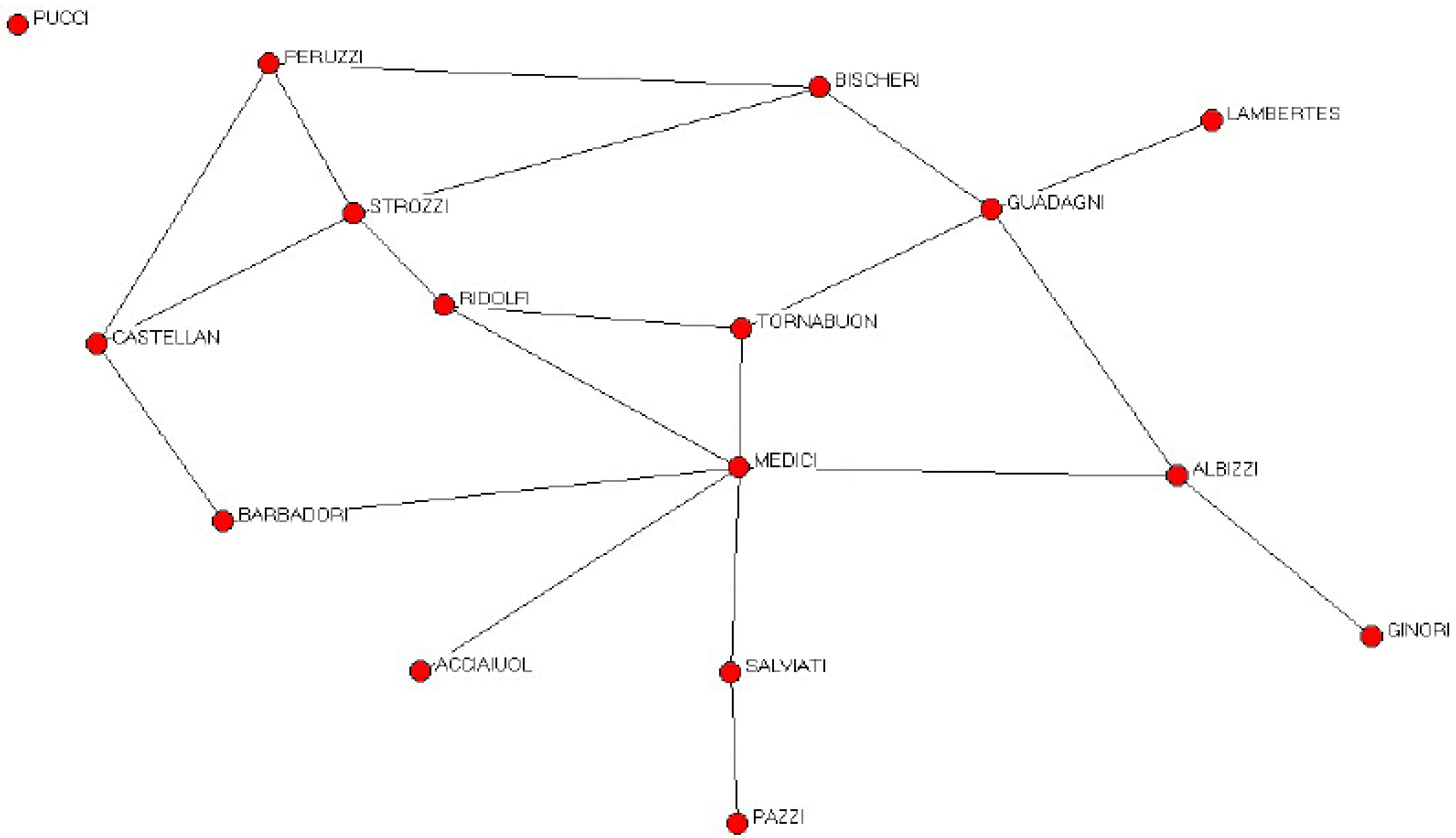
Diffusion and Strategic Interaction on Social Networks

Leeat Yariv

Summer School in Algorithmic Game Theory, Part1, 8.6.2012

Why Networks Matter

- 15th Century Florentine Marriages (Padgett and Ansell, 1993)



Why Networks Matter – Florence

- Why are the Medici ("godfathers of the Renaissance") so strong?
- Prior to the 15th century, Florence was ruled by an oligarchy of elite families
- Notably, the Strozzi had greater wealth and more seats in the state legislature, and yet were eclipsed by the Medici

Why Networks Matter – Florence

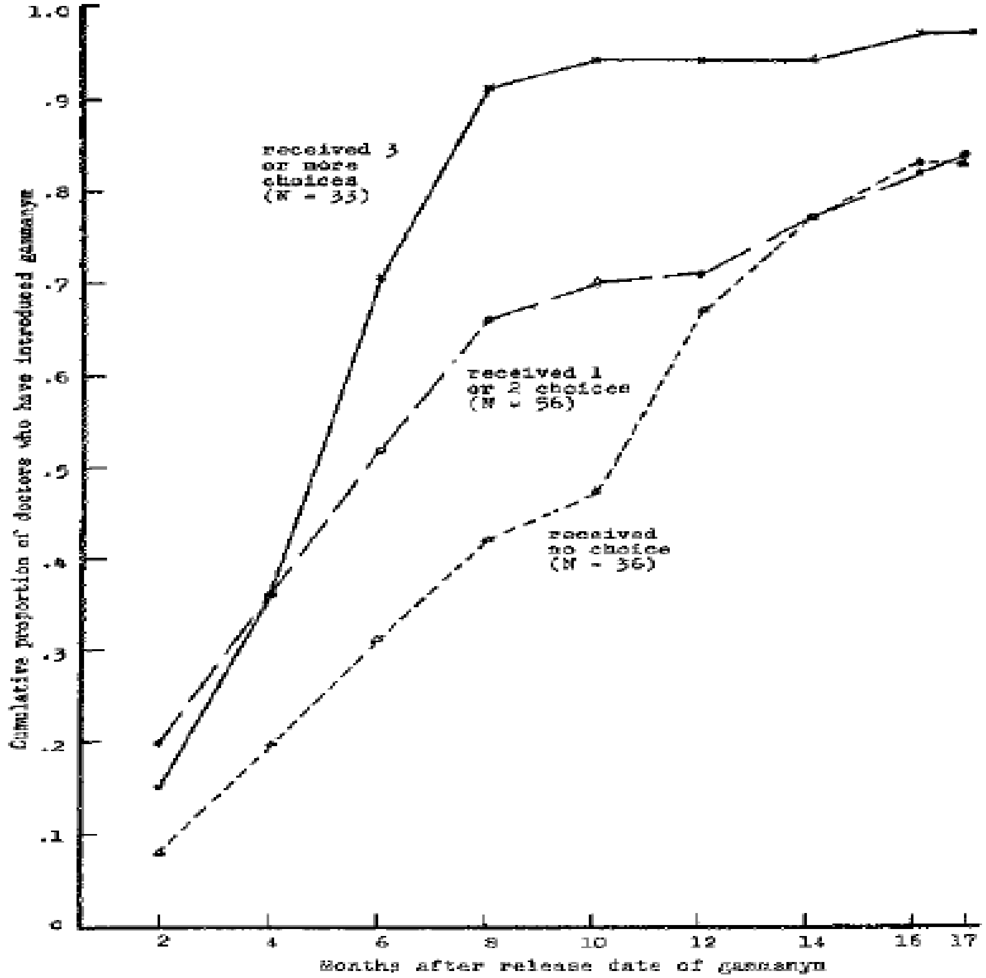
- Several notable characteristics of the marriage network (drawn for 1430):
  - High *degree*, number of connected families, but higher only by 1 relative to Strozzi or Guadagni.
  - Let P(i,j) denote the number of shortest paths between families i and j and let $P_k(i,j)$ the number of these that include k.
    - Note that the Medici are key in connecting Barbadori and Guadagni.
  - To get a general sense of importance, can look at an average of this *betweeness* calculation. Standard measure:

$$\sum_{i \neq j, k \notin \{i,j\}} \frac{P_k(i,j)/P(i,j)}{(n-1)(n-2)/2}$$

- Medici – 0.522, Strozzi – 0.103, Guadagni – 0.255.

