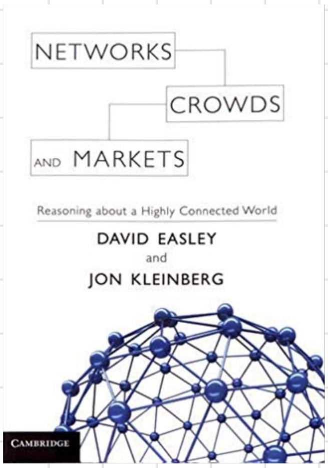
Lecture 7

Network Science

Networks in Their Surrounding Contexts

Today's topics

- Homophily
- Selection and Social Influence
- Affiliation
- Link Formation in Online Data
- A Spatial Model of Segregation



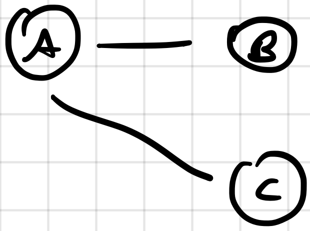
Chapter 4

"Networks in their Surrounding Context"

Surrounding Context

structural properties
link formation processes

personal characteristics
similarities between individuals



Surrounding context

factors that
exist outside the
nodes and edges of
a network
but that can have
an effect on the
evolution of the network

Homophily

the principle that we tend to be similar to our friends

your friends are not statistically significant as a random sample of the population

<u>similarities</u>

- age
- ethnic dimensions
- beliefs
- occupations
- interest
- opinions

immutable characteristics → (age, ethnic dimensions)

"birds of a feather flock together"

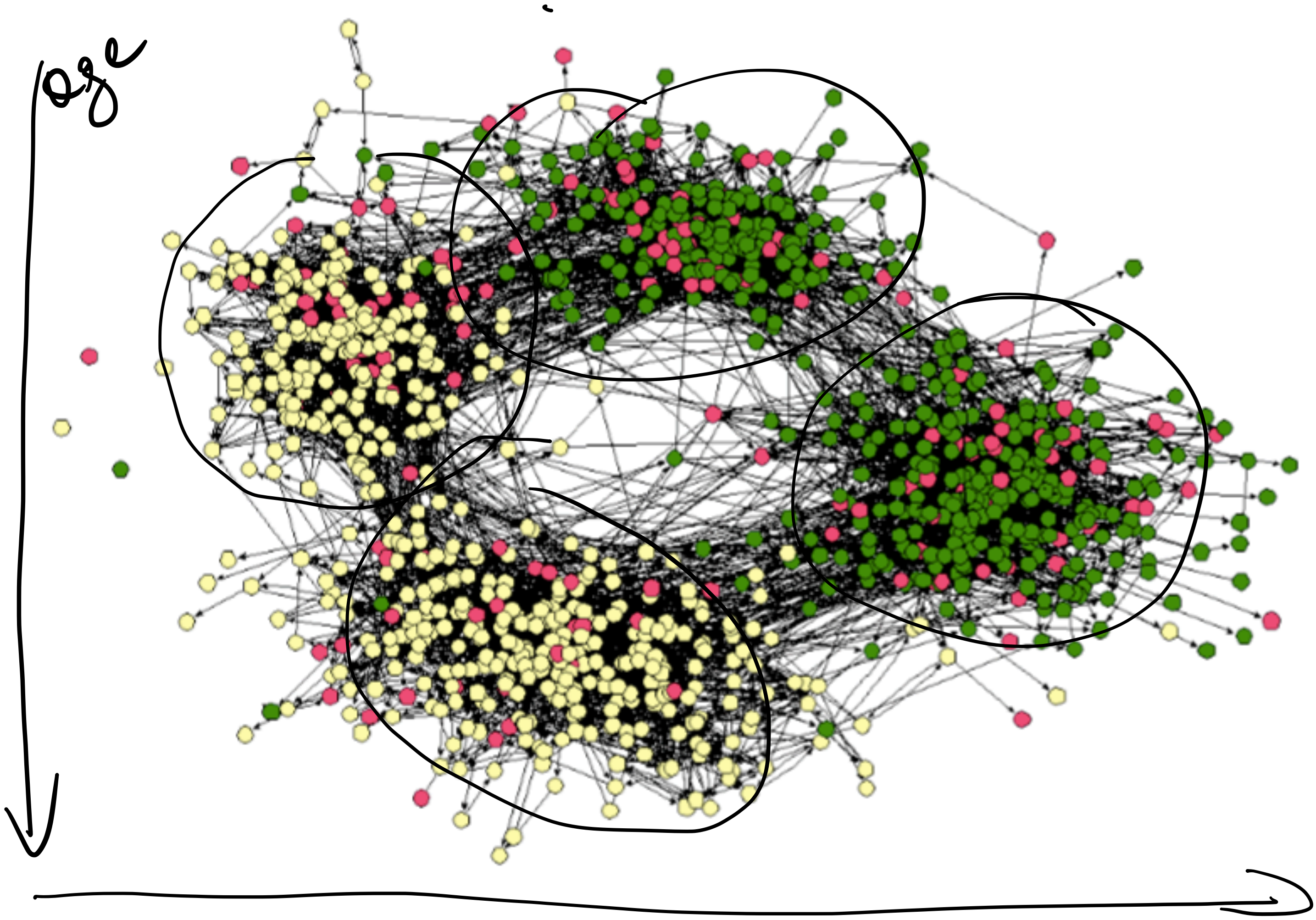


Figure 4.1: Homophily can produce a division of a social network into densely-connected, homogeneous parts that are weakly connected to each other. In this social network from a town's middle school and high school, two such divisions in the network are apparent: one based on race (with students of different races drawn as differently colored circles), and the other based on friendships in the middle and high schools respectively [304].

intrinsic triadic closures

and

contextual characteristics that influence similarities and that shape the network

