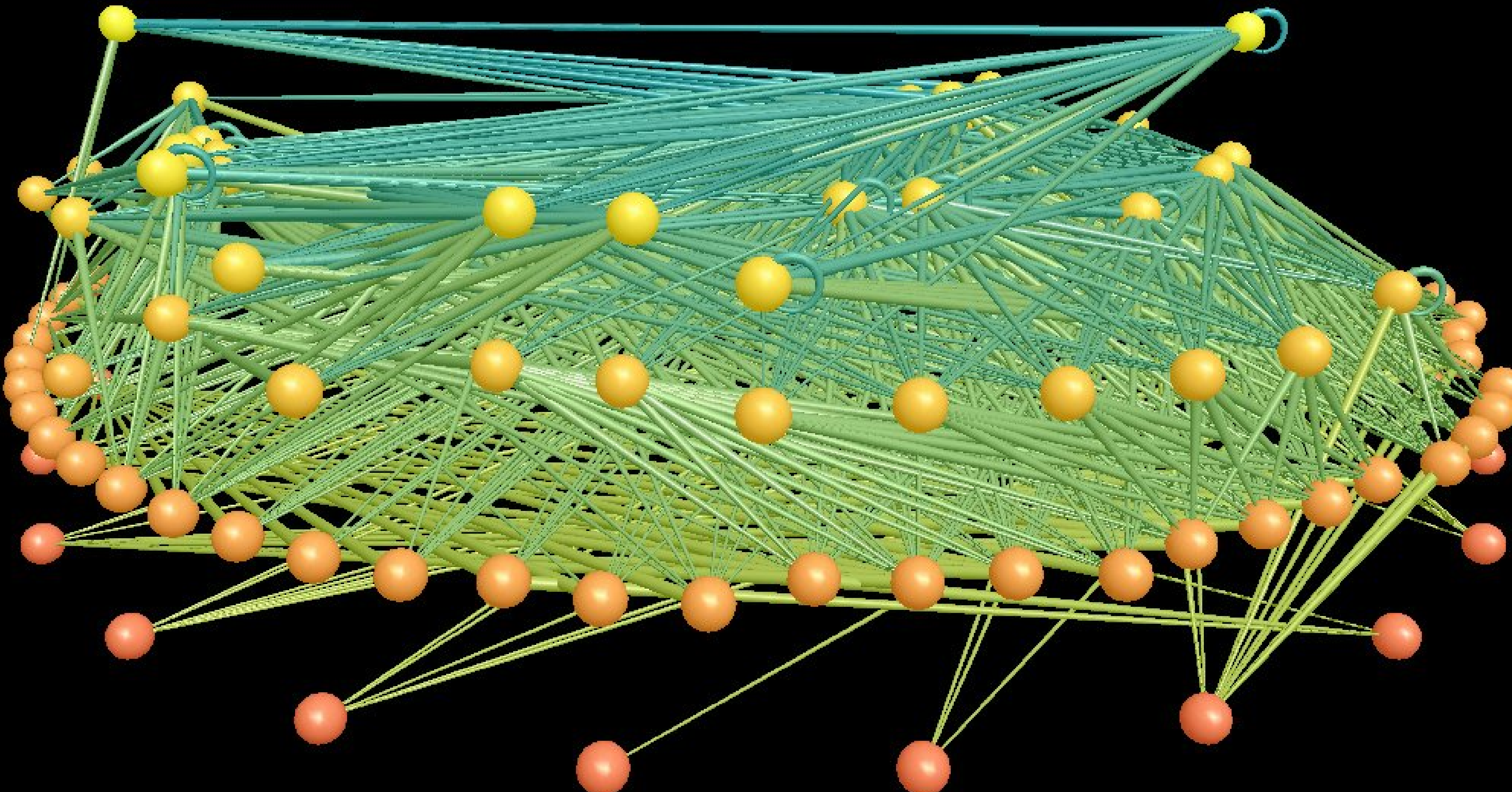
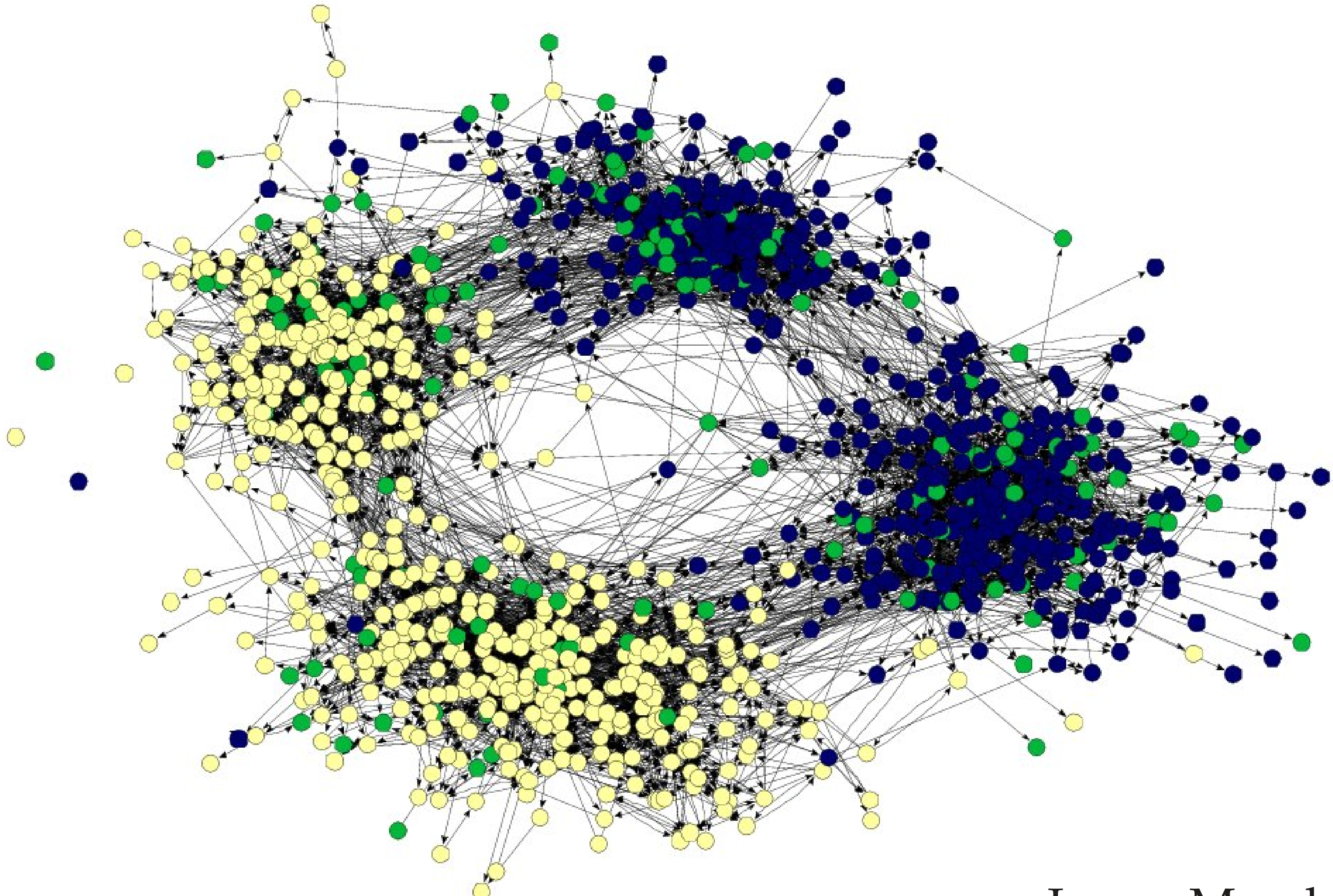
Dynamics on Networks and Message Passing Algorithms

