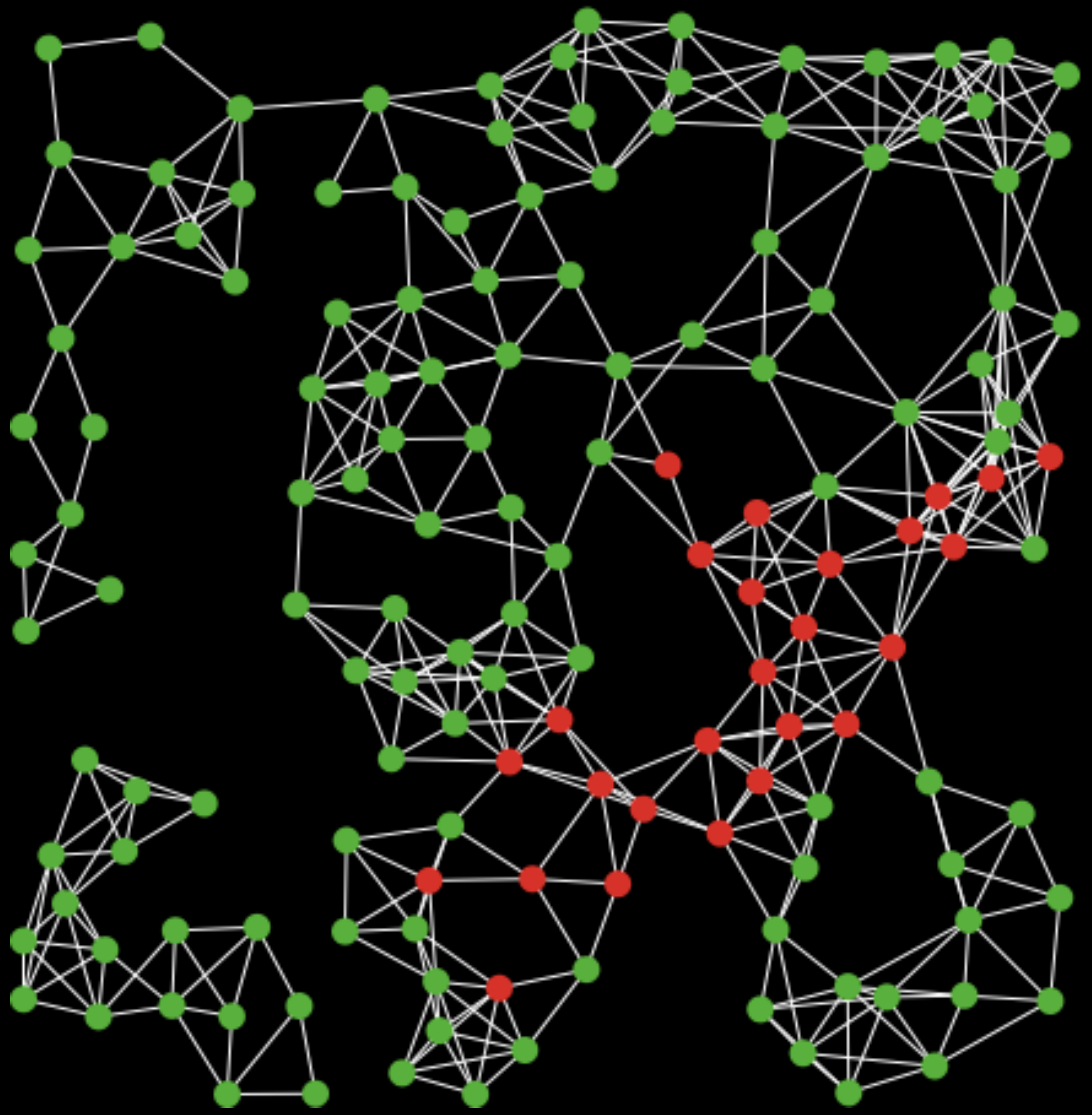
RAVI SRINIVASAN
DEPARTMENT OF STATISTICS AND DATA SCIENCES
THE UNIVERSITY OF TEXAS AT AUSTIN

Random thoughts on random walks: networks, centrality, and tracking the spread of disease

networks

- fancy word for graphs: **nodes** = vertices, **connections** = edges
- typically directed (Twitter) or undirected (Facebook)
- connectedness/giant component, irreducibility



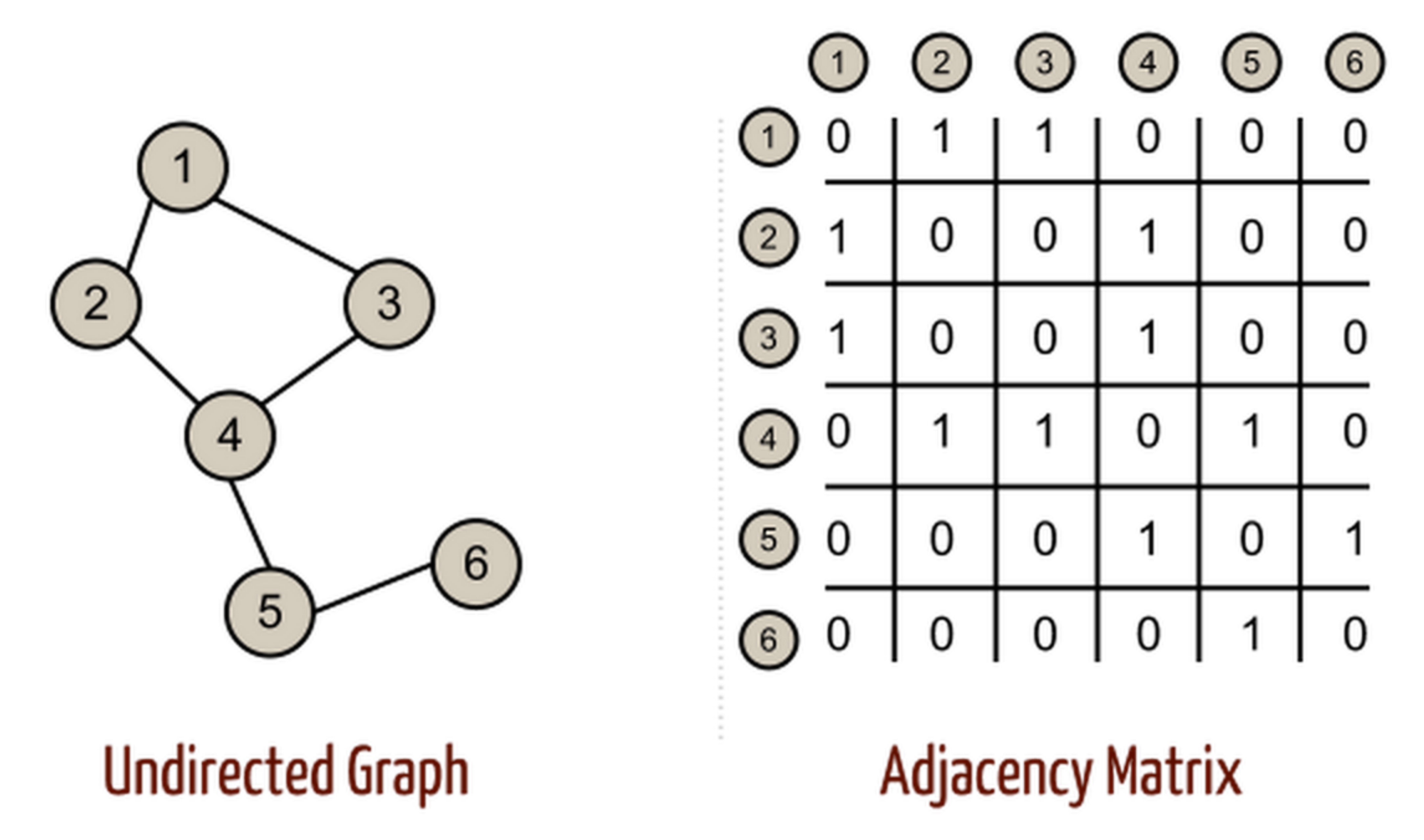
adjacency matrix

- 1 if edge from *i* to *j* (0 otherwise)
- symmetric for undirected graphs
- makes counting easier
  - degree of node *i*: $d_i = \sum_j a_{ij}$
  - # of walks of length *t* starting from node *i*:

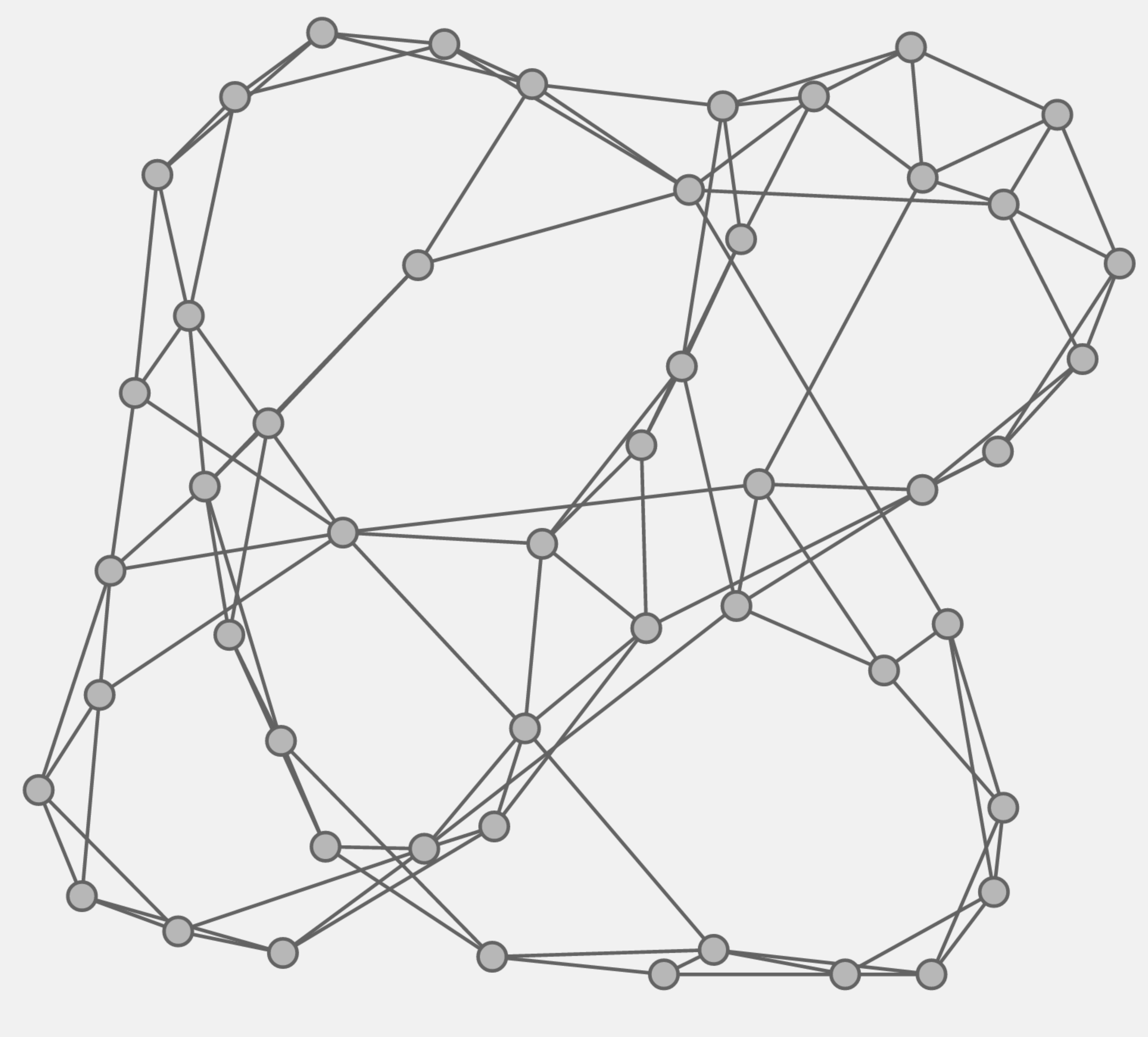
$$M_i(t) = \sum_{i_1, i_2, \ldots, i_t} a_{ii_1} a_{i_1 i_2} \cdots a_{i_{t-1} i_t}$$

  - # of 3-cycles in entire network (triple counting):

$$C^{(3)} = \sum_i (A^3)_{ii} = Tr(A^3)$$



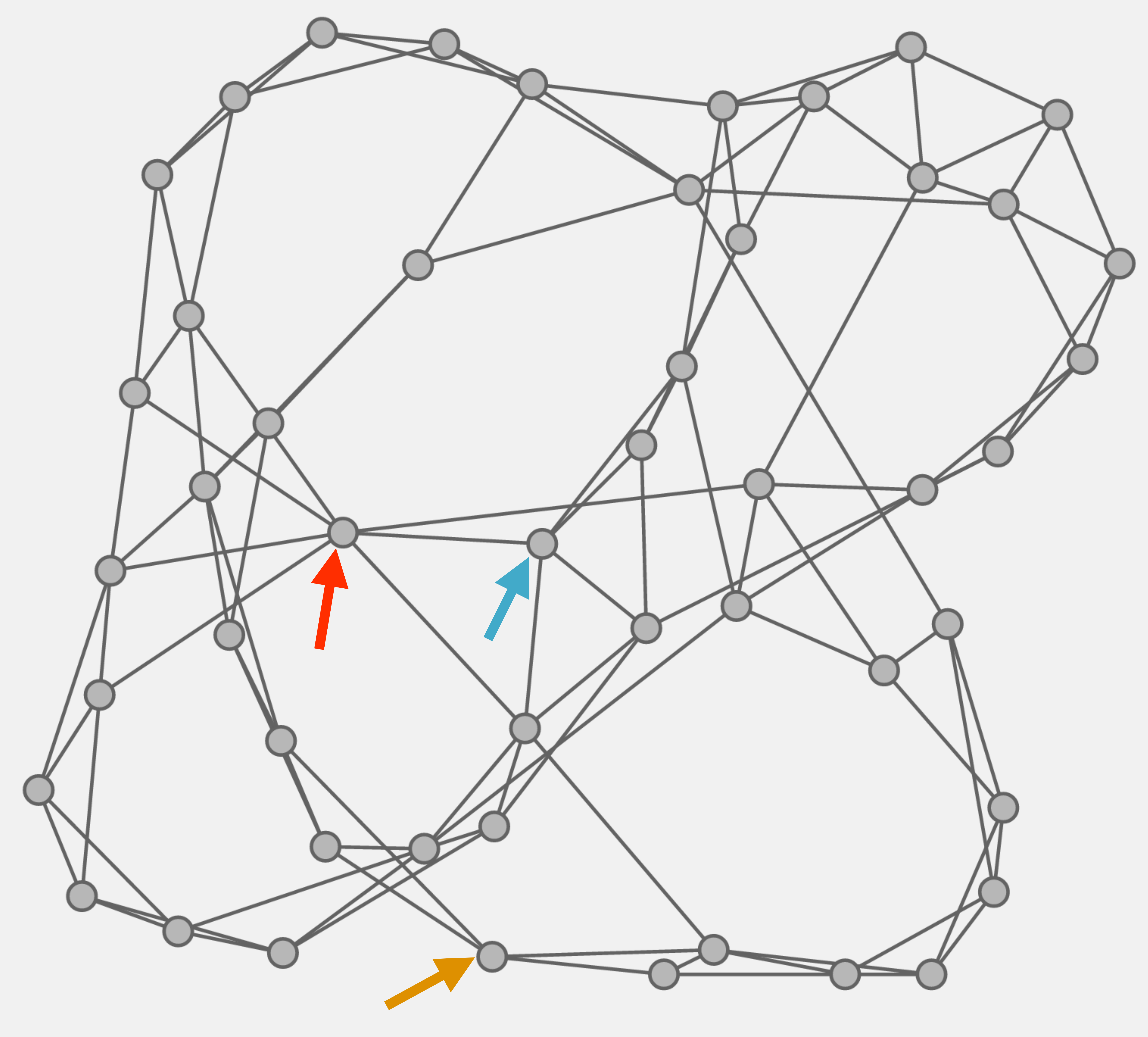
centrality



- what does it mean to be **central** within a network? context-dependent!
- create your own…