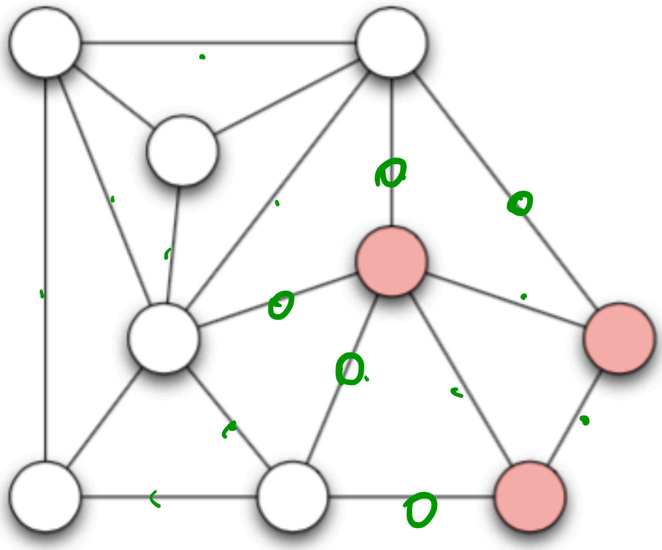
Measuring Homophily

simple test

$p = \frac{2}{3}$

$q = \frac{1}{3}$

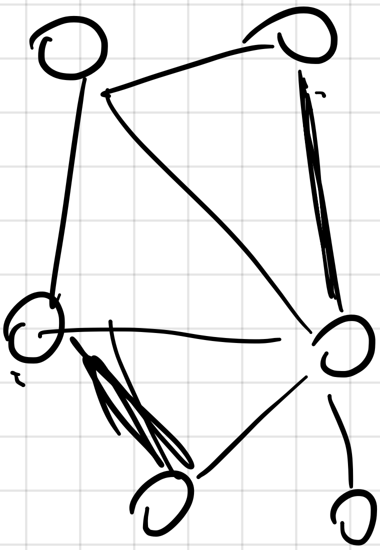
$2pq = \frac{4}{9}$



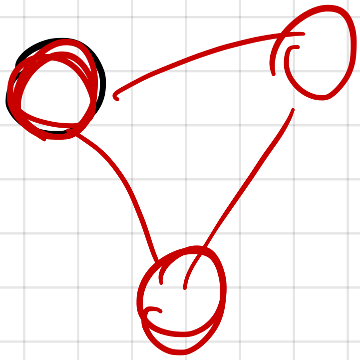
$\frac{5}{18}$

Figure 4.2: Using a numerical measure, one can determine whether small networks such as this one (with nodes divided into two types) exhibit homophily.

1. let's assign randomly a gender (color) to each node
2. count # of cross-gender edges
3. compare numbers with actual network



"perfect" homophily

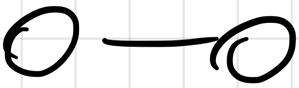


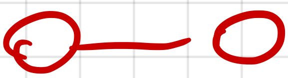
p : fraction of males

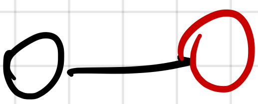
$\frac{6}{9} = \frac{2}{3}$

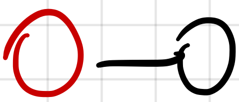
q : fraction of females

$\frac{1}{3}$

 : p^2

 : q^2

 : $p\,q$

 : $p\,q$

$\Big\}\ 2pq$

$$2pq = \frac{2 \cdot 4}{2 \cdot 9} > \frac{5}{18}$$

$2pq >$ # cross groups edges

we have homophily

Test: if the fraction of cross-types edges is "significantly less" than $2pq$ then there is a signal of homophily

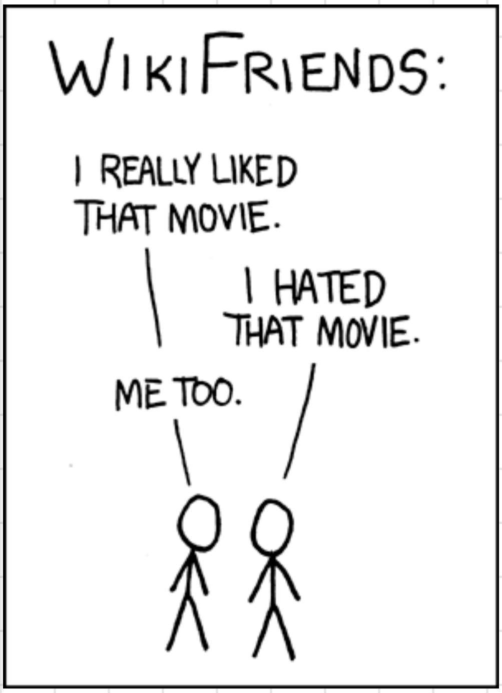
Selection and Social Influence

underlying mechanisms of homophily?

selection

tendency of individuals with similar characteristics to create a link between each other

individual char. ⇒ link



social influence

people who modify their behaviors to achieve a better alignment with their friends

link ⇒ indiv. characteristics change

The interplay of Selection and Social Influence

longitudinal methodology

- observe our networks for a long period of time
- observe both factors in action
- how do we quantify the impact?

ex.) drug dependency: social circle?

peer-pressure

Christakis and Fowler

dataset

12.000 people

obesity status

social network structure

obese vs non obese

there is a tendency toward clustering

homophily test : passed

Why

1. selection ?

2. homophily that correlates with smthg else ?

3. social influence ?

CONTAGION

Affiliation

Surrounding context by mean of network?

FOCI : Focal points

FOCUS

↳ activity

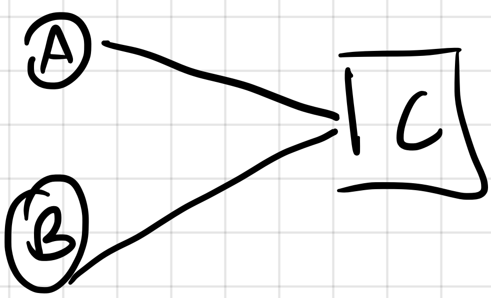
Bipartite Graphs

$$G = (U, V, E)$$

$$\forall (u,v) \in E : \; u \in U, \; v \in V$$

U : persons

V : Foci



Affiliation Networks

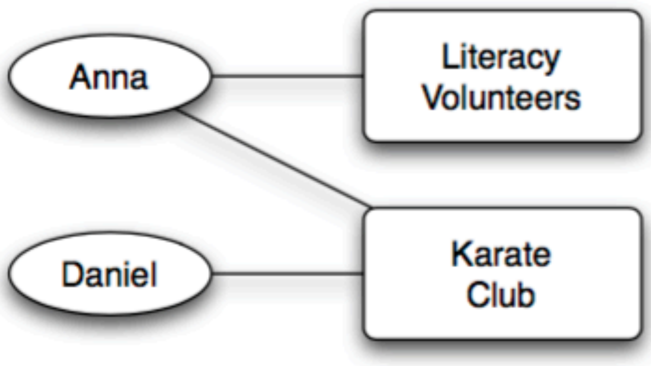


Figure 4.3: An affiliation network is a bipartite graph that shows which individuals are affiliated with which groups or activities. Here, Anna participates in both of the social foci on the right, while Daniel participates in only one.

can I complement
my affiliation netw. with
a social
netw ?

