CSC200: Lecture 7

- Today
  - Wrap up: studies of various types of closure (4.4)
  - Schelling model of spatial segregation
  - Netlogo simulations of Schelling model
- Wednesday and next few lectures
  - Game theory: Ch.6., Ch.8, Ch.9
- Announcements
  - I will discuss Quiz 1 and how it will be graded
  - Assignment 1 is will available and due October 16
  - I am not available this Tuesday, 2-3 for office hours but will be available Friday 3-4 PM
  - PROF IN A PROP : Donation to UNICEF

Quiz 1 and going forward

- I can see to some extent why the performance in quiz 1 was not what I expected or hoped for.
- In hindsight, I see that we never gave any calculated examples of embeddedness and neighbourhood overlap.
- It also seems that many students misinterpreted the second question.
- I am effectively making the second question a bonus question by counting the quiz out of 20 instead of 30.
- We hope to have the quiz graded by this weeks tutorial. A sample solution has been posted.
- Reminder about 20% rule and need to state “I do not know how to answer this question”.

Quiz 1 and going forward continued

- The quiz is only 2% of final grade so plenty of opportunity to do well in this course.
- I may give the option of counting the best say 6 of 8 quizes. The quizes are incentive to keep up in course.

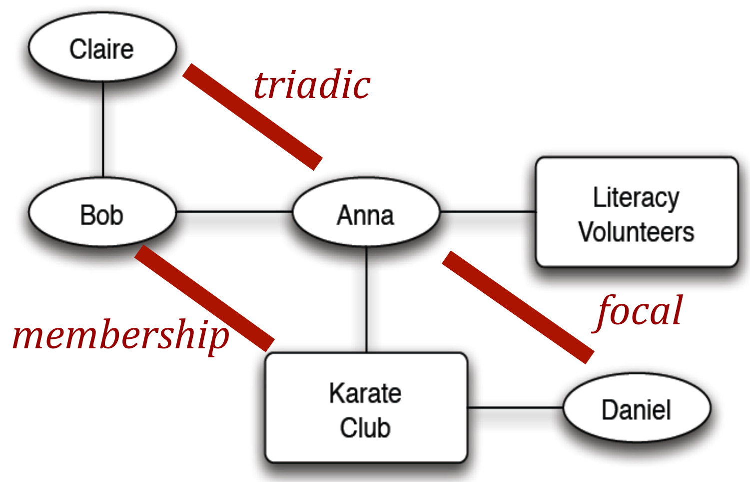
  Quizes at end of tutorial?
- The TAs tell me that you want more examples. I do tend to use the lectures for concepts and the "big picture" but welcome questions. If you have a question, good chance many others have the same or a related question.
- Having questions is a good thing. It's what we do!
- If you prefer to ask questions in tutorial that is fine and that is a good place to ask for more examples. But still you eventually have to work out examples for yourself .

We pause briefly for a "public service announcement"

- I have volunteered to participate in the Unicef "Prof in a Prop " donation plea. See next slide.
- If successful in obtaining average $1/enrolled student on Wednesday October 7, then next Wednesday October 14, I will lecture in costume.
- We now return to our regularly scheduled course material.

Recap of Last Time

- Affiliation networks
  - encode factors that might explain homophily (both social influence and selection)
- Three forms of closure
  - triadic closure
  - focal closure
  - membership closure
- Studies (e.g. Kossinetts and Watts) exploring formation of closing links
  - let's continue with this



[E&K, Ch.4, Fig. 4.5]

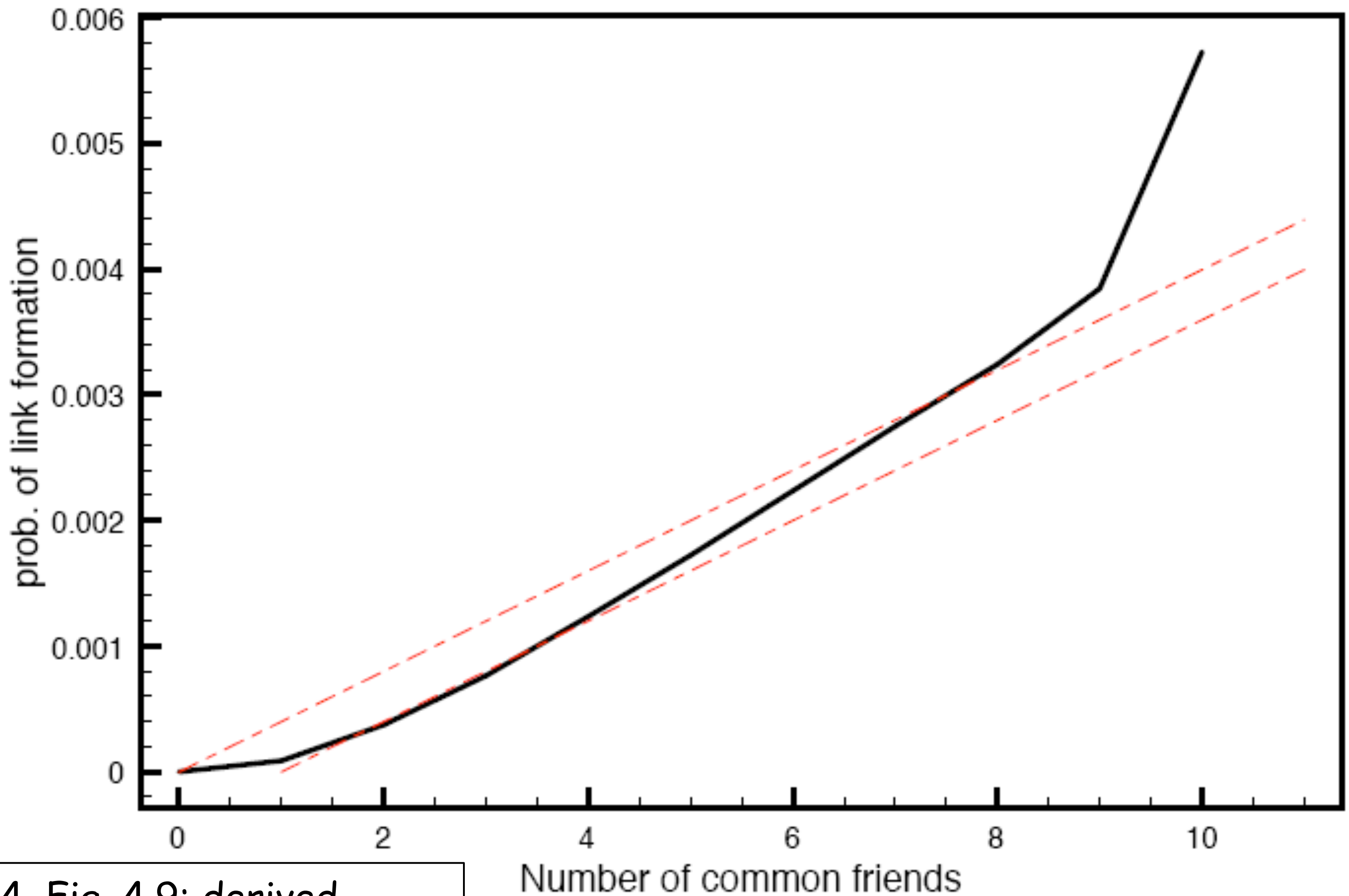
How do we measure these processes?
(review from last class)

- Closure is inherently *dynamic*
  - So we should take ***snapshots*** of the network at different times to see how the relationships evolve and to what extent each form of closure occurs
  - If *common friends* or *common interests* are causing new links (i.e., closures) then the more friends or interests in common, the more we should see this effect.
- We'll look at a couple of studies (briefly) afforded by online interactions, but realize the limitations
  - usual modeling limitations: may be many missing factors
  - when one takes snapshots may influence results
  - these studies look at link formation, but not link dissolution
    - what would the world look like if links formed but never dissolved?