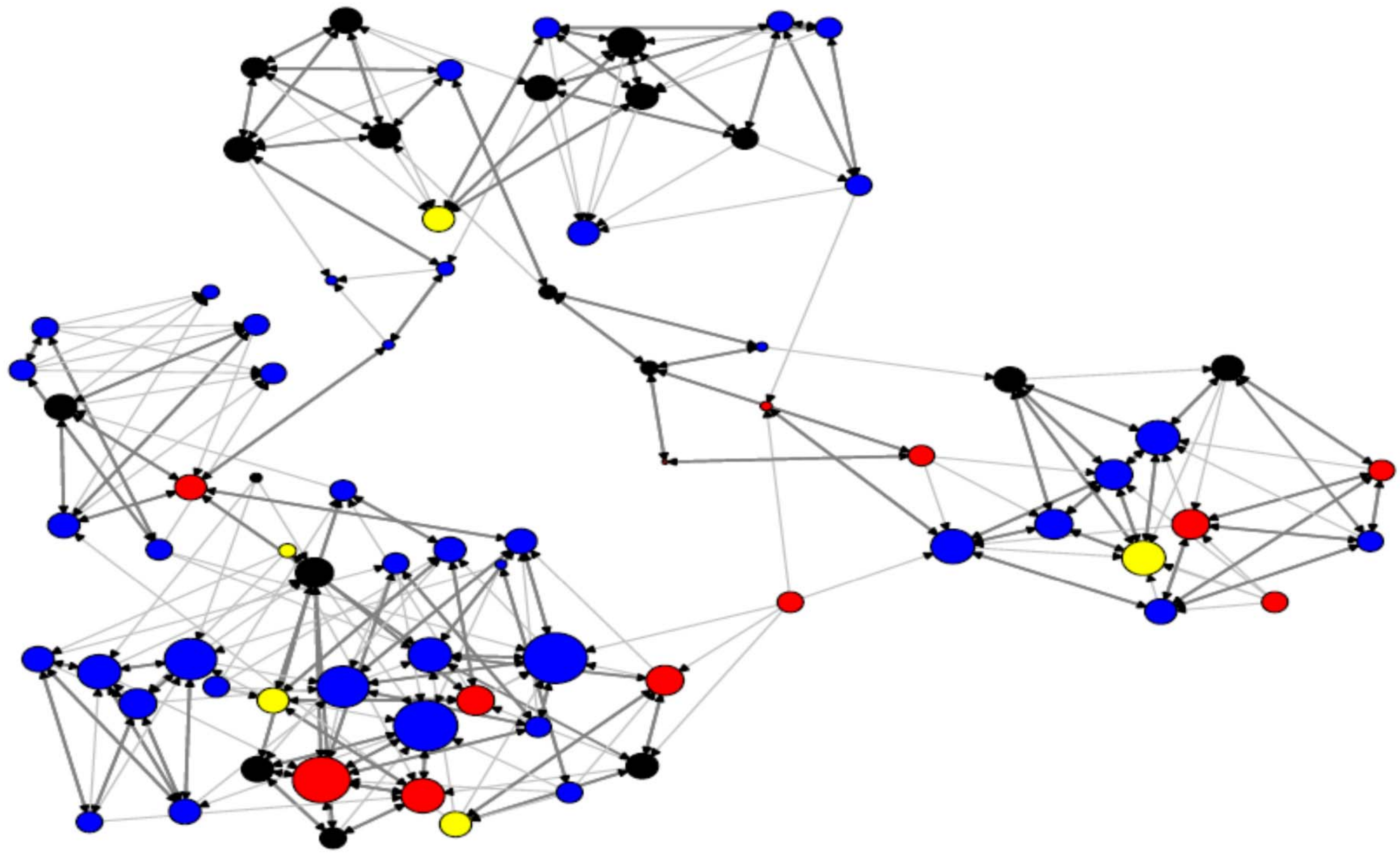
Why Networks Matter

- Diffusion, e.g., Tetracycline adoption (Coleman, Katz, and Menzel, 1966) :