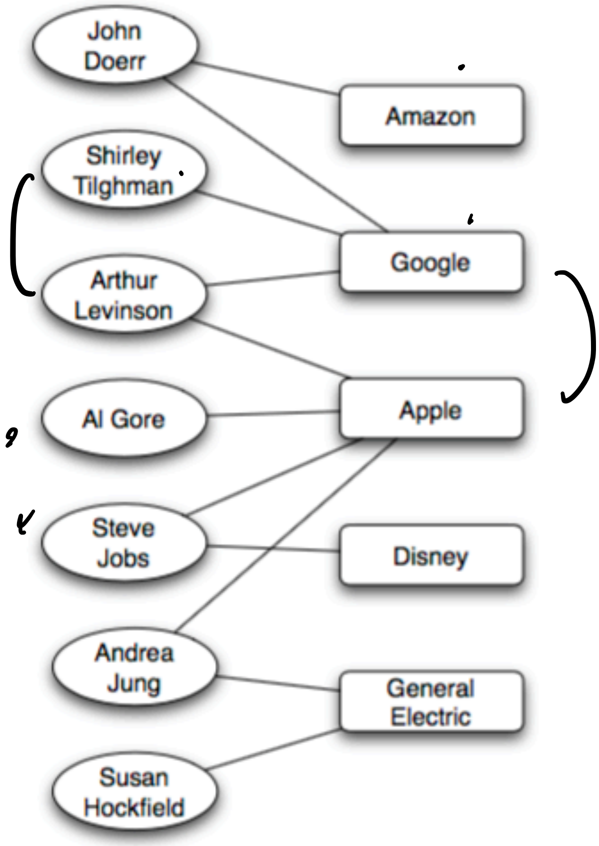


Figure 4.4: One type of affiliation network that has been widely studied is the memberships of people on corporate boards of directors [301]. A very small portion of this network (as of mid-2009) is shown here. The structural pattern of memberships can reveal subtleties in the interactions among both the board members and the companies.

projections

ex. Al Gore — Steve Jobs

ex. Google — Apple

Coevolution of Social and Affiliation Networks

change over time!

we want to merge the information from both networks

Can we predict new links formation?

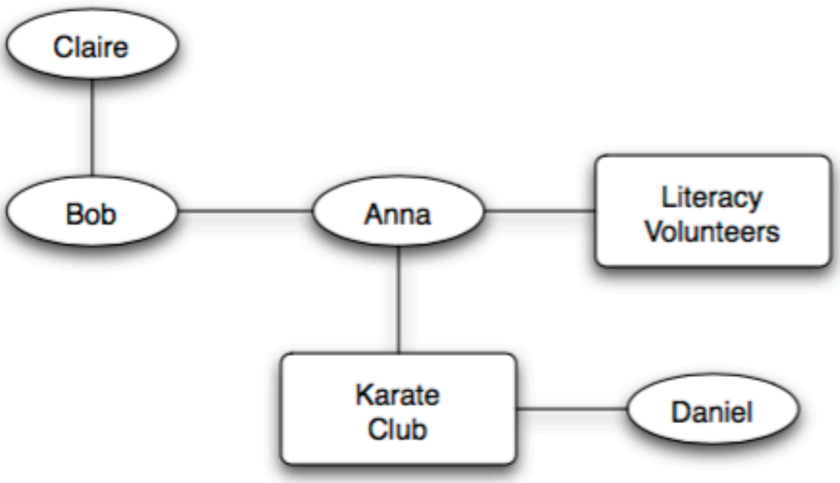
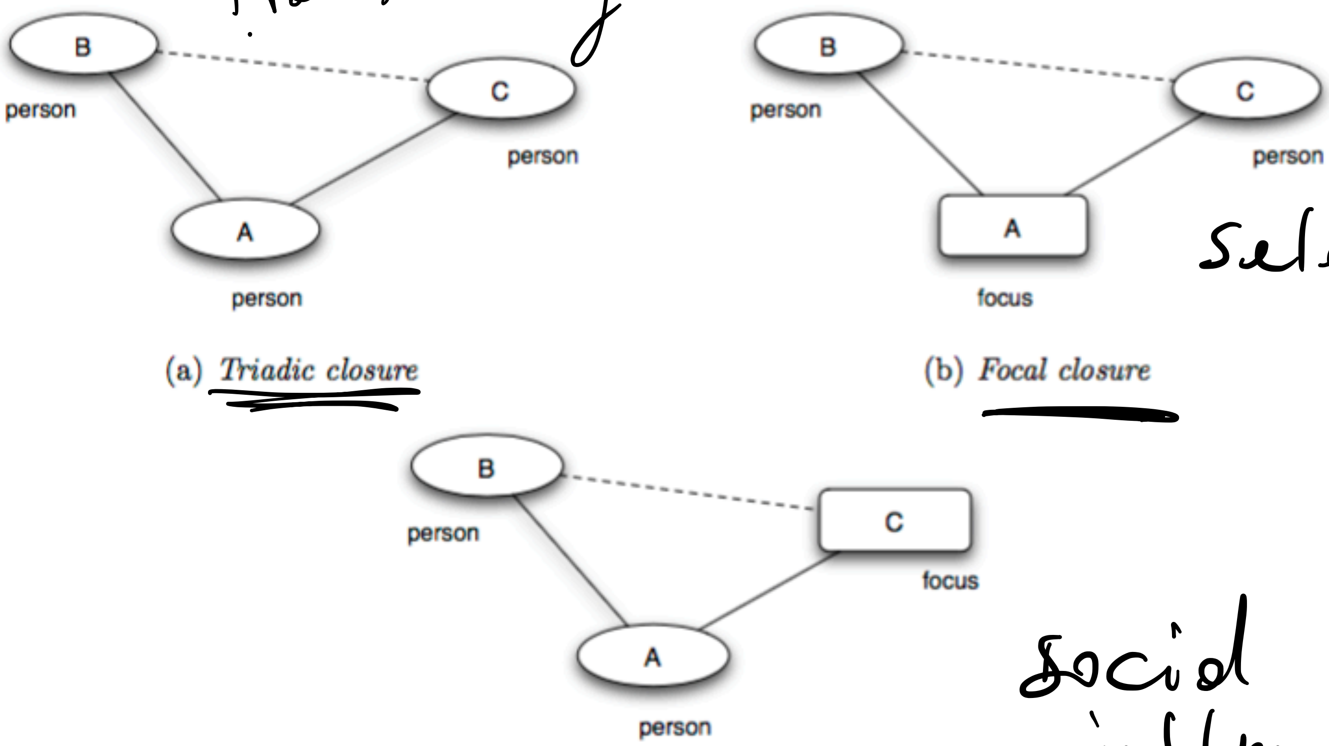