Triadic closure: dependence on # of mutual friends
(review from last class)

- Student email use, large US university **[Kossinets, Watts 2006]**
- "Friends" defined as two-way email communication in window (previous 60 days) of time.
- Goal: measure probability *T(k)* of a *new friendship* emerging between a pair of students as a function of the number *k* of mutual friends
  - probability of it happening in any given day (averaging over all such pairs and days)
- Compare data (black) with baseline theoretical model (red)
  - *baseline:* assume any single mutual friend will generate a new friendship with probability *p* and that this will happen independently for each common friend: $T(k) = 1 – (1-p)^k$
    - *think of each common friend introducing a pair to each other and succeeding at forming the friendship with probability p (independently)*

Probability (per-day) of triadic closure as a function of the number of common friends



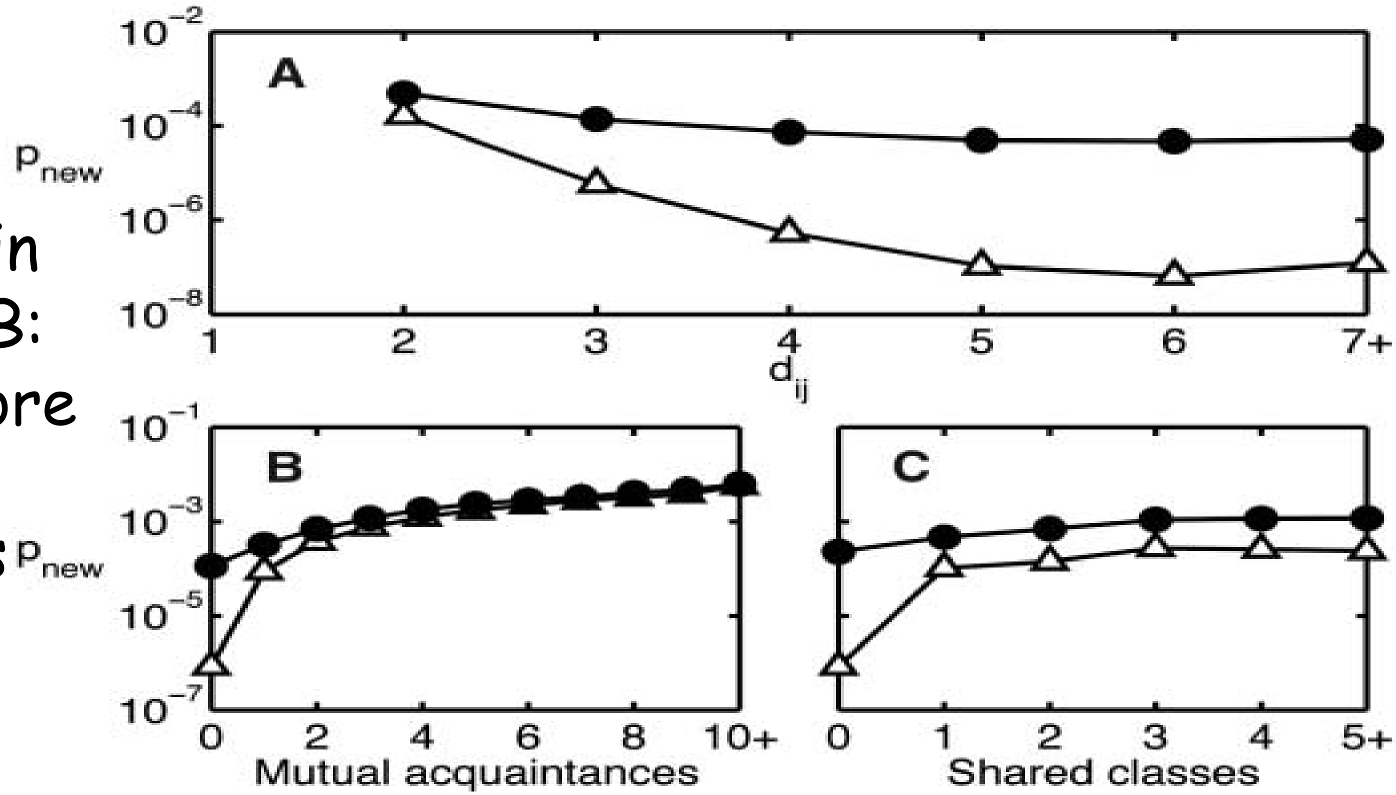
[E&K, Ch.4, Fig. 4.9; derived from Kossinets and Watts, 2006]

Observations

- Data does not show much more propensity for friendship when going from zero to one mutual friend.
  - The second dashed red line shifts the curve over by one friend so as to better compare the actual data and baseline model.
  - Why? (Consider “strength of evidence” argument?)
- Increasing from 1 to 9 friends shows “roughly linear” curve (greater slope than baseline)
  - *Baseline not really linear though close for small probability*
- A *sharp* difference going beyond 9 friends
  - theoretical baseline model (and its assumption of independence) no longer supported; NOTE: sparse data for such large k.
  - is there some threshold of mutual friends which *escalates* the pressure for triadic closure?
- ***Exercise: translate per-day probability into per-month/year***

Kossinetts and Watts Figure 1

Circle in A and B: 1 or more shared classes



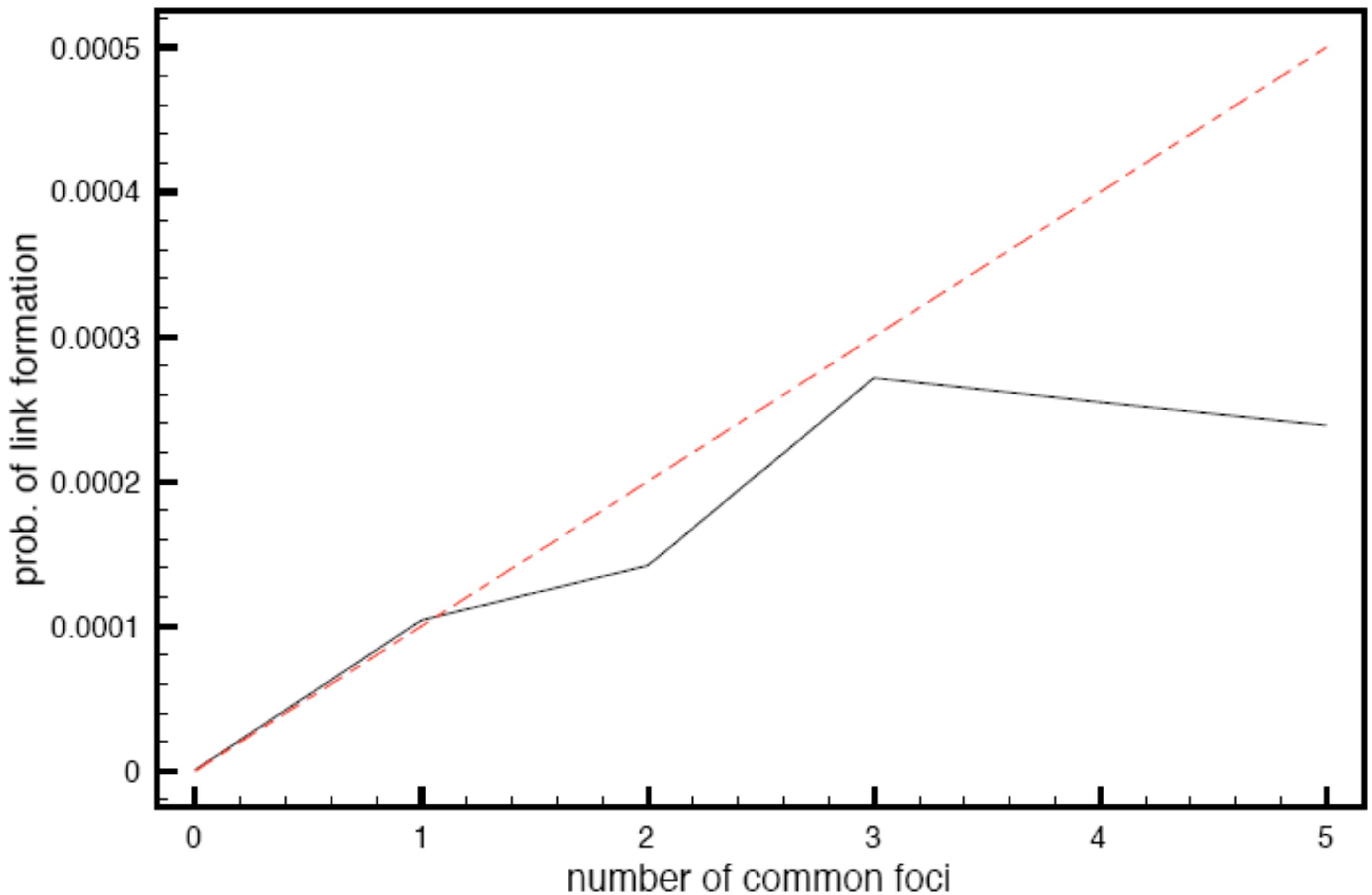
Circle in C: 1 or more mutual "friends"

Probability to form new link: A as function of min Distance; B as function of number of mutual friends; C as function of number of shared classes. Triangle indicates no shared classes (resp. mutual friends) in A,B (resp. C).