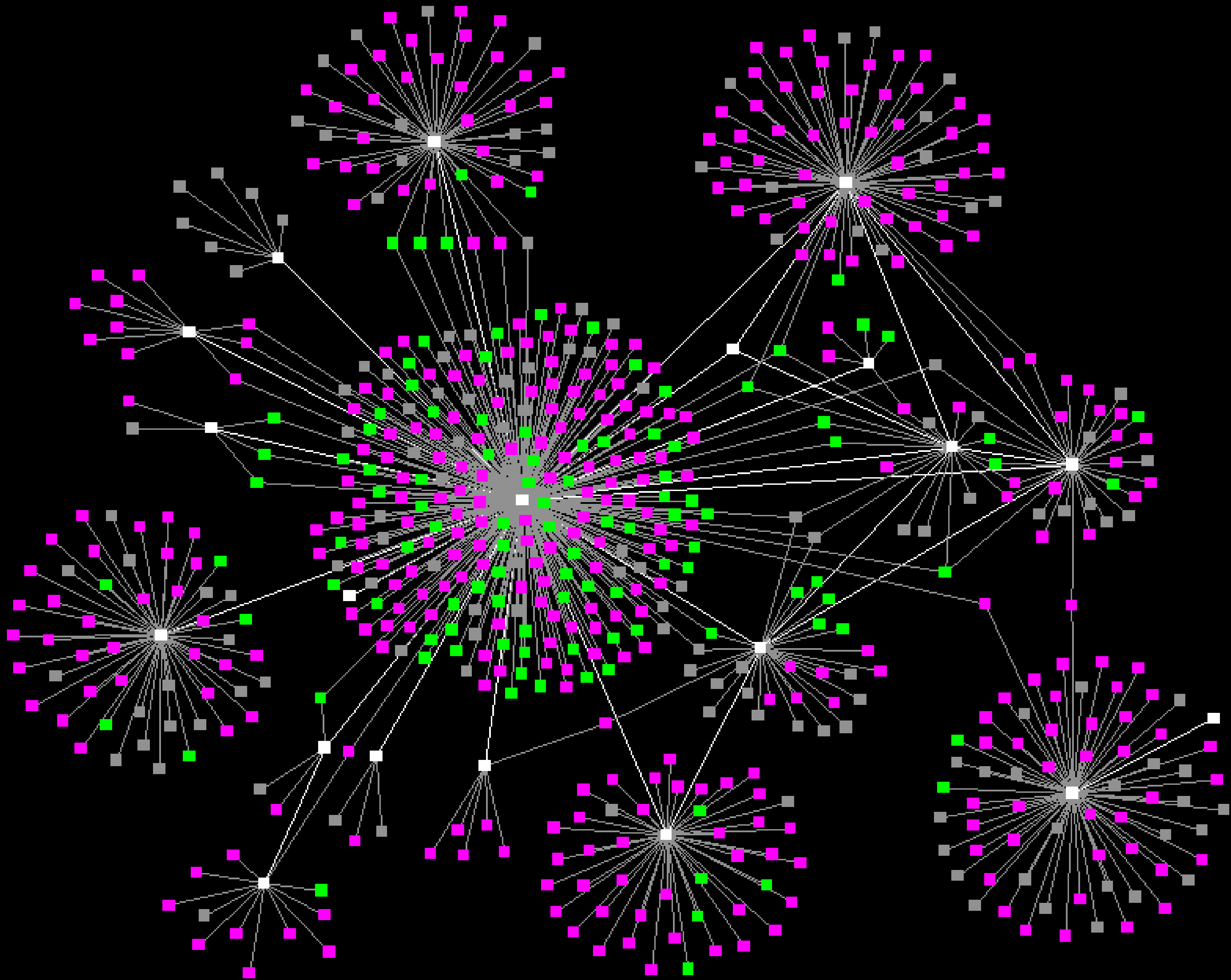
Mark Newman
University of Michigan



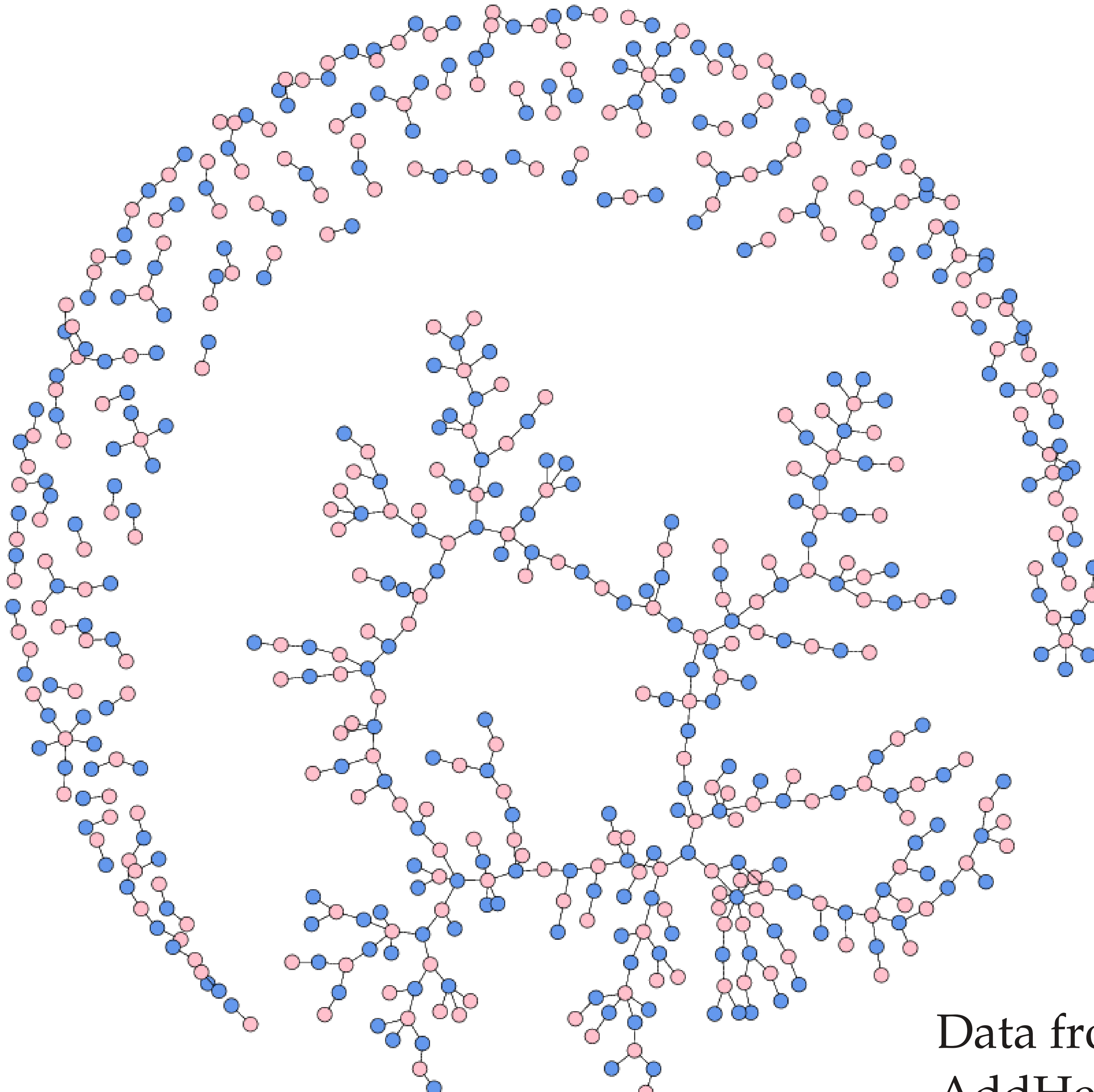
Neo Martinez and Rich Williams



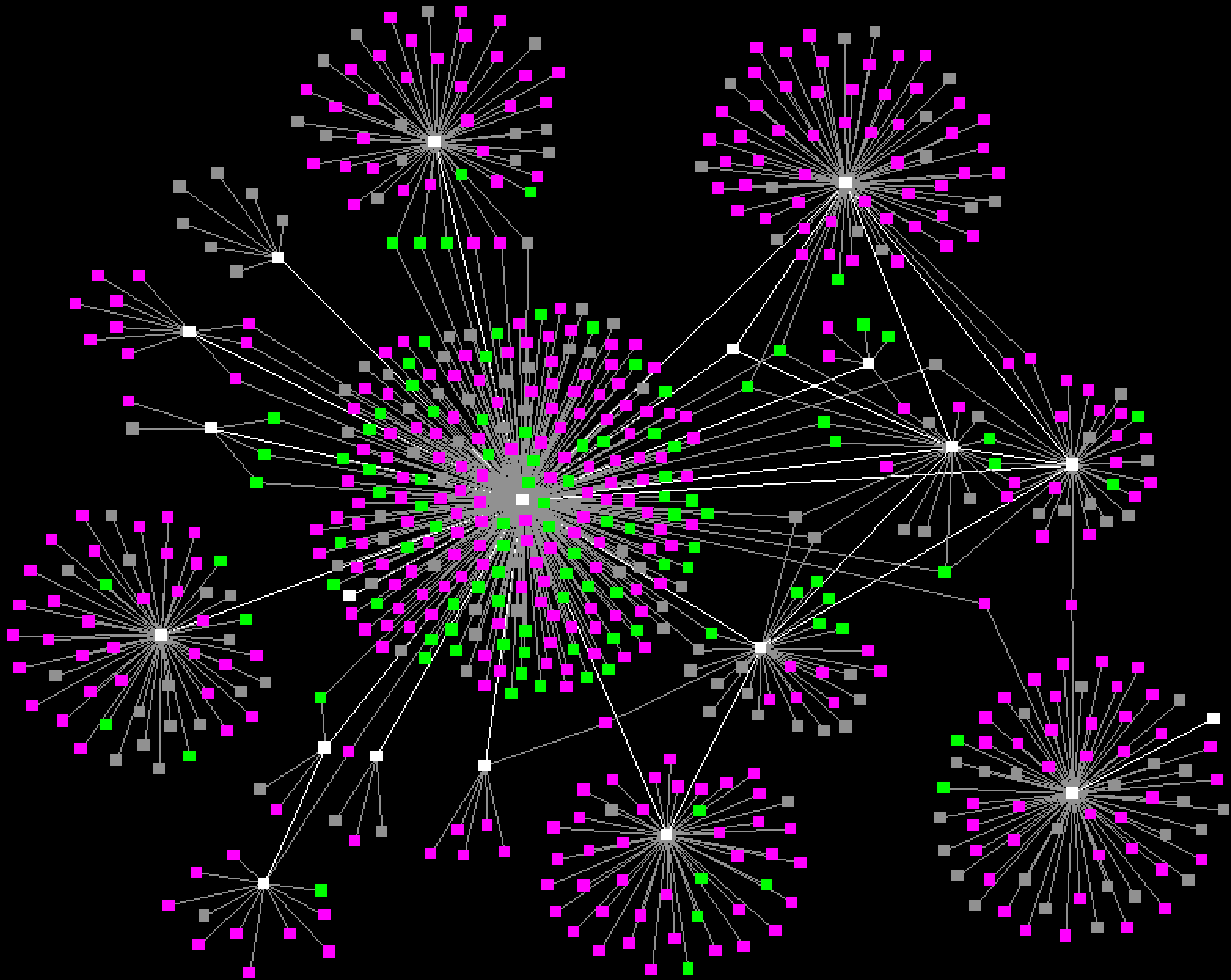
James Moody



Valdis Krebs

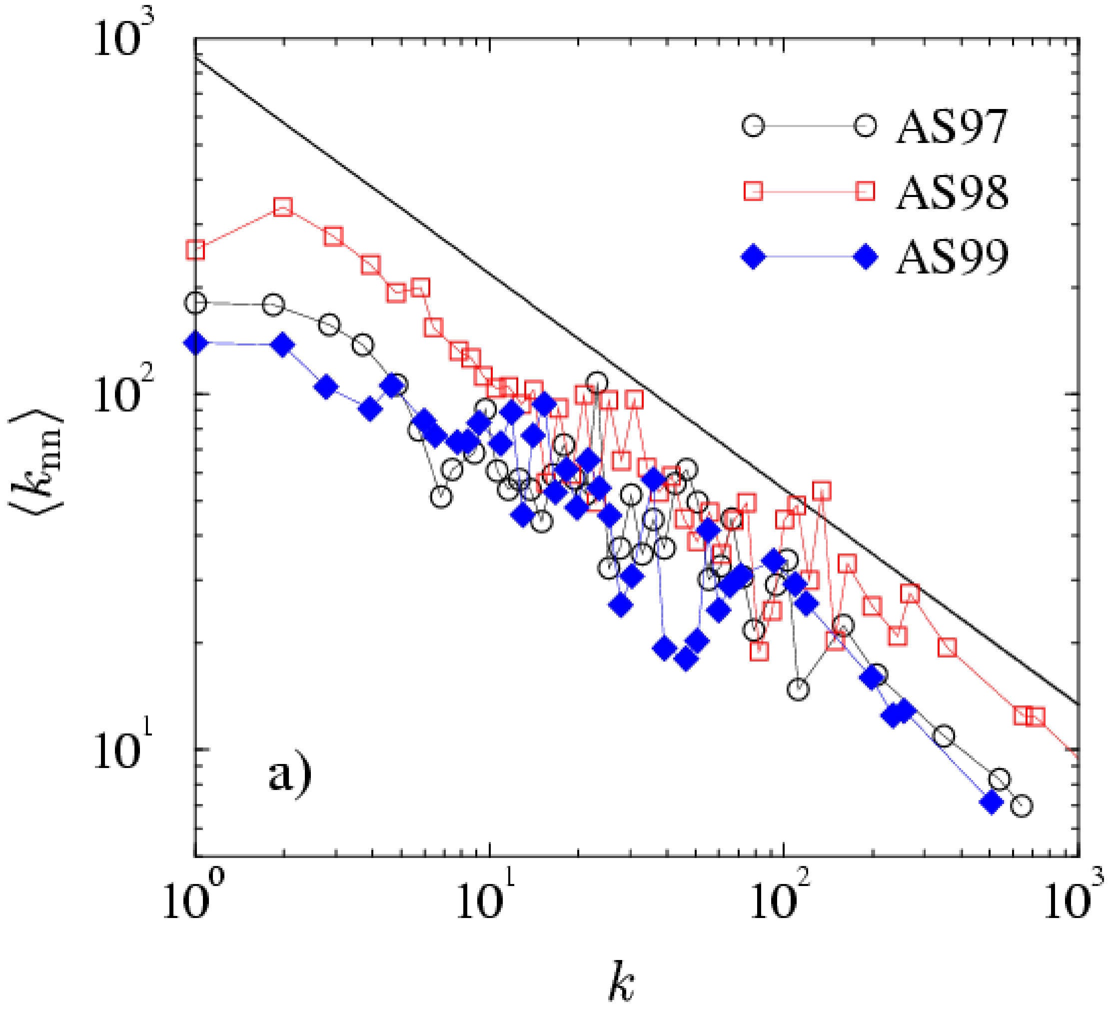