Figure 4.5: A social-affiliation network shows both the friendships between people and their affiliation with different social foci.

Closures

"transitivity"



(a) *Triadic closure*

(b) *Focal closure*

selection

(c) *Membership closure*

social influence

Figure 4.6: Each of triadic closure, focal closure, and membership closure corresponds to the closing of a triangle in a social-affiliation network.

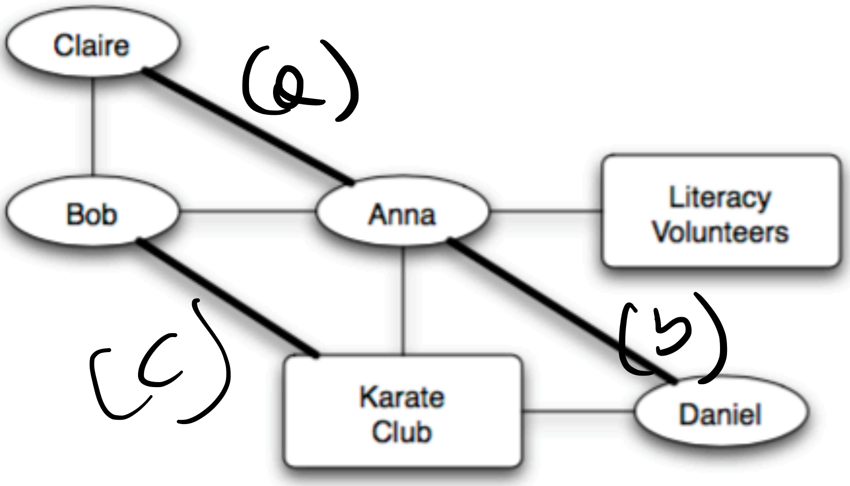


Figure 4.7: In a social-affiliation network containing both people and foci, edges can form under the effect of several different kinds of closure processes: two people with a friend in common, two people with a focus in common, or a person joining a focus that a friend is already involved in.

(a) : it is likely that friendship transitivity played a role

(b) : focal closure due to selection

(c) : membership closure due to social influence

Link Formation in Online Data

on-line platforms

Caveat: we can extrapolate from digital interactions inform. to other interactions that are not computer. mediated.

Triadic Closure

numerical questions:

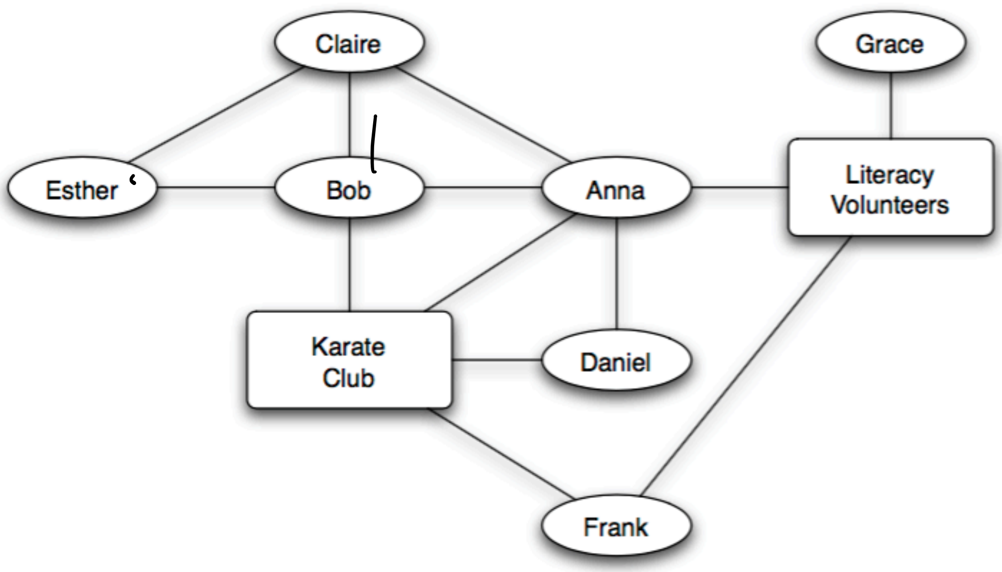


Figure 4.8: A larger network that contains the example from Figure 4.7. Pairs of people can have more than one friend (or more than one focus) in common; how does this increase the likelihood that an edge will form between them?

1. how much is more likely for a link to form if it has the effect of closing a triangle

2. how much more likely is a link to form if it closes MANY triangles?