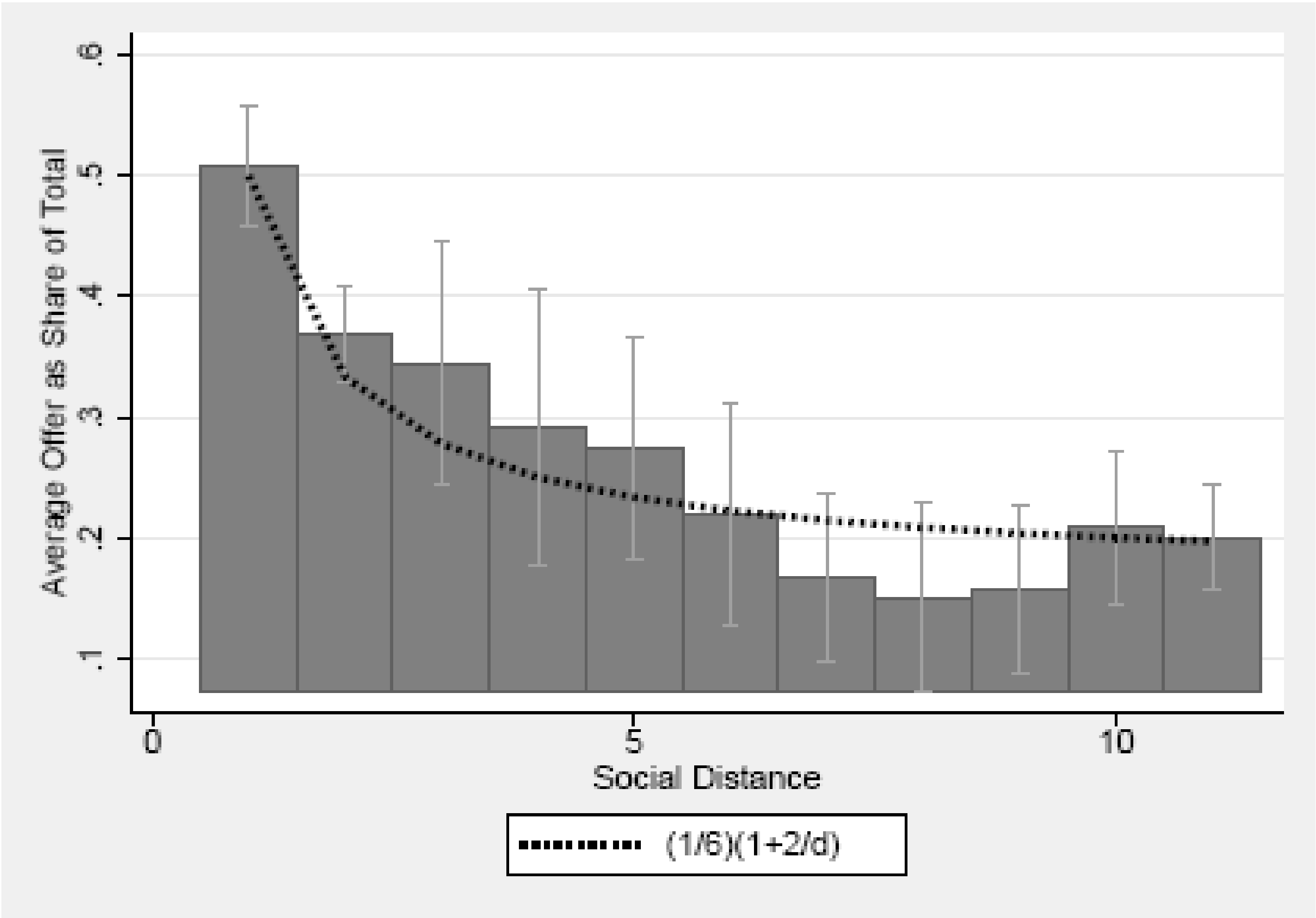


Why Networks Matter

- Giving behavior (Goeree, McConnell, Mitchell, Tromp, Yariv, 2009)



1/d Law of Giving



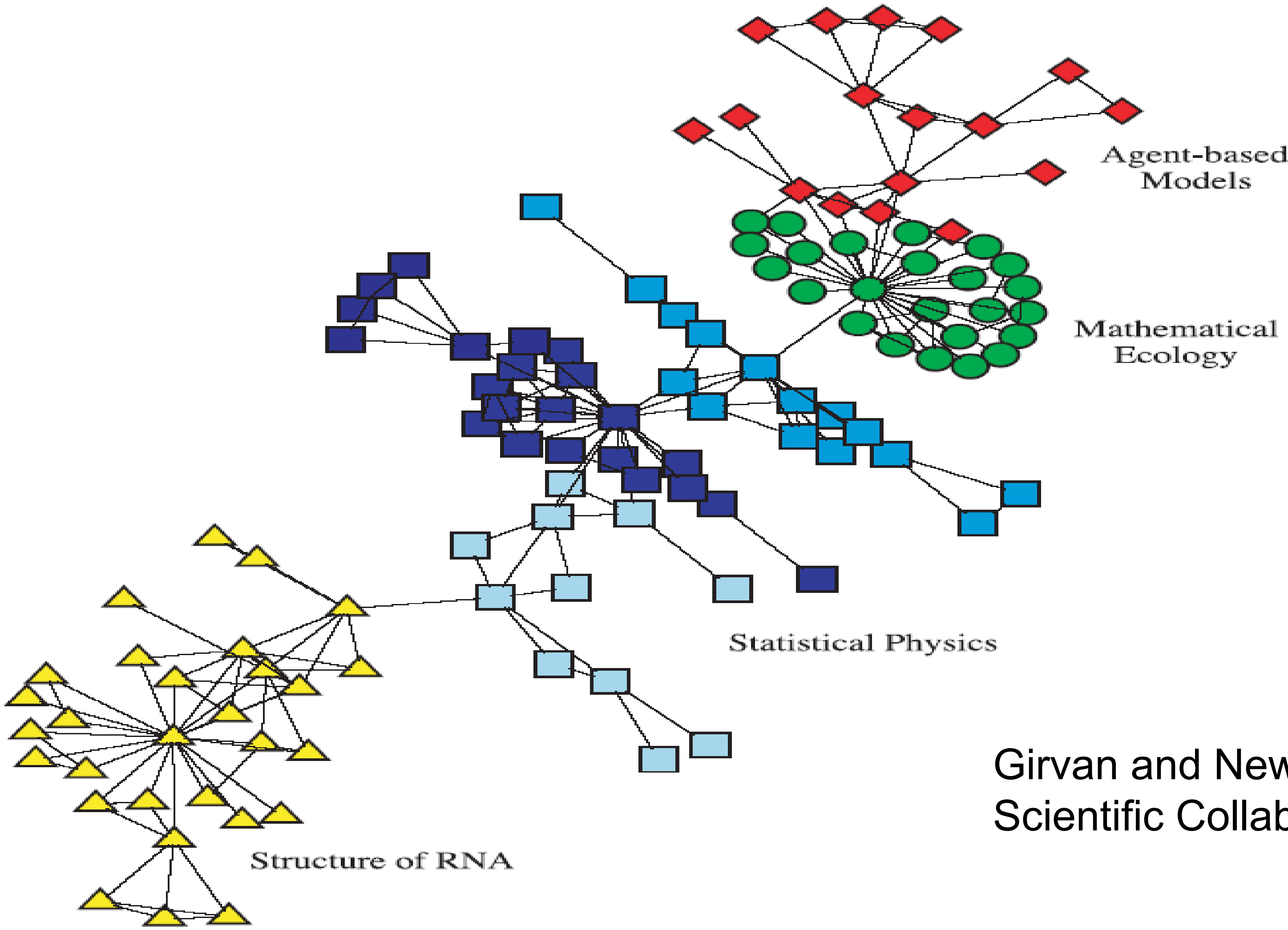
Why Networks Matter

- Matching with Network Externalities – dorms and students, faculty and offices, firms and workers, etc.
- Epidemiology – whom to vaccinate, what populations are more fragile to an epidemic, etc.
- Marketing – whom to target for advertizing, how do products diffuse, etc.
- Development – how to design micro-credit programs utilizing network information.