Focal closure: dependence on # of mutual foci

- Same study: measures focal closure probability as a function of number common classes **[Kossinets, Watts 2006]**
- “Friends” defined by email communication (as before)
- Again: measure probability *T(k)* of a *new friendship* emerging between a pair of students as a function of the number *k* of mutual classes
- Compare data (black) with baseline theoretical model (red)
  - baseline: as before, but now assume probability of friendship forming increases with k = number of classes, each class providing an independent attempt at forming the friendship: $T(k) = 1 - (1-p)^k$
- Observations somewhat similar, but important differences
  - Probabilities much lower (order of magnitude)
  - Significant tailing off from linear once 3 classes reached: why?

Probability (per-day) of focal closure as a function of the number of common classes



[E&K, Ch.4, Fig. 4.10;
from Kossinets and Watts, 2006]

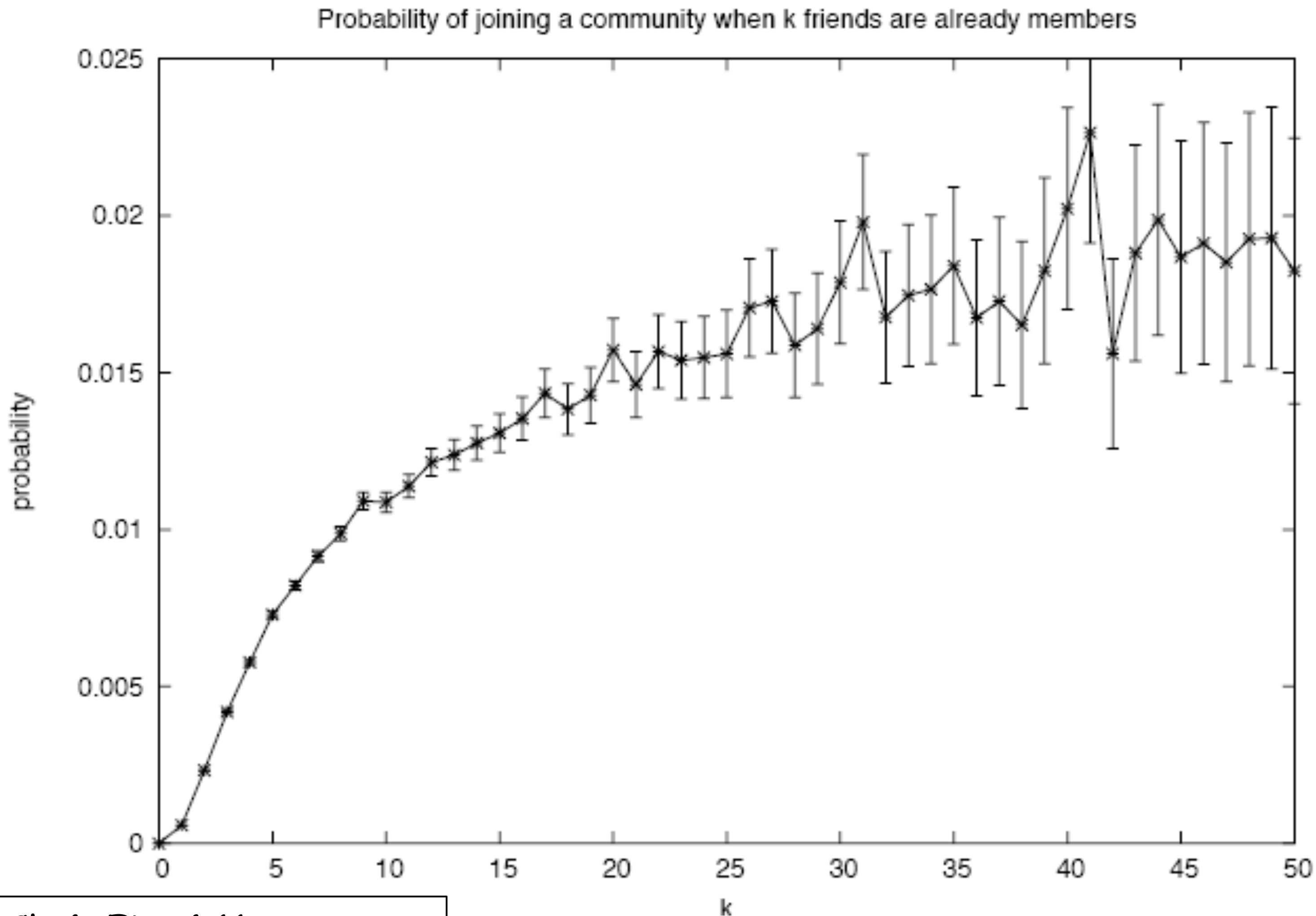
Other observations in Kossinets-Watts

- The effect of a single shared class is roughly equivalent with a single shared mutual ``friend''.
- But having additional friends has a greater effect than having additional foci (i.e. classes).
- Triadic closure (expectedly) increases with tie strength.
- Individual attributes such as status, gender, age and time at University appear to play a weaker role than expected. Similarity effect of attributes is not as influential in friendship formation as triadic closure and focal closure.
- Discussion of extent of dependence on time window.
- Question raised as to whether these findings are generic or specific to this network. But methodology is generic.

Membership Closure

- **Membership closure**
  - See the text's discussion of the LiveJournal and Wikipedia
  - We'll briefly look at the LiveJournal data **[Backstrom et al, 2006]**
    - Blogging site with social networking (friends) component
    - Communities are focal points
  - Trend: sharp increase in probability from 1-2 friends, growth tails off a bit after 5-6 friends, and significantly after 10 or so
    - Notice "error bars" grow with *k*: Why?

