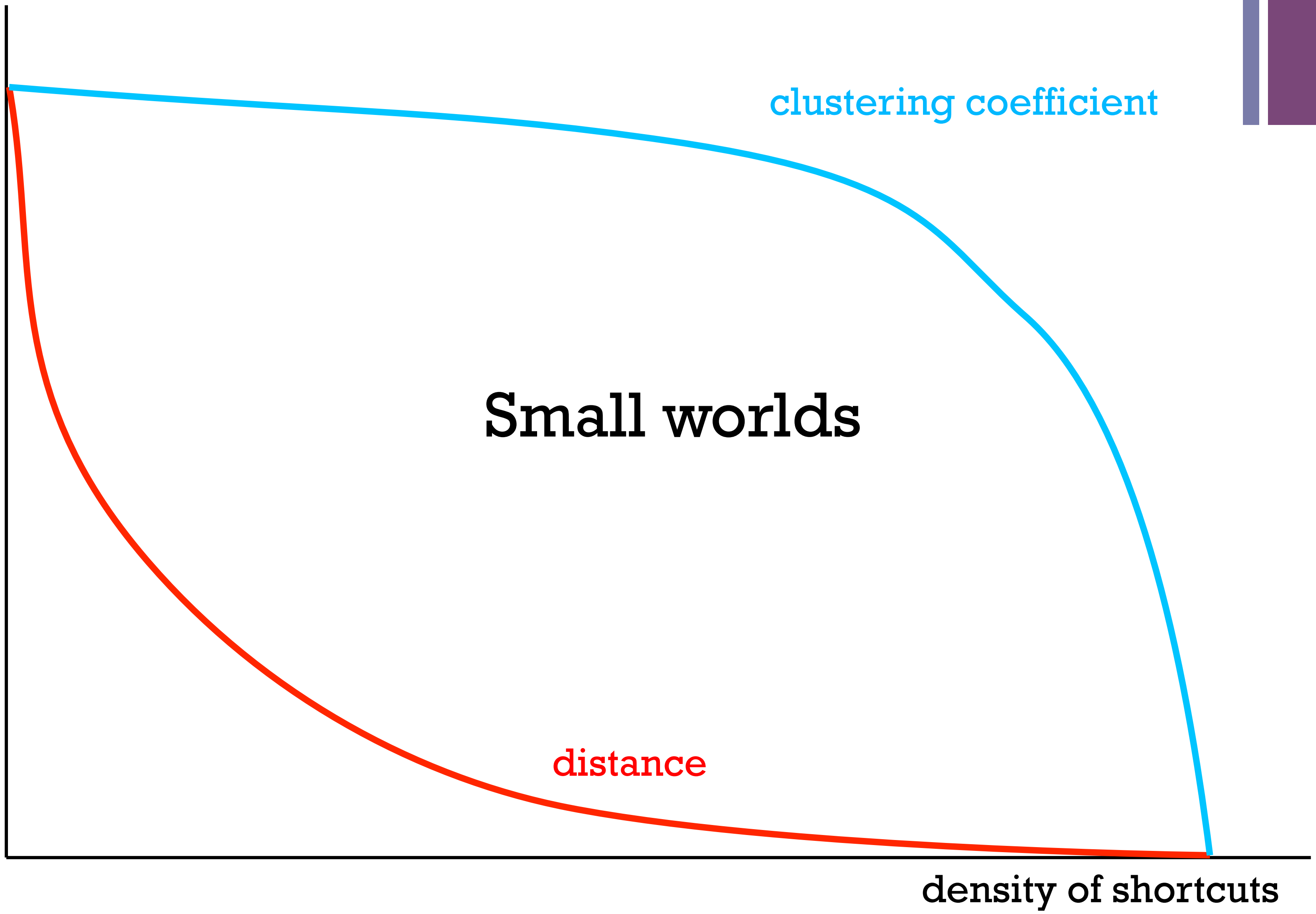
Networks Out of Control: Real-World Networks 2

Social Network

- Nodes are people. (Undirected) edges are connections representing friendships, acquaintances, business relationships, etc.
- Properties:
  - (Small) Diameter
  - Clustering
  - Navigability
  - Homophily

