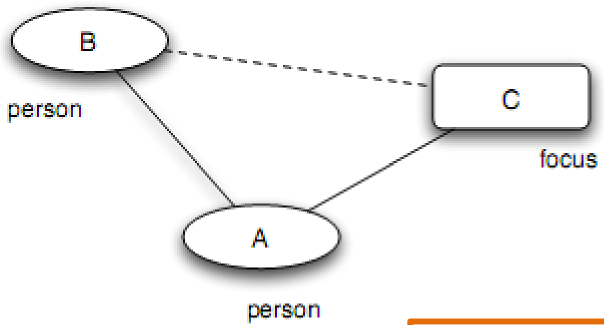


Qualitative similar, but no independent assumption

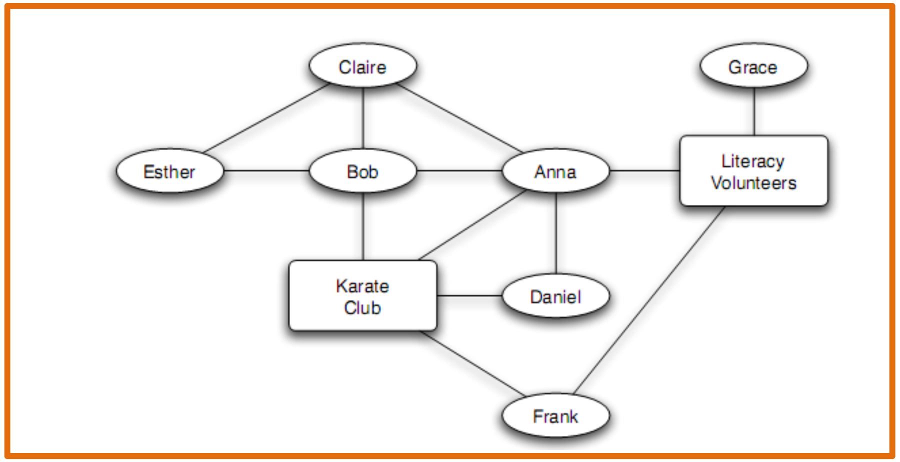
Tracking Link Formation in Online Data: focal and membership closure

Focal closure: what is the probability that two people form a link as a function of the *number of foci* that are jointly affiliated with





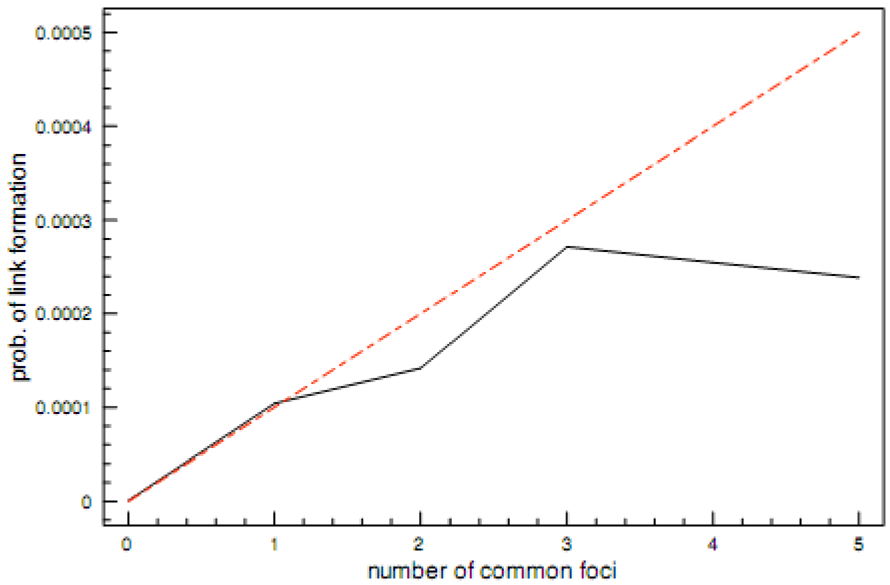
Membership closure: what is the probability that a person becomes involved with a particular focus as a function of the *number of friends* who are already involved in it?