Probability of Membership Closure

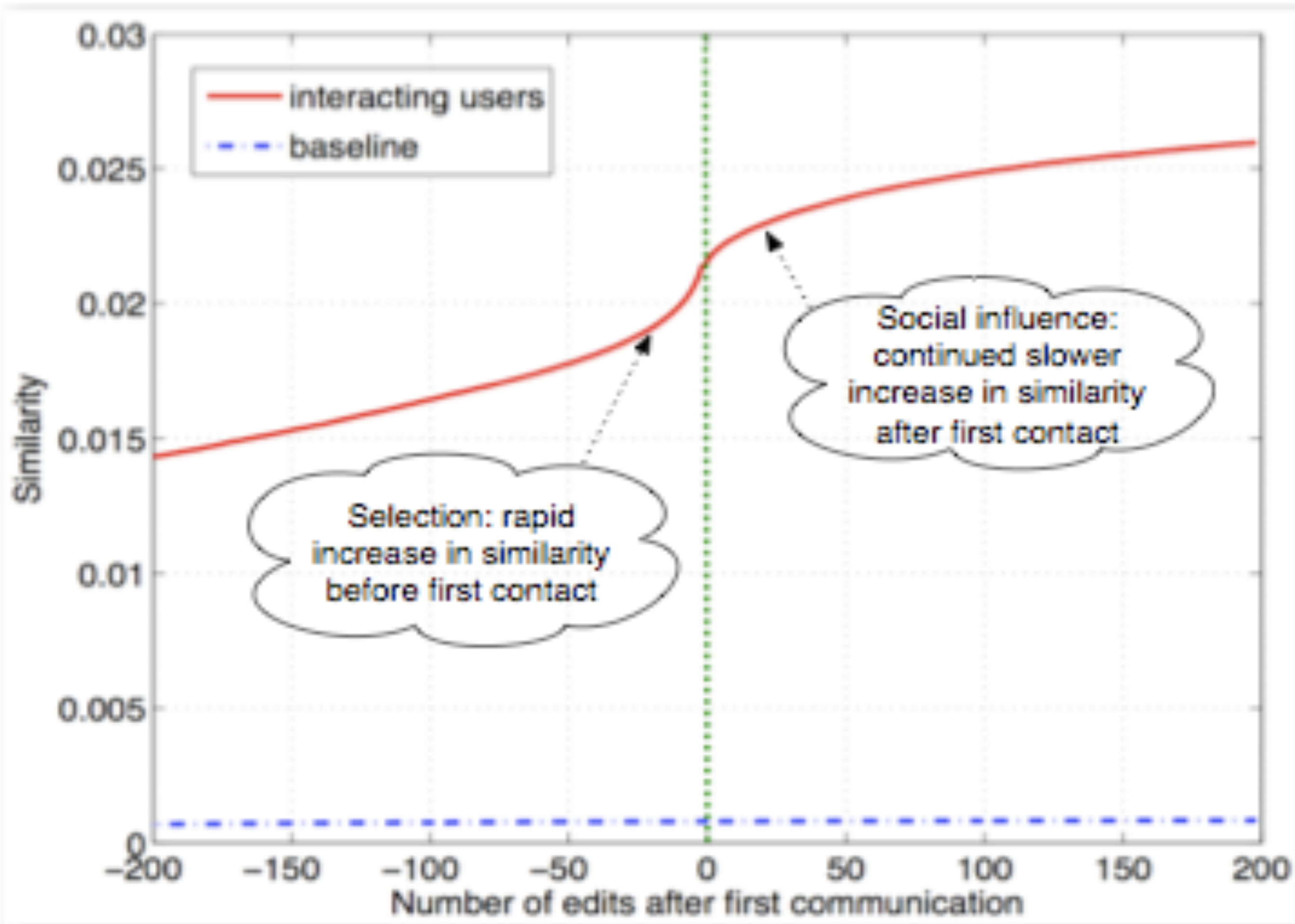


[E&K, Ch.4, Fig. 4.11;
from Backstrom et al., 2006]

Interplay of Selection and Social Influence

- We'll not discuss much further, but lets briefly consider the discussion of Wikipedia editor relationships in Ch.4.4
  - Studying impact of selection vs. social influence in Wikipedia social-affiliation network. Nodes are users with talk page; links when user communicates by writing on another talk page; affiliation in activity when someone edits article (the activity)
  - *similarity* of (A,B) measured by ratio :
    # articles edited by both A,B / #edited by at least A or B
  - *selection:* among editors who become friends at some point, initially similarity in interests (as measured by articles edited in common) seems to drive editors to "become friends" (as measured by direct communication/talk at "time 0").
  - *social influence:* once they become "friends" similarity continues to increase, albeit at a reduced rate

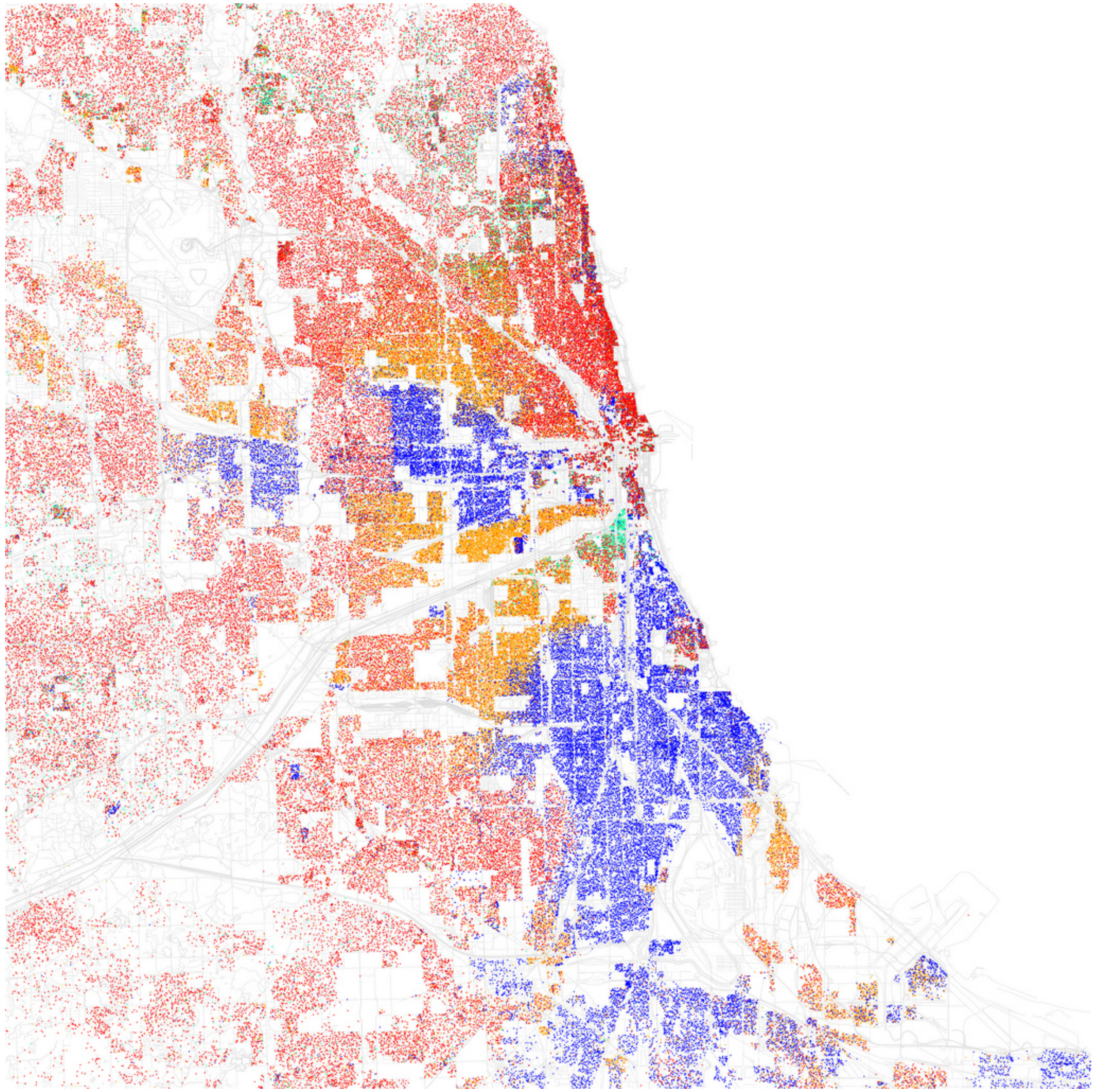
Selection and social choice



Spatial Model of Segregation

- Ethnic and racial segregation common in many cities
  - spatial segregation may be evidence of selection based homophily
  - explainable by *immutable* characteristics?
    - race, ethnicity, native country, first language?
  - explainable by (strongly correlated) *mutable* characteristics?
    - cultural affiliation, restaurants, business preferences, etc?
- Numerous studies show this (lots of press in US)
  - see discussion of Mobius, Rosenblatt 2001 study in text
  - recent visualizations by Eric Fischer using 2010 census
    - http://www.flickr.com/photos/walkingsf/sets/72157626354149574/detail/
    - (are trends in Toronto similar? different?)