Why Networks Matter

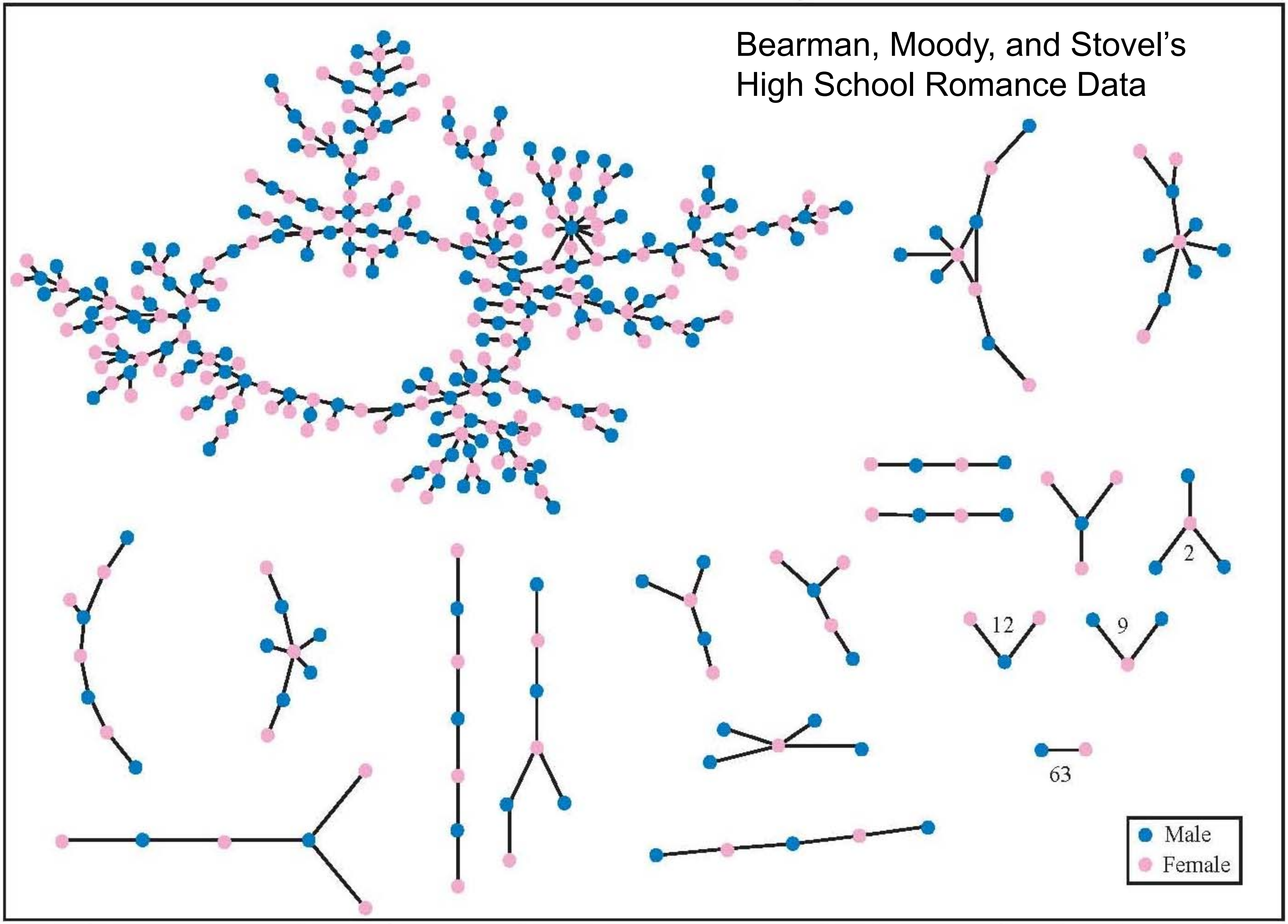
- Matching with Network Externalities – dorms and students, faculty and offices, firms and workers, etc.
- Epidemiology – whom to vaccinate, what populations are more fragile to an epidemic, etc.
- Marketing – whom to target for advertizing, how do products diffuse, etc.
- Development – how to design micro-credit programs utilizing network information.
- [[Social = "Social", agents can stand for individuals, computers, avatars, etc.]]

Networks have very different structures



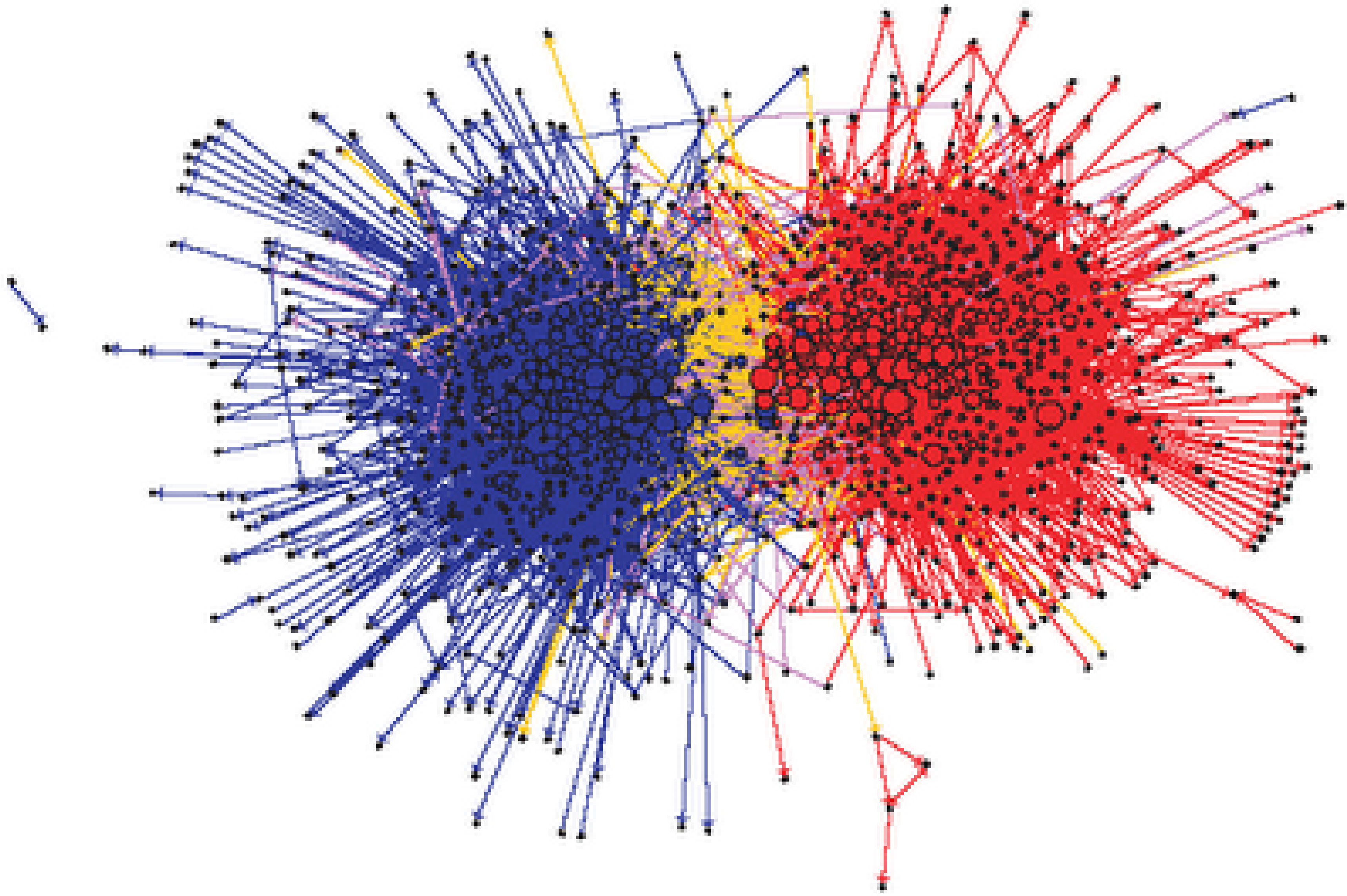
Girvan and Newman's
Scientific Collaboration Data

The Structure of Romantic and Sexual Relations at "Jefferson High School"



Bearman, Moody, and Stovel's High School Romance Data

Political Blogosphere (Adamic and Glance, 2005)