$T(k)$ = # triangles have been closed in t'' w.r.t. $N(t')$

t', t''

$N(t')$: pair of nodes have friends in common at t'

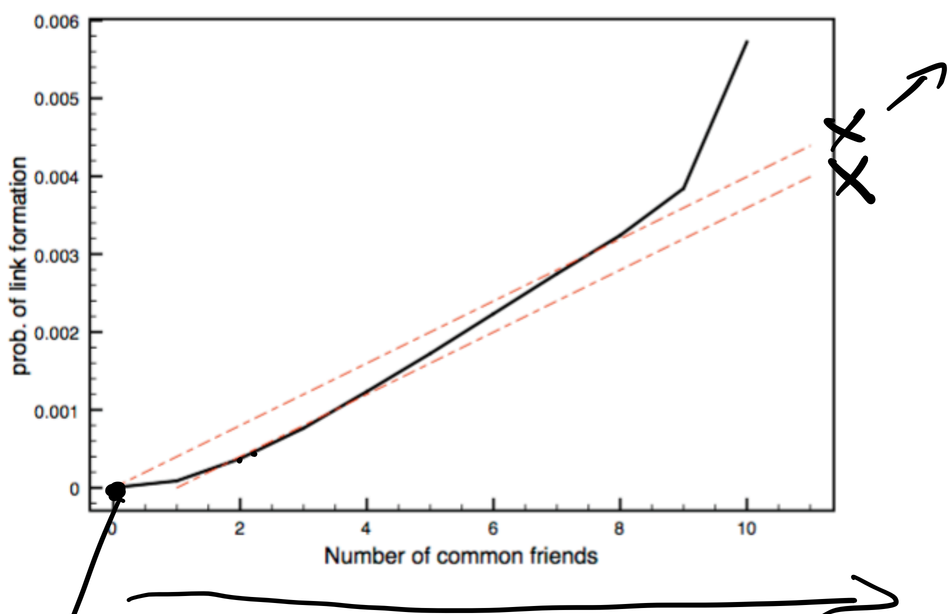


Figure 4.9: Quantifying the effects of triadic closure in an e-mail dataset [259]. The curve determined from the data is shown in the solid black line; the dotted curves show a comparison to probabilities computed according to two simple baseline models in which common friends provide independent probabilities of link formation.

$T(0)$

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

$$1 - (1-p)^{k-1}$$

Common friends

assumption

⇓

too simple

there is "something"

more than this

Focal Closure

Kossinets and Watts

email dataset
+
information of class schedules for each student

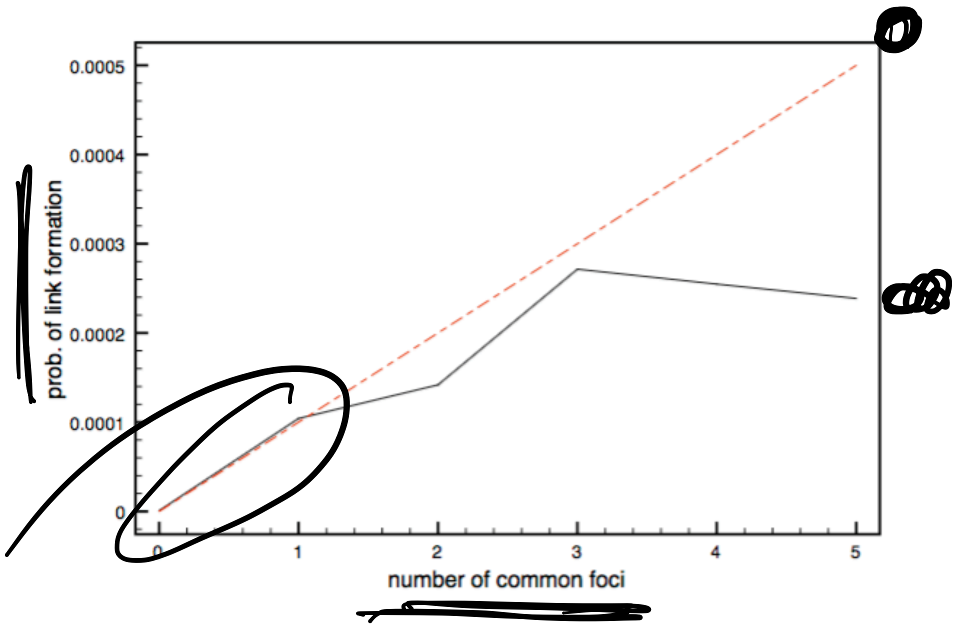


Figure 4.10: Quantifying the effects of focal closure in an e-mail dataset [259]. Again, the curve determined from the data is shown in the solid black line, while the dotted curve provides a comparison to a simple baseline.

T(0) → T(1)

empirical evidence

membership Closure

wikipedia

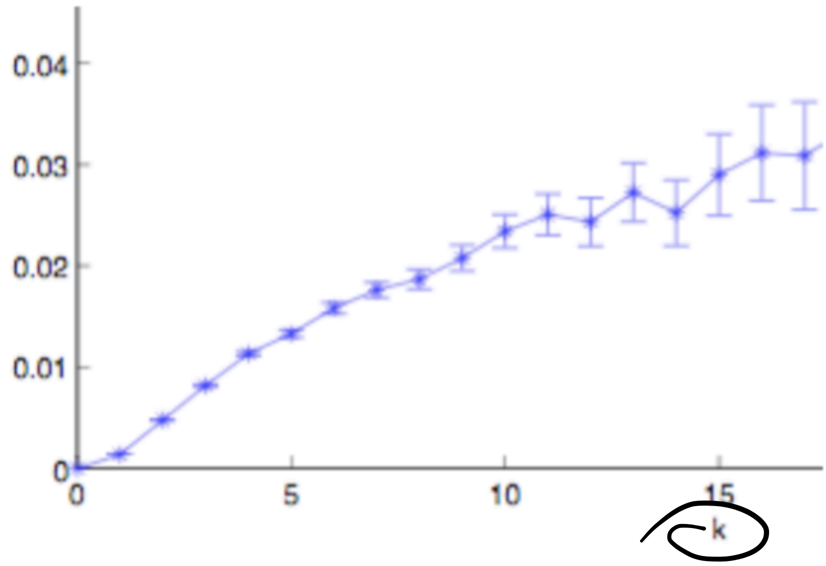


Figure 4.12: Quantifying the effects of membership closure in a large online dataset: The plot shows the probability of editing a Wikipedia articles as a function of the number of friends who have already done so [122].