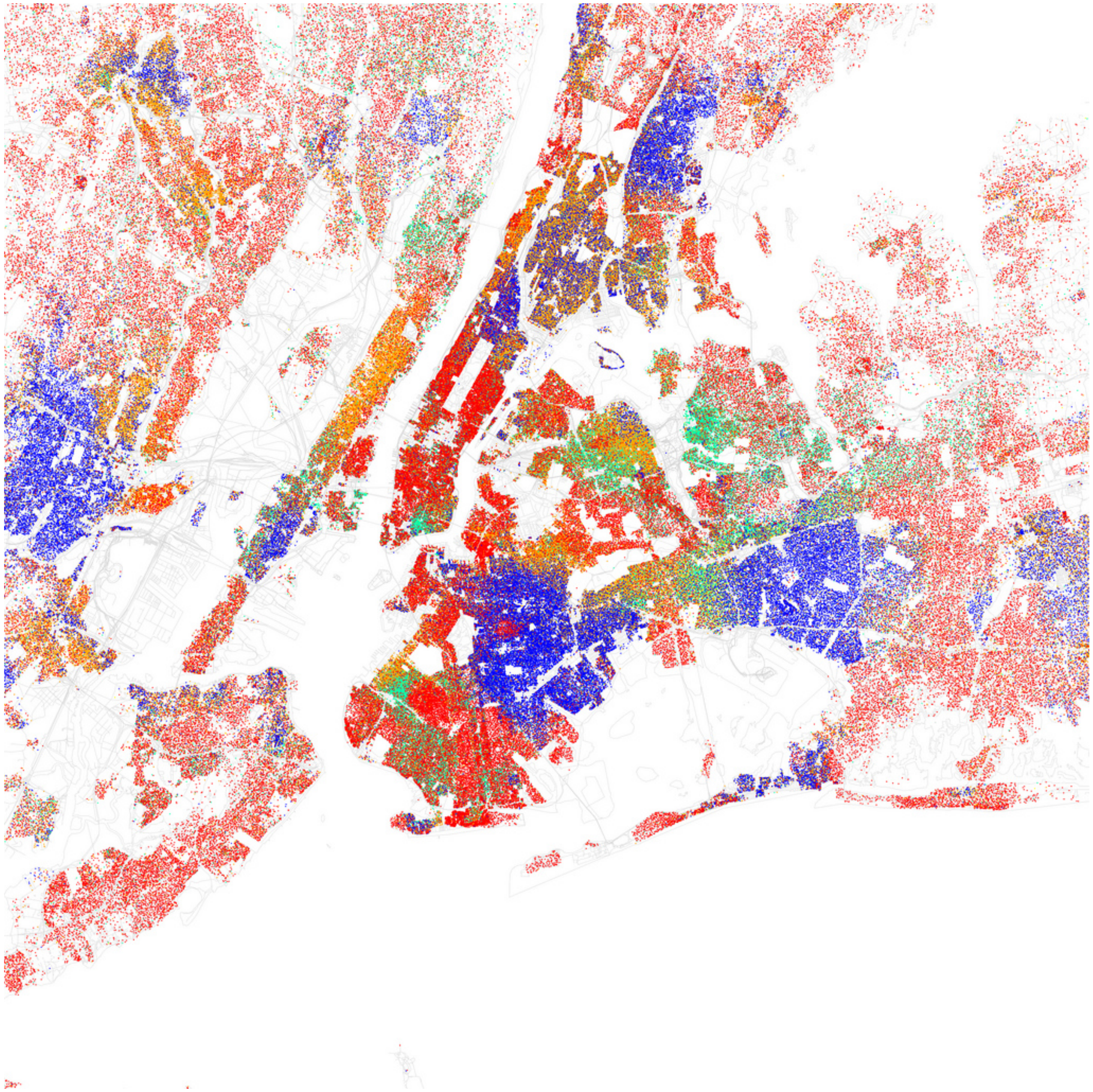
Chicago



Red: white
Blue: black
Green: Asian
Orange: Hispanic
Each dot=25 people

See: http://www.flickr.com/photos/walkingsf/sets/72157626354149574/detail/

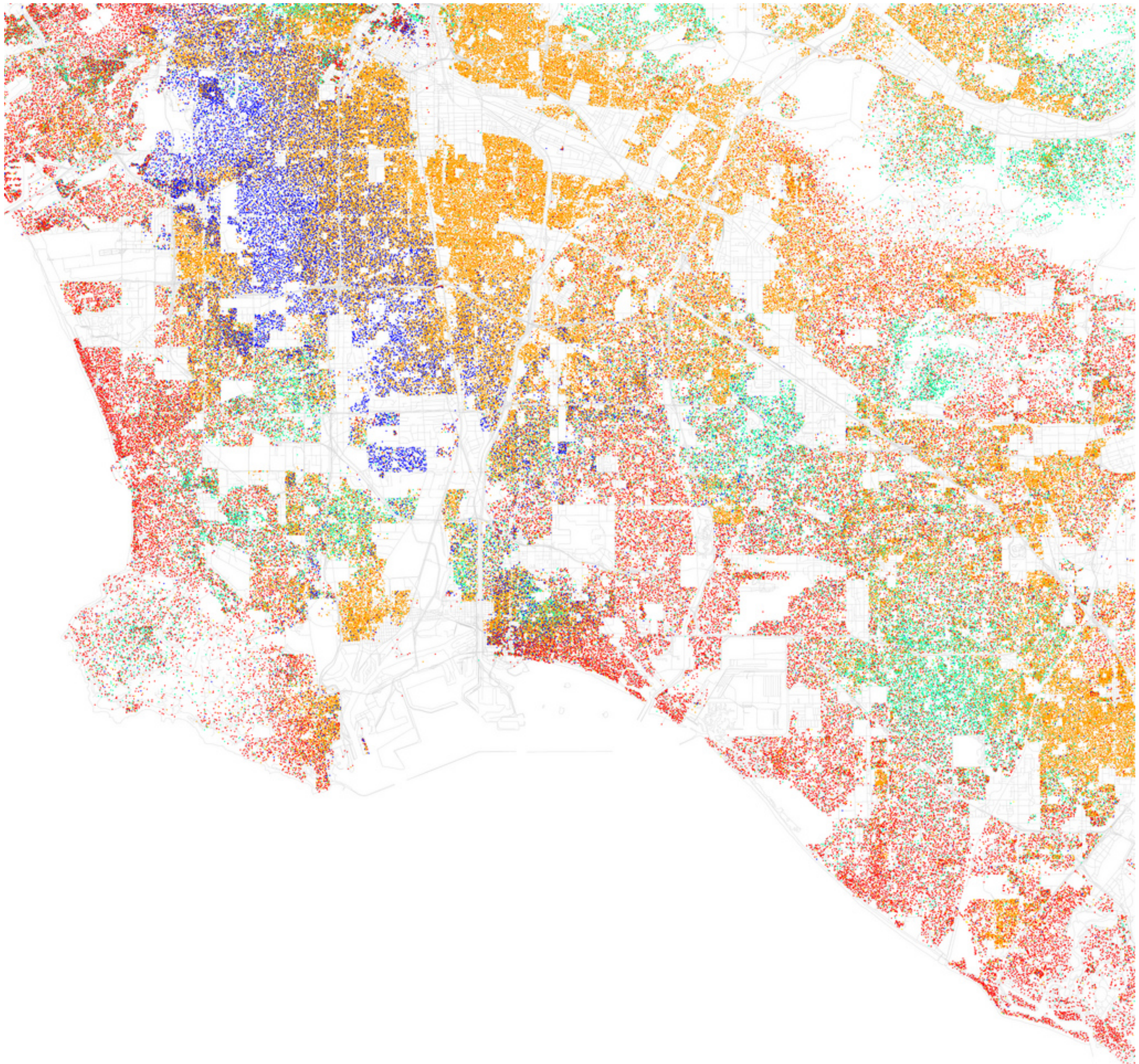
New York



Red: white
Blue: black
Green: Asian
Orange: Hispanic
Each dot=25 people

See: http://www.flickr.com/photos/walkingsf/sets/72157626354149574/detail/

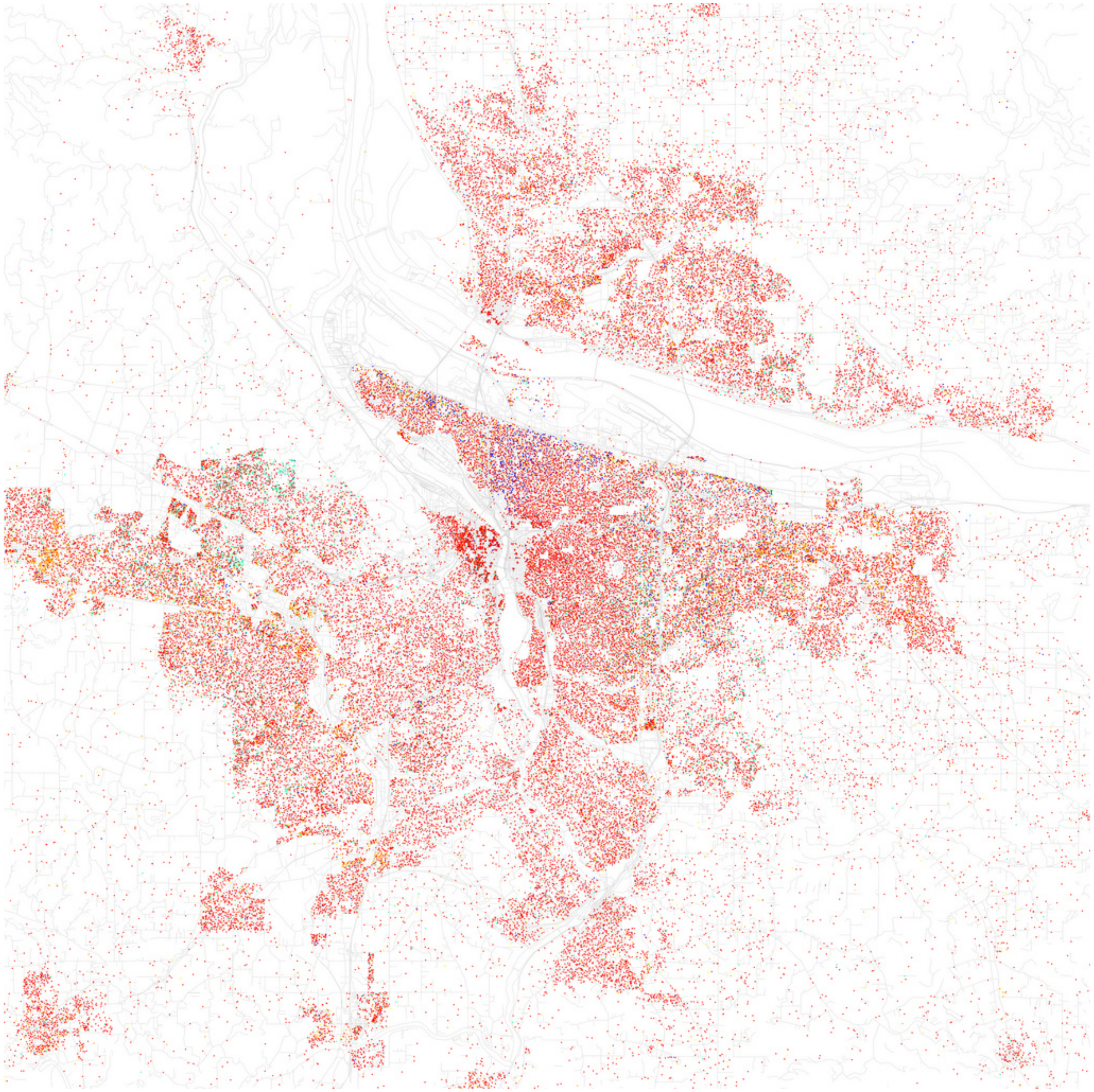
Long Beach



Red: white
Blue: black
Green: Asian