Data from
AddHealth study

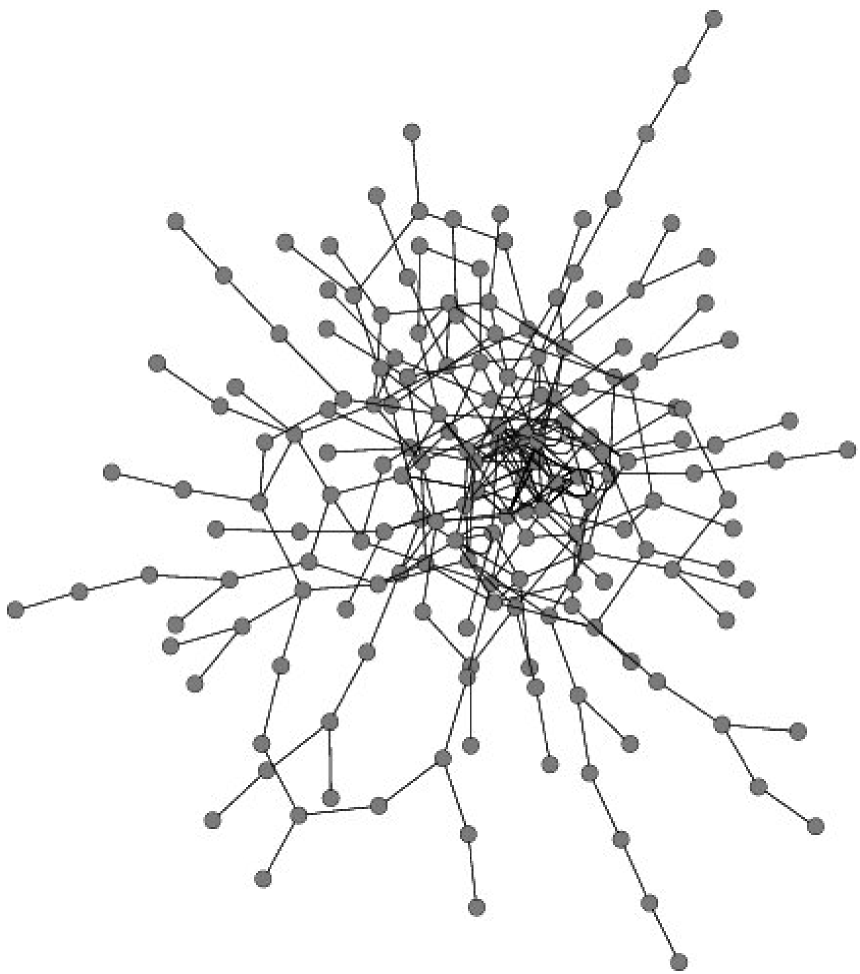


Valdis Krebs

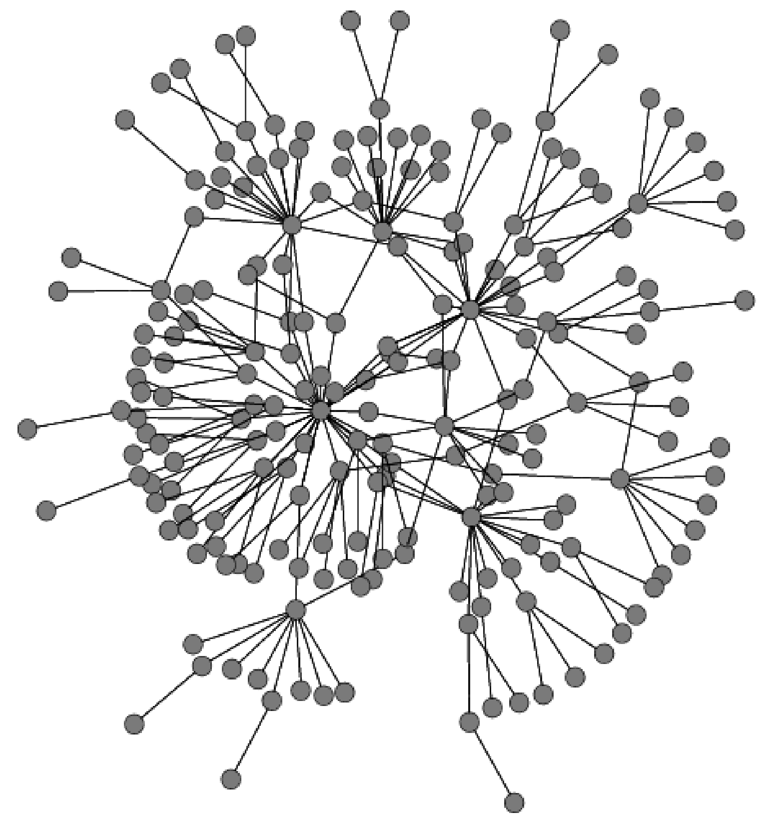
Degree correlations



Pastor-Satorras and Vespignani 2001



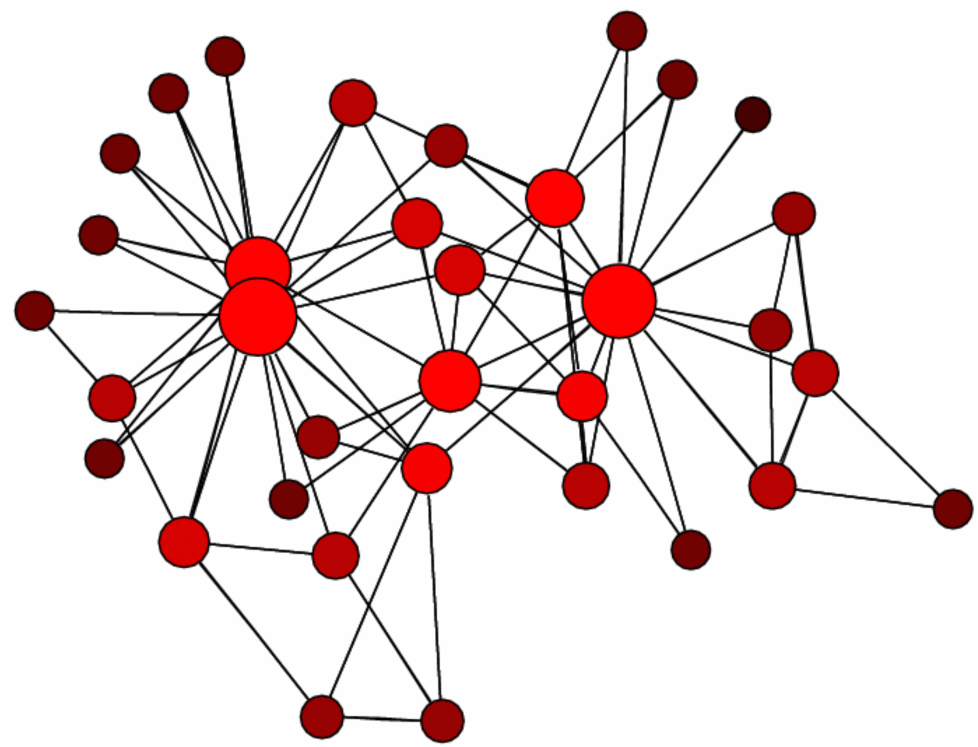
Positive correlations