Discussion

diminishing effect over k

multiple effects that operate simultaneously on the formation of a link

Homophily suggests that friends tend to have similar characteristics

- triadic closure
- focal closure
- membership closure

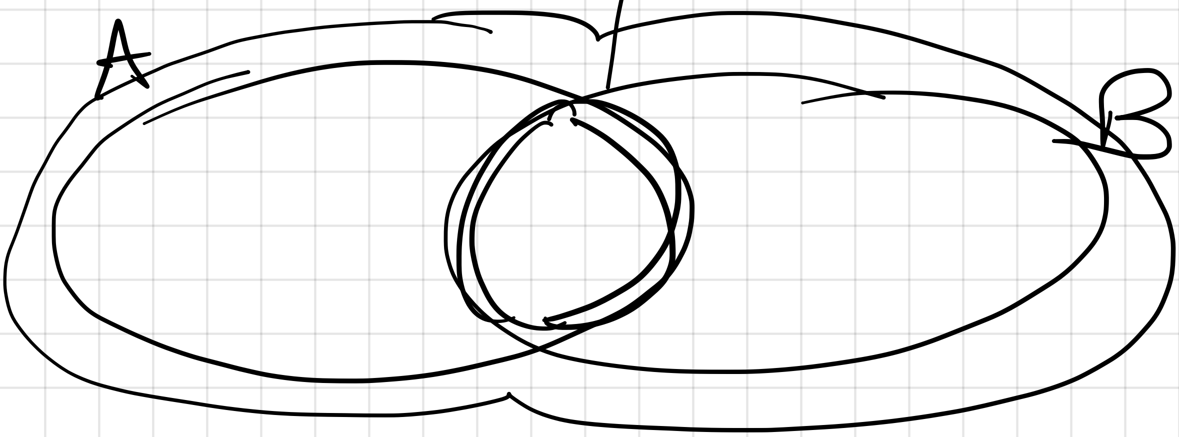
→ Correlation is not causation

→ forget linear explanations of cause-effects

Quantify The interplay between Selection and Social Influence

A, B are editors

$$sim(A, B) = \frac{\text{\# of articles edited by both}}{\text{\# of articles edited by A or B}}$$



wikipedia editors

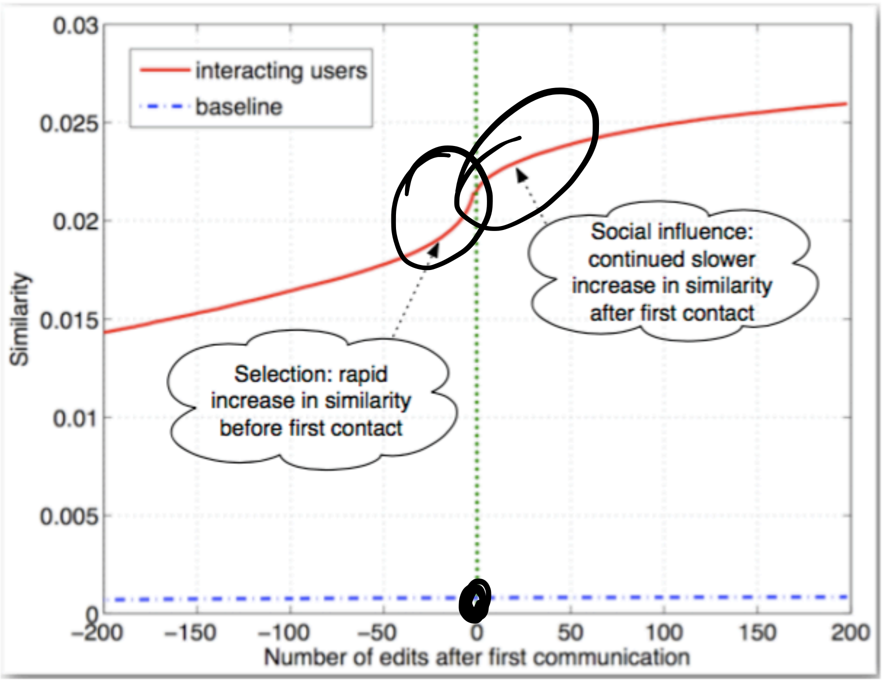


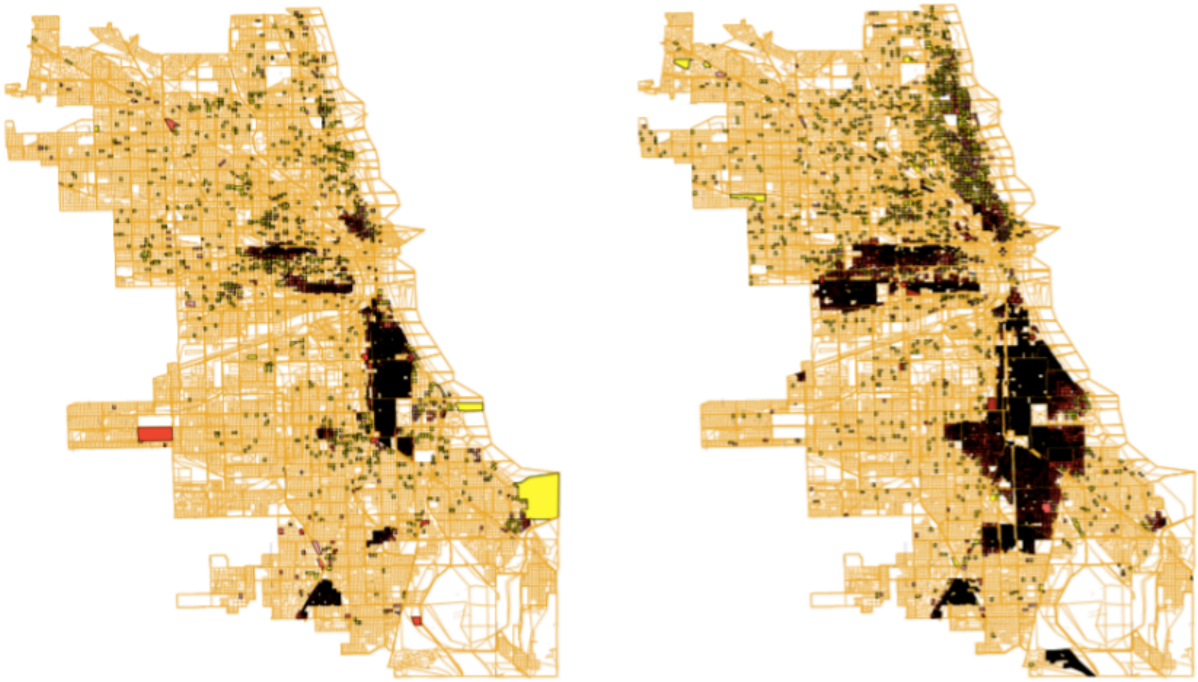
Figure 4.13: The average similarity of two editors on Wikipedia, relative to the time (0) at which they first communicated [122]. Time, on the x-axis, is measured in discrete units, where each unit corresponds to a single Wikipedia action taken by either of the two editors. The curve increases both before and after the first contact at time 0, indicating that both selection and social influence play a role; the increase in similarity is steepest just before time 0.

both effects are at work!