Orange: Hispanic
Each dot=25 people

See: http://www.flickr.com/photos/walkingsf/sets/72157626354149574/detail/

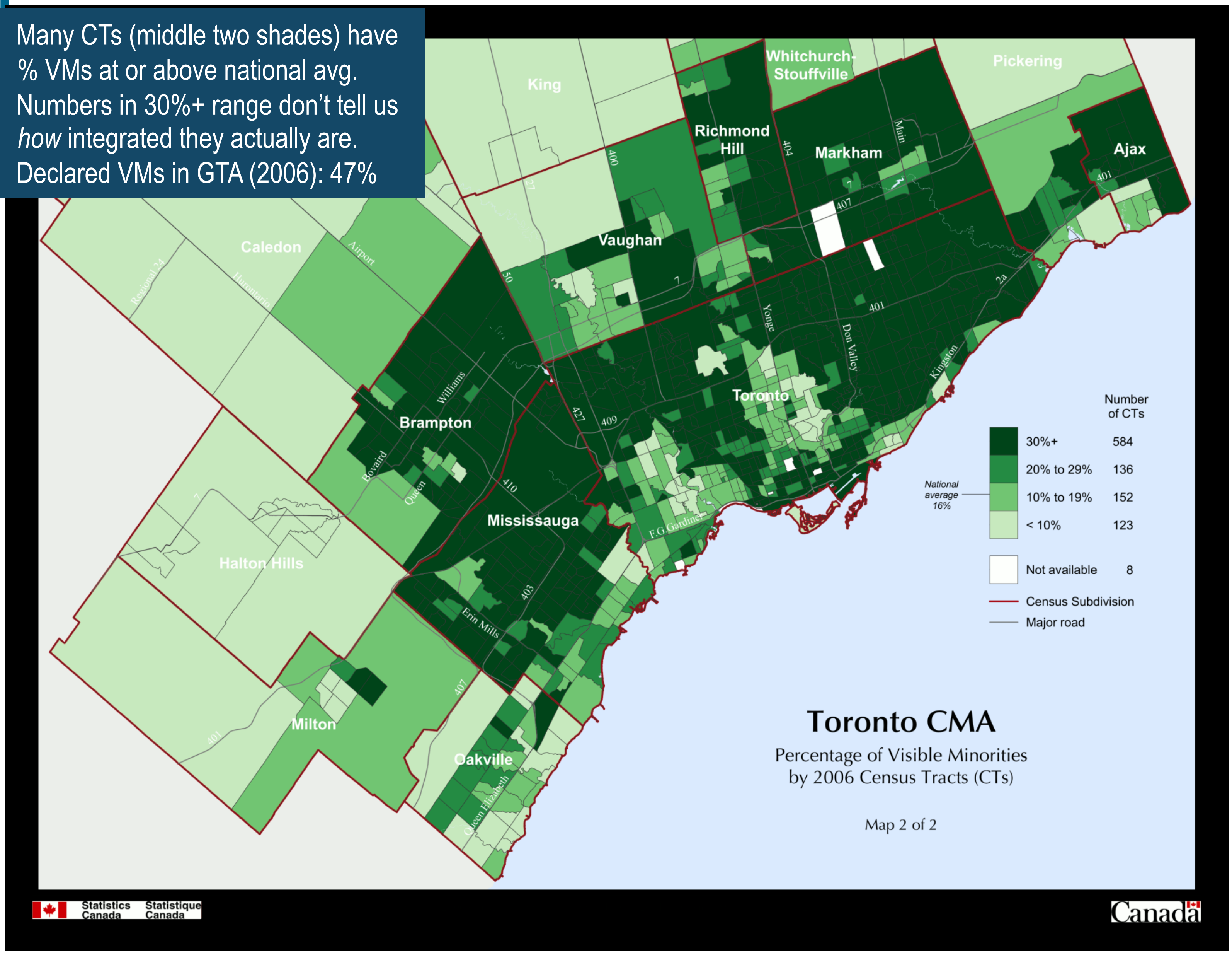
Portland



Red: white
Blue: black
Green: Asian
Orange: Hispanic
Each dot=25 people

See: http://www.flickr.com/photos/walkingsf/sets/72157626354149574/detail/

- Many CTs (middle two shades) have % VMs at or above national avg.
- Numbers in 30%+ range don't tell us *how* integrated they actually are.
- Declared VMs in GTA (2006): 47%