Clustering and Distance in SW Network



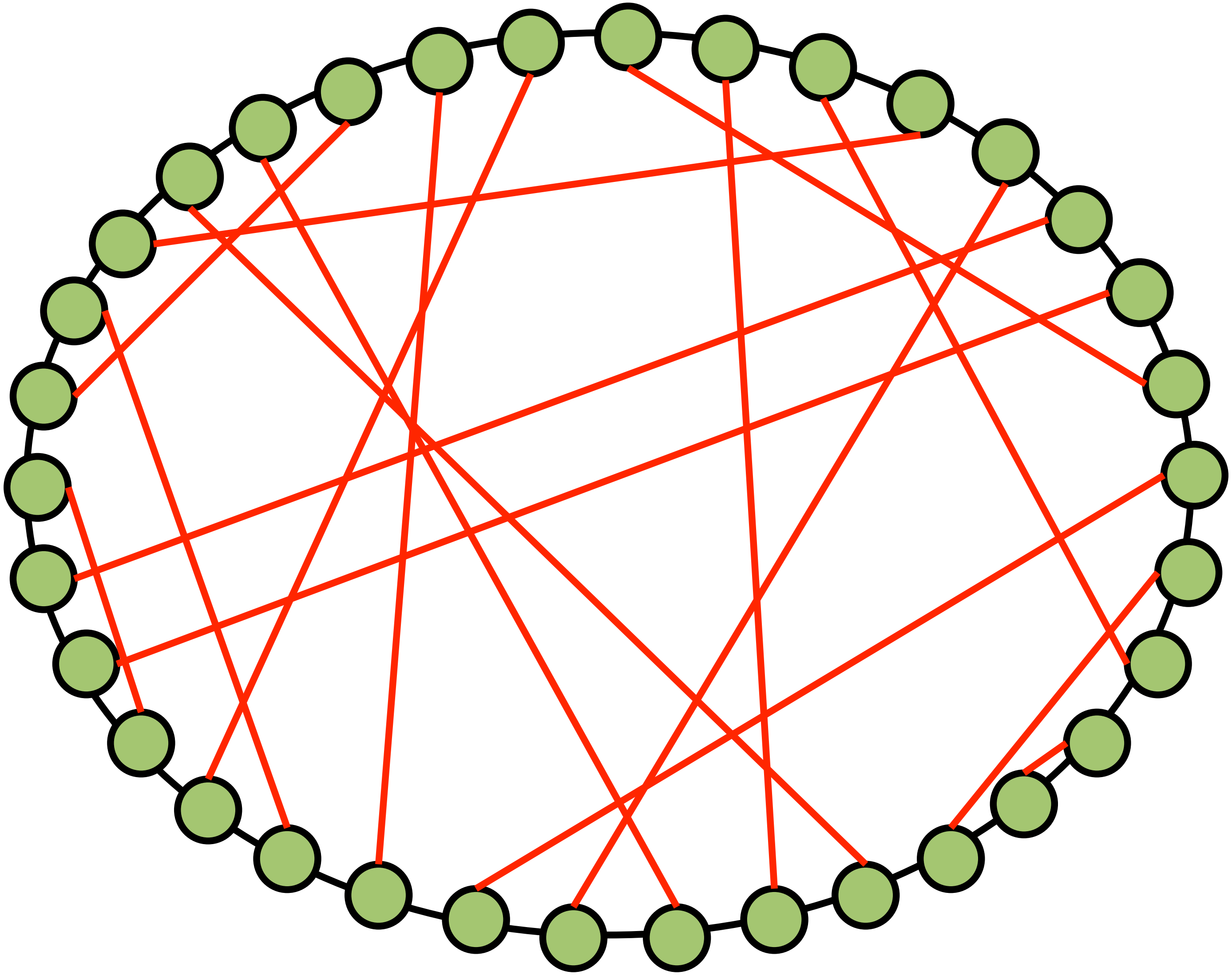
Small Diameter

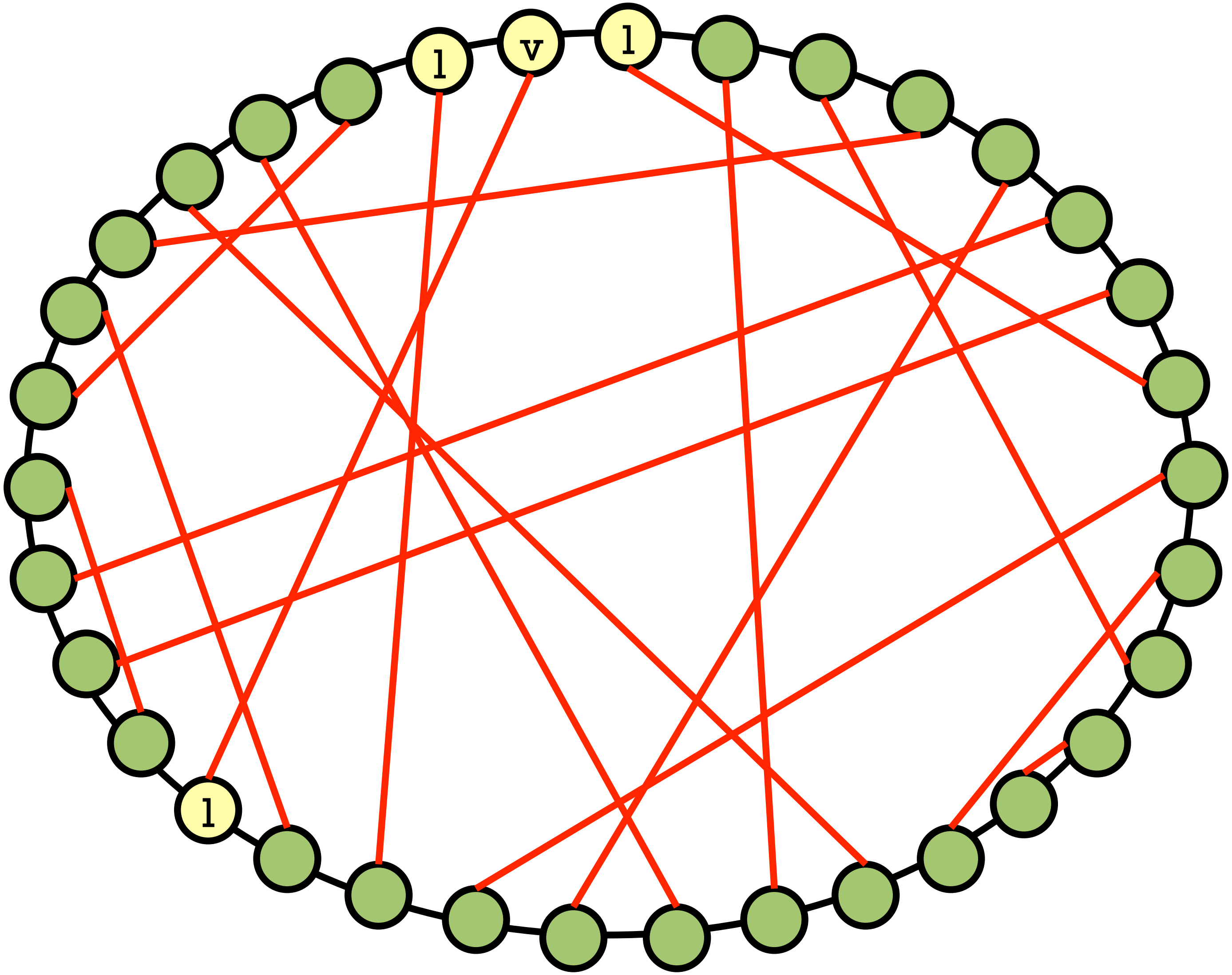
+ Small Diameter in Social Networks

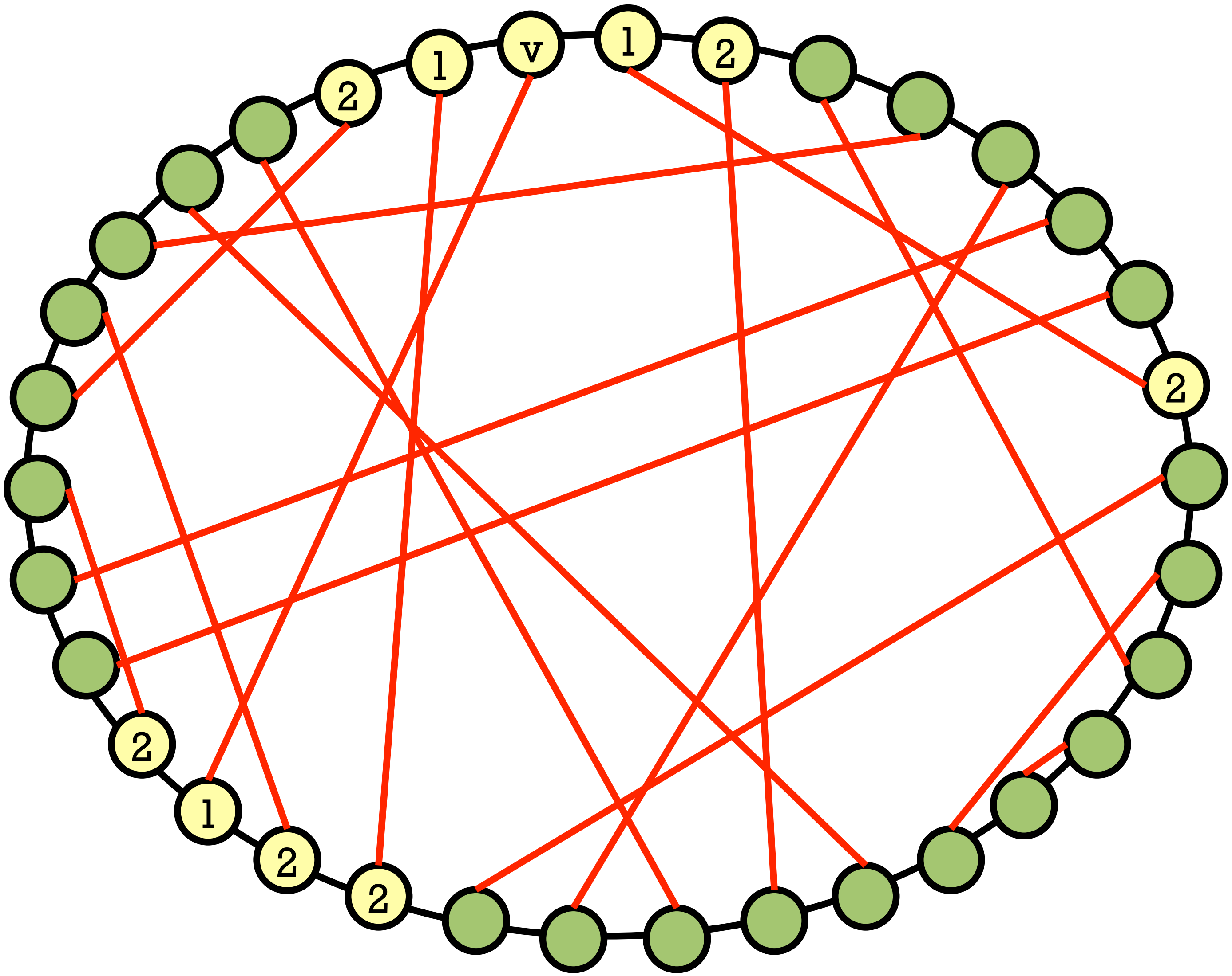
- To make our proof easier, consider a variant of WS(n,1,p)
  - Let G be a cylce + a random perfect matching.
  - Call edges in the perfect matching "chords"

- We will show that this network has small diameter!
  - $\log(n) - o(1) < \text{diam}(G) < \log(n) + \log(\log(n)) + o(1)$

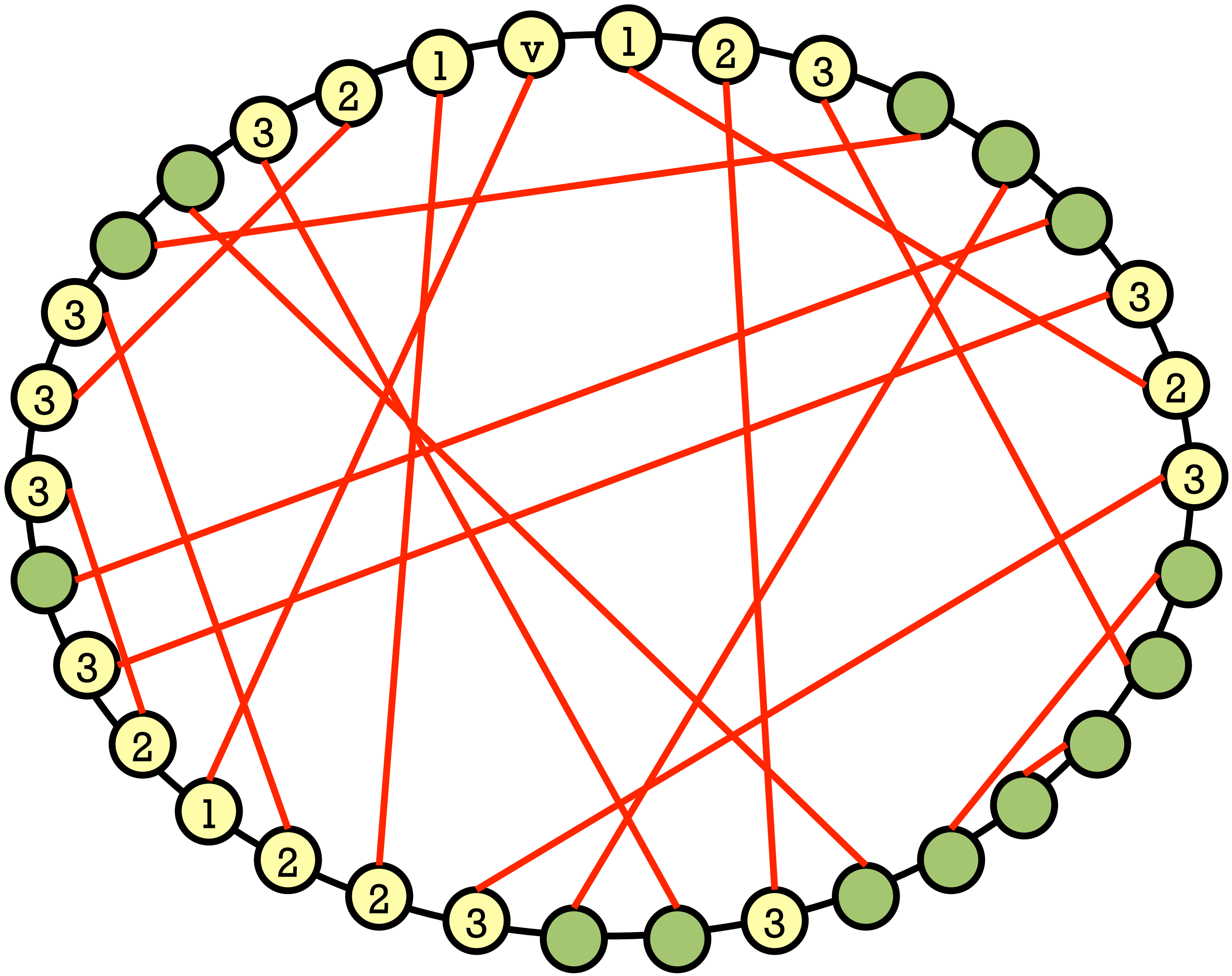
Intuition:




l
v
l
l



2
1
v
1
2
2
2
1
2
2


v
1
2
3
3
2
3
3
3
2
2
1
2
3
3
3
3
3
2
1

+

- Intuition: most chords go "far"
  - i.e., they find a new vertex, that is sufficiently far from any vertex discovered thus far.
- Proof:
  - Step 1: Look at "short" distances (1/5)log(n), and show that "local chords" are rare – i.e., most edges go somewhere new.
  - Step 2: Look at "long" distances up to (3/5)log(n), and show that "local chords" are still relatively rare
  - Use above to show that expanding from any two vertices, the process "collides" after few steps.