A Spatial Model of Segregation

Formation of homogeneous (according to some "type" or "class") homogeneous neighbors in cities

A natural spatial "signature"



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

Figure 4.14: The tendency of people to live in racially homogeneous neighborhoods produces spatial patterns of segregation that are apparent both in everyday life and when superimposed on a map — as here, in these maps of Chicago from 1940 and 1960 [302]. 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 Shelling Model

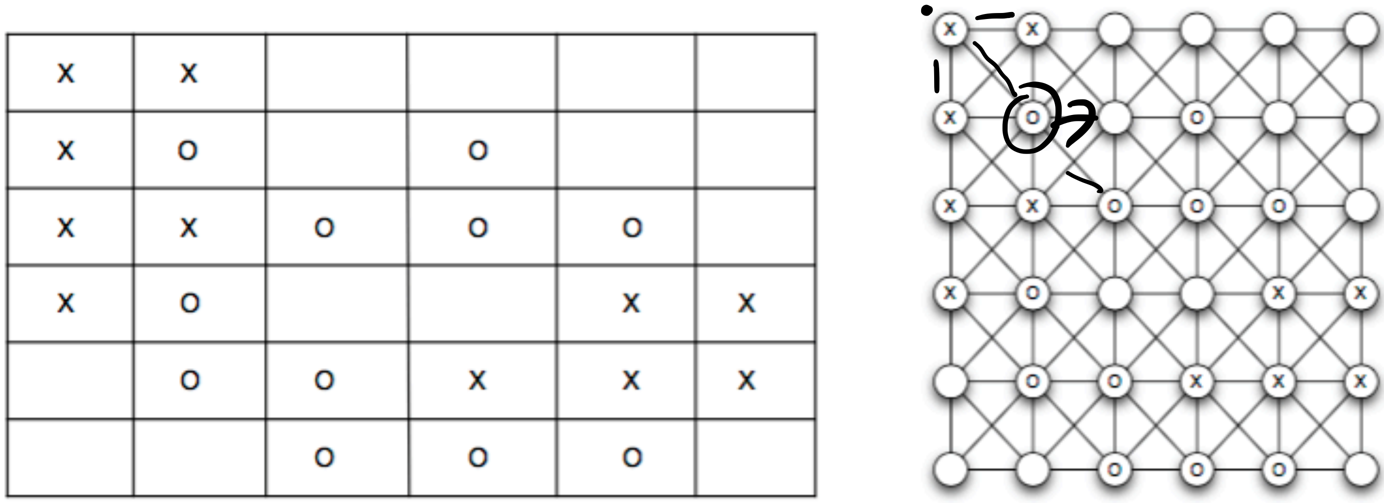
can spatial segregation arise from the effect of homophily operating at a local level?

Assumption: no individuals want segregation explicitly

Agents: X O
immutable characteristics

Agents reside in a cell of a grid
some cells contain agents
some other cells are unpopulated

Neighbors: 8 other cells "touching" an agent



(a) *Agents occupying cells on a grid.* (b) *Neighbor relations as a graph.*

Figure 4.15: In Schelling's segregation model, agents of two different types (X and O) occupy cells on a grid. The neighbor relationships among the cells can be represented very simply as a graph. Agents care about whether they have at least some neighbors of the same type.

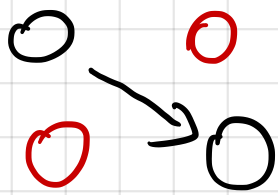
Each agent want to have at least t % neighbors of its own type

If an agent find < t neighbors are of the same type, then they are unsatisfied.

If unsatisfied, they want to move

The Dynamics of Movement

Agents move in a sequence of **rounds**



what if there is no empty cells around an unsatisfied agent:

- move the agent randomly
- just leave the agent there

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.*

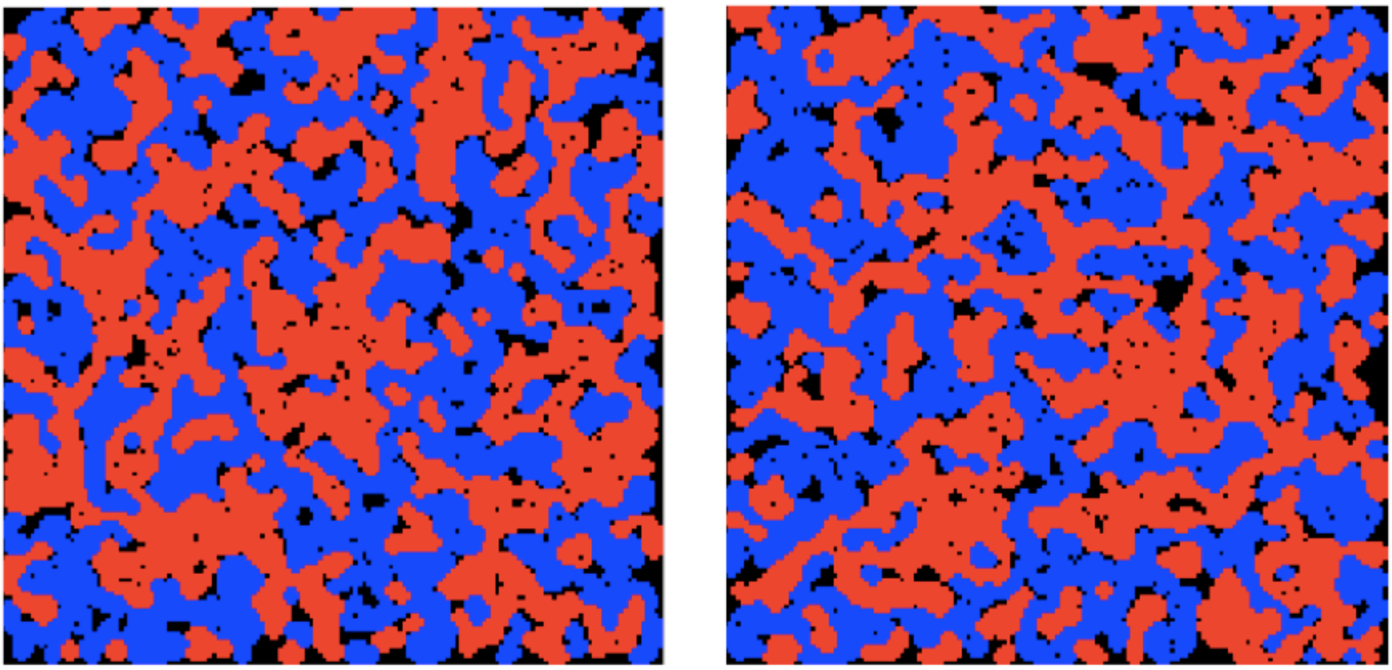
Figure 4.16: After arranging agents in cells of the grid, we first determine which agents are *unsatisfied*, with fewer than t other agents of the same type as neighbors. In one round, each of these agents moves to a cell where they will be satisfied; this may cause other agents to become unsatisfied, in which case a new round of movement begins.