http://vax.herokuapp.com/game

centrality



- what does it mean to be **central** within a network? context-dependent!
- create your own…

http://vax.herokuapp.com/game

centrality

- {in-, out-, undirected-} **degree centrality**
  "How many {followers, followees, friends} do I have?"
- **eigenvector centrality**
  "How influential am I?"
- **Katz centrality/PageRank**
  "If my friends are important, doesn't that make me important?"
- **closeness centrality**
  "How close am I to everyone else, on average?"
- **betweenness centrality**
  "How many chains of connections include me?"
- …

degree centrality

- $c_i^{deg} = d_i / \sum_i d_i$
- normalization doesn't matter
- defined using only local information (up to scaling)

eigenvector centrality

- defined recursively/iteratively: requires global knowledge of network

$$c'_i \longleftarrow \sum_j a_{ij} c_j$$

- equivalent to dominant eigenvector of adjacency matrix (hence the name)

$$A\mathbf{v} = \kappa \mathbf{v}$$

PageRank

- similar to eigenvector centrality, but using normalized adjacency matrix

$$p_{ij} = \frac{a_{ij}}{d_i}$$

- defined using global information (as before)

$$P\pi = \pi$$

- add a damping factor to ensure everyone ends up with at least some centrality
  - i.e., interpolate original graph with complete graph

Page et. al (1999)

(a disservice to) Markov chains

- transition probabilities of going from *i* to *j*

$$\begin{array}{c} \\ s_1 \\ \vdots \\ s_r \end{array} \begin{array}{c} \begin{array}{ccc} s_1 & \cdots & s_r \end{array} \\ \begin{pmatrix} p_{11} & \cdots & p_{1r} \\ \vdots & & \vdots \\ p_{r1} & \cdots & p_{rr} \end{pmatrix} \end{array}$$

- stationary distribution satisfies $\sum_i \pi_i p_{ij} = \pi_j$ (i.e., $\pi P = \pi$)
- rate of convergence given by second largest eigenvalue of transition matrix
- example: simple random walk on finite set of states

PageRank + random walks

- interpretation: proportion of time random walker spends at each node, assuming at each step neighbor selected at random
  - i.e., stationary distribution of Markov process on graph with uniform transition probabilities
  - technical details: irreducibility, aperiodicity, ergodicity
  - damping factor ensures these properties hold
- note for later: only local information used at each step in walk…

degree centrality + random walks

- for undirected networks, PageRank (w/o damping) is the simple random walk
- has degree centrality as its stationary distribution if network is connected
  - check this yourself! show that

$$\sum_j p_{ij} \left(\frac{d_j}{\sum_i d_i} \right) = \frac{d_i}{\sum_i d_i}$$

degree centrality + random walks

- moments of degree distribution: $\langle d^m \rangle = \frac{1}{N} \sum_i d_i^m = \sum_d d^m P(d)$
- if we sample a node uniformly at random:

$$P(X = i) = 1/N \Longrightarrow E(deg(X)) = \langle d \rangle$$

- instead, after performing random walk "long enough", we are sampling nodes according to their degree centralities:

$$P(X = i) = d_i / \sum_i d_i \qquad \Longrightarrow \qquad E(deg(X)) = \frac{\langle d^2 \rangle}{\langle d \rangle} = \langle d \rangle + \frac{Var(d)}{\langle d \rangle}$$

- intuition: In the limit, every edge is traversed same proportion of times. So, we are sampling a node at the end of an *edge* chosen uniformly at random!

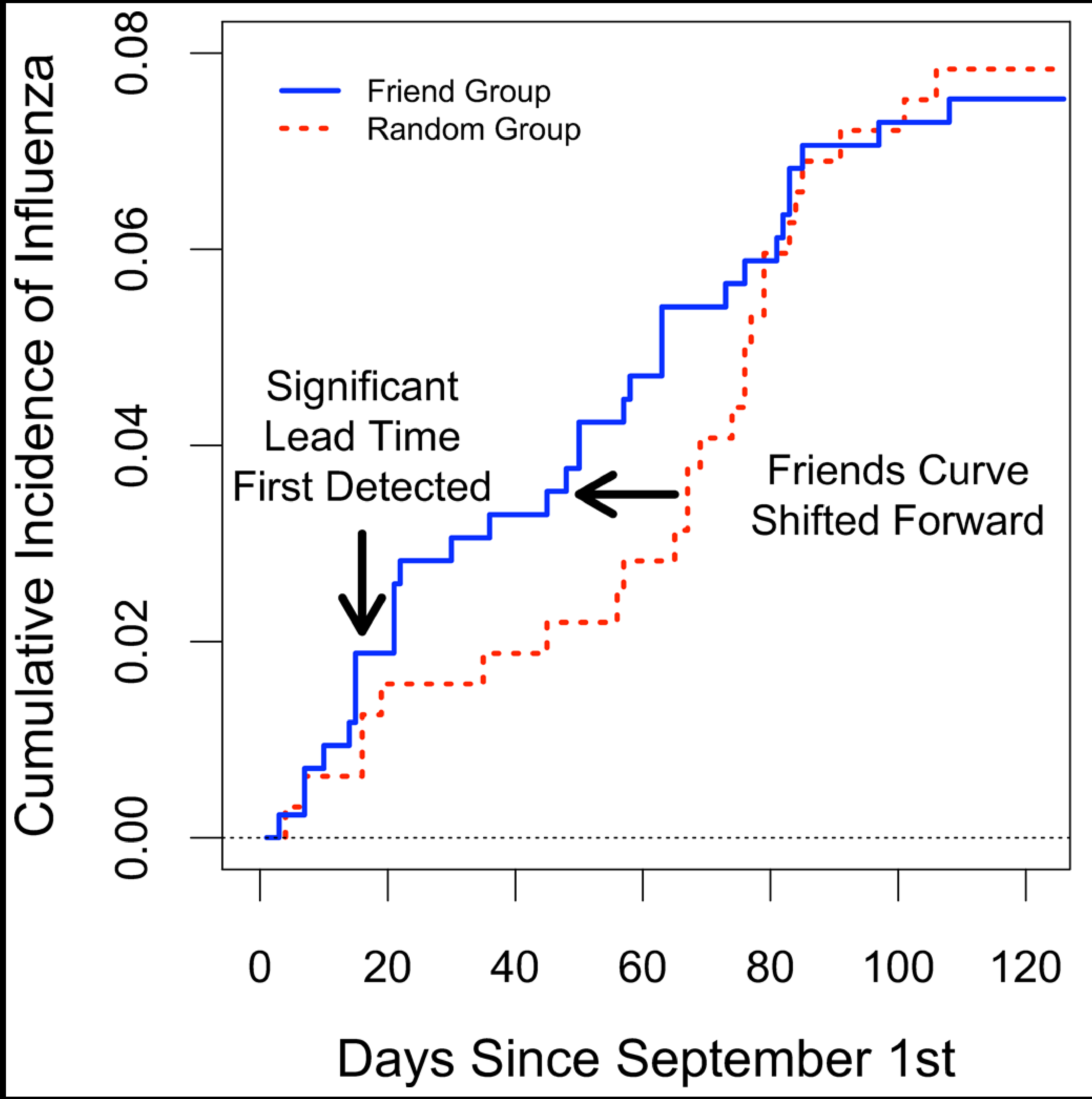
Lovasz (1993)

sampling using random walks

- constructive: If we want to sample according to degree centrality, take any node and select neighbor uniformly at random… rinse, repeat…
- principle behind Markov Chain Monte Carlo (MCMC)
  - provides way to generate samples from distribution that can’t be sampled directly
  - need to specify transition probabilities such that stationary distribution of Markov chain is the desired sampling distribution

real-world implications

- Christakis & Fowler (2010)
  - flu outbreak among Harvard student population from Sep.-Dec. '09
  - 2009 H1N1 pandemic: ~60M infected in US, ~300K hospitalizations
  - two groups: random vs. *friends* of random (FoR)
  - epidemic curve in FoR tracked progression in random group ~14d in advance
  - deal with it: on average, your friends have more friends than you

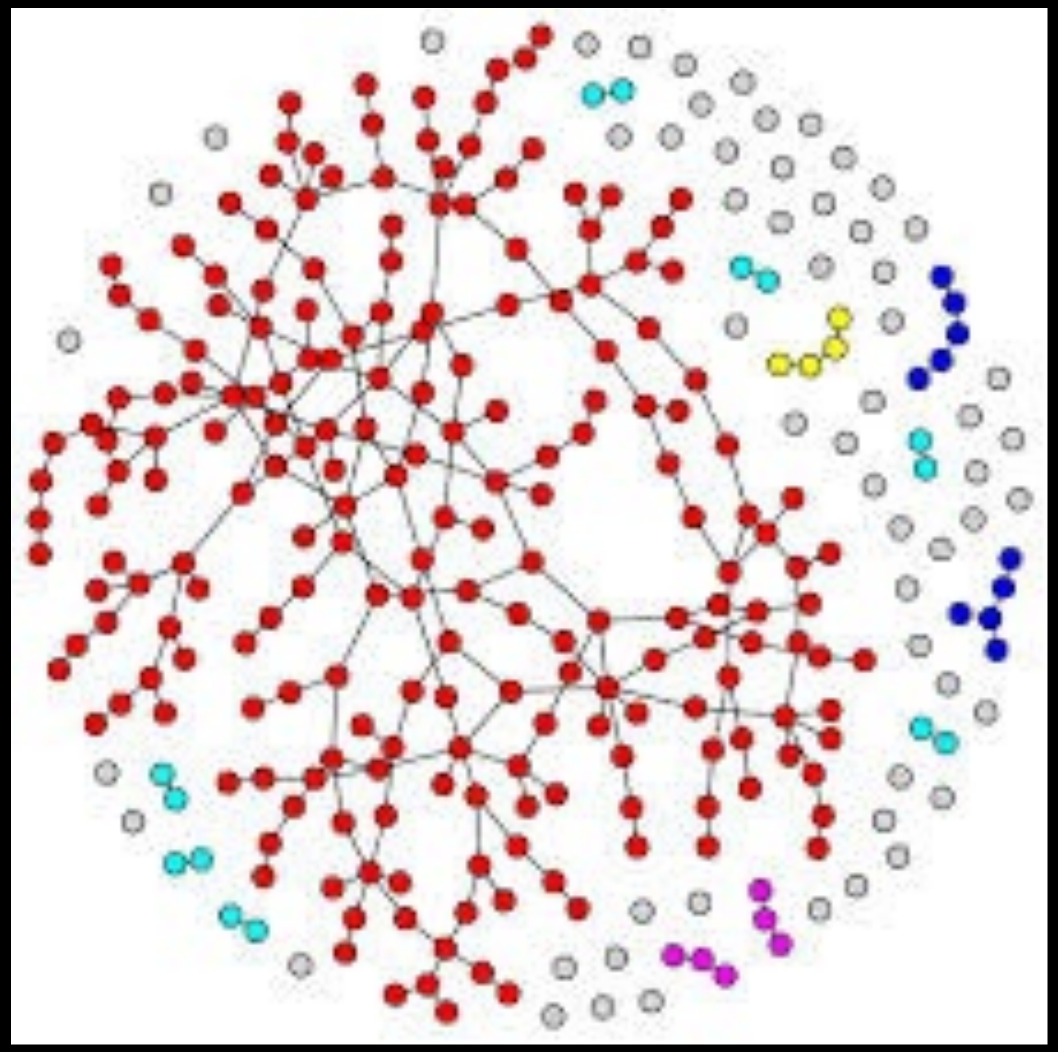


Strogatz, New York Times (2012)

an aside

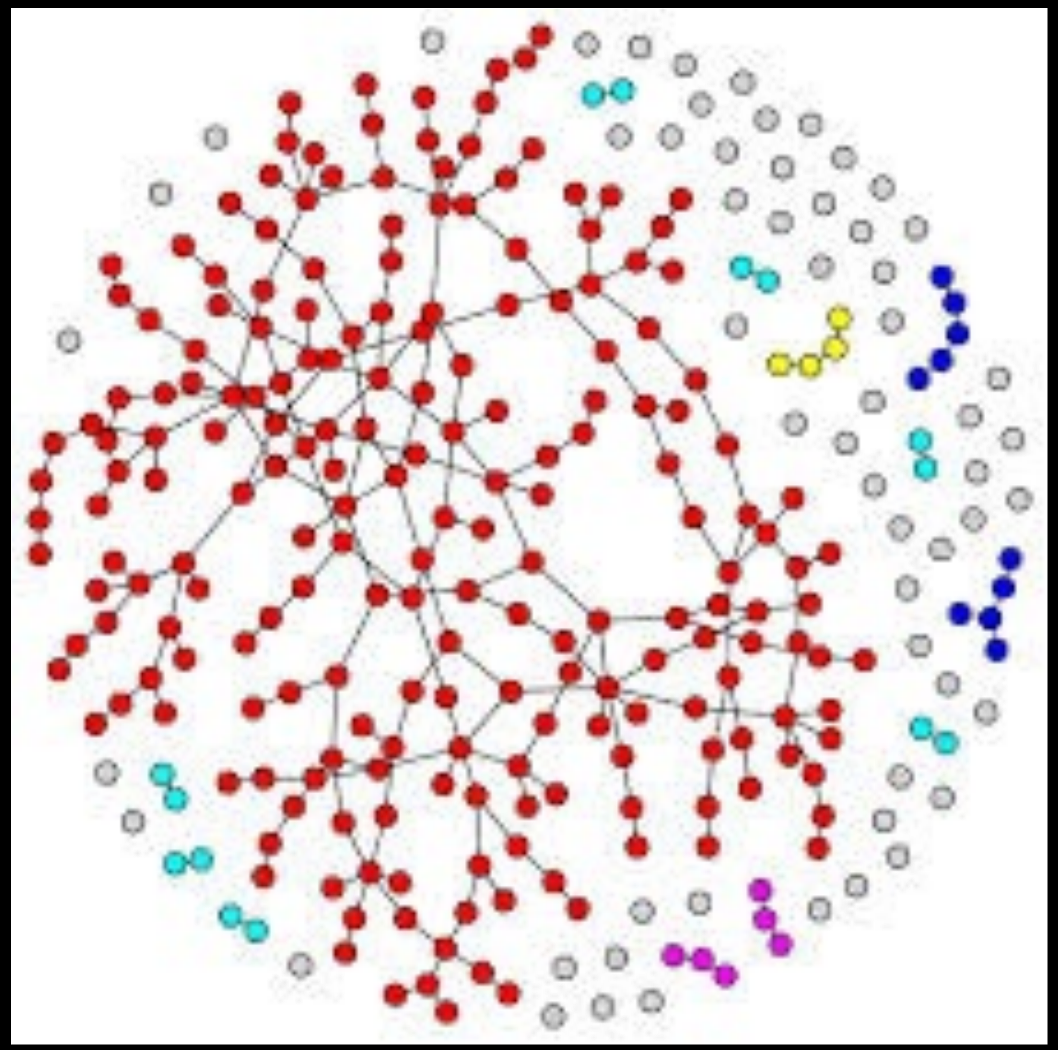
- long history of using (random) walks to infer properties of the social graph
- Milgram’s small-world experiment (1967)
  - Watts-Strogatz model: shortcuts in highly clustered networks
  - Kleinberg model: distribution of shortcut lengths and efficient routing

an aside



Erdos-Renyi (tree-like)