Tracking Link Formation in Online Data: focal and membership closure

For focal closure

E-mail ("who-talks-to-whom" dataset type)
Use the class schedule of each student
Focus: class (common focus – a class together)



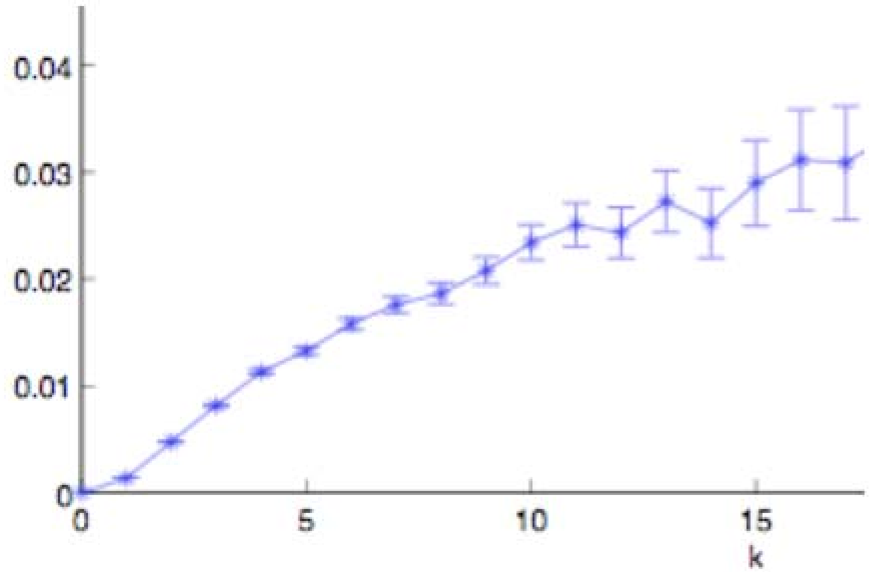
Subsequent shared classes after the first produce a diminishing effect

Tracking Link Formation in Online Data: focal and membership closure

For membership closure

Wikipedia
Node: Wikipedia editor who maintains a use account and user talk page
Link: if they have communicated with one writing on the user talk page of the other
Focus: Wikipedia article
Association to focus: edited the article

An initial increasing effect: the probability of editing a Wikipedia article is more than twice as great when you have two connections into the focus than one



✓Also, multiple effects can operate simultaneously

Tracking Link Formation in Online Data: interplay between selection and social influence

- Underlying social network
- Measure for behavioral similarity

Wikipedia
Node: Wikipedia editor who maintains a use account and user talk page
Link: if they have communicated with one writing on the user talk page of the other

Editor's behavior: set of articles she has edited

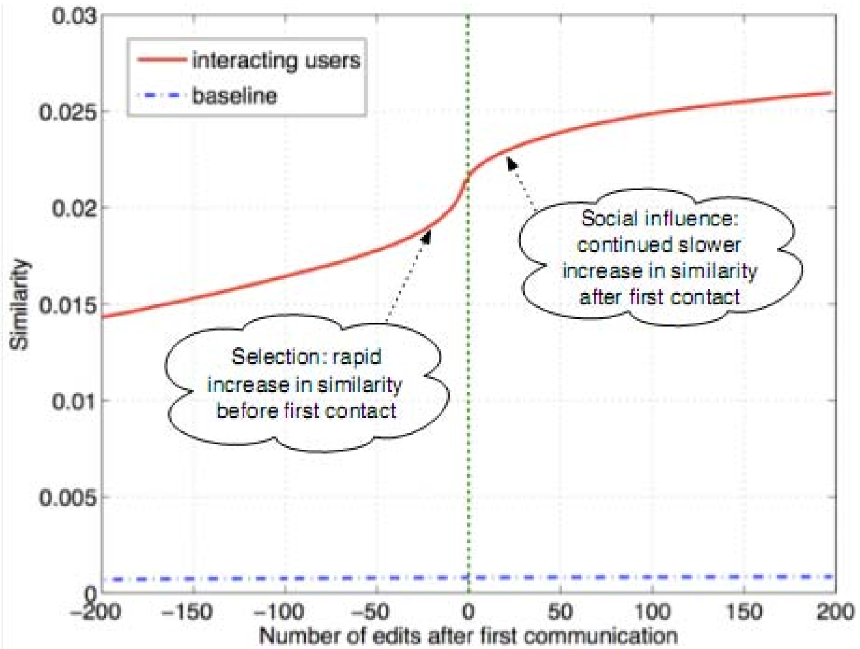
$$\frac{|N_A \cap N_B|}{|N_A \cup N_B|}$$

Neighborhood overlap in the bipartite affiliation network of editors and articles consisting only of edges between editors and the articles they have edited

Wikipedia editors who have communicated are significantly more similar in their behavior than pairs of Wikipedia editors who have not (homomphily)
Selection (editors form connections with those have edited the same articles) vs Social Influence (led to the articles of people they talk to)