Small Diameter in Random Graphs

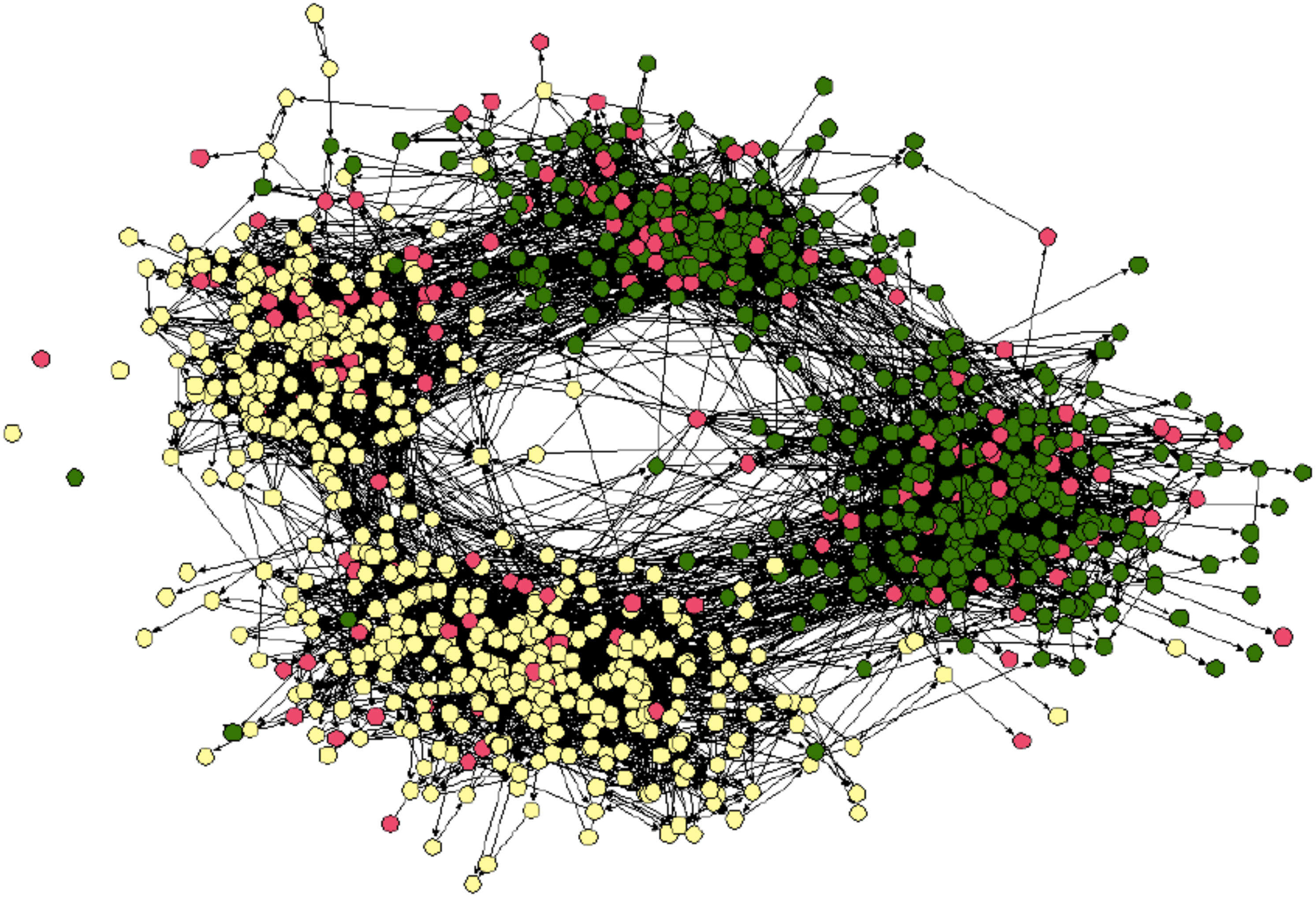
- Similar results (with similar, but more involved, proofs) hold for G(n,p), G(n,r), and WS(n,k,p).

- Diam{G(n,p)} = (1 + o(1) log(n) / log(np) a.a.s.
- Diam{G(n,r)} \in (1 – eps, 1+ eps) log(n) / log(r-1) a.a.s.
- AvgDist{WS(n,k,p)} = log(n/k) / log(k) a.a.s.

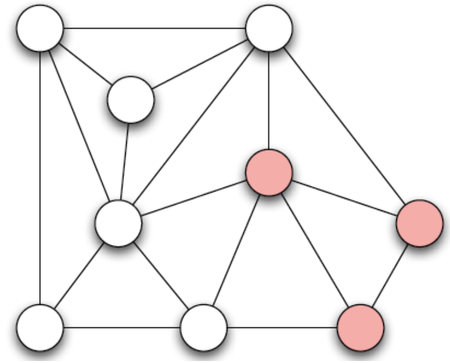
Homophily

Homophily

- We tend to be similar to our friends.



+ Homophily



- Can we measure homophily?
  - For two "types", let p be the fraction of type A, and q be the fraction of type B in the network.
  - Select an edge uniformly at random. If there is no homophily:
    - The probability that we selected an A-A edge:
    - The probability that we selected a B-B edge:
    - The probability that we selected a A-B or B-A edge:
- We say that there is homophily if the percentage of A-B or B-A edges is "significantly" less than 2pq.
  - Here we mean statistical significance as some deviation is expected just due to randomness.
  - E.g., in the small example: q = 1/3, p = 2/3, so 2pq = 8/18, but we only observe 5 cross-edges.
  - Can also have *inverse homophily* (or *heterophily*).
  - Easily generalizable to more than two types.

Homophily and Clustering

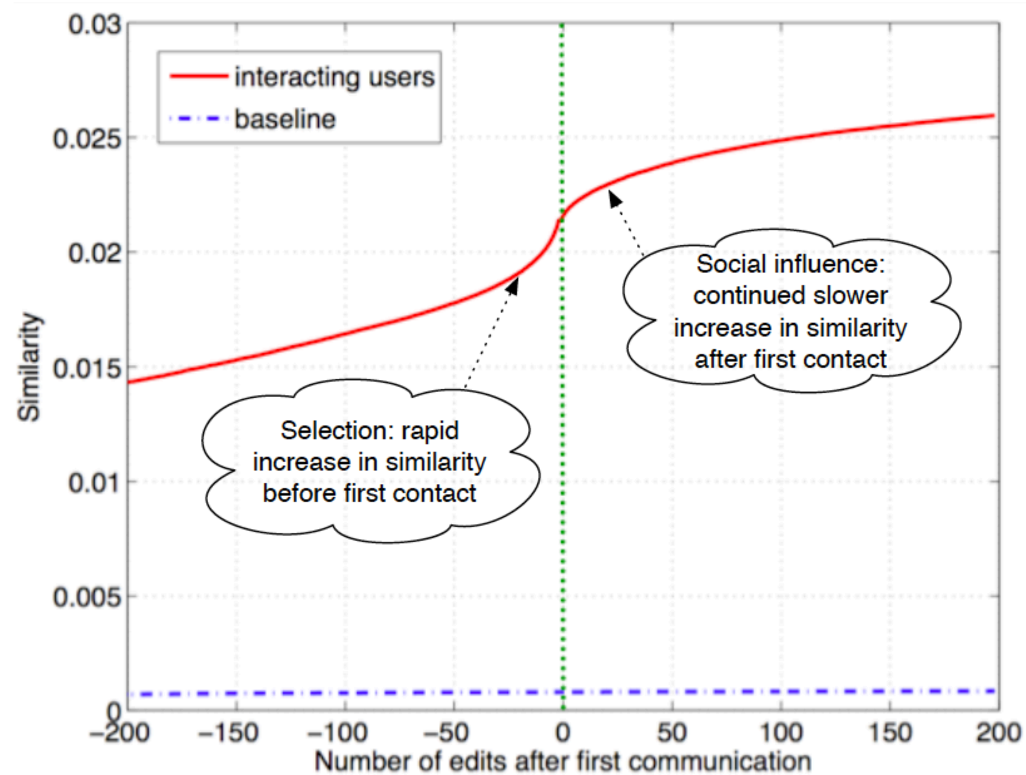
- Clustering is the *observed* result of homophily.
  - Take the extreme case where there are no A-B edges.
    - Then A-A edges and B-B edges are naturally more dense -> higher clustering.