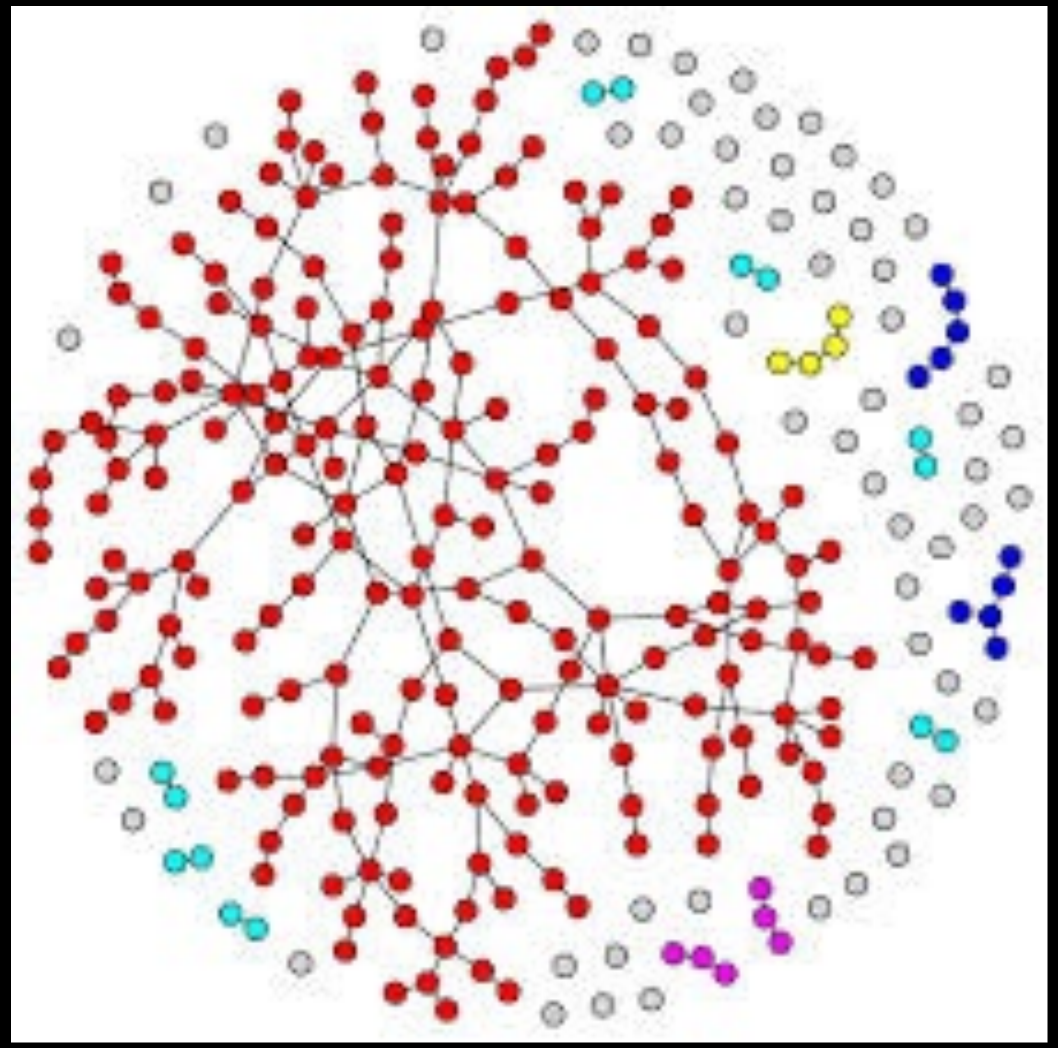
an aside



Erdos-Renyi (tree-like)

small diameter, no clustering

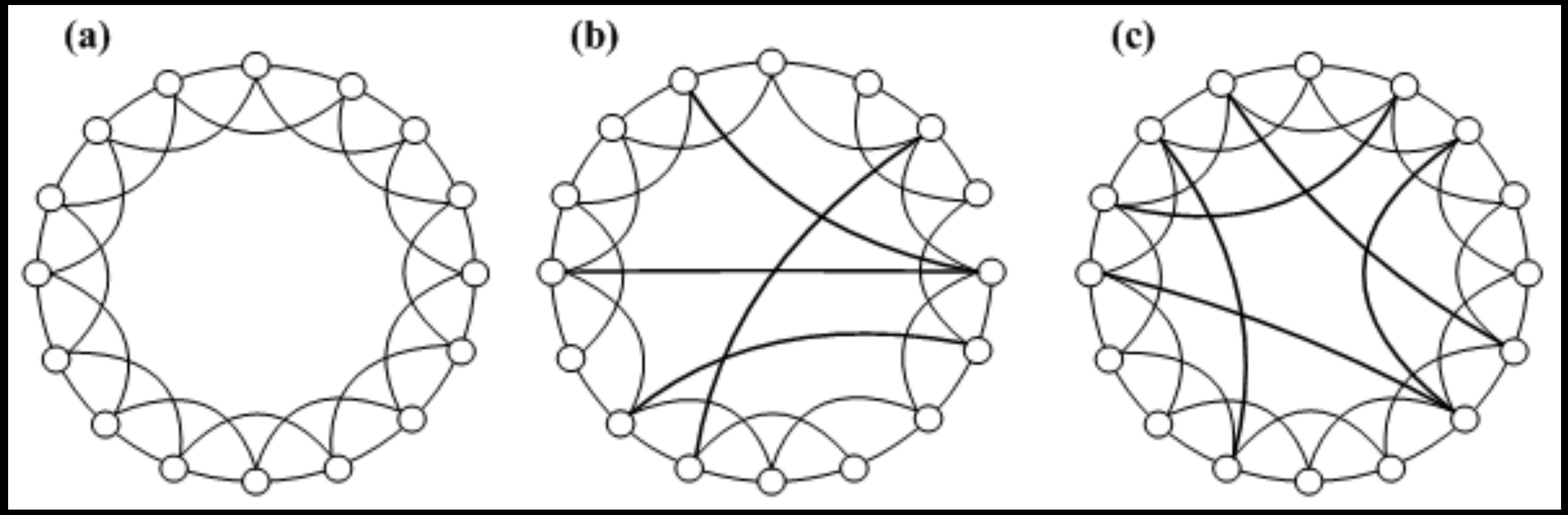
an aside



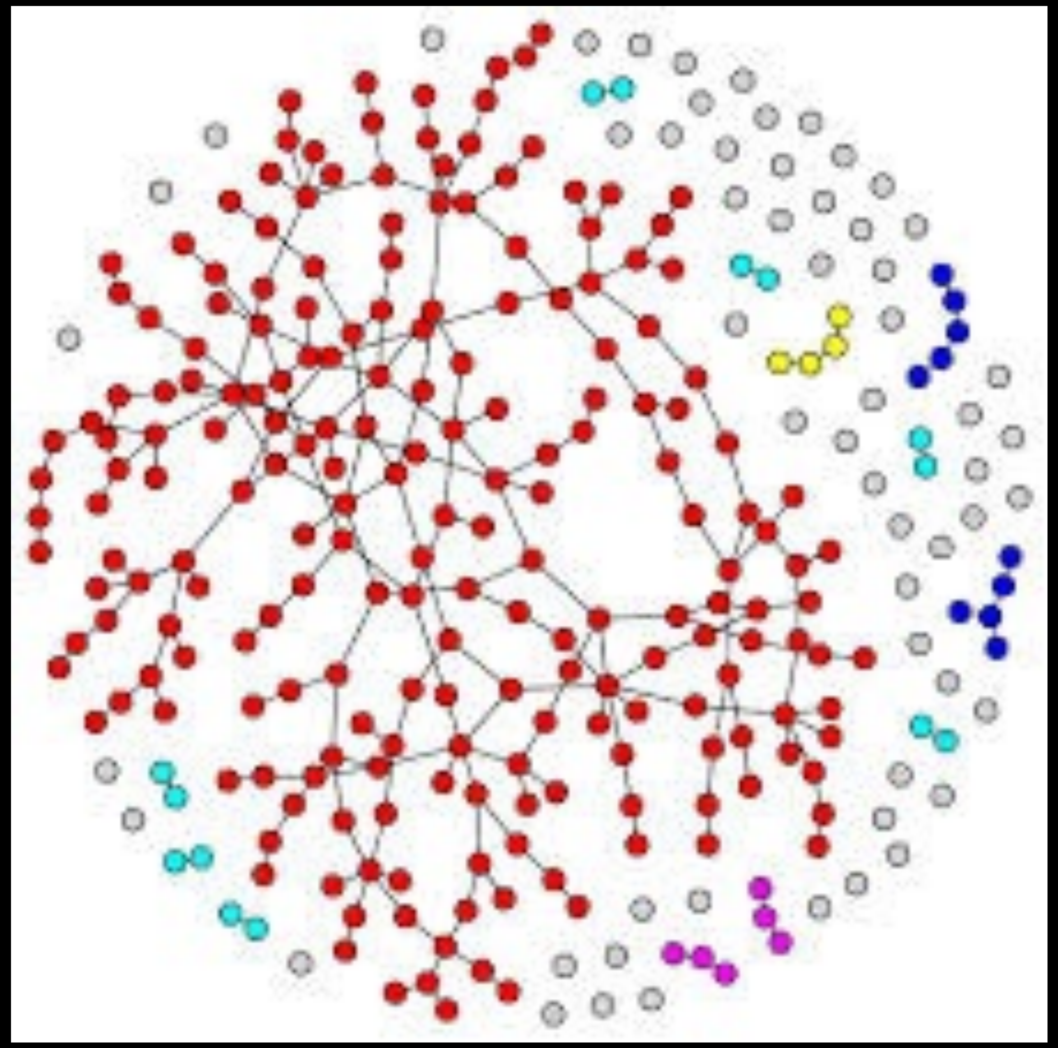
Erdos-Renyi (tree-like)

small diameter, no clustering

Watts-Strogatz (lattice + shortcuts)



an aside

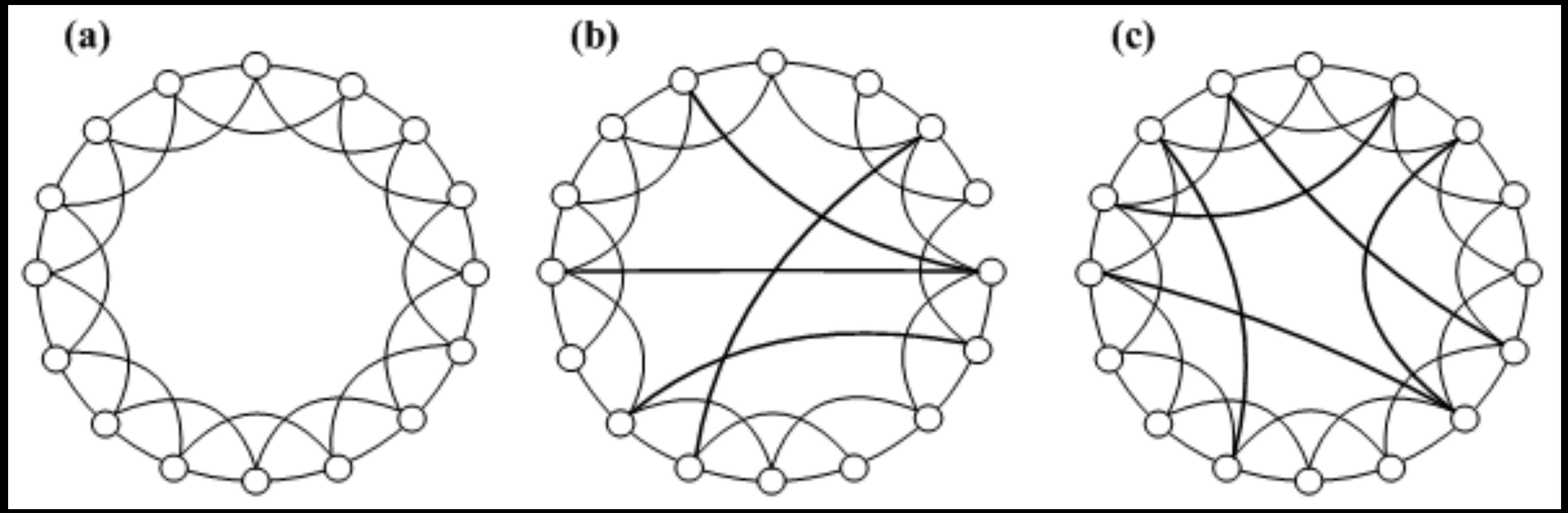


Erdos-Renyi (tree-like)