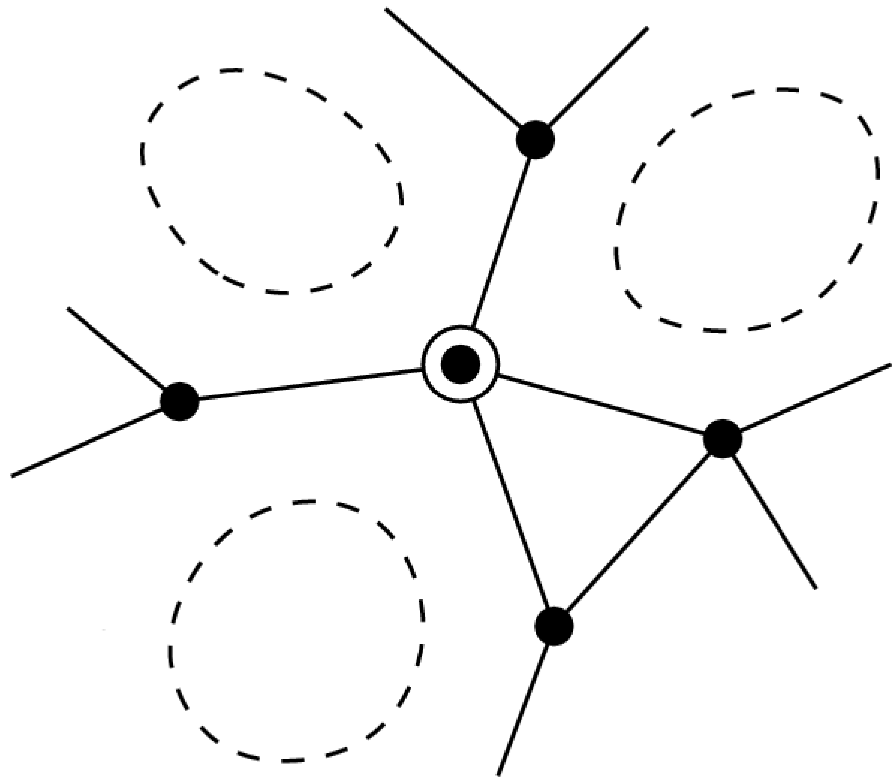
Negative correlations

Transitivity or Clustering

- If A knows B and B knows C, then probably A knows C
  - "the friend of my friend is also my friend"
  - *or* two of my friends are likely also to be friends of each other