+ Homophily

- Why does it exist?
  - Selection – the tendency of people to form friendships with others who are like them.
    - Different scales and levels of intentionality, includes both
      - active selection: *becoming friends with a classmate who is also interested in sports,* and
      - passive selection: *having friends in the same socio-economic class because you live in the same neighborhood.*
    - Characteristics Drive Links
  - Social Influence – the tendency of people to become more like their friends.
    - For example, learning to ski because your friends already ski.
      - Related to Affiliation Networks and Cascades (future lectures).
    - Links Drive Characteristics

+
Homophily

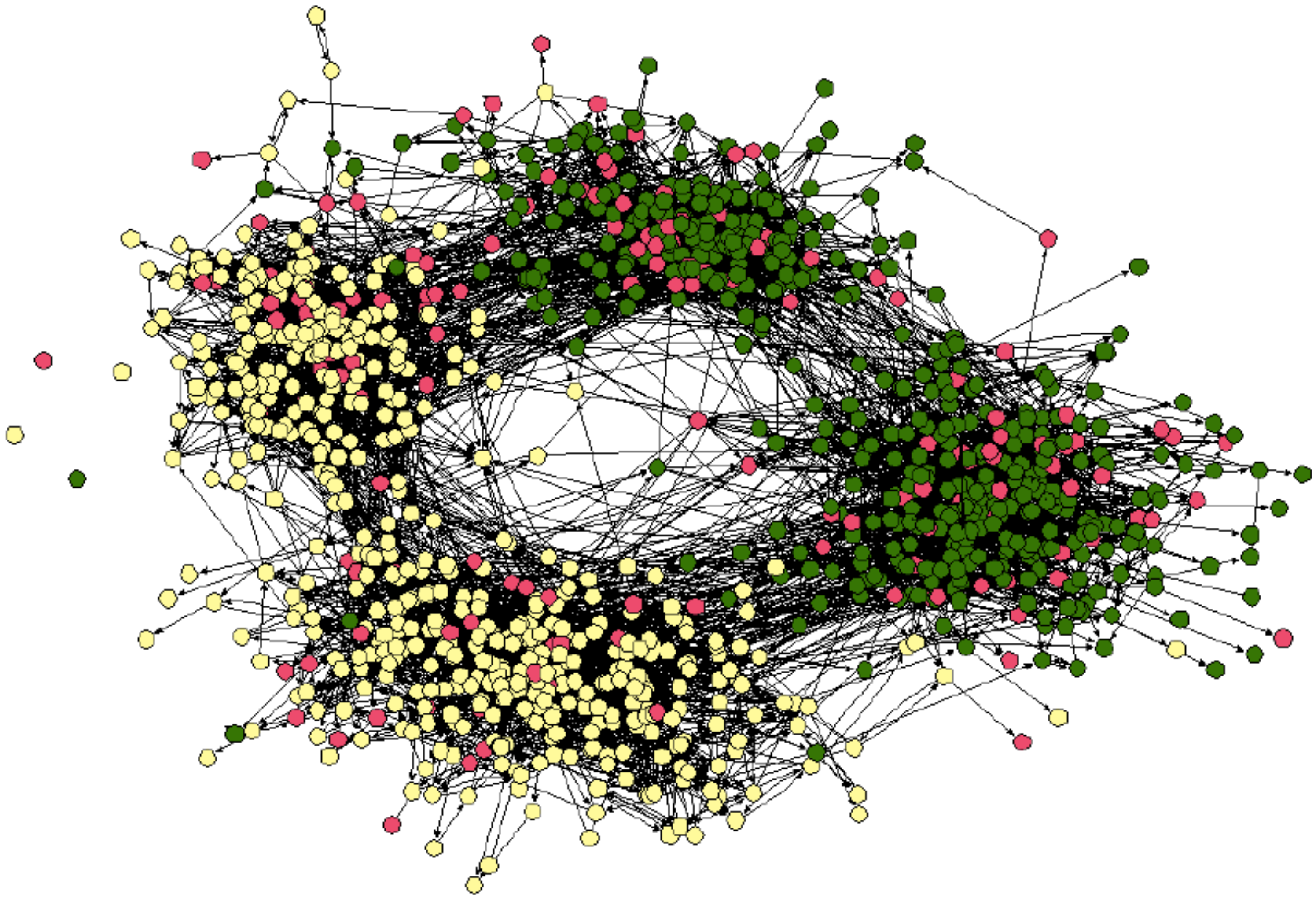
- When is homophily is due to selection vs social influence?
  - It is not possible to tell from a single snapshot of the network – must use a *longitudinal study* in which social behaviors and network ties tracked over a long period of time.
    - Allows us to see if behavioral changes occur *before* or *after* a social tie is formed.



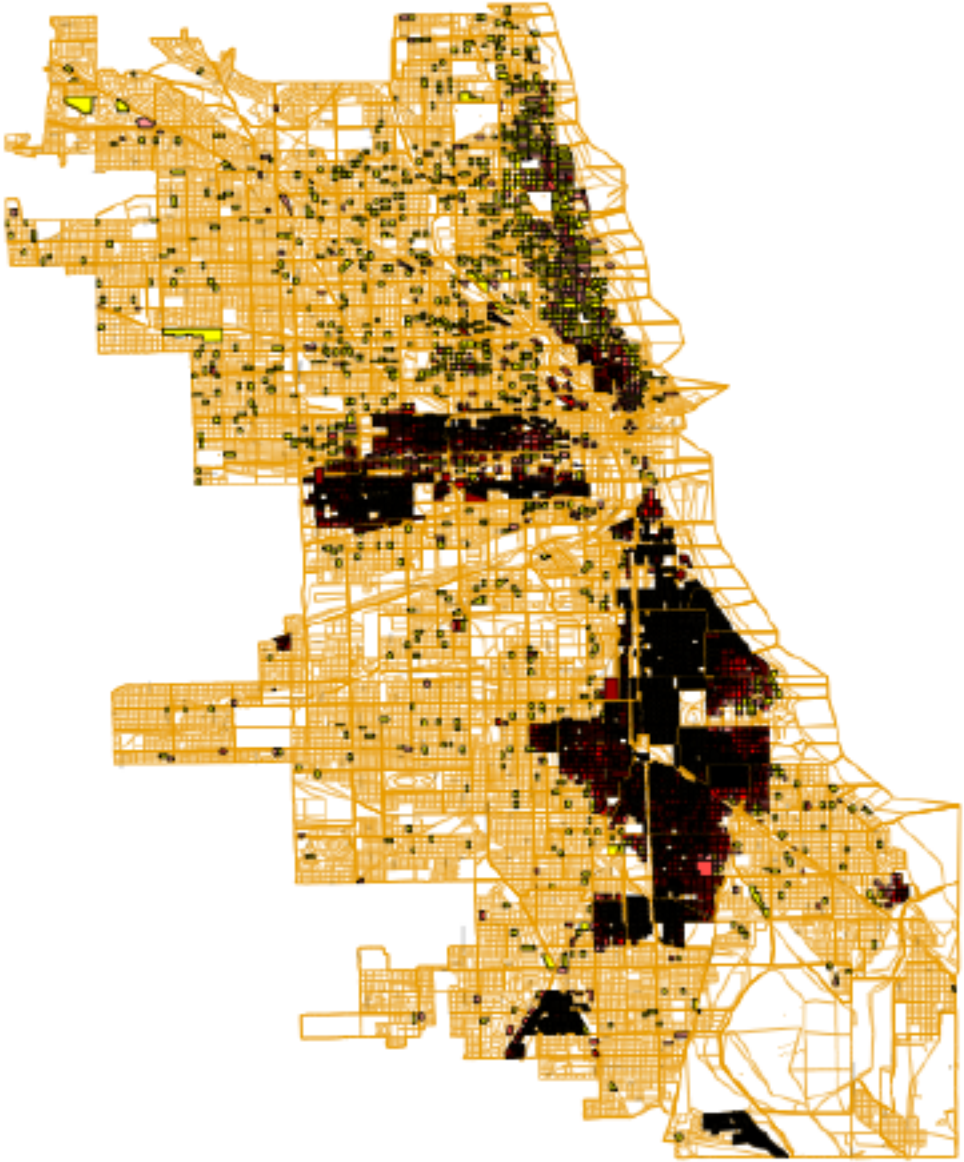
Homophily

- Example studies:
  - Teenager drug use:
    - Selection comparable or greater than social influence!
    - Implications on realistic interventions.
  - Longitudinal study on obesity over 32 years:
    - Found homophily when classify types by obese and non-obese.
      - Social influence comparable or greater than selection!
      - Implications on realistic interventions