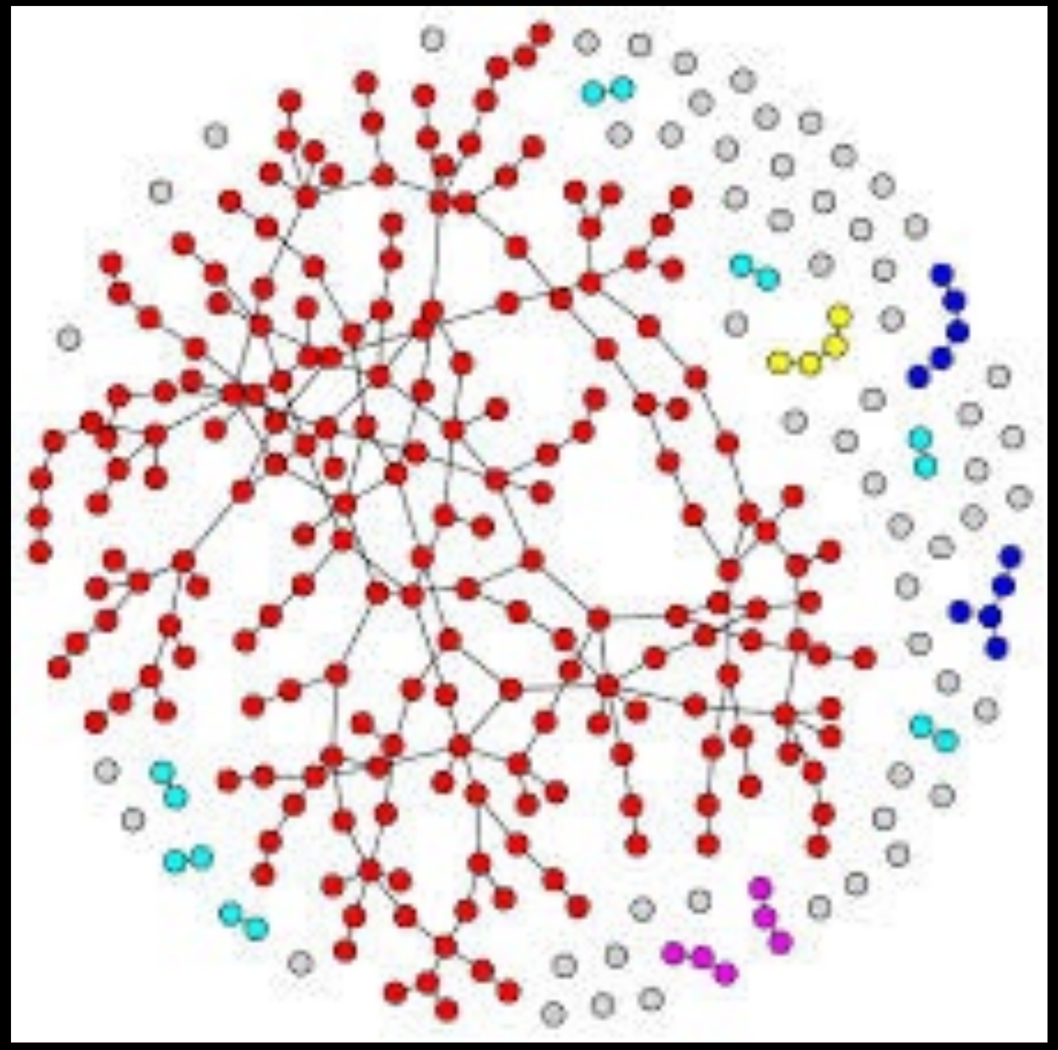
small diameter, no clustering

small diameter, high clustering

Watts-Strogatz (lattice + shortcuts)



an aside

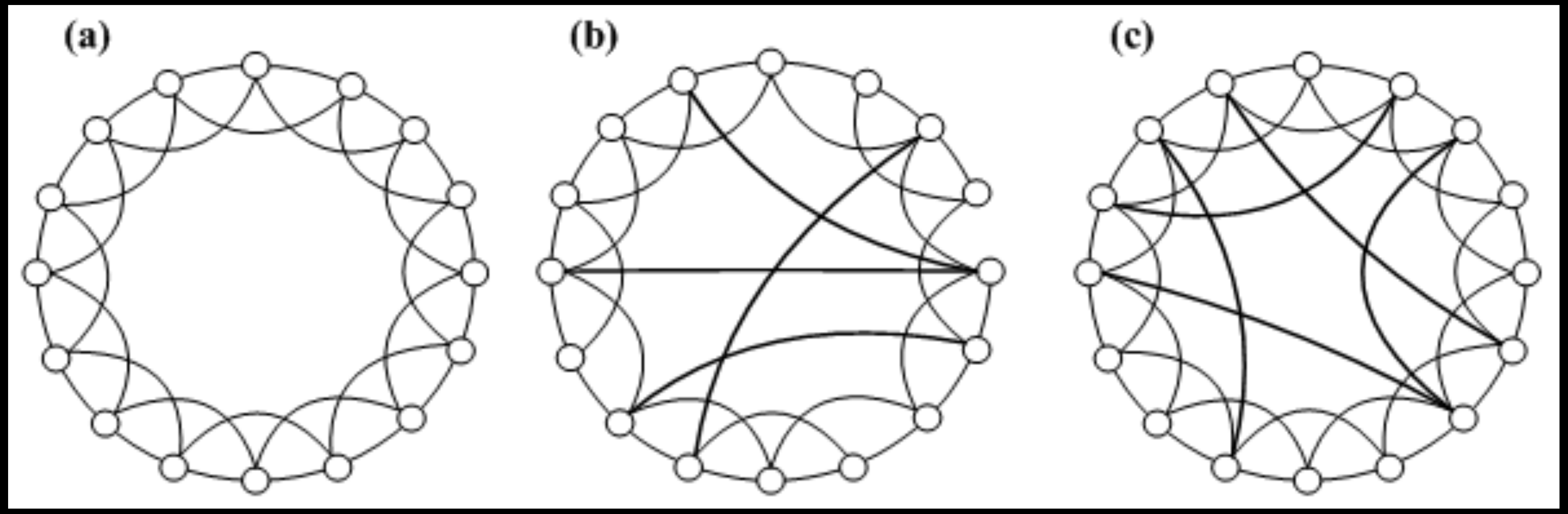


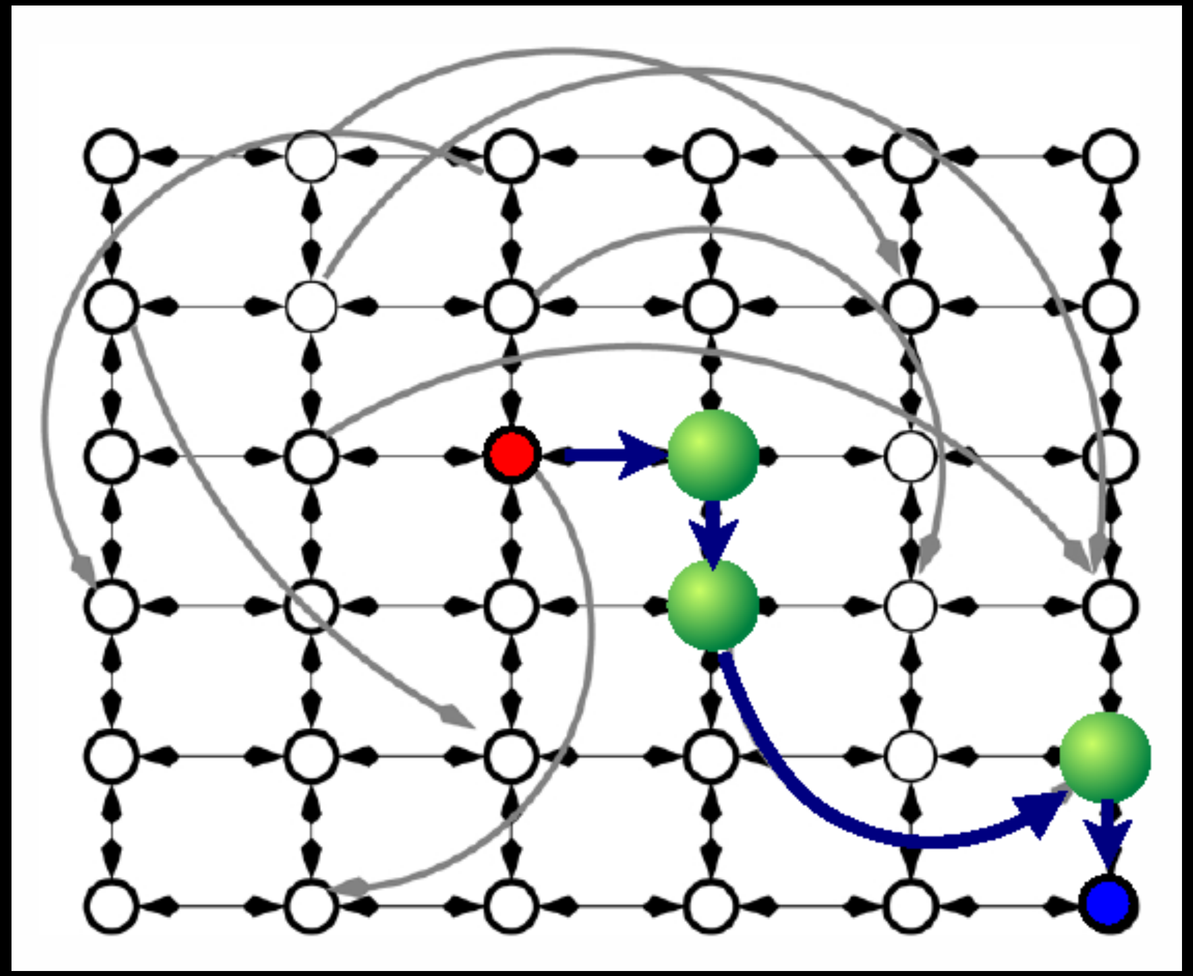
Erdos-Renyi (tree-like)

small diameter, no clustering

small diameter, high clustering

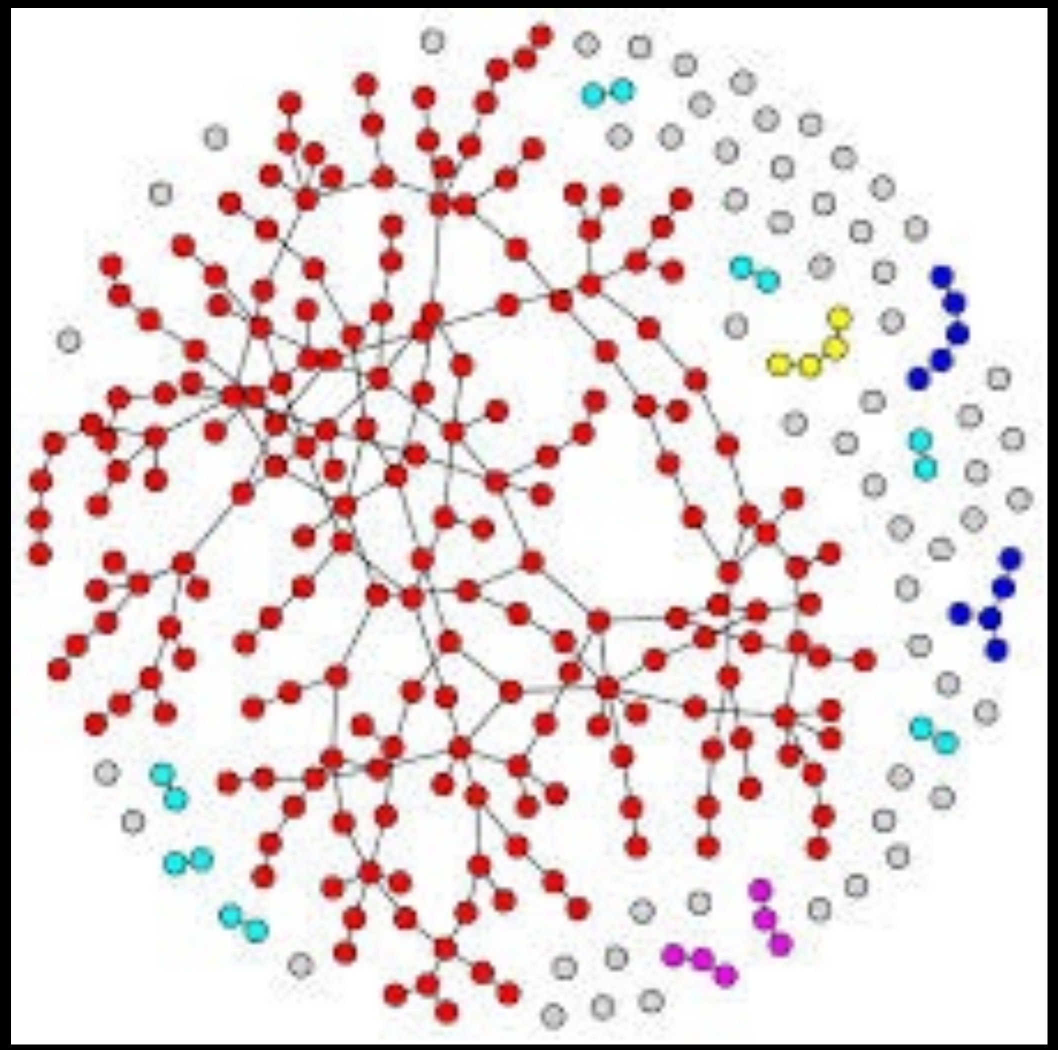
Watts-Strogatz (lattice + shortcuts)