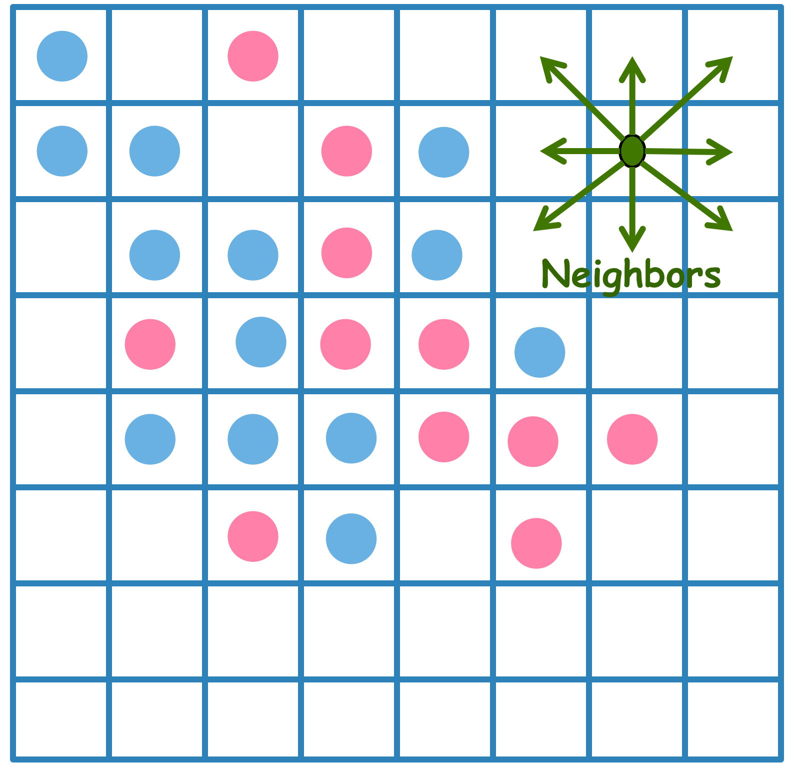
http://geodepot.statcan.gc.ca/2006/13011619/04152314121501 04_20080513012009031 3011619_05-eng.jsp?theme=minorities&serie=CMA&geo=Toronto&callingName=2008051301200903130 11619_05-eng.jsp&fileName=&Submit=Next

Schelling Model

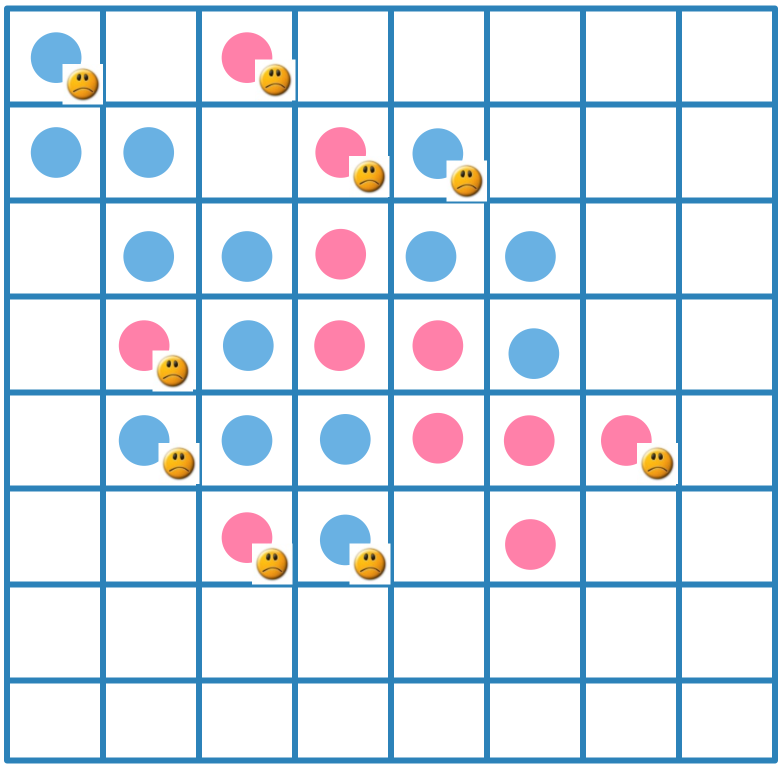
- Model offers one perspective on spatial segregation
  - assume some degree of desired homophily w.r.t. physical neighbors
    - i.e., you'd like to live near *some* people who are like you
  - *even if degree of homophily is mild, dramatic segregation can emerge*
- Simple mathematical model of movement (in a city, in a classroom, etc.)
  - people keep moving until they are happy with their neighbors
- We're interested in what happens when the process converges or stabilizes (or reaches *equilibrium*)

Schelling Model

- Simple model of movement
  - agents have one of two *types* (X/O; blue/red; boy/girl)
  - place agents on a grid, each with 8 *potential neighbors* in adjacent cells (some cells may be empty)
  - an agent is *happy (satisfied)* where they are if they have at least *t* neighbors who are like them; *unhappy* otherwise
    - threshold *t*: inverse of degree of “tolerance”
  - unhappy agents take turns moving to a better (empty) cell