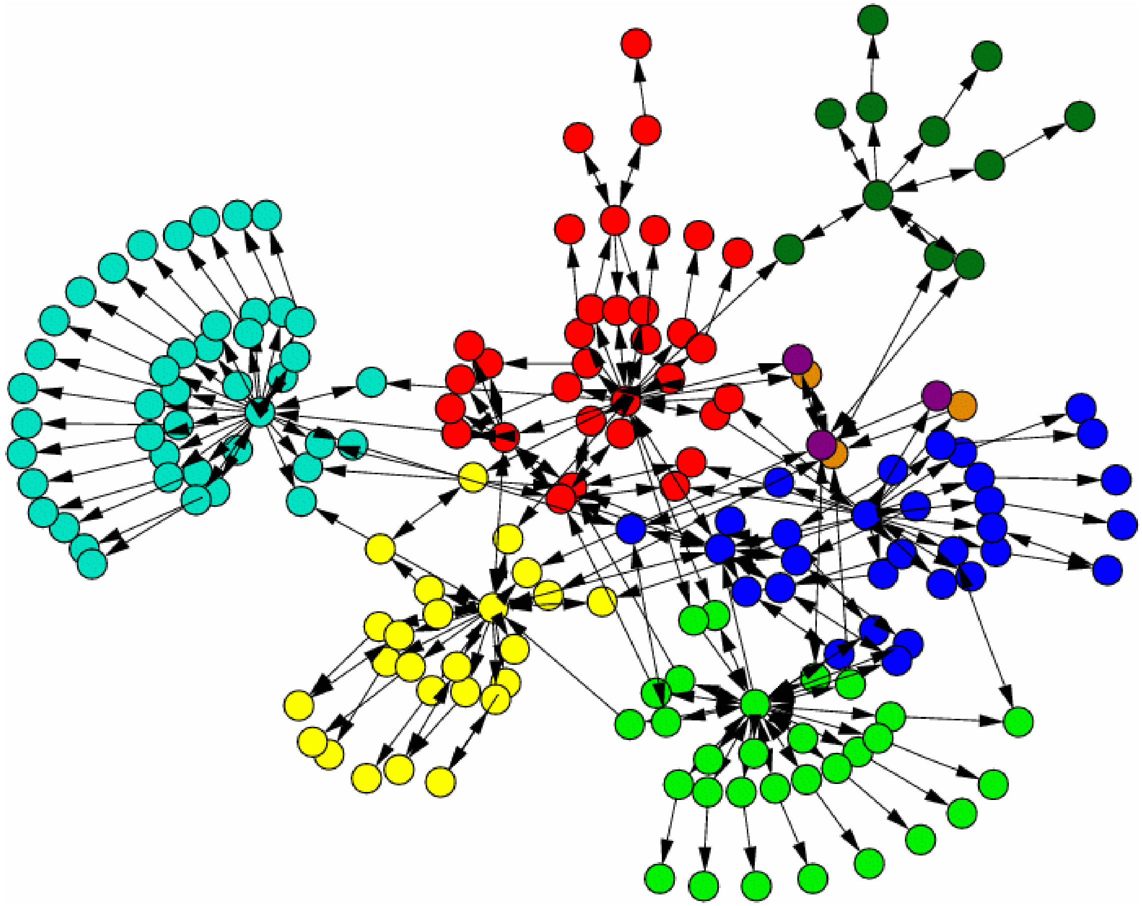


Structural holes



- Many structural holes = low clustering coefficient
  - Used in sociology as a measure of power or influence, i.e., a centrality measure

Reciprocity

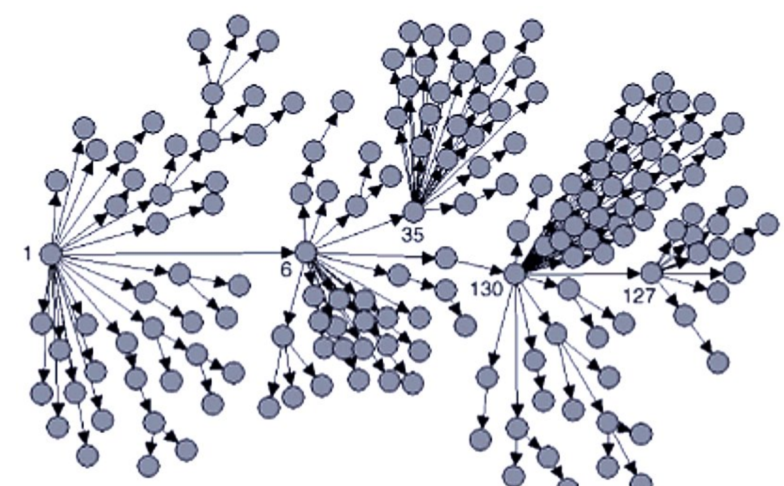


- Reciprocity values:
  - Web ~55%, Email ~25%, Friendship network ~40%

Dynamic processes

- Most networks are interesting for their dynamics:
  - Packets on the Internet
  - Surfing on the web
  - Metabolic reactions in a metabolic network
  - Energy or carbon flows in ecological networks
  - Information, news, rumors, fads, fashions, gossip, or disease on social networks