Kleinberg (distribution of shortcuts)

an aside

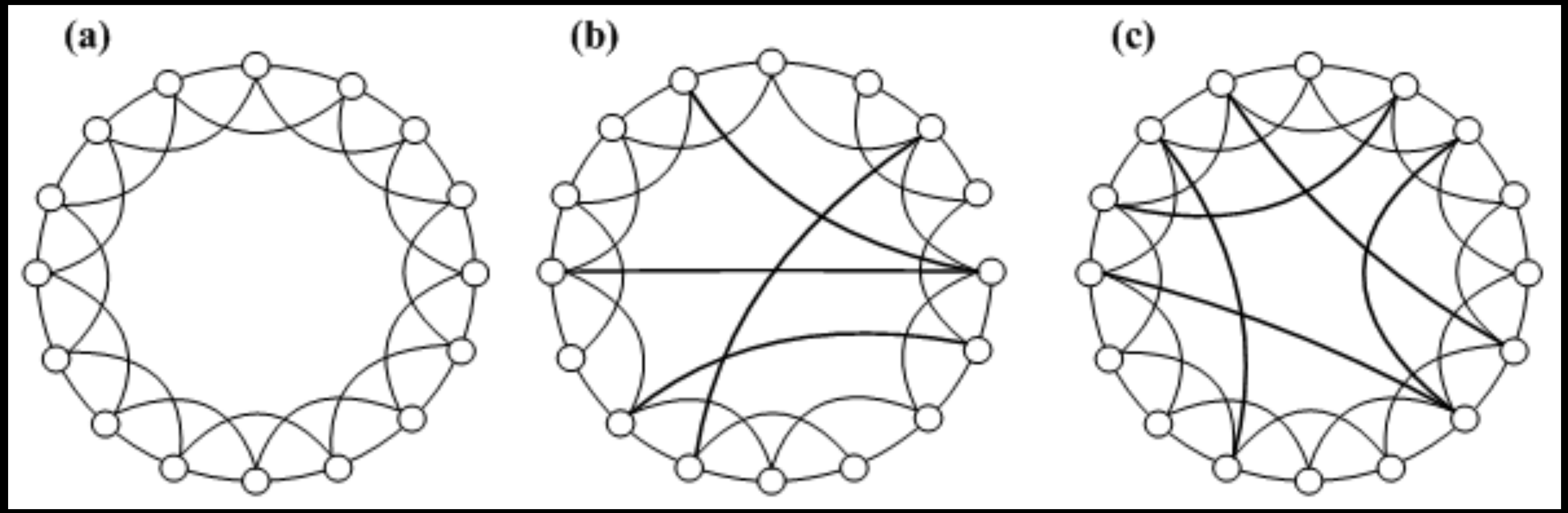


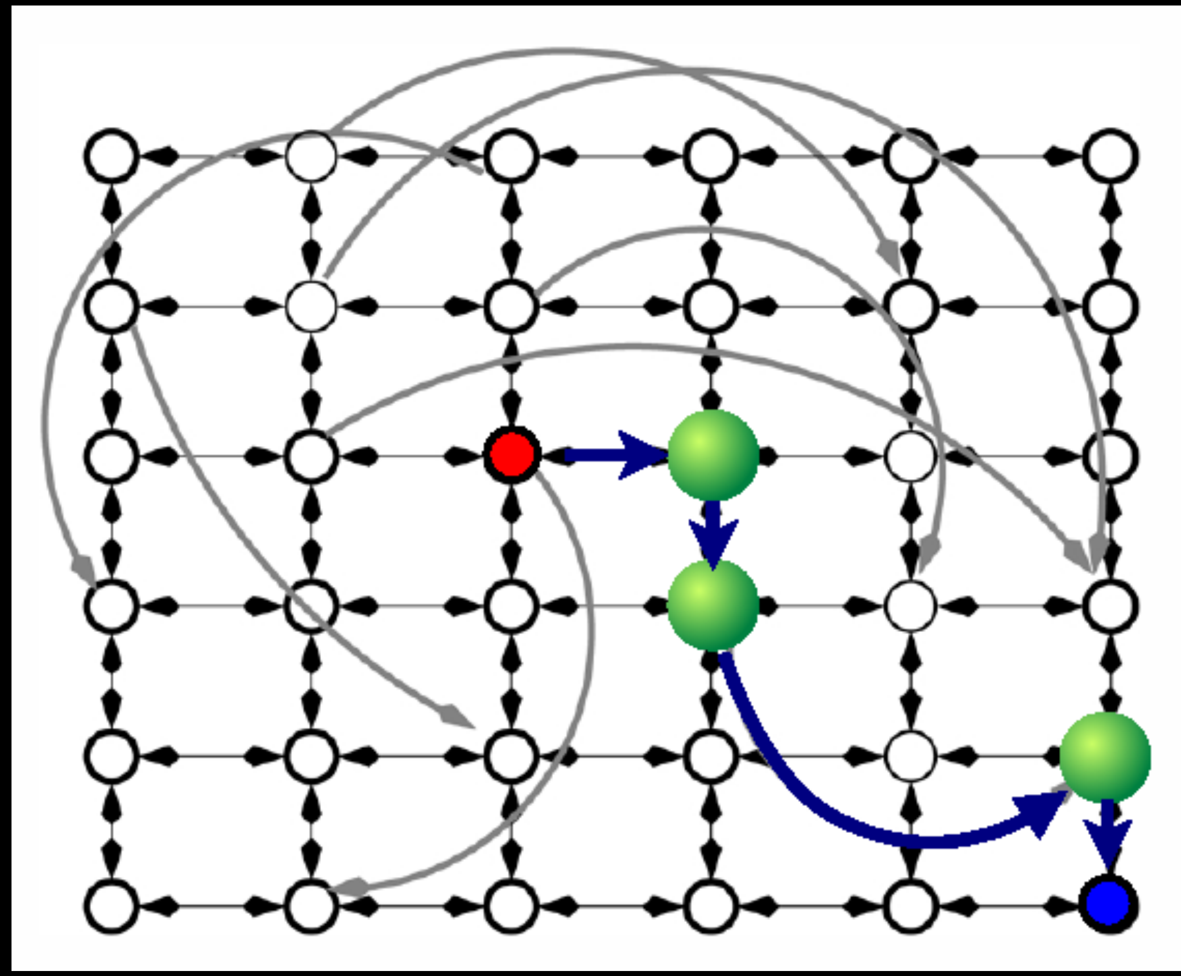
Erdos-Renyi (tree-like)

small diameter, no clustering

small diameter, high clustering

Watts-Strogatz (lattice + shortcuts)

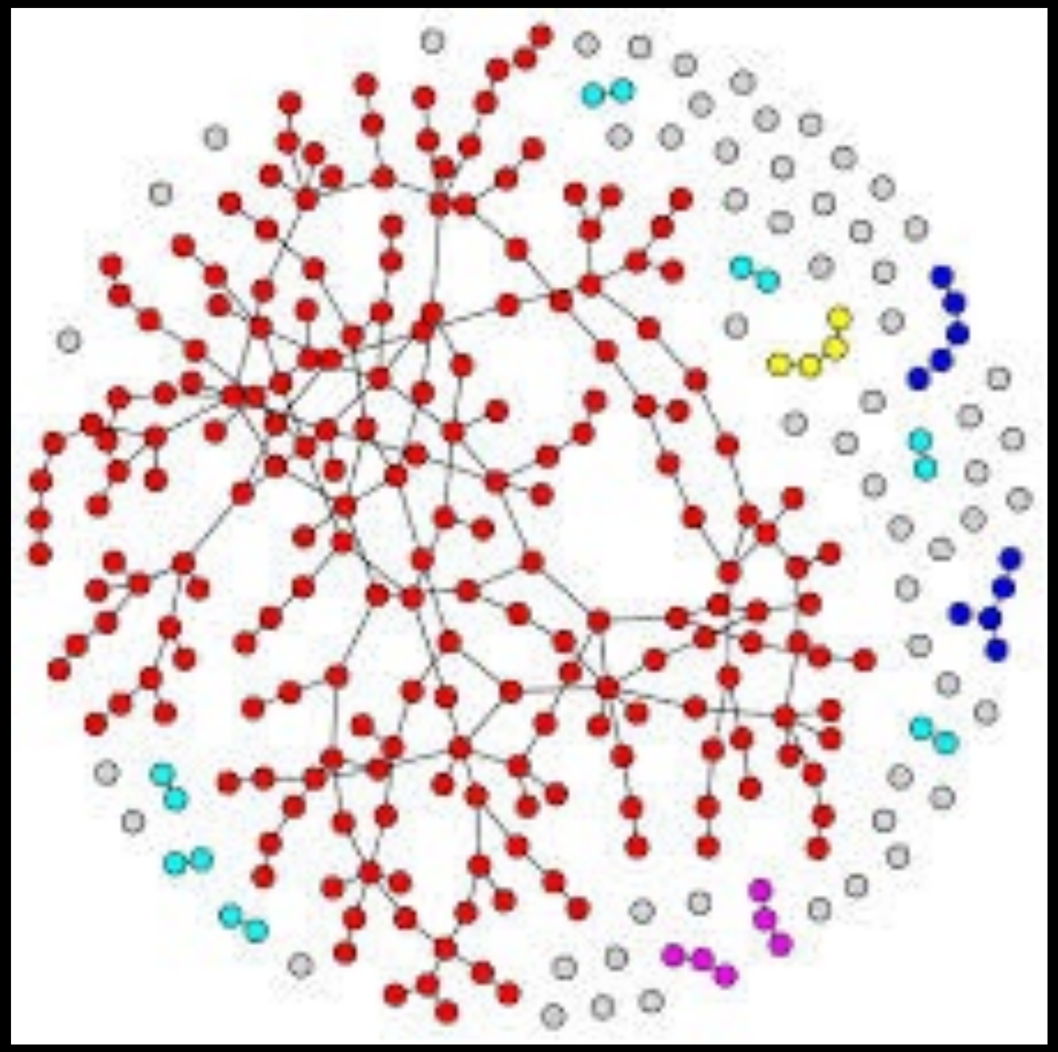




Kleinberg (distribution of shortcuts)

greedy routing finds shortcuts!

an aside

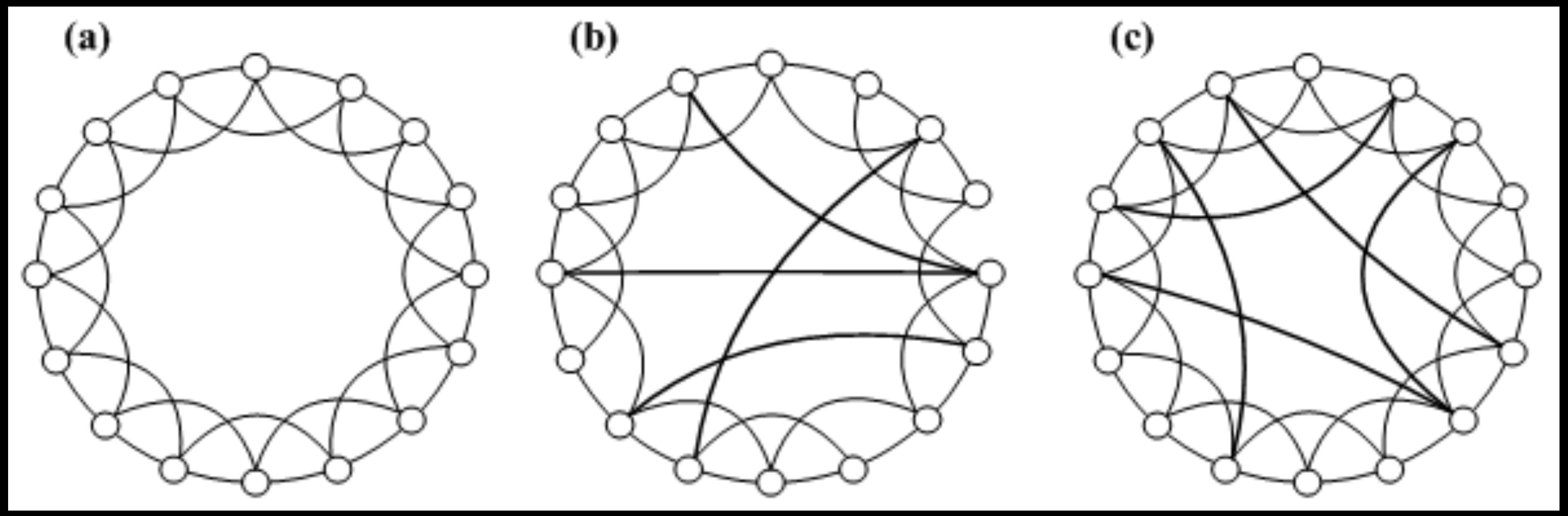


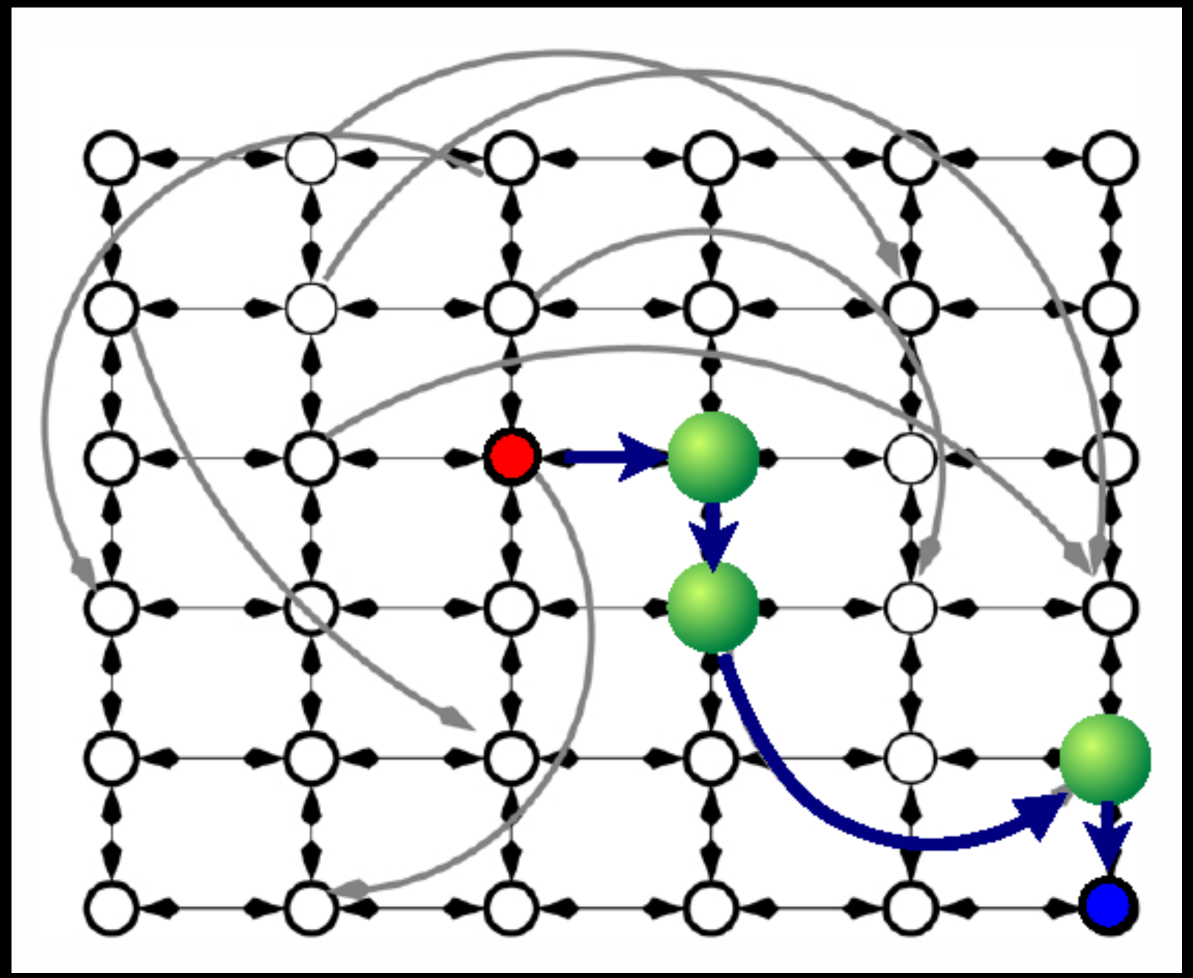
Erdos-Renyi (tree-like)

small diameter, no clustering

small diameter, high clustering

Watts-Strogatz (lattice + shortcuts)