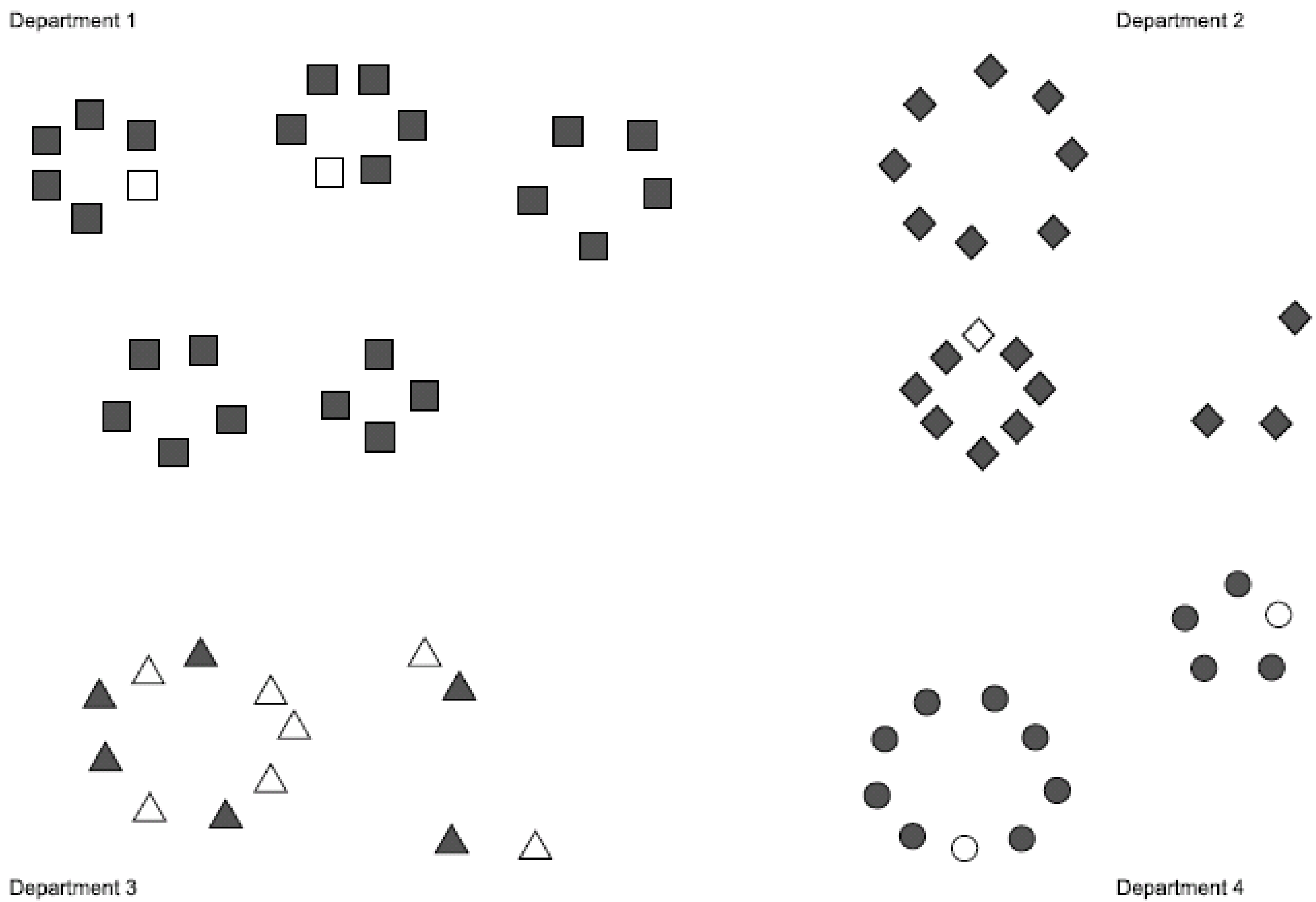


Networks have very different structures

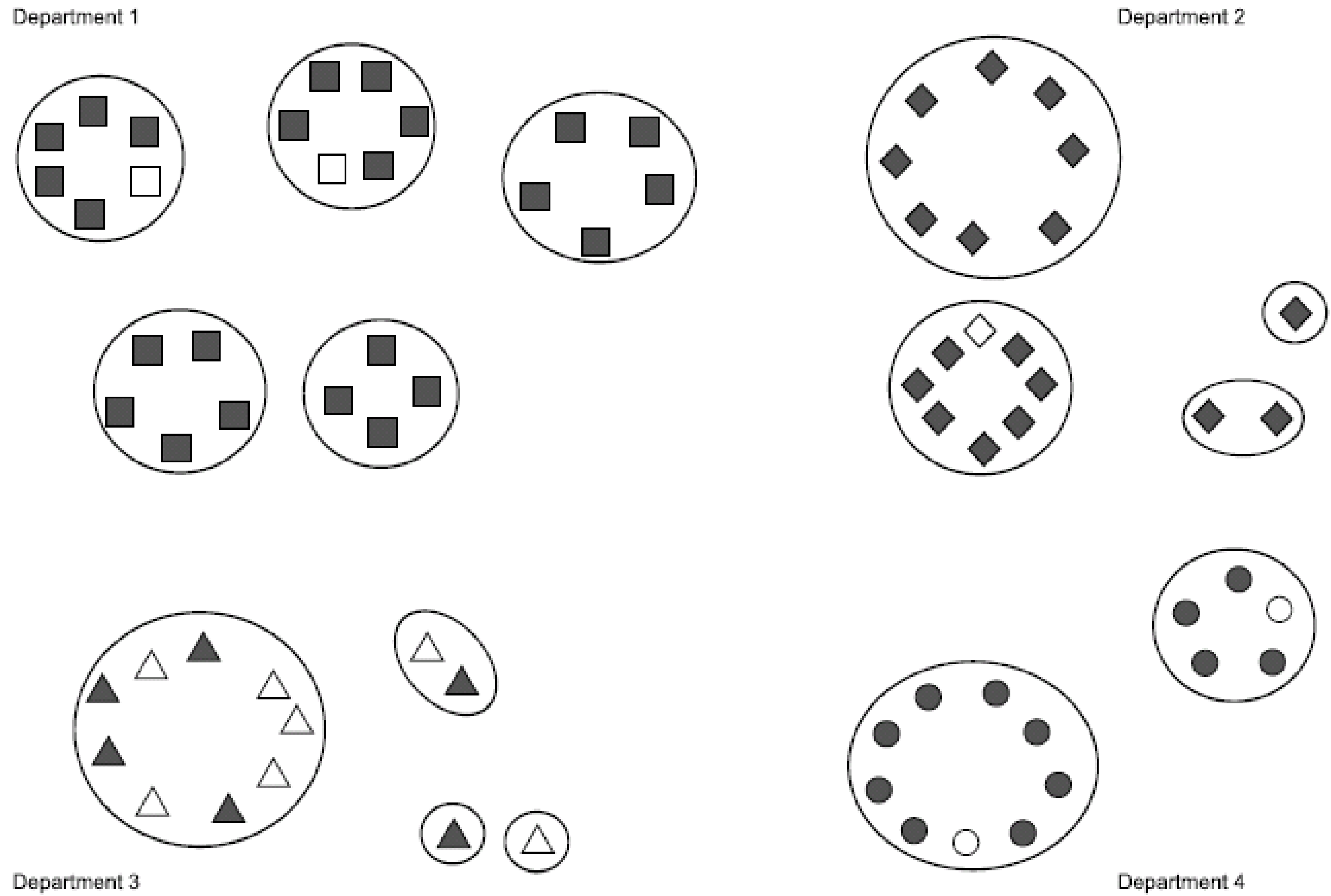
- Depending on which layer we look at

- Consider faculty at a professional school in the U.S. (Baccara, Imrohoroglu, Wilson, and Yariv, 2012):
  - Institutional
  - Social
  - Co-authorship