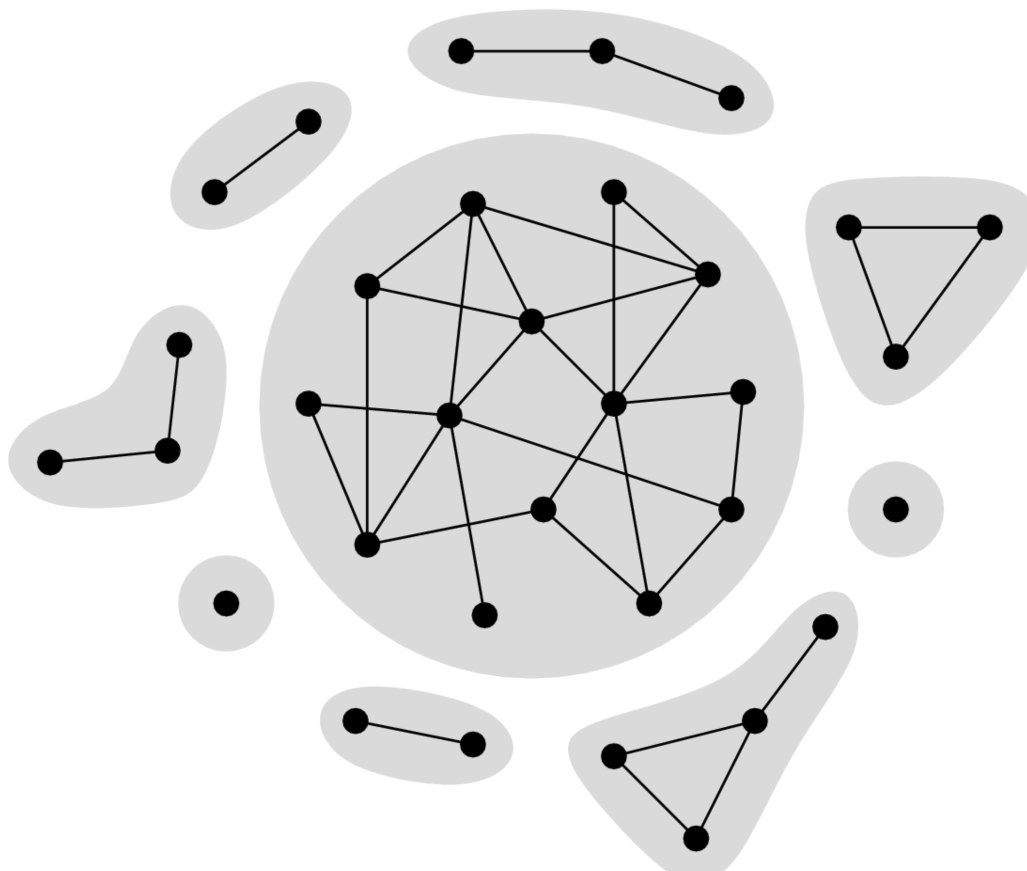
Brockmann 2011



Borgatti 2003

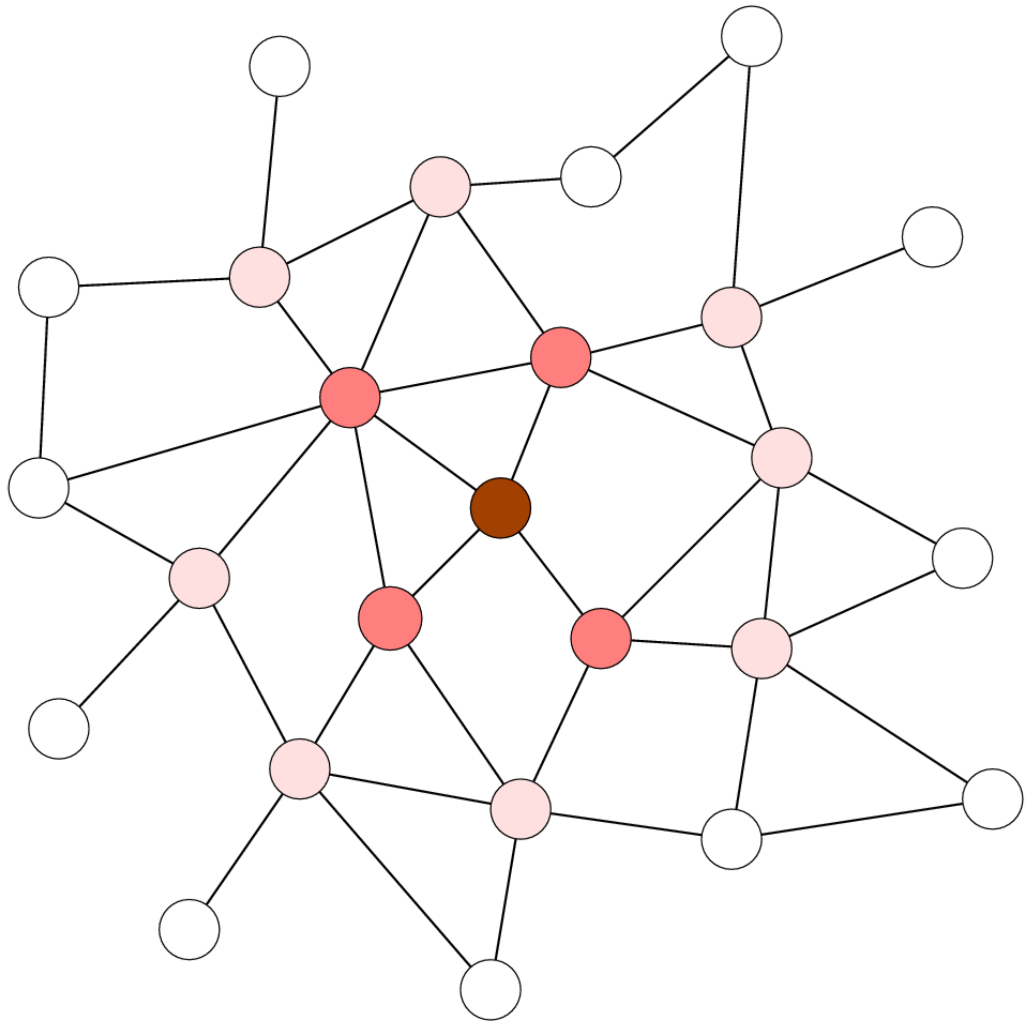
Spread of disease

- In the simplest case, we can just assume a disease, starting with a single carrier, spreads to everyone it can reach



Breadth-first search

- The simplest way to do it is breadth- or depth-first search:

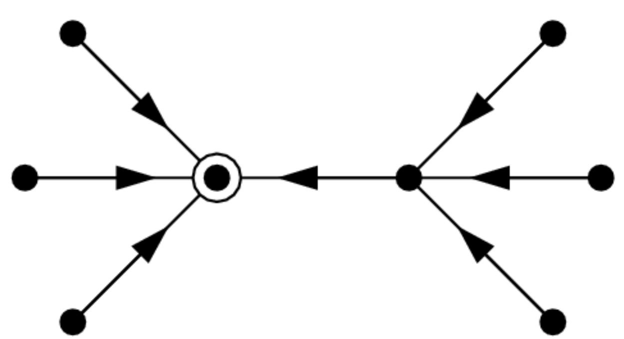


Message passing

- Let u_i be the probability that node i is **not** in the giant component
- Node i is not in the giant component if none of its neighbors are:

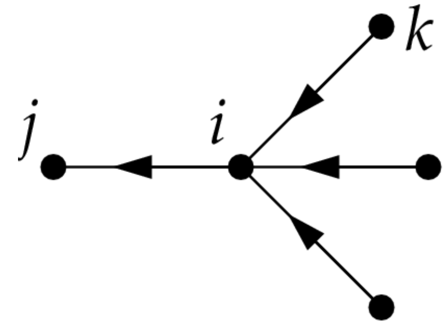
$$u_i = \prod_{j \in \mathcal{N}(i)} u_j$$

- But we need to exclude connections via the node itself:



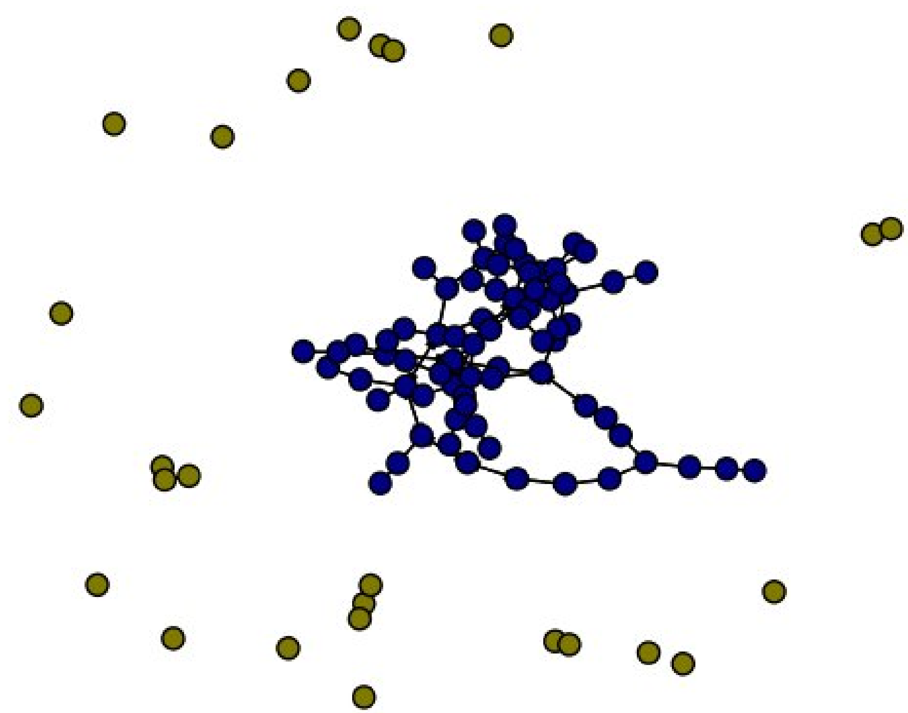
Cavity method

- Define a separate message $u_{i \to j}$ for each edge in the network:



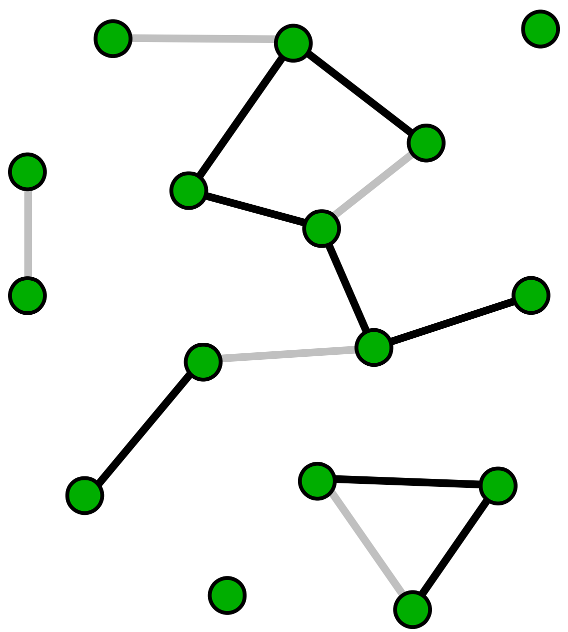
$$u_{i \to j} = \prod_{\substack{k \in \mathcal{N}(i) \\ k \neq j}} u_{k \to i}$$