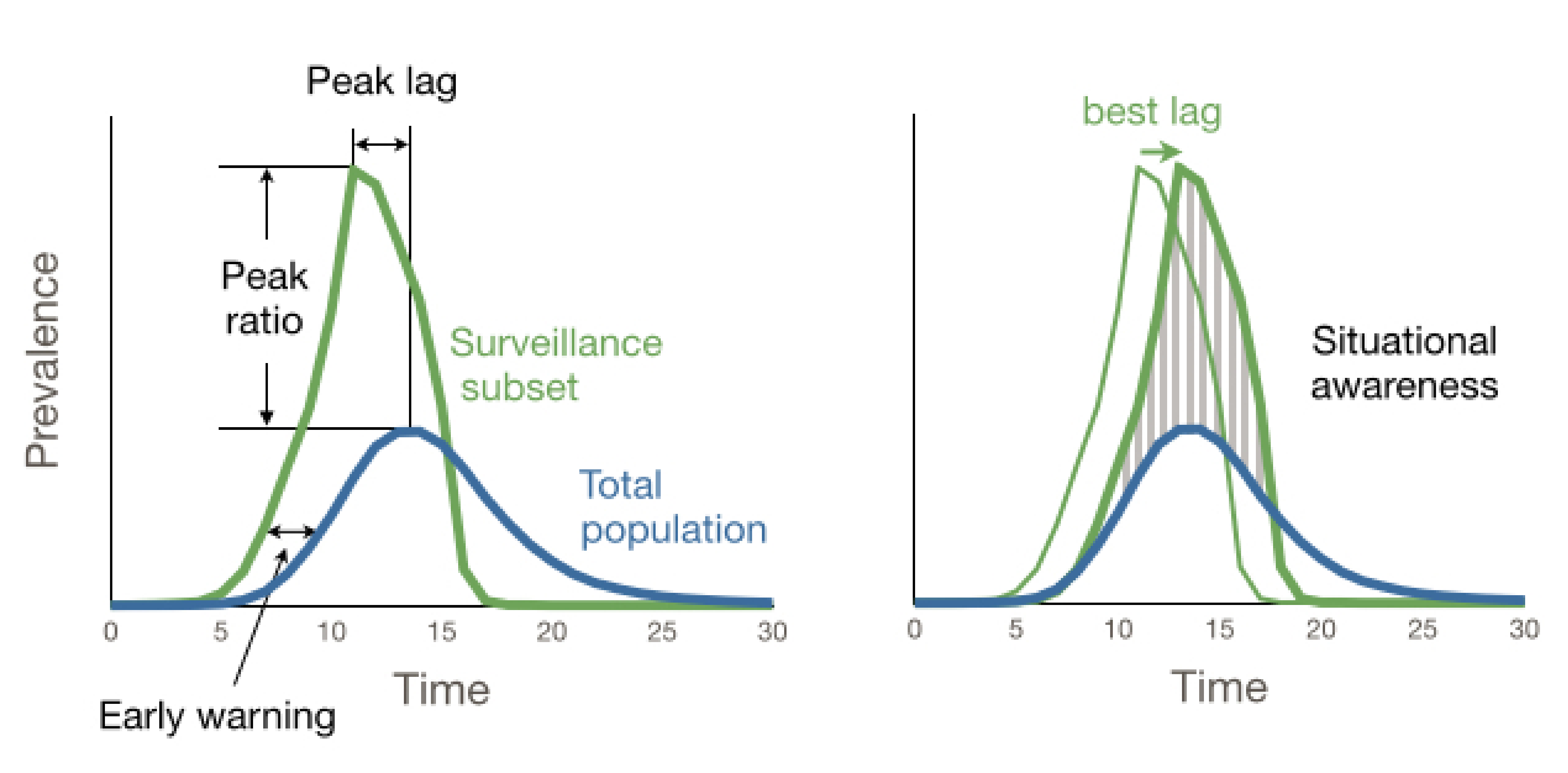
Kleinberg (distribution of shortcuts)

greedy routing finds shortcuts!

Watts-Strogatz (1998); Kleinberg (2000)

disease surveillance

- what do public health officials and CDC care about?
  - situational awareness
  - early detection of epidemic onset
  - peak timing and intensity
- practical reasons
  - vaccine supply and distribution
  - allowing hospitals to run at high capacity and prepare for large influx of patients
- difficulties both mathematical/statistical and practical



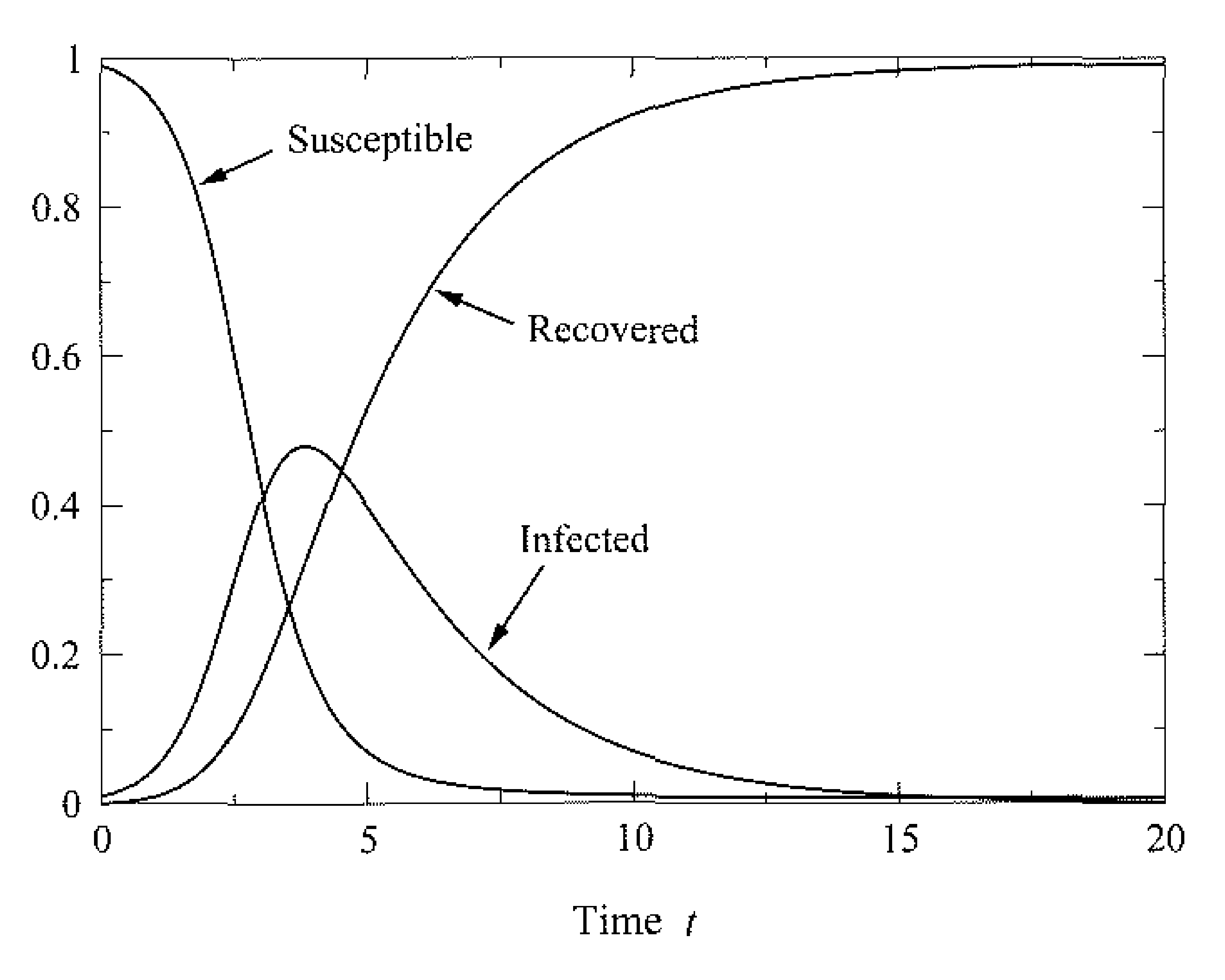
let's do a little math

- SIR model on a graph (recall Andy's talk!)

$$\frac{ds_i}{dt} = -\beta \sum_j a_{ij} s_i x_j$$

$$\frac{dx_i}{dt} = \beta \sum_j a_{ij} s_i x_j - \gamma x_i$$

$$\frac{dr_i}{dt} = \gamma x_i$$



Newman (book); message passing: Karrer-Newman (2013)

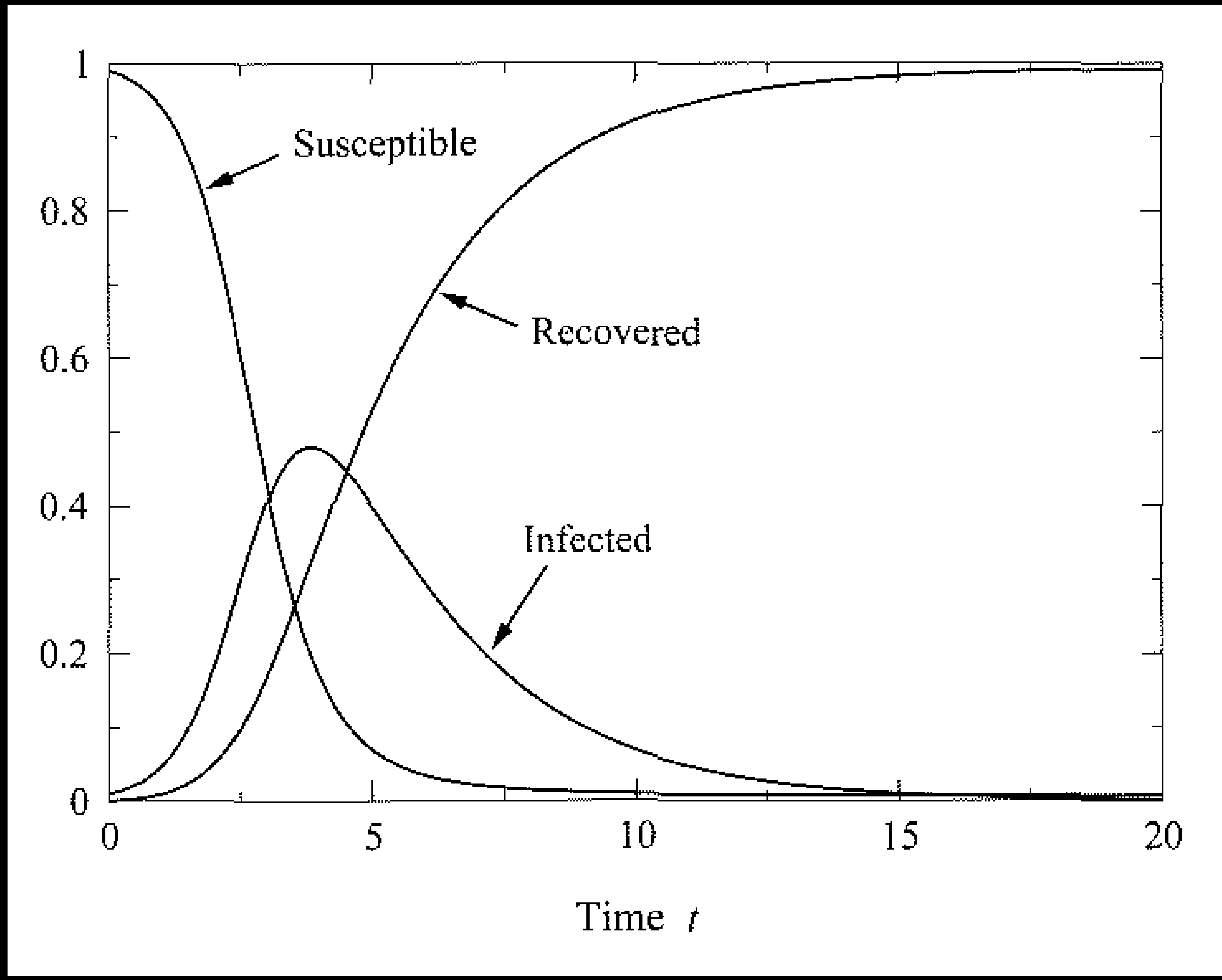
let's do a little math

- SIR model on a graph (recall Andy's talk!)

$$\frac{ds_i}{dt} = -\beta \sum_j a_{ij} s_i x_j$$

$$\frac{dx_i}{dt} = \beta \sum_j a_{ij} s_i x_j - \gamma x_i$$

$$\frac{dr_i}{dt} = \gamma x_i$$



Newman (book); message passing: Karrer-Newman (2013)