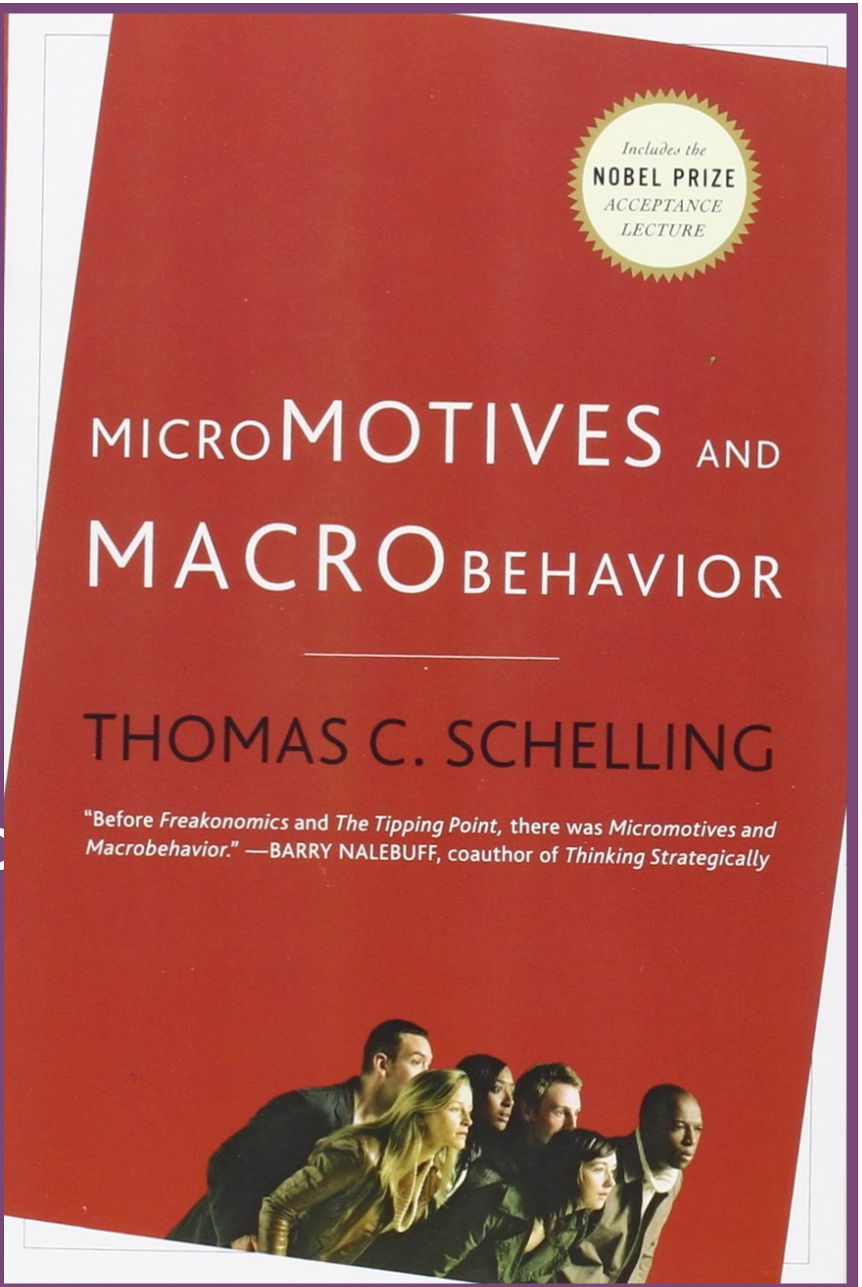
Homophily



+ Homophily



Schelling's Neighborhoo



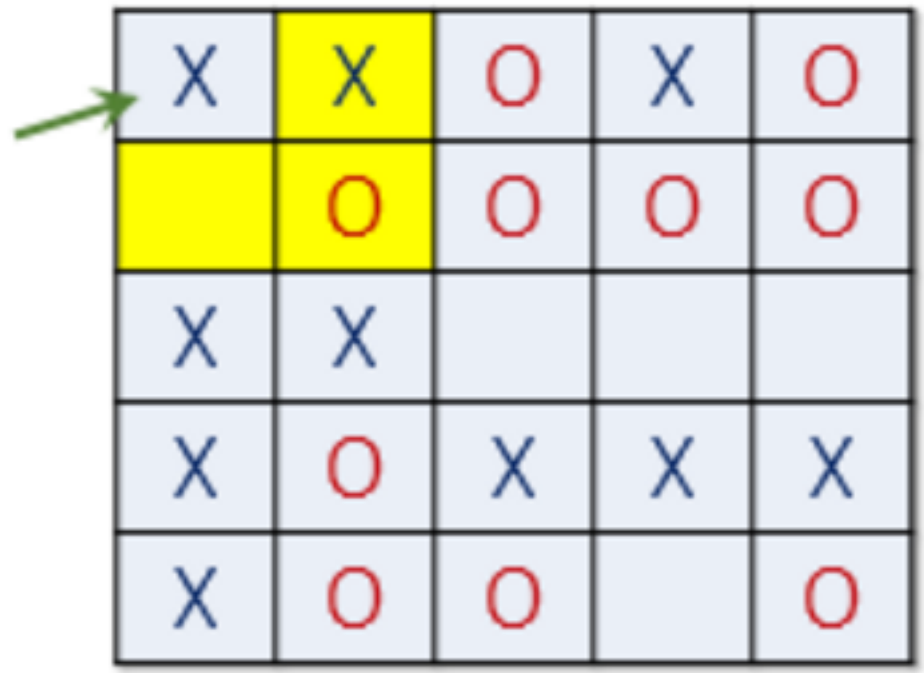
+ Neighborhood Model

- There are two *types* of individuals, living in a grid.
- An individual's *neighborhood* consists of the (up to) 8 squares that surround it.
- An agent is *satisfied* if at least a p fraction of its neighbors are of the same type.
  - Example: p = 50%

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

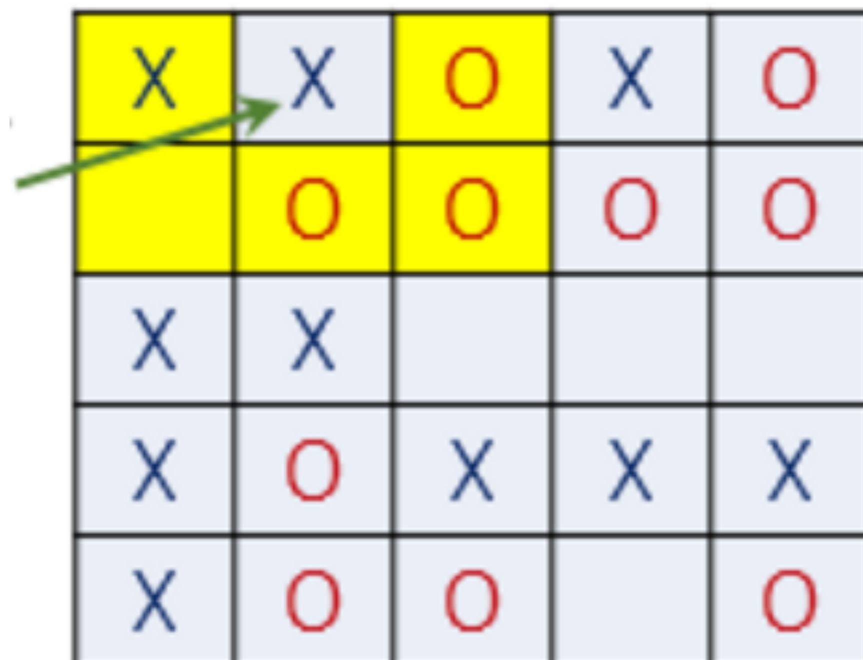
Neighborhood Model

- There are two *types* of individuals, living in a grid.
- An individual's *neighborhood* consists of the (up to) 8 squares that surround it.
- An agent is *satisfied* if at least a p fraction of its neighbors are of the same type.
  - Example: p = 50%



Neighborhood Model

- There are two *types* of individuals, living in a grid.
- An individual's *neighborhood* consists of the (up to) 8 squares that surround it.
- An agent is *satisfied* if at least a p fraction of its neighbors are of the same type.
  - Example: p = 50%



Neighborhood Model

- There are two *types* of individuals, living in a grid.
- An individual's *neighborhood* consists of the (up to) 8 squares that surround it.
- An agent is *satisfied* if at least a p fraction of its neighbors are of the same type.
  - Example: p = 50%

X	X*	O	X*	O
	O	O	O	O
X	X			
X	O*	X	X	X
X	O	O		O*

Neighborhood Model

- There are two *types* of individuals, living in a grid.
- An individual's *neighborhood* consists of the (up to) 8 squares that surround it.
- An agent is *satisfied* if at least a p fraction of its neighbors are of the same type.
  - Example: p = 50%
- All dissatisfied neighbors move to a random unoccupied cell.
- Repeat.

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

Neighborhood Model

- Are there a fixed points?
  - (Often) yes (depends on number of blank squares and p).
- Given a fixed point, we can consider its homophily – think of each square as a node, with an edge to each of the eight squares surrounding it.
- What kind of fixed points exist with respect to homophily?
  - Has extreme homophily:
    - Segregated (upper and lower triangles)
  - Has no homophily:
    - Integrated (checkered)
- If we initialize randomly, do we converge to a point with high or low homophily?

Neighborhood Model

- Similar results are observed when we modify the rules:
  - An agent is satisfied if at least k neighbors are of its same type (regardless of how many neighbors are not of the same type).
    - More "aggregation" than "segregation", but the end result is the same.
  - Different percentages (or number of) types.
    - Tends to exacerbate segregation/aggregation as "rare" types have to cluster together in order to be satisfied.

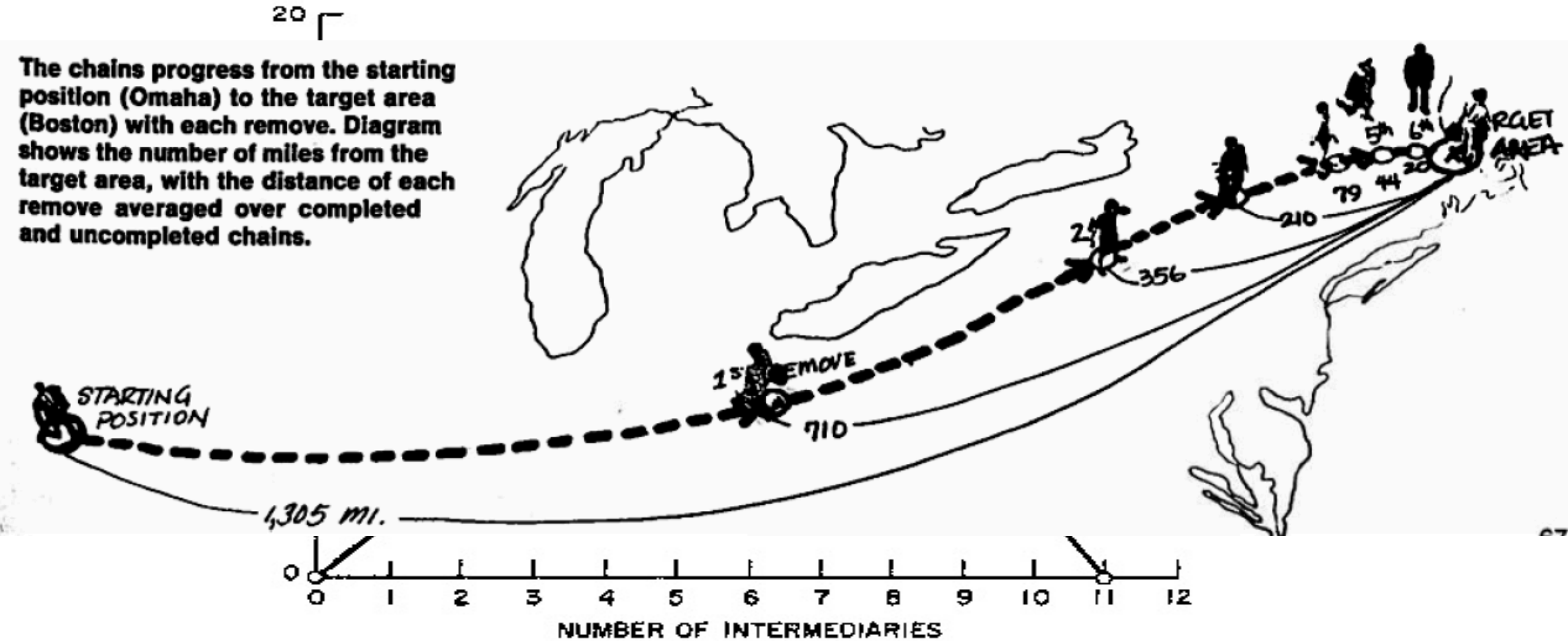
- Suggestive conclusions:
  - segregation does not require extreme negative opinions.
  - on the other hand, "positive" in-group behavior can be just as harmful on a large scale as "negative" out-group behavior.
  - before segregation, most individuals were satisfied – can incentivize them to not move?