Simulate the model

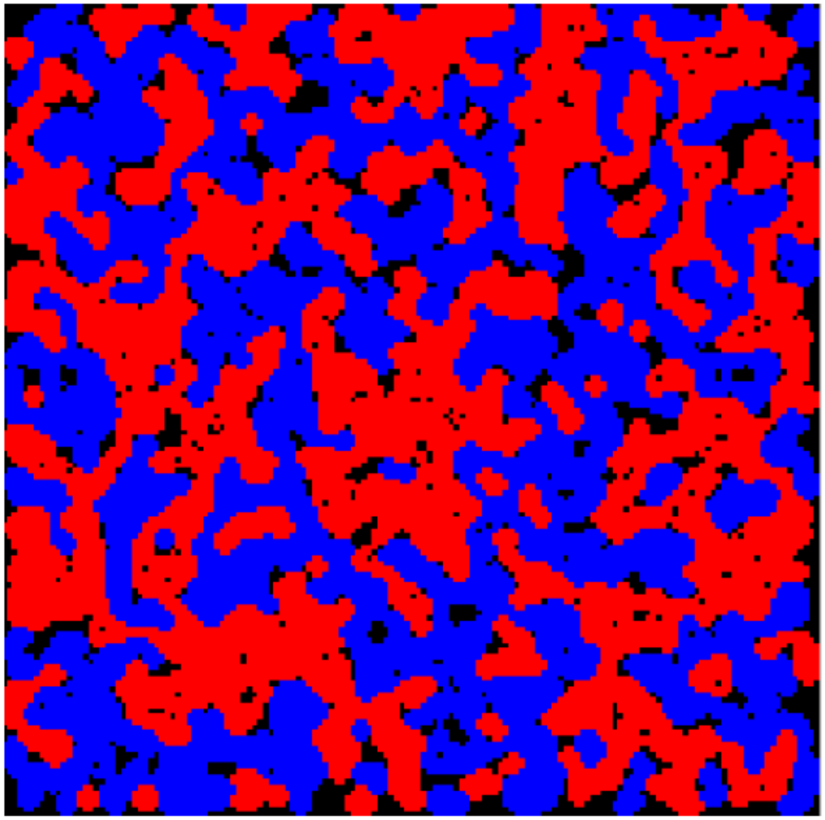


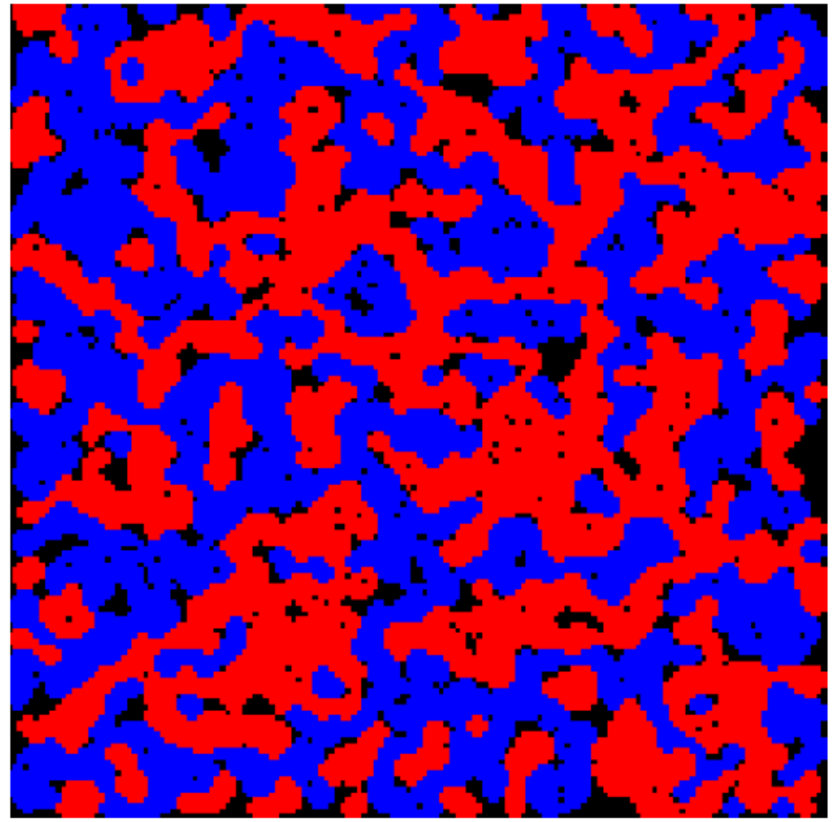
- If t=3, which agents are unhappy?
  - marked with a
- Where will unhappy agents go?
  - find an empty space that will make them happy (or move to random empty spot)
  - many variants of the model: who moves when, where they can move, etc.
- Simple model: sweep from top left (by row or column) to bottom right, letting each agent move if unhappy; move to first place in sweep order that makes you happy (if no such cell, move to place with most “like” neighbors)
  - *simulate next step in this picture*
- Moves in one round can make “currently happy” agents unhappy at next round

Let's look at a Netlogo Simulation

- See course web page (resources) for description and link
  - software available for free
  - installed on CDF (command: netlogo)
  - *(we viewed a handful of NetLogo segregation model simulations using various parameter settings)*

Schelling Model at Convergence (t=3)

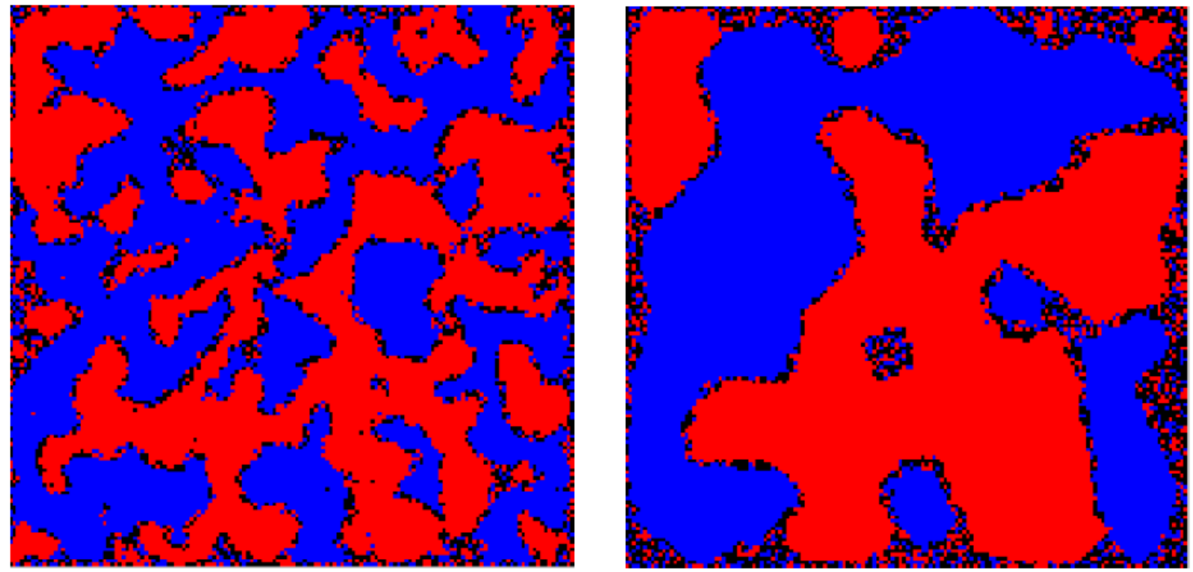




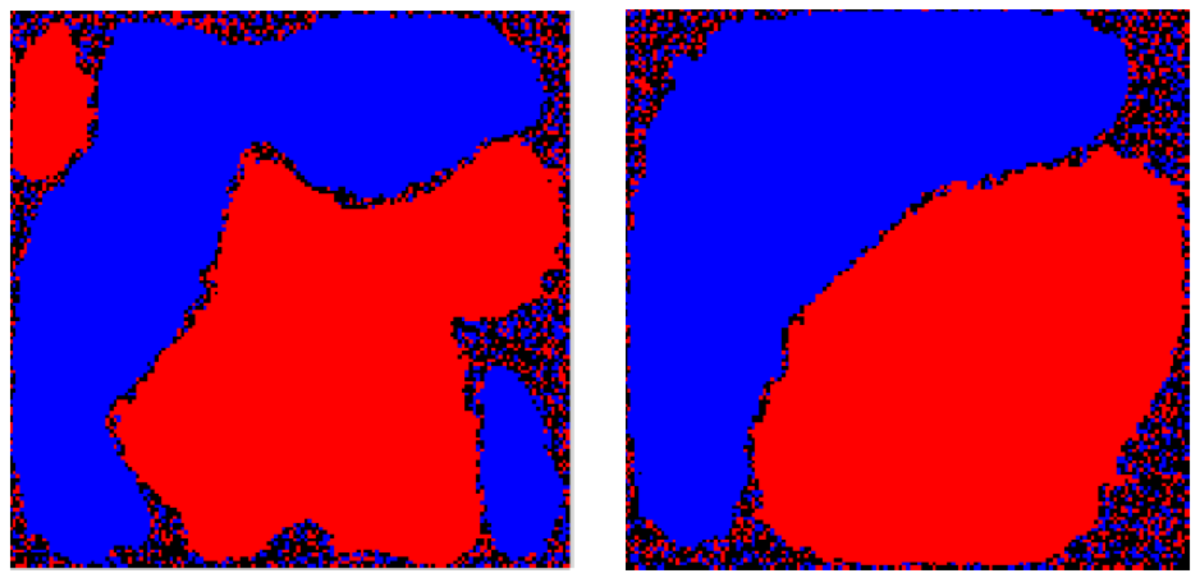
[Two runs: 150x150 grid, 10000 red, 10000 blue; threshold 3. E&K, Ch.4, Fig. 4.17]

Schelling Model: In Progress (t=4)

[One run, with placement after 20, 150, 350, 800 steps. 150x150 grid, 10000 red, 10000 blue; thresh = 4. E&K, Ch.4, Fig. 4.19]



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



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

Key points

- Start with some random assignment
  - note that rarely is a city initialized with a random assignment!
- With threshold of t=3
  - you see some integration, but largely arranged in large pockets
  - clusters form as previously happy agents are “abandoned” and are forced to move to more homogeneous neighborhoods
- With threshold of t=4
  - you see almost no integration
- Ask yourself what might happen with *mixed thresholds*
  - e.g., 20% of agents have t=2, 80% have t=5
- At larger thresholds, the simulations can take a long time to converge
- Illustrates how small amounts of homophily have a strong influence on segregation; how immutable characteristics influence mutable characteristics in unexpected ways:
  - consequences for other mutable characteristics: culture, beliefs, political opinions, etc.