Larger Examples

Computer simulations To look for patterns at larger scale

We want to run different simulations and make some comparisons

⇒ integrated pattern?



(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).

Interpretations of the Model

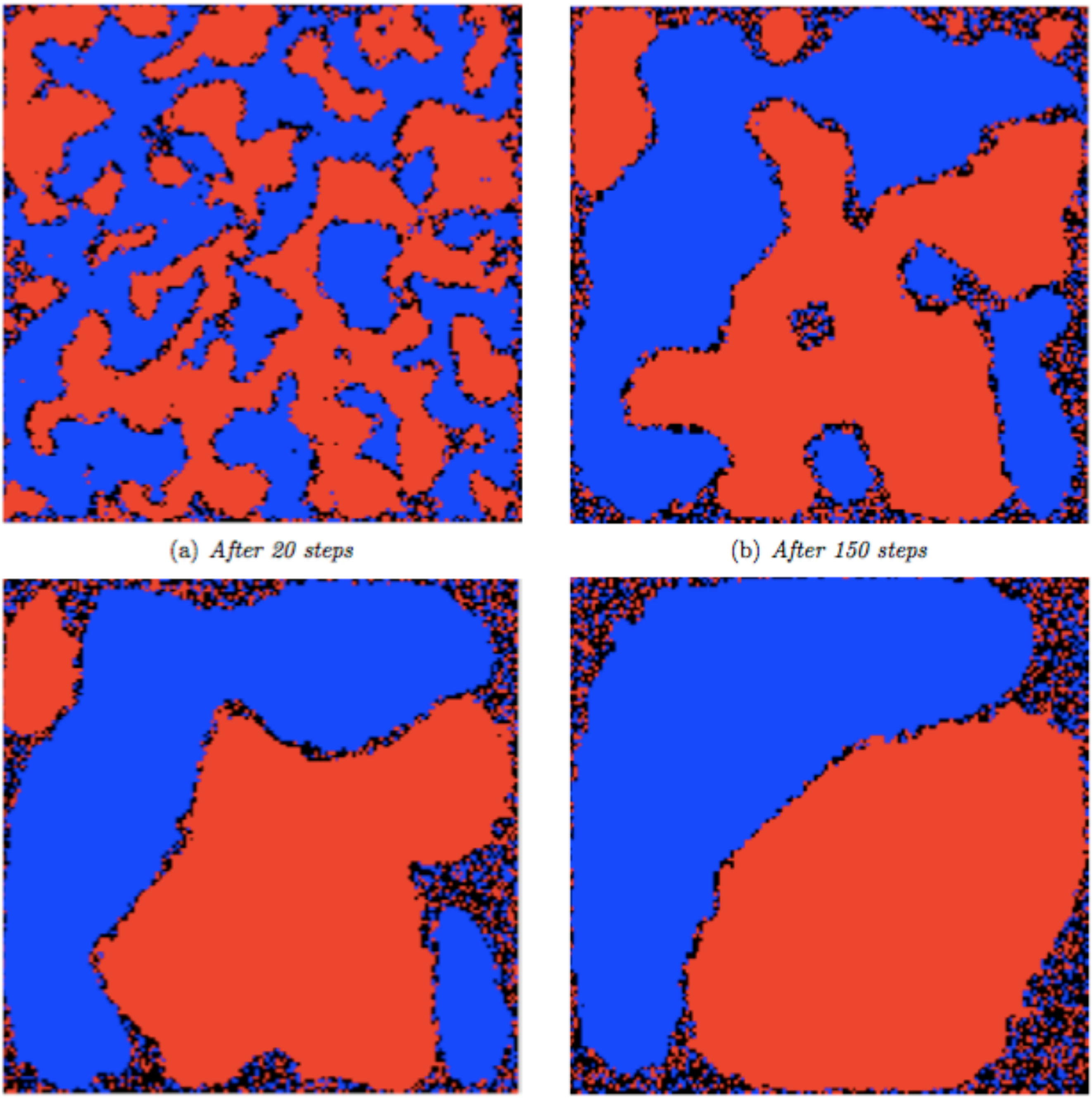
An agent would be perfectly happy to be in minority with $t=3$

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

Figure 4.18: With a threshold of 3, it is possible to arrange agents in an integrated pattern: all agents are satisfied, and everyone who is not on the boundary on the grid has an equal number of neighbors of each type.

↗ possible, but segregated realization are more likely the basic (random) reasons behind segregation are at work even in an idealized setting where agents are perfectly happy of being a minority

Of course if t is larger segregation is amplified



(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.

Take Home Message

homophily

selection and
social influence

not enough

surrounding context
affiliation and
social networks

not enough

contagion phenomena
are complex
and every factors
are in interplay

also:

segregation emerges
even when weak
constraints are
imposed.