- Has two fixed points, at 0 and 1



Percolation

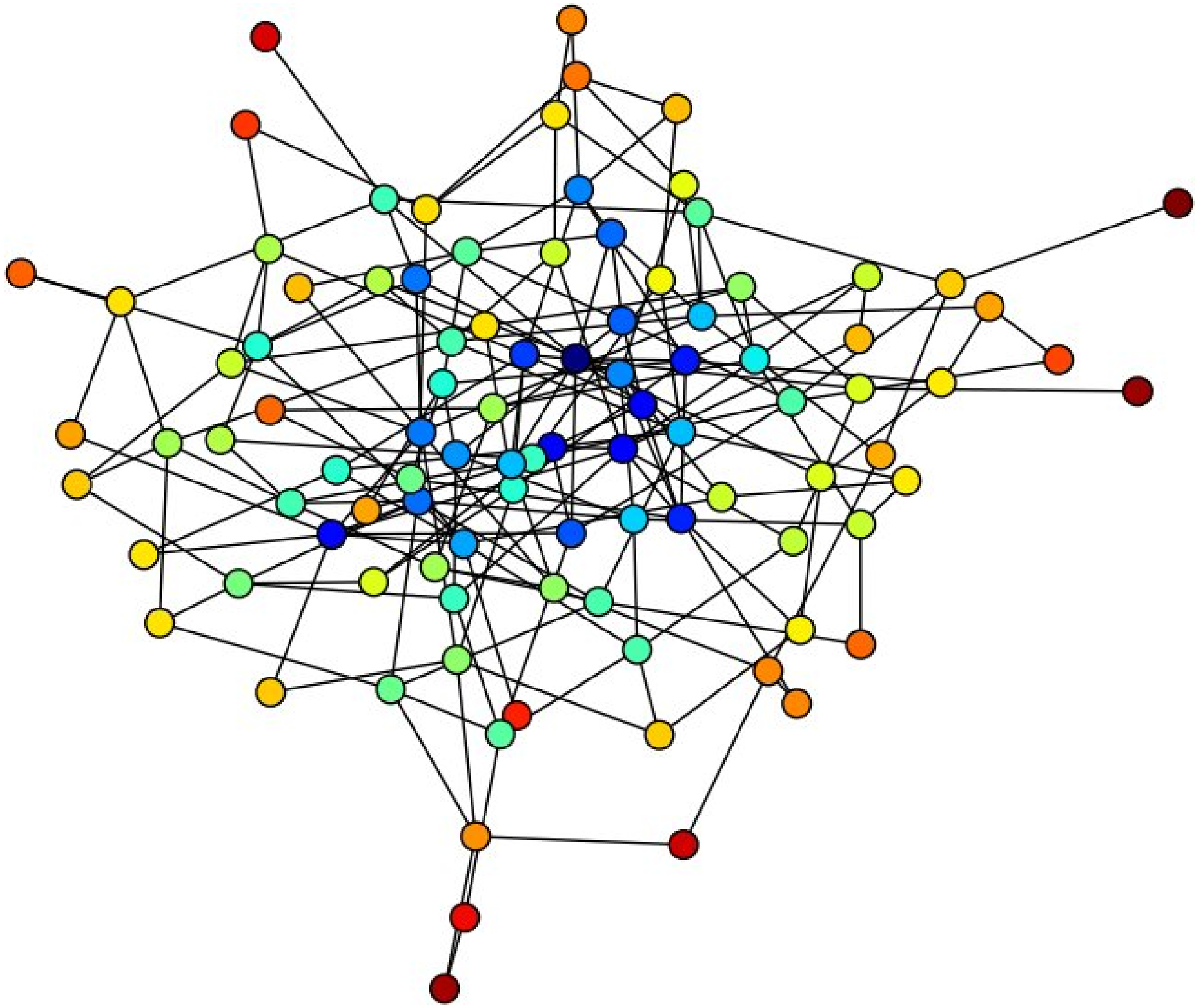
- Consider a disease that doesn't definitely spread, it only spreads with probability p
- This is equivalent to saying that each edge in the network is occupied or not with probability p, i.e., it's bond percolation
- Direct algorithmic approaches:
  - Breadth-first search again
  - Union-find algorithms (talks by Ziff and by Mertens)



Percolation

- But now our message passing approach is clearly superior
  - It can tell us about all realizations of the process at once
- There are two ways *j* can fail to get the disease from its neighbor *i*
  - Either the edge between them is unoccupied
  - Or it's occupied but *i* doesn't have the disease:

$$u_{i \to j} = 1 - p + p \prod_{\substack{k \in \mathcal{N}(i) \\ k \neq j}} u_{k \to i}$$



Percolation

- You can also average over the network structure
- Let u be the average message. Then, on a network where every node has degree $k + 1$, we have

$$u = 1 - p + pu^k$$

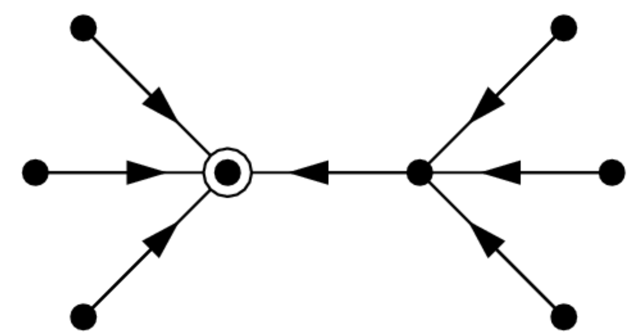
- More generally,

$$u = 1 - p + p \sum_k f_k u^k$$

- This is a well-known result for percolation on networks (Callaway, MN, Watts, and Strogatz 2000)

Coinfection

- Consider a coinfection process where there are two diseases and infection with one is dependent on infection with the other



$$g_0(z) = \sum_{k=0}^{\infty} p_k z^k, \qquad g_1(z) = \sum_{k=0}^{\infty} q_k z^k,$$

$$S_1 = 1 - g_0(1 - uT_1), \qquad u = 1 - g_1(1 - uT_1),$$

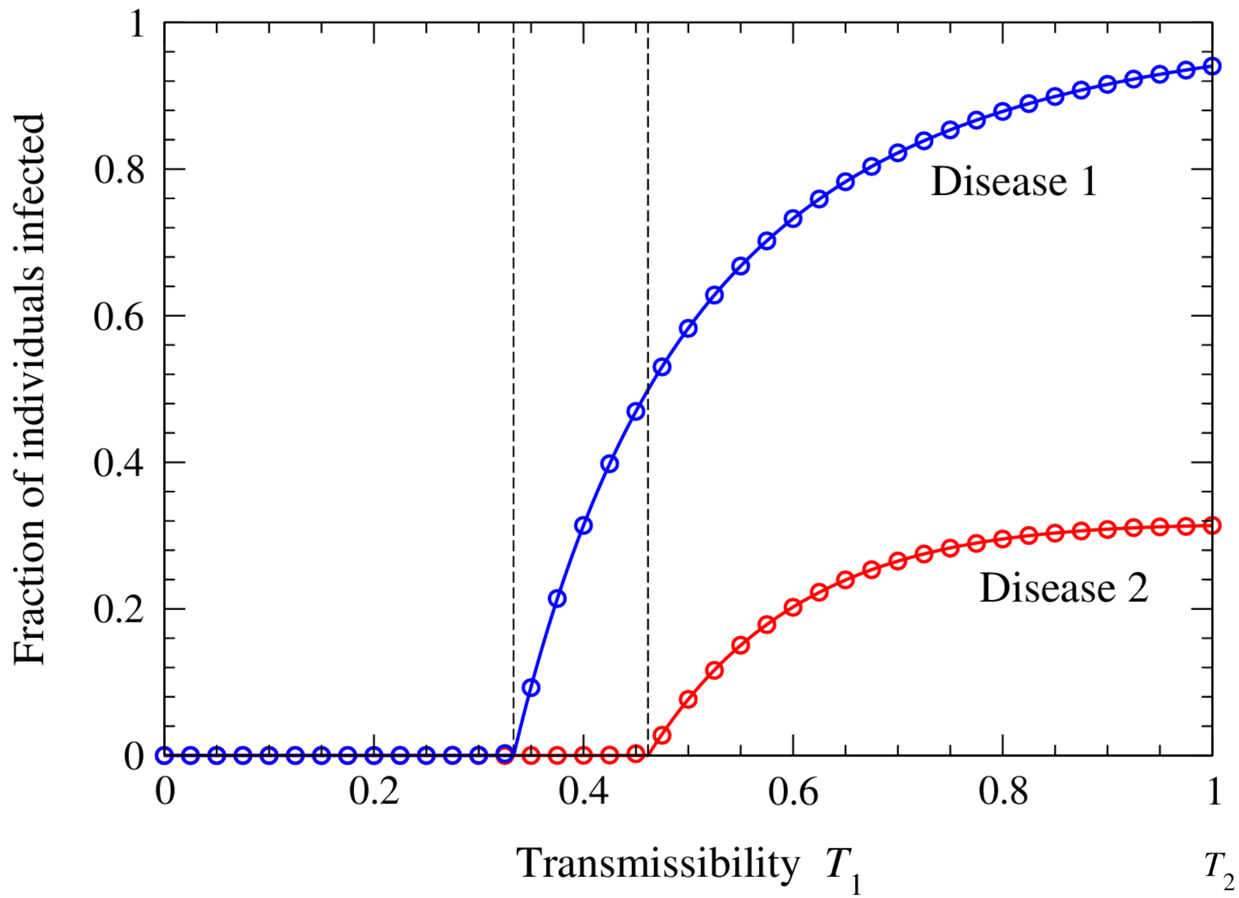
$$h_0(y,z) = \frac{1}{S_1}\Big(g_0\big[uy + (1-u)(1 - T_1 + zT_1)\big] - g_0\big[u(1-T_1)y + (1-u)(1-T_1+zT_1)\big]\Big),$$

$$h_1(y,z) = \frac{1}{u}\Big(g_1\big[uy + (1-u)(1 - T_1 + zT_1)\big] - g_1\big[u(1-T_1)y + (1-u)(1-T_1+zT_1)\big]\Big),$$