X	X*	O	X*	O
	O	O	O	O
X	X			
X	O*	X	X	X
X	O	O		O*

Navigability

Social Networks: Small-World

- [Milgram 1969] experiment to study the *average distance* between two nodes in a social network.



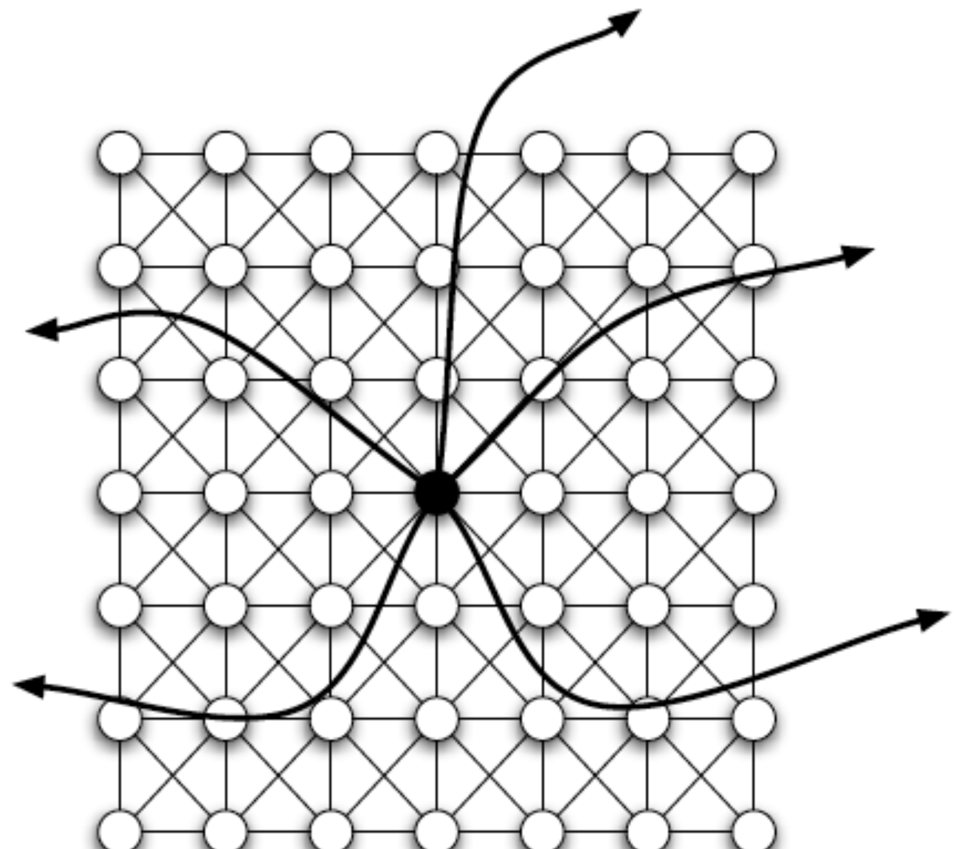
- Paths are not just short – they can be found!

+ Social Networks: Small-World

- A *decentralized routing algorithm* takes local (node-level) decisions on where to forward a message next based only on
  - the geographic location of the current node and its neighbors,
  - the geographic location of the target node, and
- What do we mean by ``geographic location''?

Watts-Strogatz on a Grid

- There are n^2 nodes arranged in a square grid in R^2, and we endow the space with the l_1 norm.
  - Nodes know the positions of themselves, the target, and their neighbors.
- Every node v connects to all nodes u such that $d(u,v) \leq r$.
- Every node has k additional edges connected to uniformly random endpoints u.



Is Watts-Strogatz on a Grid Navigable?

- Note: the distance between a (randomly selected) source and target is O(n) a.a.s.
- Our goal: Reach the target in $O(n^{\delta})$ steps for $\delta << 1$.
- Our approach:
  - Consider a ball B of nodes within some "short" distance n^{δ} to the target.
    - Within the ball, can reach the target in n^{δ} steps.
  - Can we reach the ball quickly?
    - Without shortcuts takes $O(n - n^{\delta})$ steps to reach the ball a.a.s.
    - Must make use of shortcuts, in particular, need to show that at least one of the first $O(n^{\delta})$ nodes has a shortcut to B.

+ # Is Watts-Strogatz on a Grid Navigable?

- How many vertices are there in B? $1+\sum_{j=1}^{n^\delta} 4j \le 4n^{2\delta}$
- What is the probability that a vertex v has a shortcut into B?

$$\mathbb{P}[E_v] = r\frac{|B|}{n^2} \le 4rn^{2\delta-2}$$

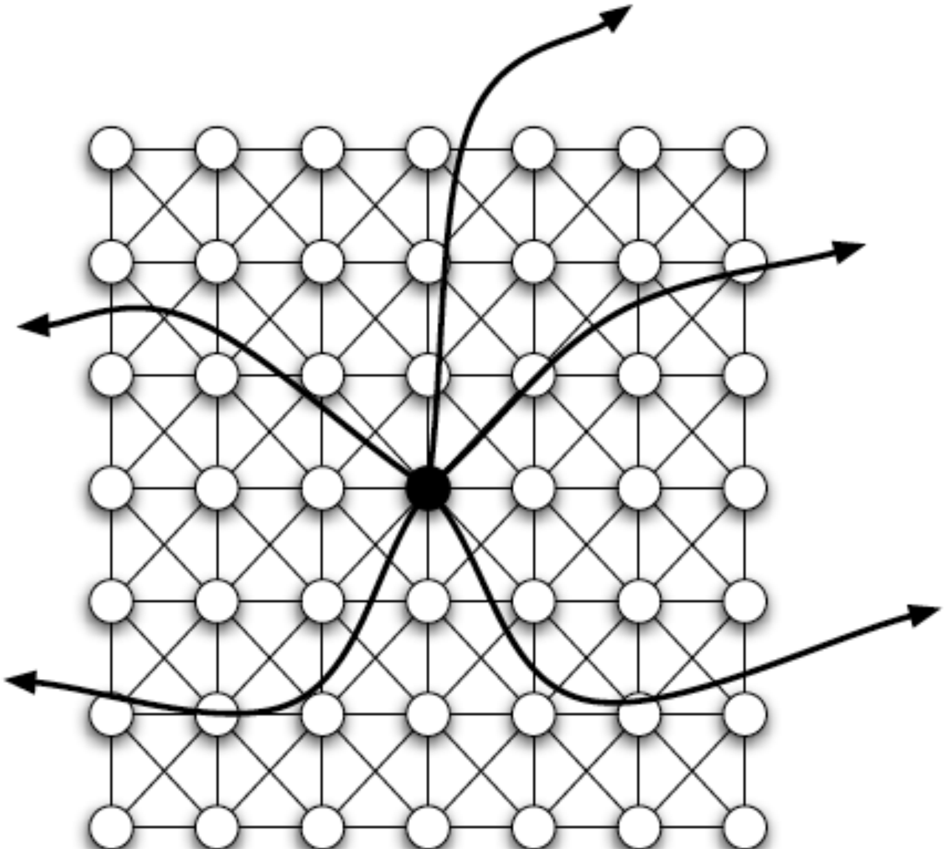
- What is the probability that any vertex in the first t ($= \lambda n^\delta$) steps has a shortcut into B?

$$\mathbb{P}[E = \cup_{1\le i\le t} E_{v_i}]$$

- If $3\delta - 2 < 0$ (i.e., $\delta < ⅔$) this probability is vanishing.
- Thus, routing takes (at least!) $n^{⅔}$ steps!

Distance-Proportional Watts-Strogatz on a Grid

- There are n^2 nodes arranged in a square grid in R^2, and we endow the space with the l_1 norm.
- Every node v connects to all nodes u such that $d(u,v) \leq r$.
- Every node has k additional edges connected i.i.d. to u **proportionally to $d(u,v)^{-\gamma}$ for constant $\gamma \geq 0$.**



$$\frac{d(u,v)^{-\gamma}}{\sum_{u \neq v} d(u,v)^{-\gamma}}$$

Is Distance-Proportional Watts-Strogatz Navigable?

- When $\gamma = 0$
  - The model is exactly the original WS model on the grid, so still not navigable!
  - In fact, a similar proof shows it is not navigable for $\gamma < 2$; the number of steps is at least n^{δ} for $\delta = (2 - \gamma)/3$.
- What is the problem?
  - Shortcuts are ``too random''
- What about $\gamma > 2$?