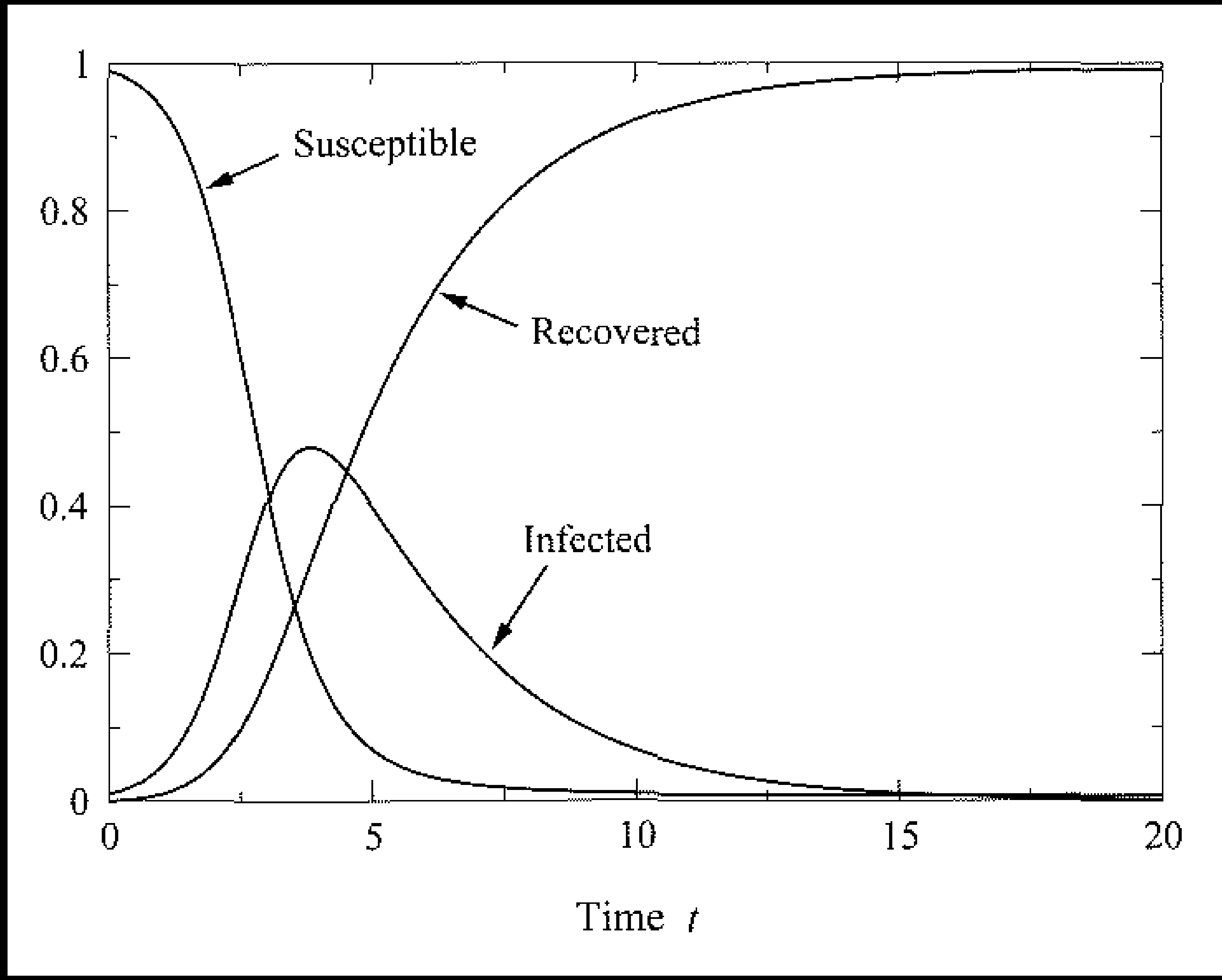
let's do a little math

- SIR model on a graph (recall Andy's talk!)

$$\frac{ds_i}{dt} = -\beta \sum_j a_{ij} \cancel{s_i} x_j$$

$$\frac{dx_i}{dt} = \beta \sum_j a_{ij} \cancel{s_i} x_j - \gamma x_i$$

$$\frac{dr_i}{dt} = \gamma x_i$$



for small times, P(*i* is S) is approximately 1

Newman (book); message passing: Karrer-Newman (2013)

linearized SIR on graph

$$\frac{d\mathbf{x}}{dt} = \beta A\mathbf{x} - \gamma\mathbf{x}$$

linearized SIR on graph

$$\frac{d\mathbf{x}}{dt} = \beta A\mathbf{x} - \gamma\mathbf{x}$$

dominant e-value of *A*

corresponding e-vector

$$\mathbf{x}(t) = \exp(t(A - \gamma I))\mathbf{x}(0) \approx e^{(\beta\kappa - \gamma)t}\mathbf{v}$$

linearized SIR on graph

$$\frac{d\mathbf{x}}{dt} = \beta A\mathbf{x} - \gamma\mathbf{x}$$

dominant e-value of *A*

corresponding e-vector

$$\mathbf{x}(t) = \exp(t(A - \gamma I))\mathbf{x}(0) \approx e^{(\beta\kappa - \gamma)t}\mathbf{v}$$

epidemic takes off if $R_0 = \frac{\beta}{\gamma} > \frac{1}{\kappa}$

early detection + eigenvector centrality

early detection + eigenvector centrality

dominant e-value of *A*

corresponding e-vector

$$\mathbf{x}(t) = \exp(t(A - \gamma I))\mathbf{x}(0) \approx e^{(\beta\kappa - \gamma)t}\mathbf{v}$$

early detection + eigenvector centrality

dominant e-value of A

corresponding e-vector

$$\mathbf{x}(t) = \exp(t(A - \gamma I))\mathbf{x}(0) \approx e^{(\beta\kappa - \gamma)t}\mathbf{v}$$

$$p = \frac{\mathbf{x}(\tau_S) \cdot \mathbf{1}_S}{M} = \frac{e^{(\beta\kappa-\gamma)\tau_S}\mathbf{v} \cdot \mathbf{1}_S}{M}$$

expected prevalence hits p in surveillance set

$$p = \frac{\mathbf{x}(\tau) \cdot \mathbf{1}}{N} = \frac{e^{(\beta\kappa-\gamma)\tau}\mathbf{v} \cdot \mathbf{1}}{N}$$

expected prevalence hits p in entire network

early detection + eigenvector centrality

dominant e-value of *A*

corresponding e-vector

$$\mathbf{x}(t) = \exp(t(A - \gamma I))\mathbf{x}(0) \approx e^{(\beta\kappa - \gamma)t}\mathbf{v}$$

$$p = \frac{\mathbf{x}(\tau_S) \cdot \mathbf{1}_S}{M} = \frac{e^{(\beta\kappa-\gamma)\tau_S}\mathbf{v} \cdot \mathbf{1}_S}{M}$$

expected prevalence hits p in surveillance set

$$p = \frac{\mathbf{x}(\tau) \cdot \mathbf{1}}{N} = \frac{e^{(\beta\kappa-\gamma)\tau}\mathbf{v} \cdot \mathbf{1}}{N}$$

expected prevalence hits p in entire network

$$\Delta\tau := \tau_S - \tau = \frac{1}{\beta\kappa - \gamma}\ln\left(\frac{c}{c_S}\right)$$

early detection + eigenvector centrality

dominant e-value of A

corresponding e-vector

$$\mathbf{x}(t) = \exp(t(A - \gamma I))\mathbf{x}(0) \approx e^{(\beta\kappa - \gamma)t}\mathbf{v}$$

$$p = \frac{\mathbf{x}(\tau_S) \cdot \mathbf{1}_S}{M} = \frac{e^{(\beta\kappa-\gamma)\tau_S}\mathbf{v} \cdot \mathbf{1}_S}{M}$$

expected prevalence hits p in surveillance set

$$p = \frac{\mathbf{x}(\tau) \cdot \mathbf{1}}{N} = \frac{e^{(\beta\kappa-\gamma)\tau}\mathbf{v} \cdot \mathbf{1}}{N}$$

expected prevalence hits p in entire network

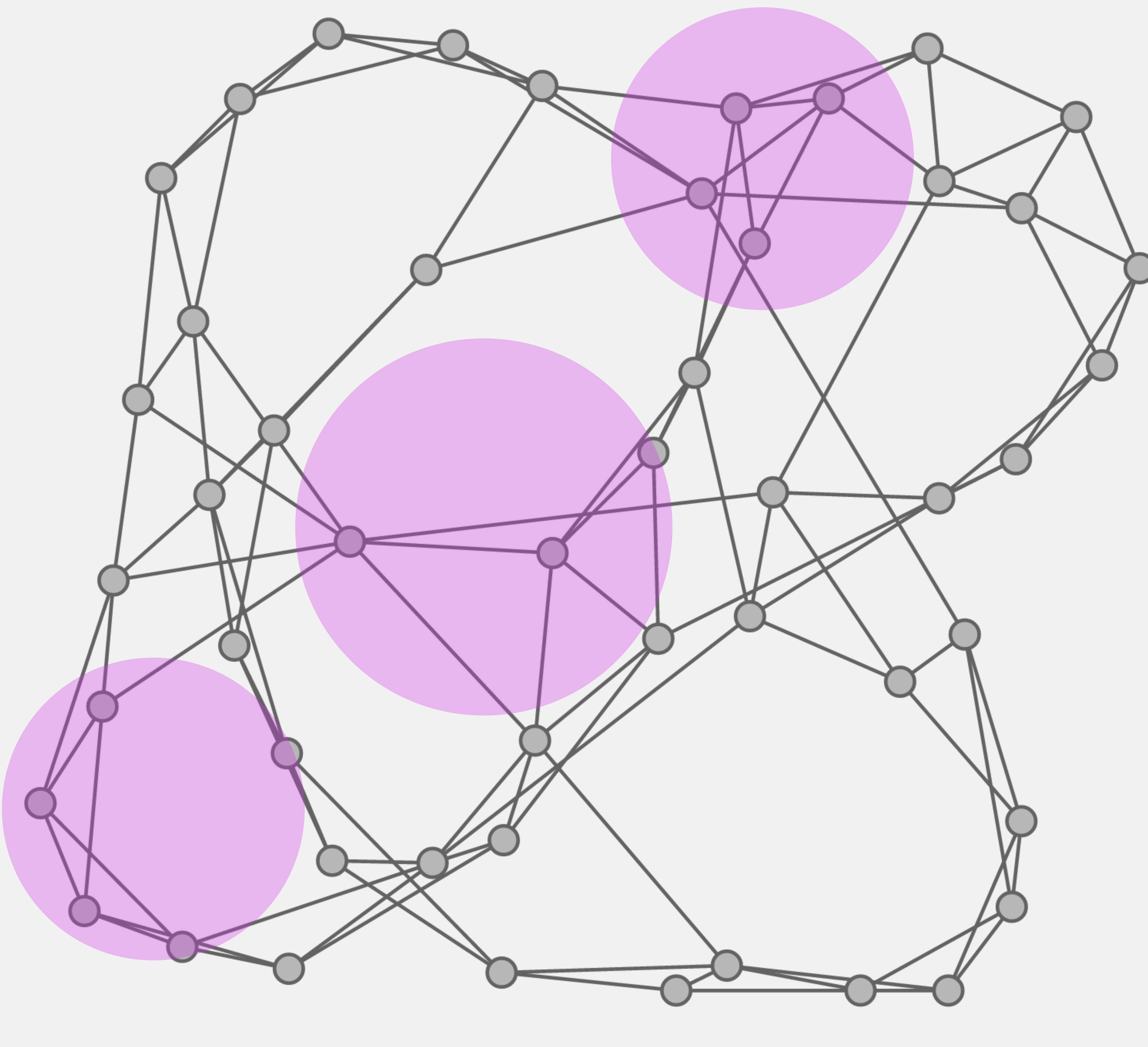
average e-vector centrality in surveillance set

$$\Delta\tau := \tau_S - \tau = \frac{1}{\beta\kappa - \gamma} \ln\left(\frac{c}{c_S}\right)$$

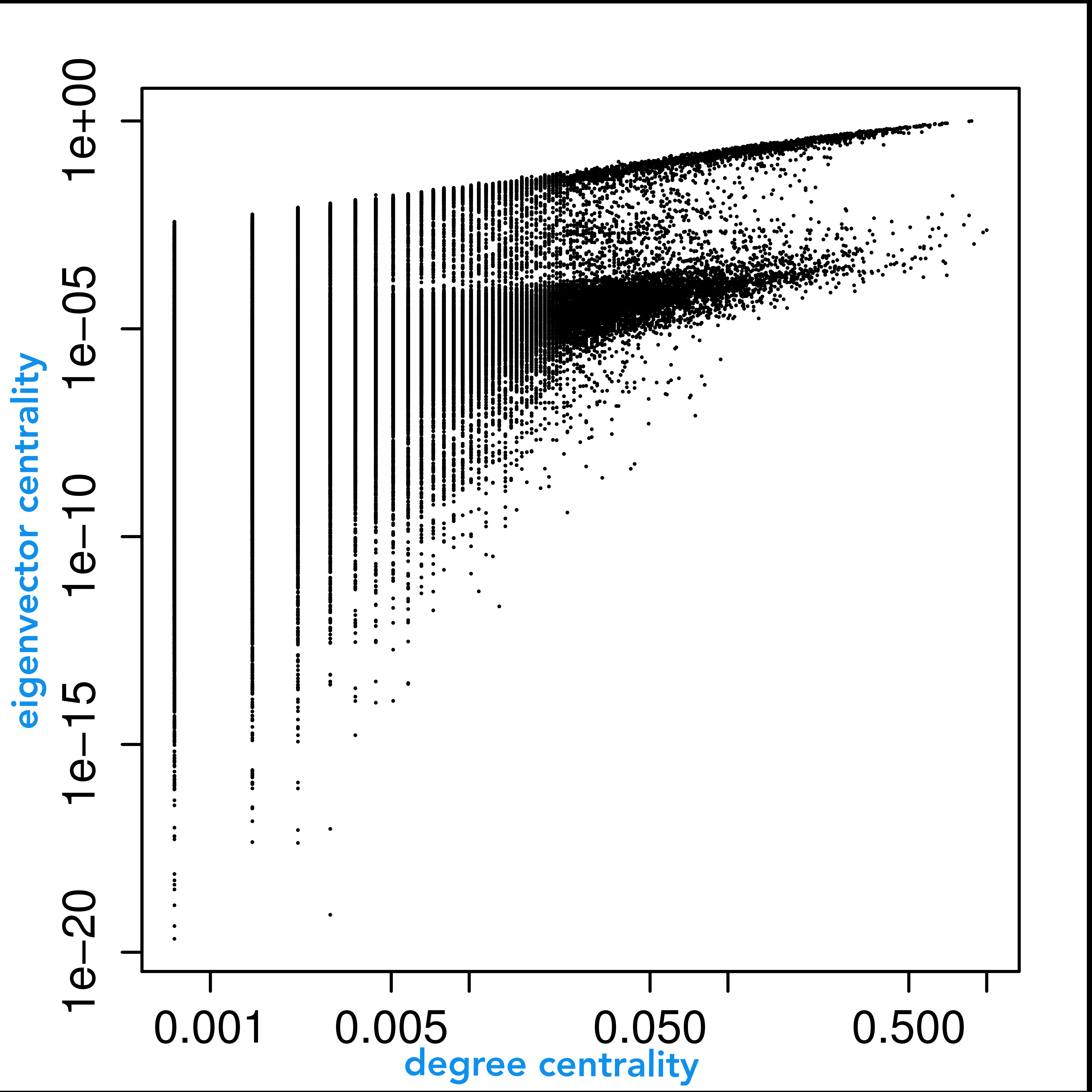
average e-vector centrality in network

early detection + eigenvector centrality

- problem: not possible* without knowing entire network to begin with!
- never really know the underlying social graph, can only infer global properties from local subnetworks
- is there a random walk whose stationary distribution is eigenvector centrality?



correlation of centrality measures



- could use degree centrality (locally recoverable) as a proxy for eigenvector centrality
- can we do better: is there a random walk whose stationary distribution is eigenvector centrality?

maximal entropy random walk (MERW)

$$p_{ij} = \frac{a_{ij} v_j}{\kappa v_i}$$

transition probabilities

Burda (2010)

maximal entropy random walk (MERW)

$$p_{ij} = \frac{a_{ij} v_j}{\kappa v_i}$$

transition probabilities

$$\psi_i = v_i^2$$

stationary distribution!