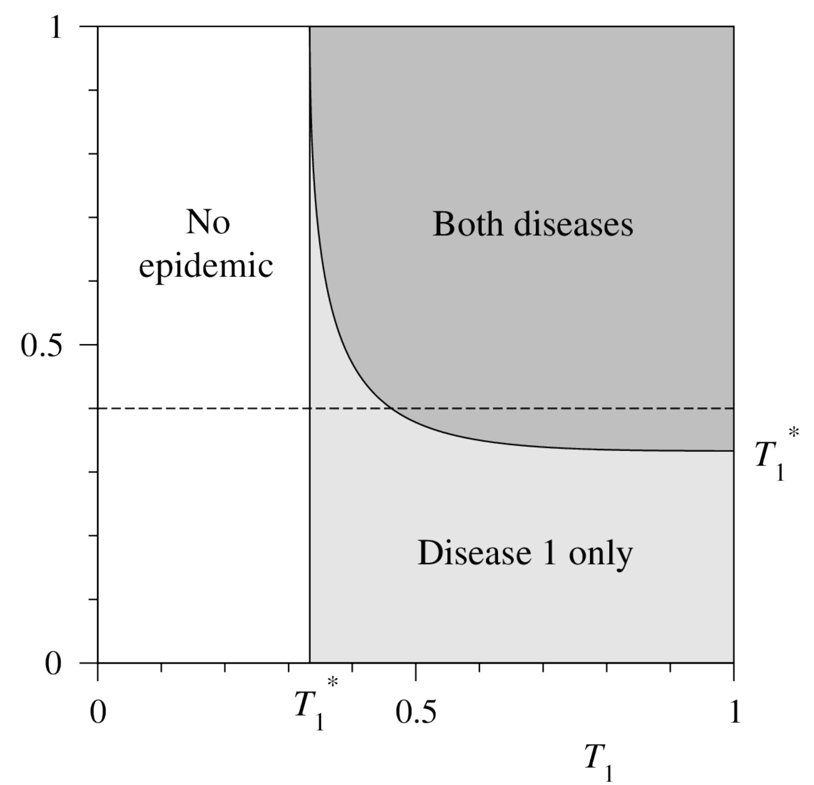
$$h_2(y,z) = \frac{g_1\big[u(1-T_1)y + (1-u)(1-T_1+zT_1)\big]}{1-u},$$

$$S_2 = 1 - h_0(1 - vT_2, 1 - wT_2),$$

$$v = 1 - h_1(1 - vT_2, 1 - wT_2), \qquad w = 1 - h_2(1 - vT_2, 1 - wT_2).$$



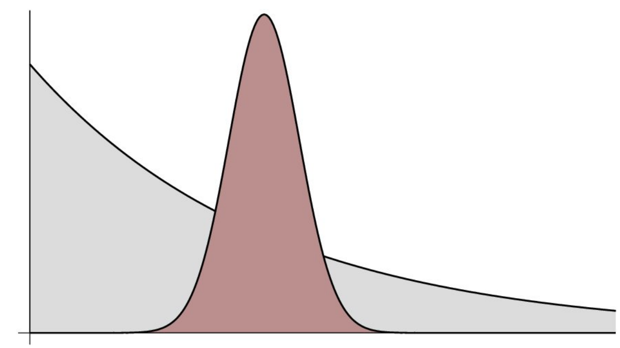
Fraction of individuals infected
Disease 1
Disease 2
0
0.2
0.4
0.6
0.8
1
Transmissibility T_1
T_2



No epidemic
Both diseases
Disease 1 only
0
0.5
1
T_1^*
T_1

Dynamics

- Let $p(\tau)$ be the probability density for the time τ after infection at which transmission occurs between an infected individual and their susceptible neighbor
  - This distribution can be anything
  - Most theories assume exponential
  - Real diseases are not exponential



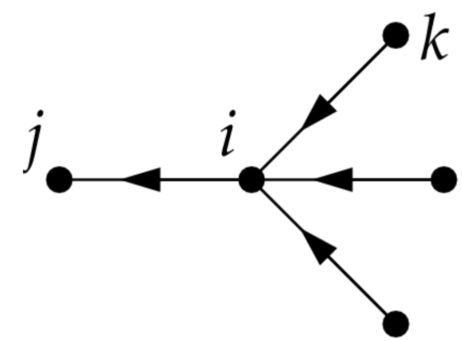
- The probability does not integrate to 1:

$$\text{Probability of transmission } p = \int_0^\infty p(\tau)\,\mathrm{d}\tau$$

Dynamics

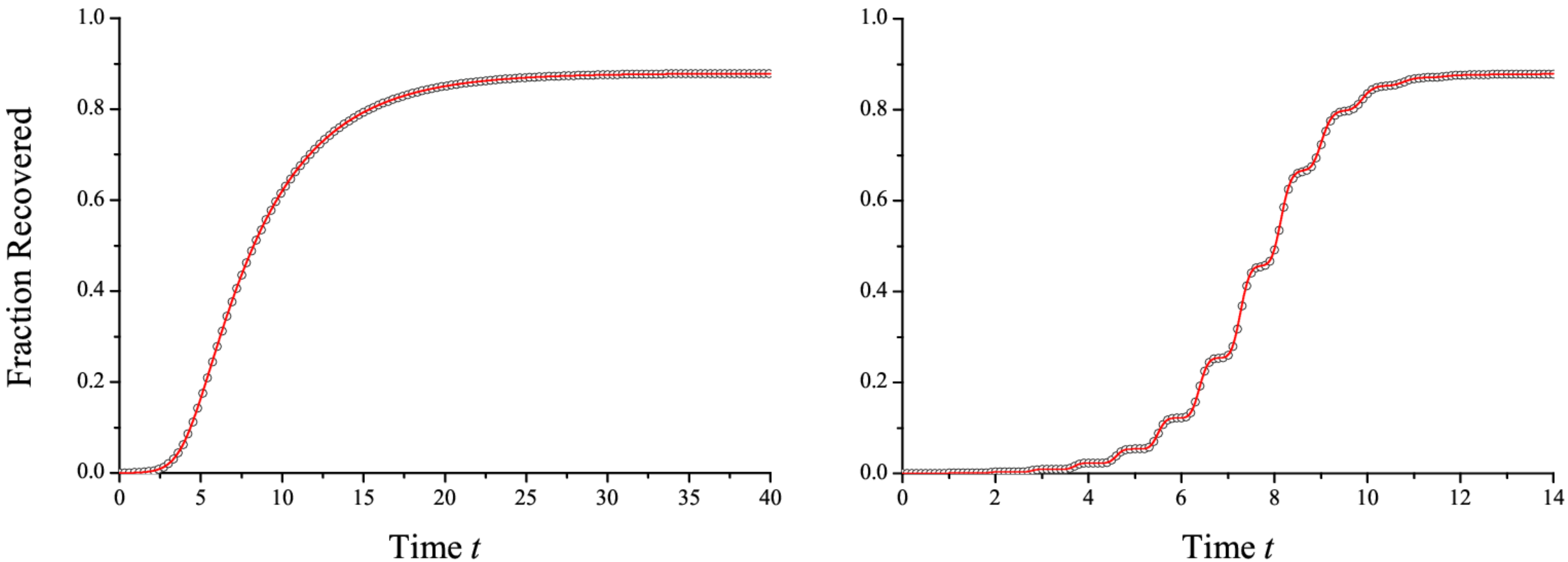
- Define a dynamic message $u_{i\to j}(t)$ equal to the probability that i does not transmit disease to j before time t

$$u_{i\to j}(t) = 1 - \int_0^t p(\tau)\,\mathrm{d}\tau + \int_0^t p(\tau) \prod_{\substack{k\in\mathcal{N}(i)\\ k\neq j}} u_{k\to i}(t-\tau)\,\mathrm{d}\tau$$



- Note that our earlier percolation process is just the special case of this equation when $t \to \infty$
- Just set $u_{i\to j} = u_{i\to j}(\infty)$ and $p = \int_0^\infty p(\tau)\,\mathrm{d}\tau$

- We can use this expression in various ways:
  - We can use it to give the behavior on a single network
  - Or we can average over networks within some ensemble



- There are some disadvantages:
  - The messages are now functions, which have to be represented somehow
  - It can be quite numerically intensive
  - It's only exact for locally tree-like networks
- Message passing has promise for all sorts of problems on networks:
  - Origin of epidemics
  - Model fitting, graphical models
  - Coloring, max-cut, and other optimization problems
  - Spin models

Thanks

Brian Karrer

- Published as B. Karrer and MEJN, *Phys. Rev. E* **82**, 016101 (2010)

- From the National Science Foundation and the McDonnell Foundation