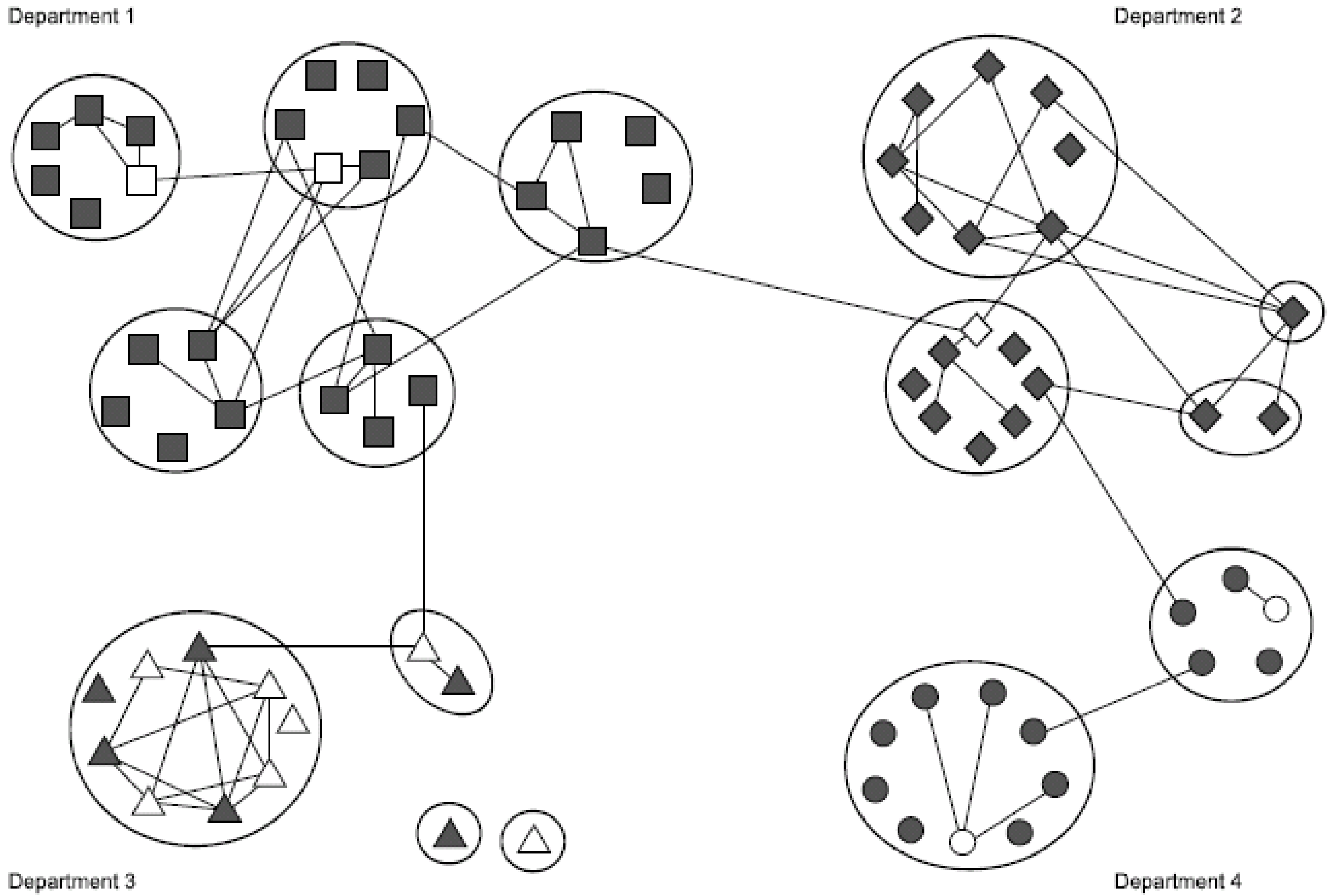
Department



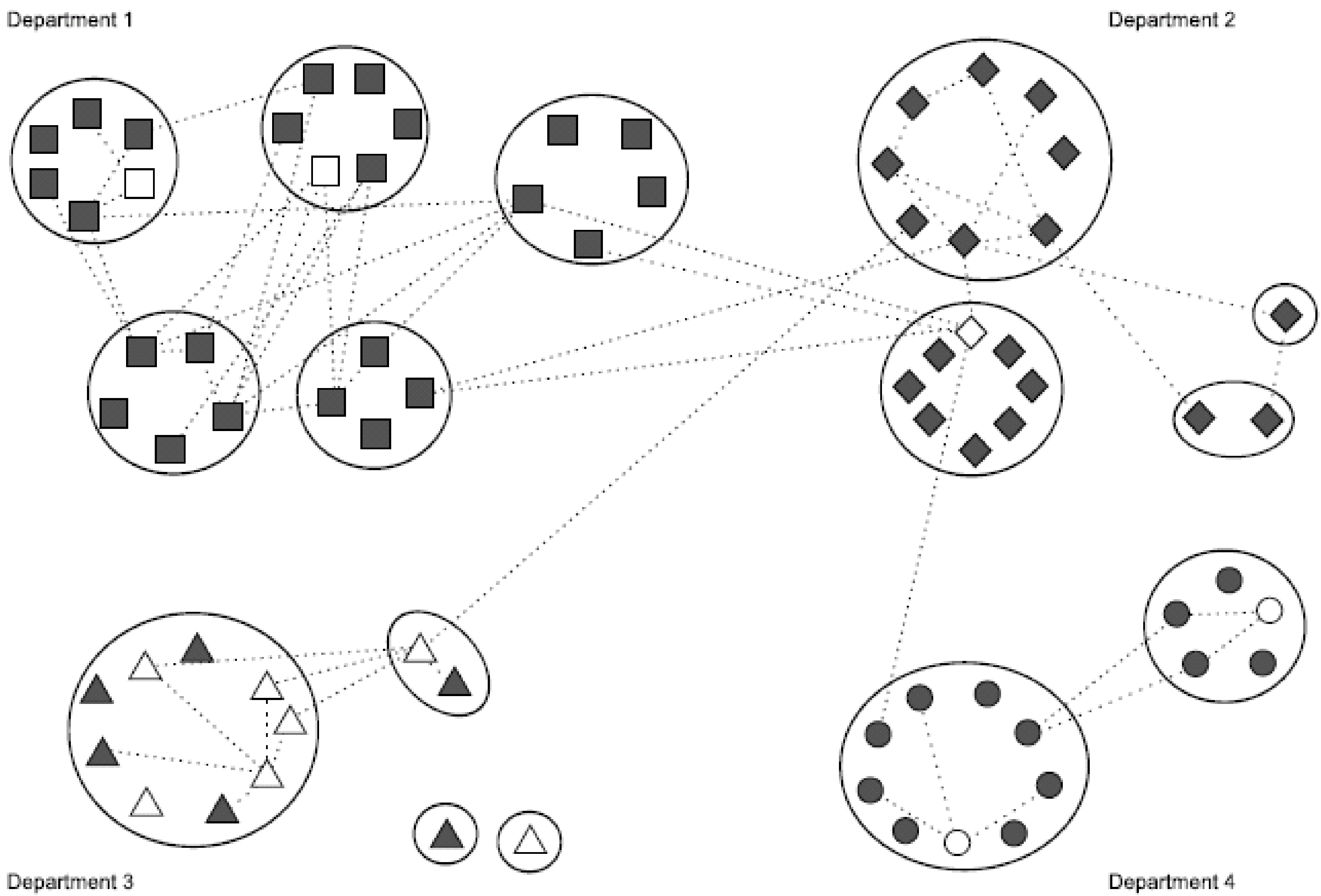
Research field



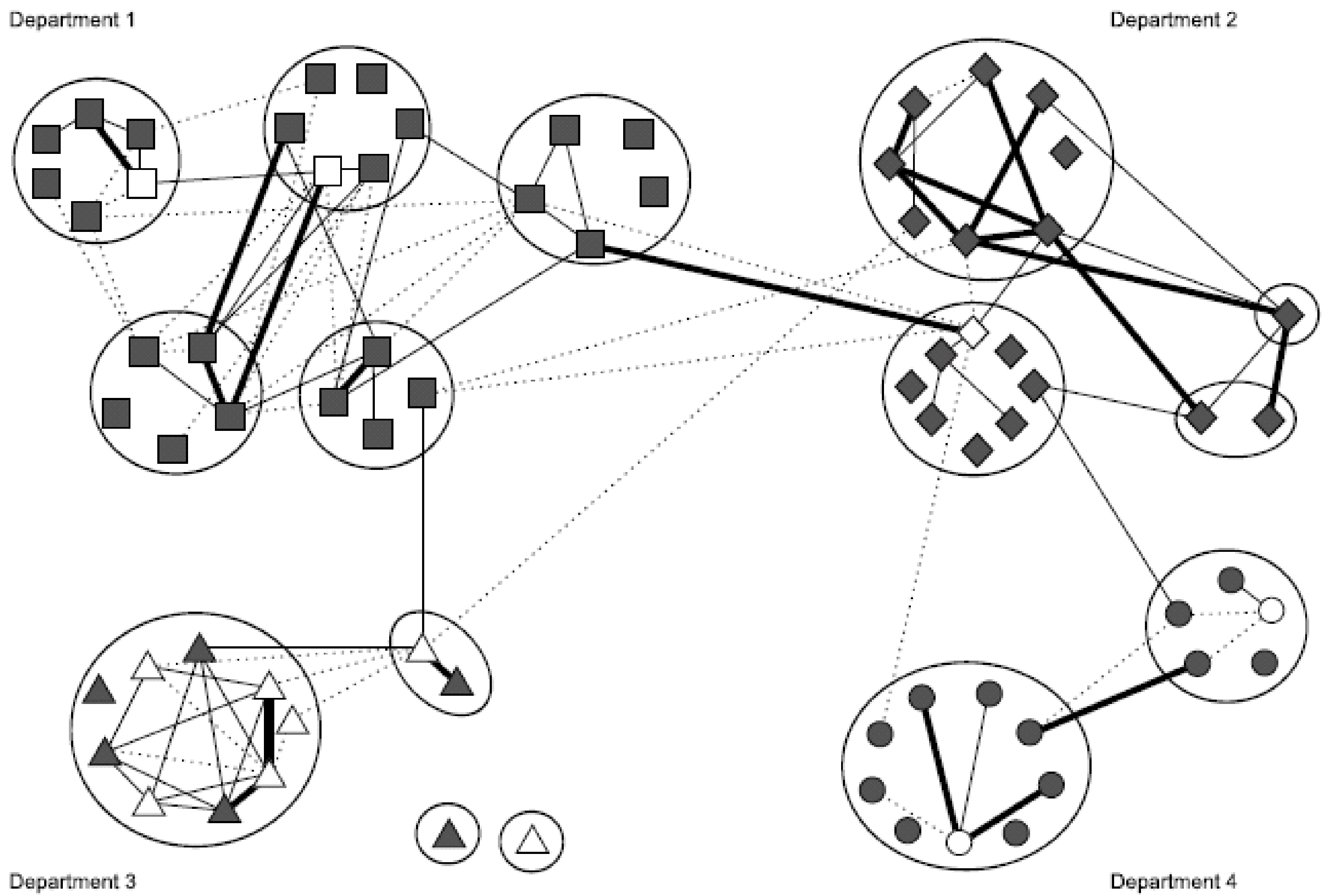
Coauthorships



Friendships



Composite



The Big Questions

- How does the structure of networks impact outcomes:
  - In different locations within the network and across different network architectures
  - Static and dynamic

- How do networks form to begin with (given the interactions that occur over them)

All that in three hours?!

- Basic notions of networks
- diffusion models for pedestrians
- More general games played on networks
- (if time) Basic group formation model

Caveats

- Talks biased toward my own work
- They are more economically oriented (we care a lot about welfare, less about complexity)
- You’re still welcome to complain and ask questions!
- A great read: Jackson (2008)

Summarizing Networks

- N={1,…,n} individuals, vertices, nodes, agents, players

Summarizing Networks

- N={1,…,n} individuals, vertices, nodes, agents, players

- g is (an undirected) network (in $\{0,1\}^{nxn}$):

$$g_{ij} = \begin{cases} 1 & i \neq j \text{ connected} \\ 0 & \text{otherwise} \end{cases}$$

Summarizing Networks

- $N=\{1,\ldots,n\}$ individuals, vertices, nodes, agents, players