Burda (2010)

maximal entropy random walk (MERW)

$$p_{ij} = \frac{a_{ij} v_j}{\kappa v_i}$$

transition probabilities

$$\psi_i = v_i^2$$

stationary distribution!

$$A\mathbf{v} = \kappa \mathbf{v} \qquad \Longrightarrow \qquad \sum_j p_{ij} \psi_j = \frac{1}{\kappa} v_i \sum_j a_{ij} v_j = v_i^2 = \psi_i$$

Burda (2010)

maximal entropy random walk (MERW)

$$p_{ij} = \frac{a_{ij} v_j}{\kappa v_i}$$

transition probabilities

$$\psi_i = v_i^2$$

stationary distribution!

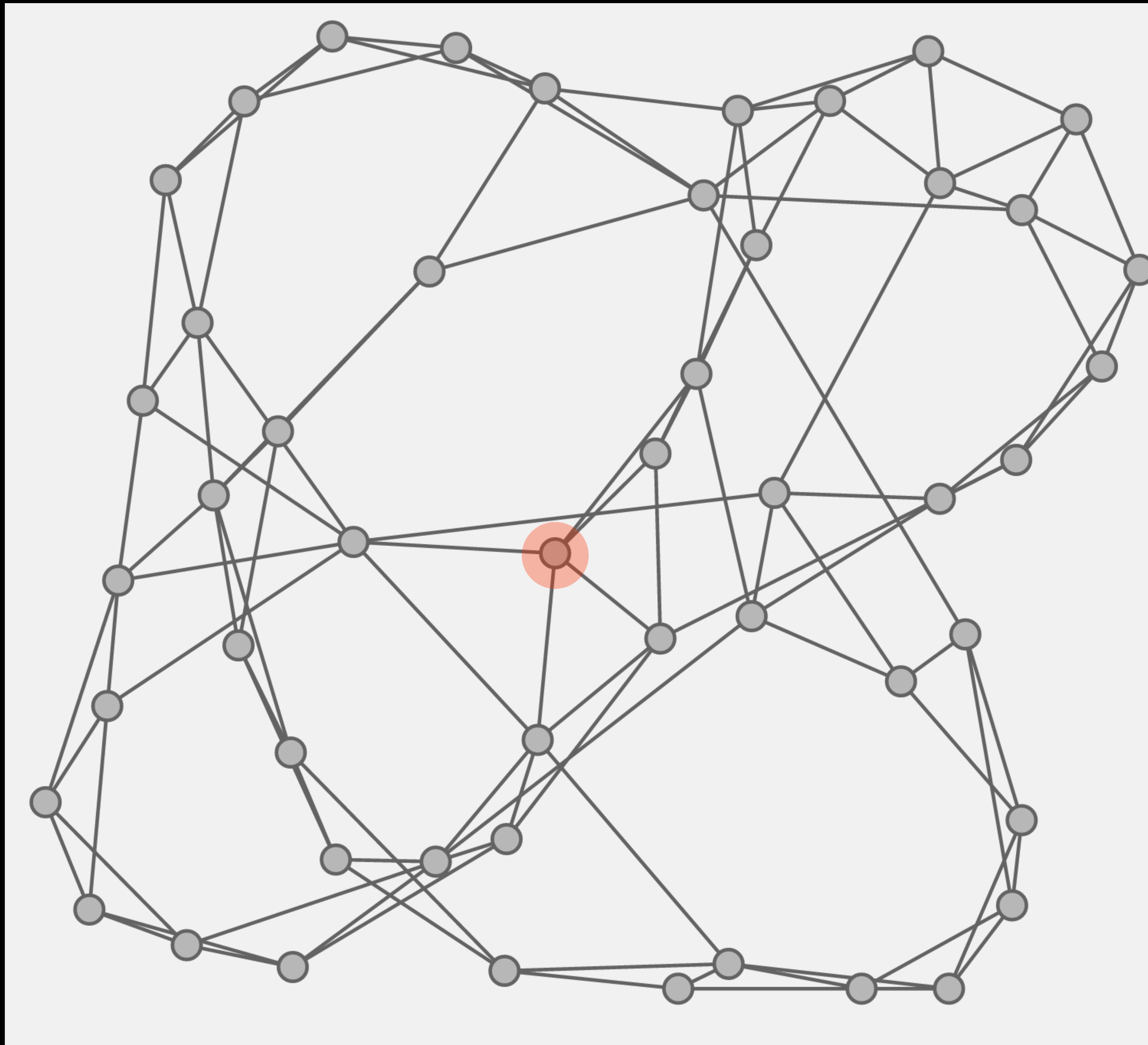
$$A\mathbf{v} = \kappa \mathbf{v} \qquad \Longrightarrow \qquad \sum_j p_{ij} \psi_j = \frac{1}{\kappa} v_i \sum_j a_{ij} v_j = v_i^2 = \psi_i$$

but transition probabilities given in terms of e-vector centralities!

Burda (2010)

maximal entropy random walk

- why **maximal entropy**?
- uniform distribution on a set has maximal entropy among all distributions on that set
- transition probabilities of MERW put equal probability on all **paths** of length *t* starting from a given node, as *t* goes to ∞



entropy rate

probability of particular path

$$S(t) = -\sum_{i_1,\ldots,i_t} p(i, i_1, \ldots, i_t) \ln p(i, i_1, \ldots, i_t)$$

Entropy of set of paths of length *t* starting at *i*

entropy rate

probability of particular path

$$S(t) = -\sum_{i_1,\ldots,i_t} p(i, i_1, \ldots, i_t) \ln p(i, i_1, \ldots, i_t)$$

Entropy of set of paths of length *t* starting at *i*

$$h = \lim_{t \to \infty} \frac{\ln S(t)}{t}$$

entropy rate

entropy rate

- minimal entropy: all probability is put on one path starting from *i:*

0 for all paths except one!

$$S(t) = -\sum_{i_1,\ldots,i_t} p(i, i_1, \ldots, i_t) \ln p(i, i_1, \ldots, i_t) = 0$$

minimal entropy

- maximal entropy: uniform probability on all paths of length *t* starting from *i:*

uniform probability on all paths

$$S(t) = -\sum_{i_1,\ldots,i_t} \frac{1}{M_i(t)} \ln\left(\frac{1}{M_i(t)}\right) = \ln M_i(t)$$

maximal entropy

$$M_i(t) = \sum_{i_1, i_2, \ldots, i_t} a_{ii_1} a_{i_1 i_2} \cdots a_{i_{t-1} i_t}$$

transition probabilities of MERW

$$p(i, i_1, \ldots, i_t) = \frac{a_{ii_1} a_{i_1 i_2} \ldots a_{i_{t-1} i_t}}{\sum_{i_1, i_2, \ldots, i_t} a_{ii_1} a_{i_1 i_2} \ldots a_{i_{t-1} i_t}}$$

uniform probability on all paths

$$\pi(i_1|i) = \lim_{t \to \infty} \frac{a_{ii_1} \sum_{i_2} a_{i_1 i_2} \ldots \sum_{i_t} a_{i_{t-1} i_t}}{\sum_{i_1} a_{ii_1} \sum_{i_2} a_{i_1 i_2} \ldots \sum_{i_t} a_{i_{t-1} i_t}}$$

corresponding transition probabilities!

Sinatra et. al (2011)

transition probabilities of MERW

$$\pi(i_1|i) = \lim_{t\to\infty} \frac{a_{ii_1} \sum_{i_2} a_{i_1 i_2} \ldots \sum_{i_t} a_{i_{t-1} i_t}}{\sum_{i_1} a_{ii_1} \sum_{i_2} a_{i_1 i_2} \ldots \sum_{i_t} a_{i_{t-1} i_t}}$$

Sinatra et. al (2011)

transition probabilities of MERW

$$\pi(i_1|i) = \lim_{t\to\infty} \frac{a_{ii_1} \sum_{i_2} a_{i_1 i_2} \ldots \sum_{i_t} a_{i_{t-1} i_t}}{\sum_{i_1} a_{ii_1} \sum_{i_2} a_{i_1 i_2} \ldots \sum_{i_t} a_{i_{t-1} i_t}}$$

$$\pi^0(i_1|i) = \frac{a_{ii_1}}{\sum_{i_1} a_{ii_1}}$$

0th-order approximation

"How many friends do I have?"

Sinatra et. al (2011)

transition probabilities of MERW

$$\pi(i_1|i) = \lim_{t\to\infty} \frac{a_{ii_1}\sum_{i_2} a_{i_1 i_2}\dots\sum_{i_t} a_{i_{t-1}i_t}}{\sum_{i_1} a_{ii_1}\sum_{i_2} a_{i_1 i_2}\dots\sum_{i_t} a_{i_{t-1}i_t}}$$

$$\pi^0(i_1|i) = \frac{a_{ii_1}}{\sum_{i_1} a_{ii_1}}$$

0th-order approximation

"How many friends do I have?"

$$\pi^1(i_1|i) = \frac{a_{ii_1}k(i_1)}{\sum_{i_1} a_{ii_1}k(i_1)}$$

1st-order approximation

"How many friends do my friends have?"

Sinatra et. al (2011)

transition probabilities of MERW

$$\pi(i_1|i) = \lim_{t\to\infty} \frac{a_{ii_1}\sum_{i_2} a_{i_1 i_2}\dots\sum_{i_t} a_{i_{t-1}i_t}}{\sum_{i_1} a_{ii_1}\sum_{i_2} a_{i_1 i_2}\dots\sum_{i_t} a_{i_{t-1}i_t}}$$

$$\pi^0(i_1|i) = \frac{a_{ii_1}}{\sum_{i_1} a_{ii_1}}$$

0th-order approximation

"How many friends do I have?"

$$\pi^1(i_1|i) = \frac{a_{ii_1}k(i_1)}{\sum_{i_1} a_{ii_1}k(i_1)}$$

1st-order approximation

"How many friends do my friends have?"

$$\begin{aligned}\pi^2(i_1|i) &= \frac{a_{ii_1}\sum_{i_2} a_{i_1 i_2}k(i_2)}{\sum_{i_1} a_{ii_1}\sum_{i_2} a_{i_1 i_2}k(i_2)} \\ &= \frac{a_{ii_1}k(i_1)k_{nn}(i_1)}{\sum_{i_1} a_{ii_1}k(i_1)k_{nn}(i_1)},\end{aligned}$$

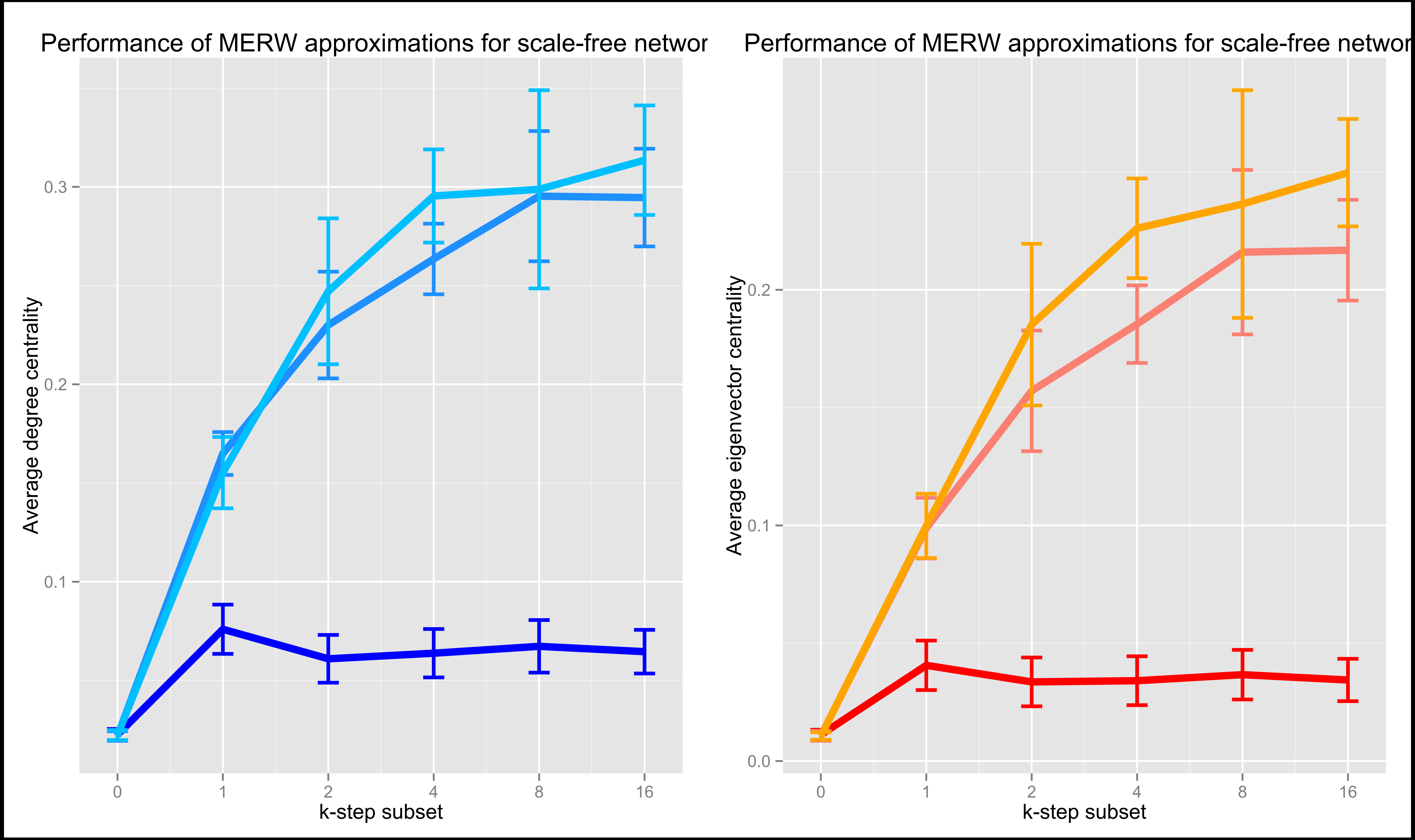
2nd-order approximation

"What is the average degree of my friends' friends?"

Sinatra et. al (2011)

approximated MERW

performance on Montreal network



recap

- can use random walks to sample nodes with desired properties
- how to incorporate into a realistic sampling methodology?
- network models are fun to work with!