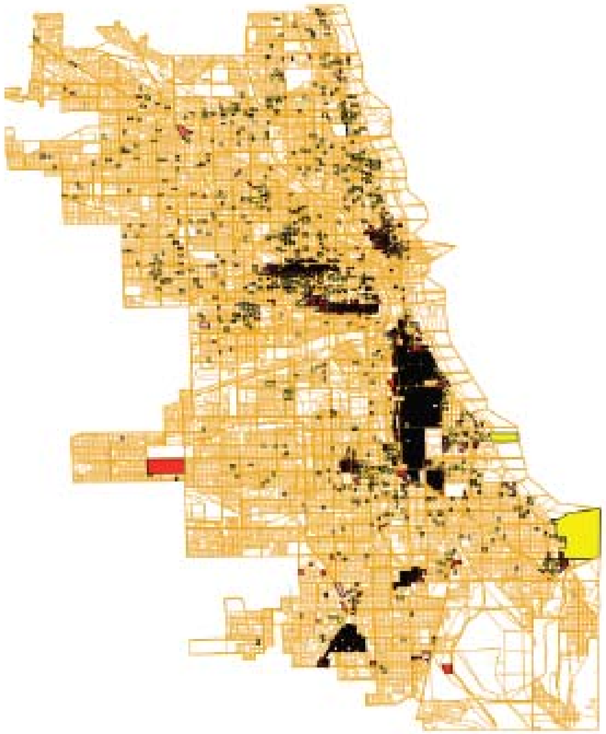
Tracking Link Formation in Online Data: interplay between selection and social influence

Actions in Wikipedia are timestamped
For each pair of editors A and B who have ever communicated record their similarity over time
Time moves in discrete units, advancing by one “tick” whenever either A or B performs an action on Wikipedia
Time 0 when they first communicated

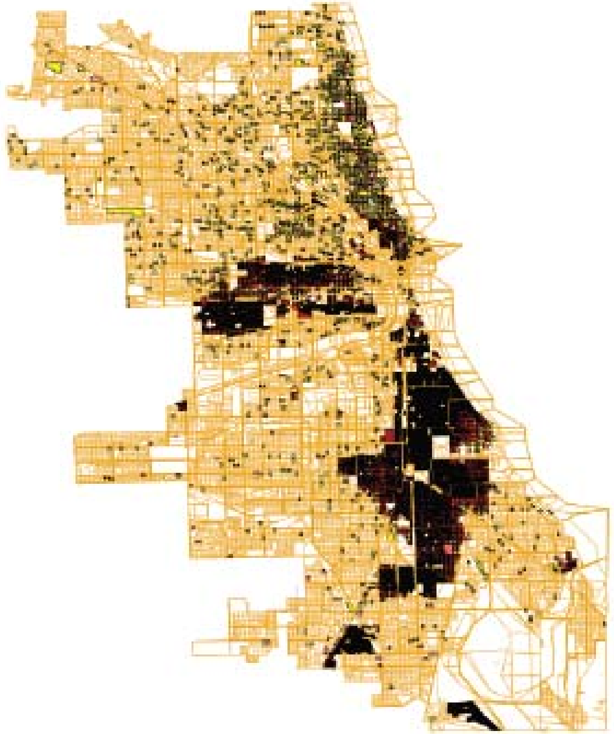


Similarity is clearly increasing both before and after the moment of first interaction (both selection and social influence)
Not symmetric around time 0 (particular role on similarity)
Significant increase before they meet
Blue line shows similarity of a random pair (non-interacting)

A Spatial Model of Segregation



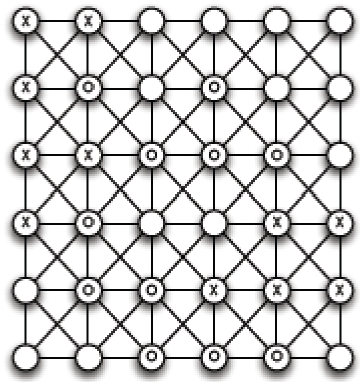
(a) *Chicago, 1940*



(b) *Chicago, 1960*

A Spatial Model of Segregation: The Schelling Model

Neighbors (including diagonal)



A threshold t: if an agent discovers that fewer than t of its neighbors are of the same type of itself, then it has an interest to move to a new cell
Unsatisfied