- g is (an undirected) network (in $\{0,1\}^{n\mathrm{x}n}$):

$$g_{ij} = \begin{cases} 1 & i \neq j \text{ connected} \\ 0 & \text{otherwise} \end{cases}$$

- $N_i(g)$ i's neighborhood,

$$N_i(g) = \{ j \mid g_{ij} = 1 \}$$

Summarizing Networks

- N={1,…,n} individuals, vertices, nodes, agents, players

- g is (an undirected) network (in $\{0,1\}^{n\times n}$):

$$g_{ij} = \begin{cases} 1 & i \neq j \text{ connected} \\ 0 & \text{otherwise} \end{cases}$$

- $N_i(g)$ i's neighborhood,

$$N_i(g) = \{ j \mid g_{ij} = 1 \}$$

- $d_i(g)=|N_i(g)|$ i's degree

Examples

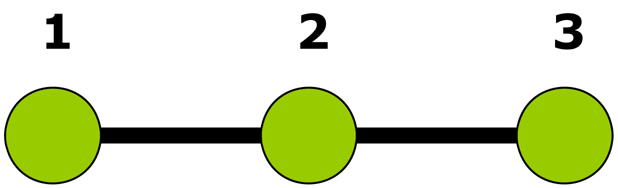
- The line

$$g = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

Examples

- The line

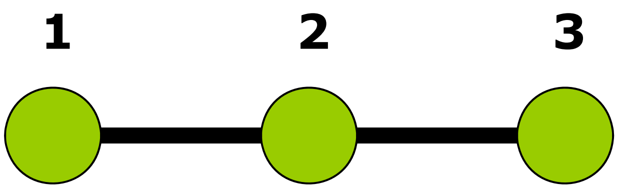
$$g = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$



Examples

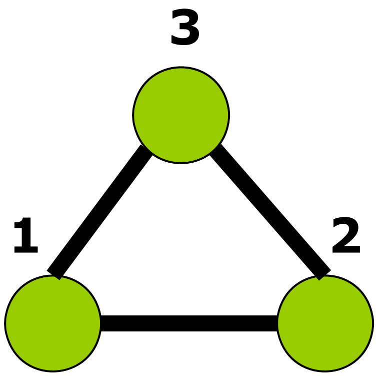
- The line

$$g = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$



- The triangle (special case of a circle...)

$$g = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$



Degree Distributions

- P(d) – frequency of degree d nodes

- **Examples**:

  1. **Regular network** – $P(k)=1$, $P(d)=0$ for all $d \neq k$.

  2. **Complete network** – $P(n)=1$.