+ Is Distance-Proportional Watts-Strogatz Navigable?

$$\mathbb{P}\{d(u,v) > d\} \leq$$

- Let E_{vi} denote the event that, at step i the vertex v_i has a shortcut of length at least $n^{1-\beta}$, and E the event that this occurs in the first λn^{β} steps.

$$\mathbb{P}[E] \leq \sum_{i=1}^{\lambda n^{\beta}} \mathbb{P}[E_{v_i}]$$

- For this to not be vanishing, β must be at least $(\gamma-2)/(\gamma-1)$.
- Thus, the message can at best find shortcuts of distance less than $n^{1-\beta}$ in the first λn^{β} steps, for a total progress O(n), so at least $O(n^{\beta})$ steps are required.

Is Distance-Proportional Watts-Strogatz Navigable?

- When $\gamma < 2$, shortcuts are ``too random''
- When $\gamma > 2$, shortcuts are ``too short''
- Is there a sweet-spot at $\gamma = 2$?

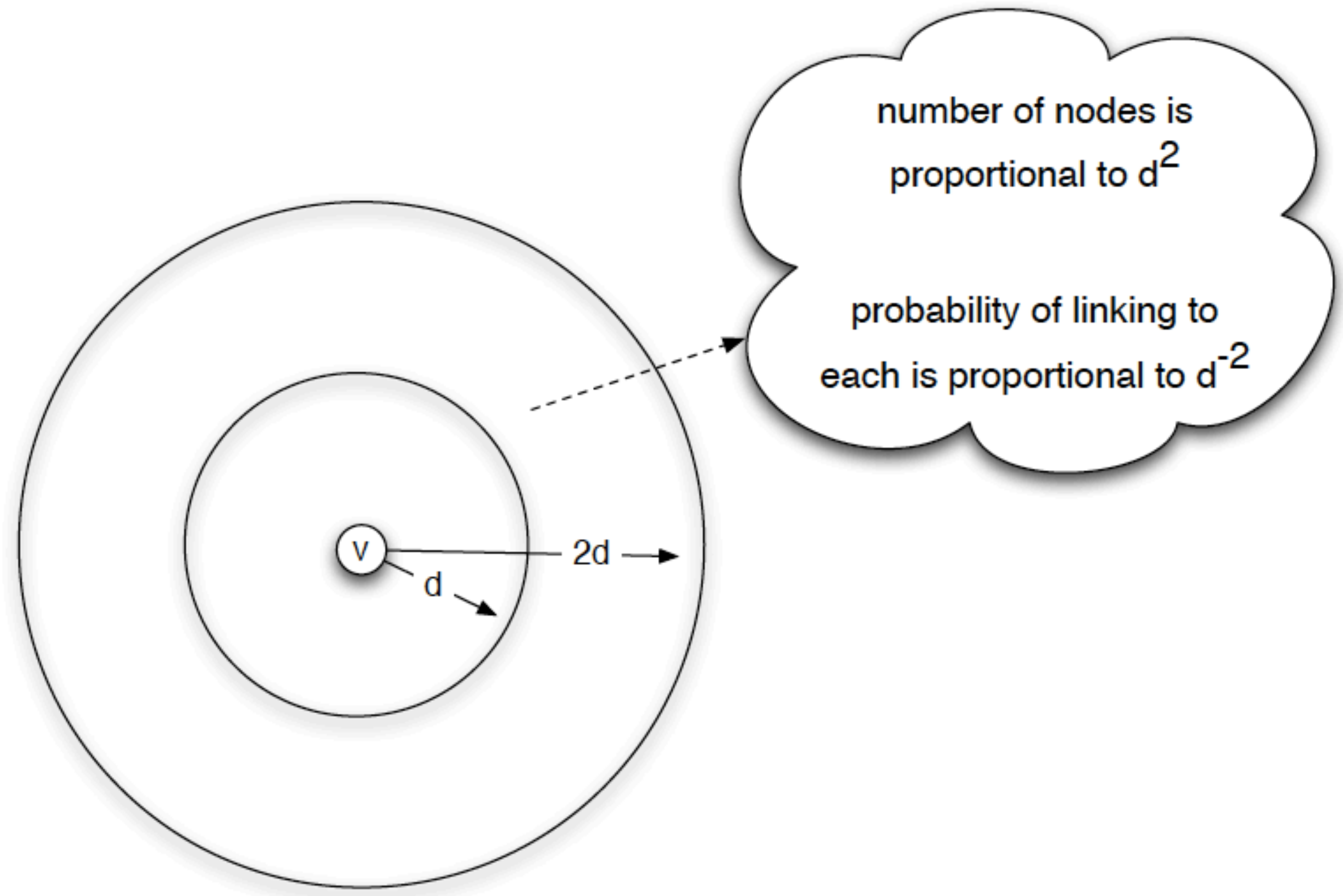
Is Distance-Proportional Watts-Strogatz Navigable?

- Show constructive proof for $r=1$ and $k=1$ (this is the worst case) with steps $O((\log n)^2)$.
- At each time step, send to the neighbor that is closest to the target.
  - Note: this always terminates as progress is made at every step, if only through the lattice.
- Definitions:
  - The annuli U_j is the set of nodes at lattice distance in $[2^j+1,\ 2^{j+1}]$ from the target.
  - The ball B_i is the union of all U_j with $j < i$.
  - The algorithm is in phase j when the message is in U_j
  - Note: there are at most log n phases.

Is Distance-Proportional Watts-Strogatz Navigable?

- Approach: Show that we progress quickly from phase to phase.
- If we are in phase j at node u, then to end the phase, we must pass the message to a node in B_j.
  - The probability of connecting to B_j is at least:
    $$|B_j| \cdot \frac{\max_{v \in B_j} d(u,v)^{-\gamma}}{\sum_{u \neq v} d(u,v)^{-\gamma}}$$
  - The size of B_j: $|B_j| \geq 2^{2j-1}$
  - The maximum distance: $\max_{v \in B_j} d(u,v)^{-\gamma} \leq 2^{j+1} + 2^j < 2^{j+2}$
  - The normalizing constant is:

Is Distance-Proportional Watts-Strogatz Navigable?



Is Distance-Proportional Watts-Strogatz Navigable?

- Approach: Show that we progress quickly from phase to phase.
- If we are in phase j at node u, then to end the phase, we must pass the message to a node in B_j.
  - The probability of connecting to B_j is at least: $|B_j| \cdot \frac{\max_{v \in B_j} d(u,v)^{-\gamma}}{\sum_{u \neq v} d(u,v)^{-\gamma}}$
  - The size of B_j: $|B_j| \geq 2^{2j-1}$
  - The maximum distance: $\max_{v \in B_j} d(u,v)^{-\gamma} \leq 2^{j+1} + 2^j < 2^{j+2}$
  - The normalizing constant is: $\sum_{u \neq v} d(u,v)^{-\gamma} \leq 4\ln(6n)$

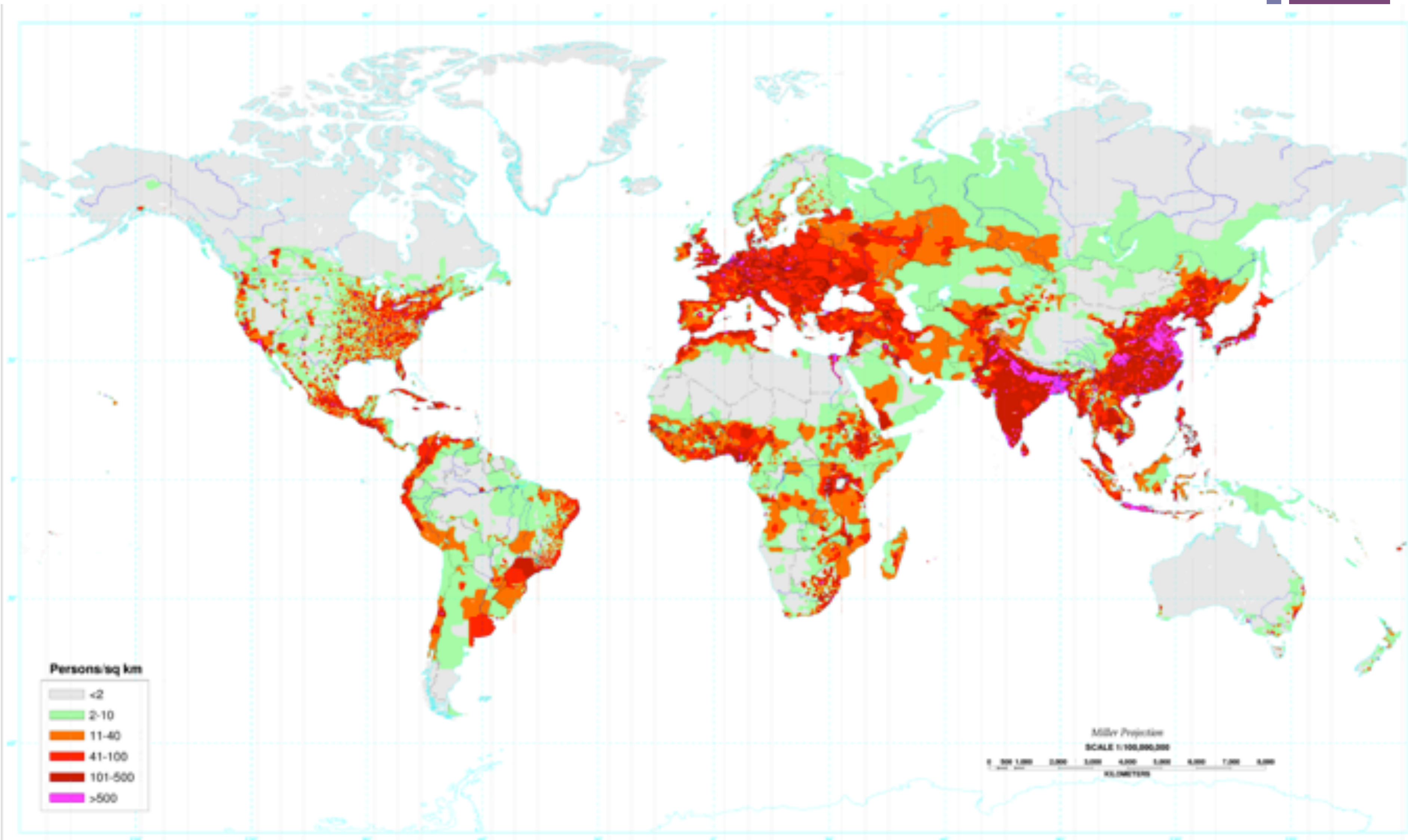
Is Distance-Proportional Watts-Strogatz Navigable?