x	x				
x	o		o		
x	x	o	o	o	
x	o			x	x
	o	o	x	x	x
		o	o	o	

t = 3

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

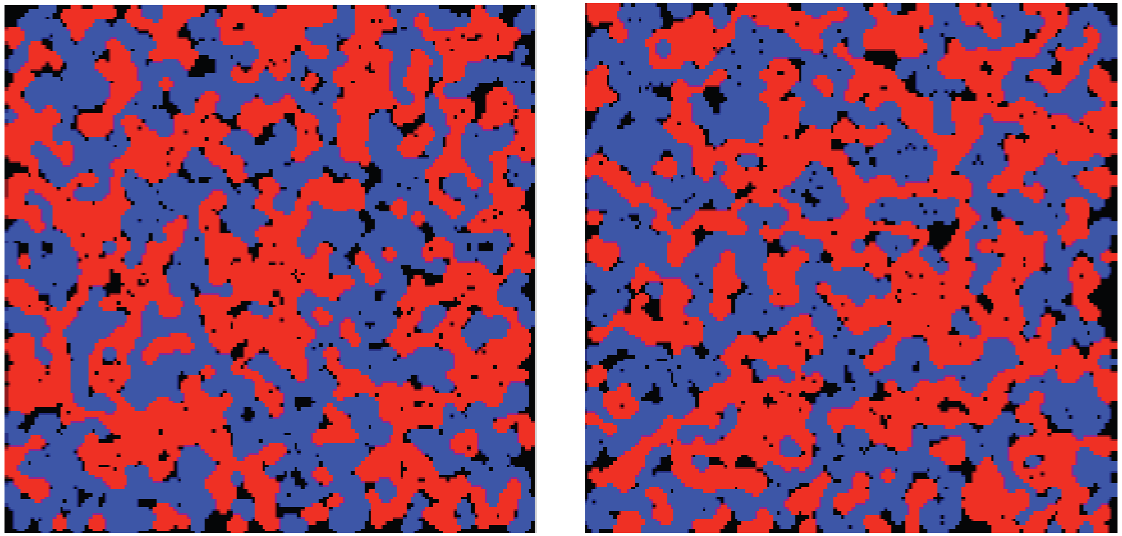
A Spatial Model of Segregation: The Schelling Model

Agents move in rounds: consider unsatisfied agents at some order
How to move? (in a random order, downwards?) what if no empty position?

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*		

A Spatial Model of Segregation: The Schelling Model



(a) *A simulation with threshold 3.* (b) *Another simulation with threshold 3.*

Figure 4.17: Two runs of a simulation of the Schelling model with a threshold t of 3, on a 150-by-150 grid with 10,000 agents of each type. Each cell of the grid is colored red if it is occupied by an agent of the first type, blue if it is occupied by an agent of the second type, and black if it is empty (not occupied by any agent).

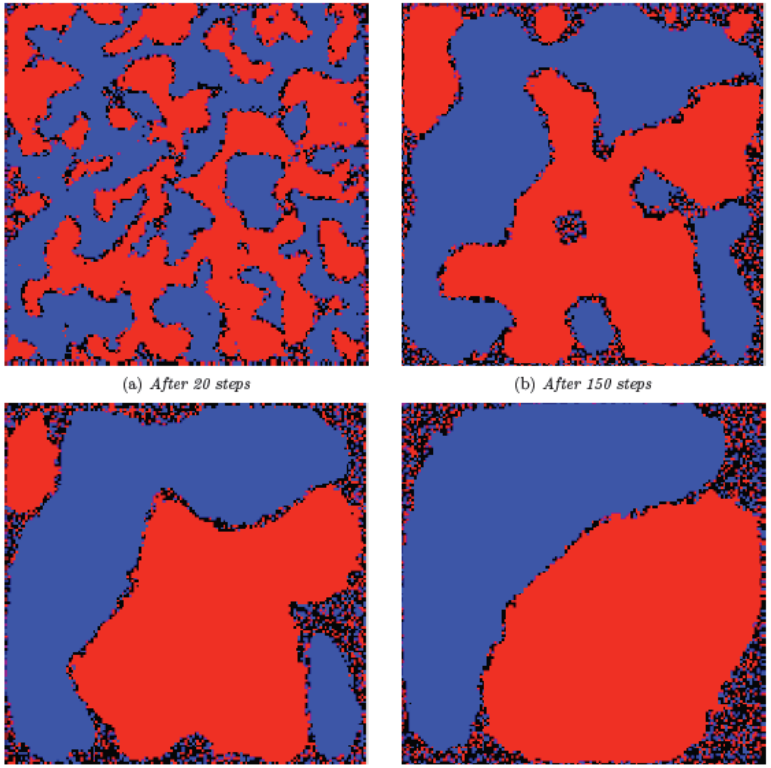
A Spatial Model of Segregation: The Schelling Model

Satisfied even in the minority among its neighbors
Requirements not globally incompatible

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

If we start from a random configuration, attach to clusters

A Spatial Model of Segregation: The Schelling Model



(a) *After 20 steps* (b) *After 150 steps*

(c) *After 350 steps* (d) *After 800 steps*

Figure 4.19: Four intermediate points in a simulation of the Schelling model with a threshold t of 4, on a 150-by-150 grid with 10,000 agents of each type. As the rounds of movement progress, large homogeneous regions on the grid grow at the expense of smaller, narrower regions.