- Therefore, we change phases with probability on the order of 1/log(n).
- In expectation, it takes O(log(n)) steps to change phases, and as we noted before, there are log(n) phases, hence we have an efficient decentralized routing algorithm that requires $O((\log n)^2)$ steps!

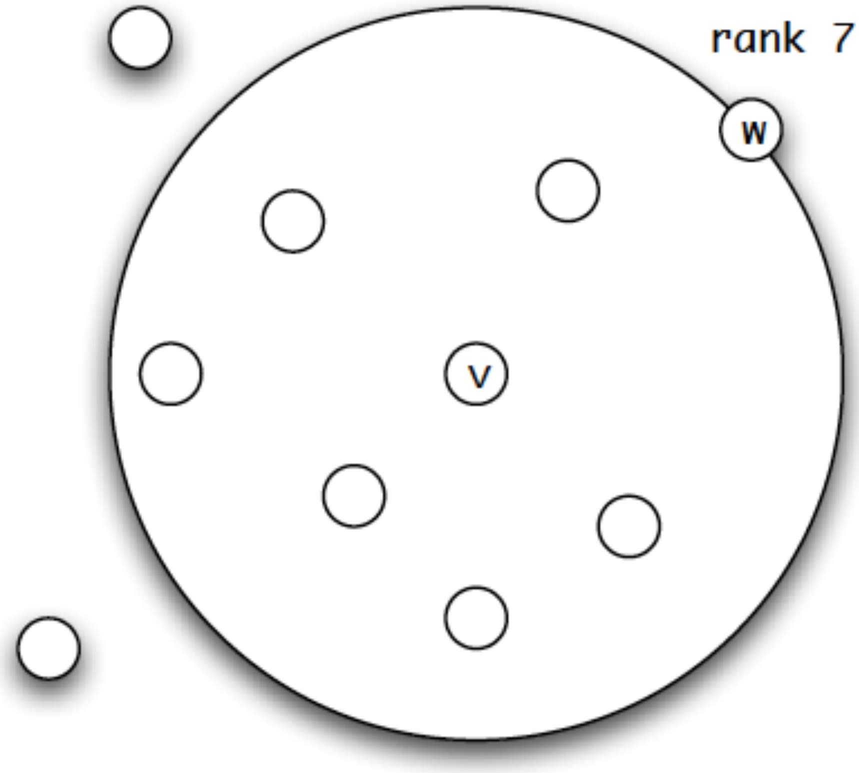
+
Is Distance-Proportional Watts-Strogatz Navigable?

- Navigability:
  - When $\gamma < 2$, shortcuts are ``too random''
  - When $\gamma > 2$, shortcuts are ``too short''
  - There a sweet-spot at $\gamma = 2$!
- In general, can have a d-dimensional lattice, and have a similar phase transition at $\gamma = d$.
- Can also take other underlying topologies (e.g., see a case for trees in the notes).

Are real-world networks Navigable?

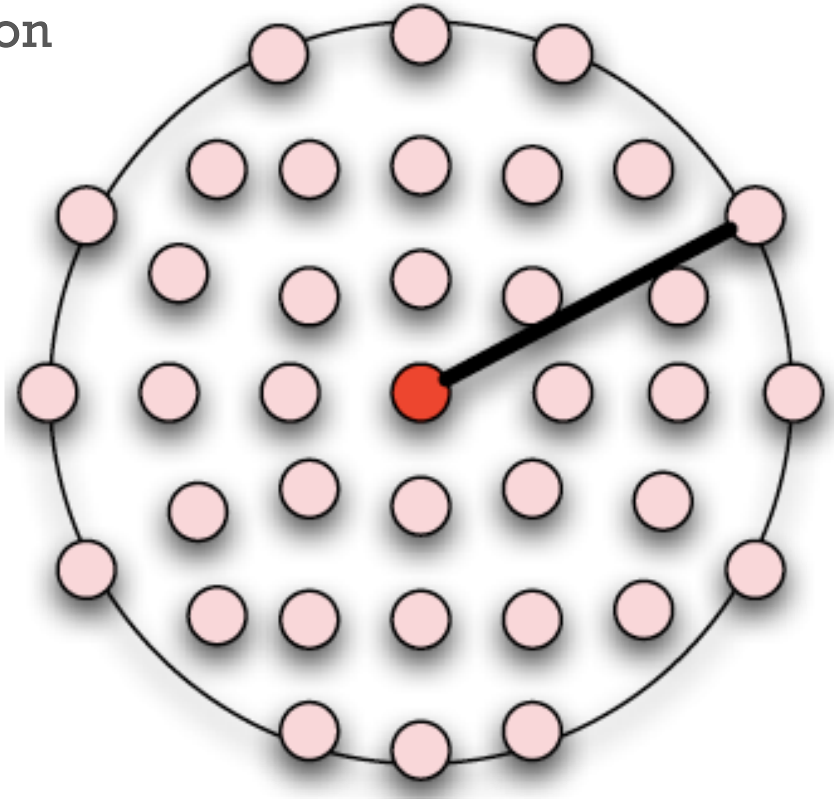


Are real-world networks Navigable?



Are real-world networks Navigable?

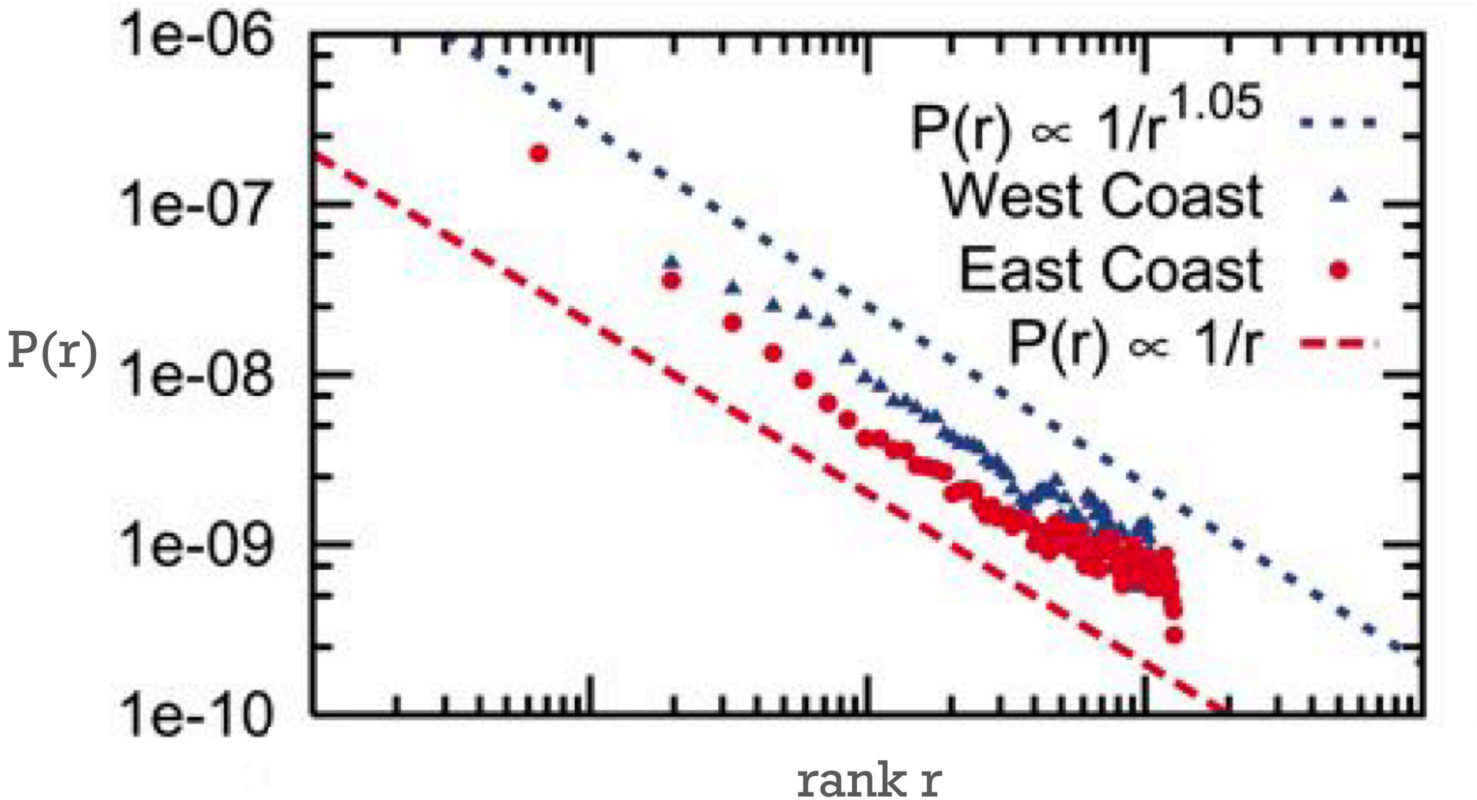
- Phase transition at $\gamma = 1$ with respect to the rank!



distance d

Are real-world networks Navigable?

- LiveJournal data:



Are real-world networks Navigable?

- Does this mean we can reach any target in a social network via decentralized search?
- Attempts to replicate Milgram's experiment have had mixed results.
- In particular, completion rates vary dramatically:
  - Highest for individuals with high social visibility, e.g., professors and journalists.
- In our models, the networks were (effectively) symmetric – this need not be the case in general!

Are real-world networks Navigable?