Erdos-Renyi (or Poisson) Networks

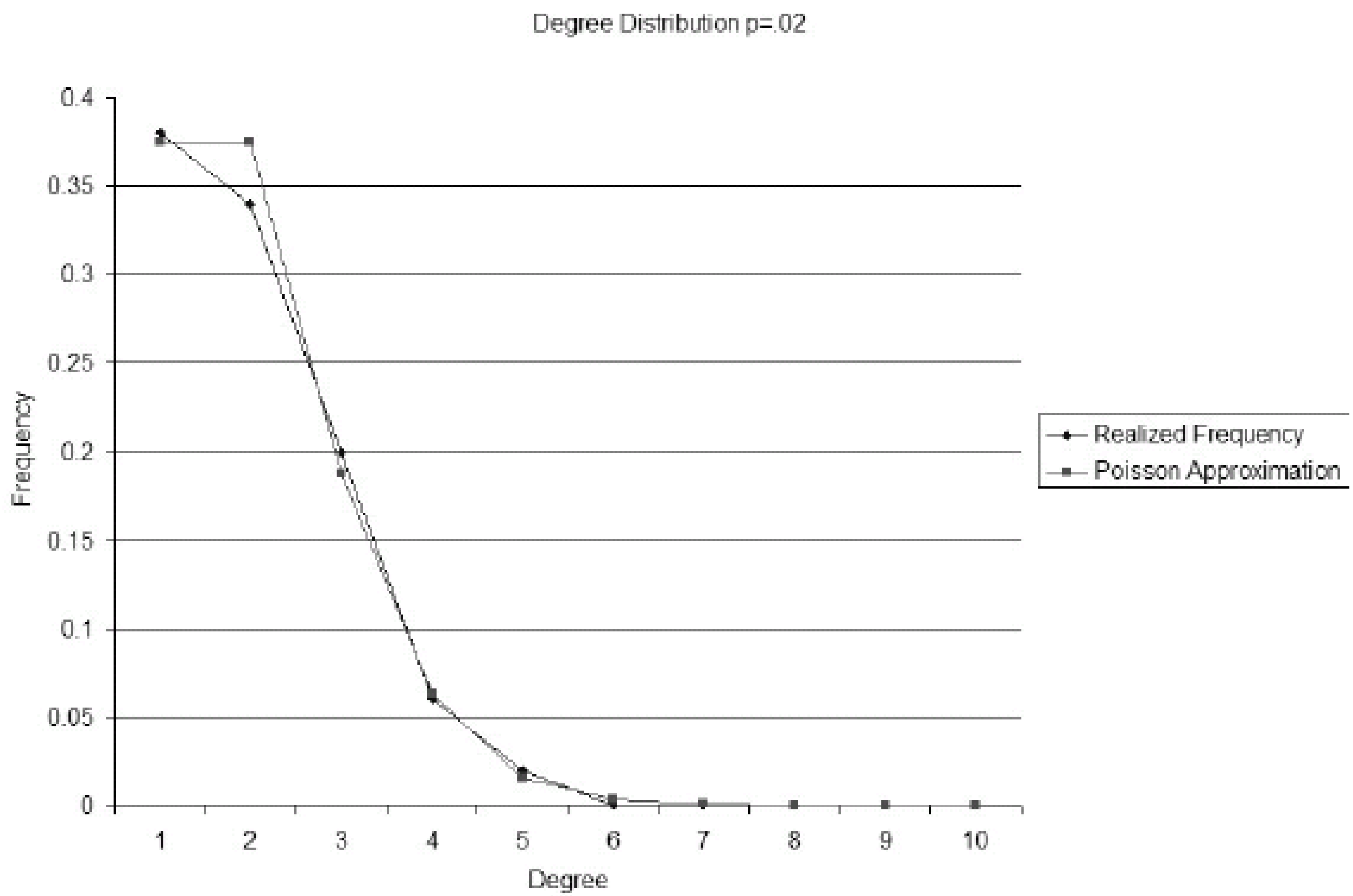
- Erdos and Renyi (1959, 1960, 1961) – some of the first to discuss random networks.

- Each link is formed with probability p

$$P(d) = \binom{n-1}{d} p^d (1-p)^{n-1-d}$$

- For large n,

Poisson Network



“Phase Transitions” in Poisson Networks

-
- Pick parameters so that only one isolated node (with degree 0) on average:

$$e^{-(n-1)p} = \frac{1}{n} \quad \leftrightarrow \quad p(n-1) = \ln(n)$$

- For example, n=50 → $p = \frac{\ln(50)}{49} = 0.0798$

Poisson Network

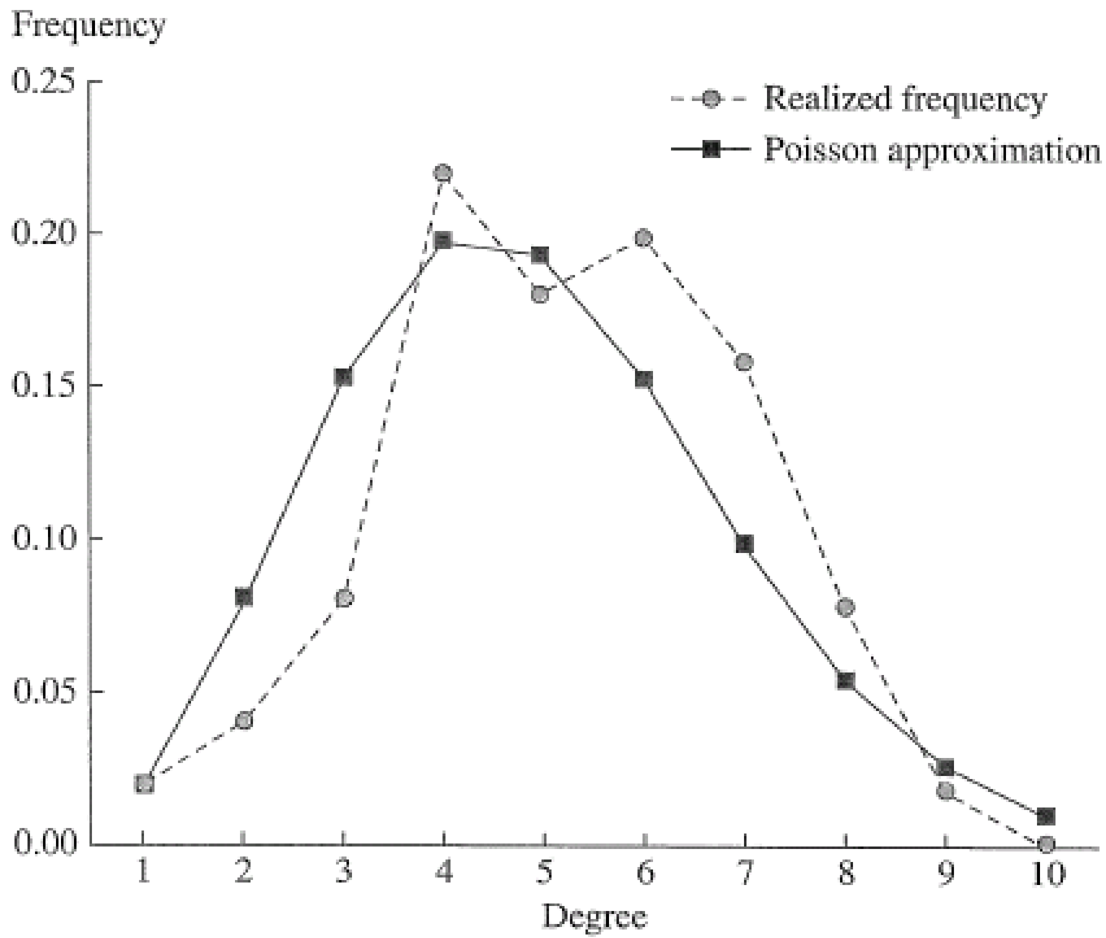


FIGURE 1.7 Frequency distribution of a randomly generated network and the Poisson approximation for a probability of .08 on each link.

Coauthorships and Poisson

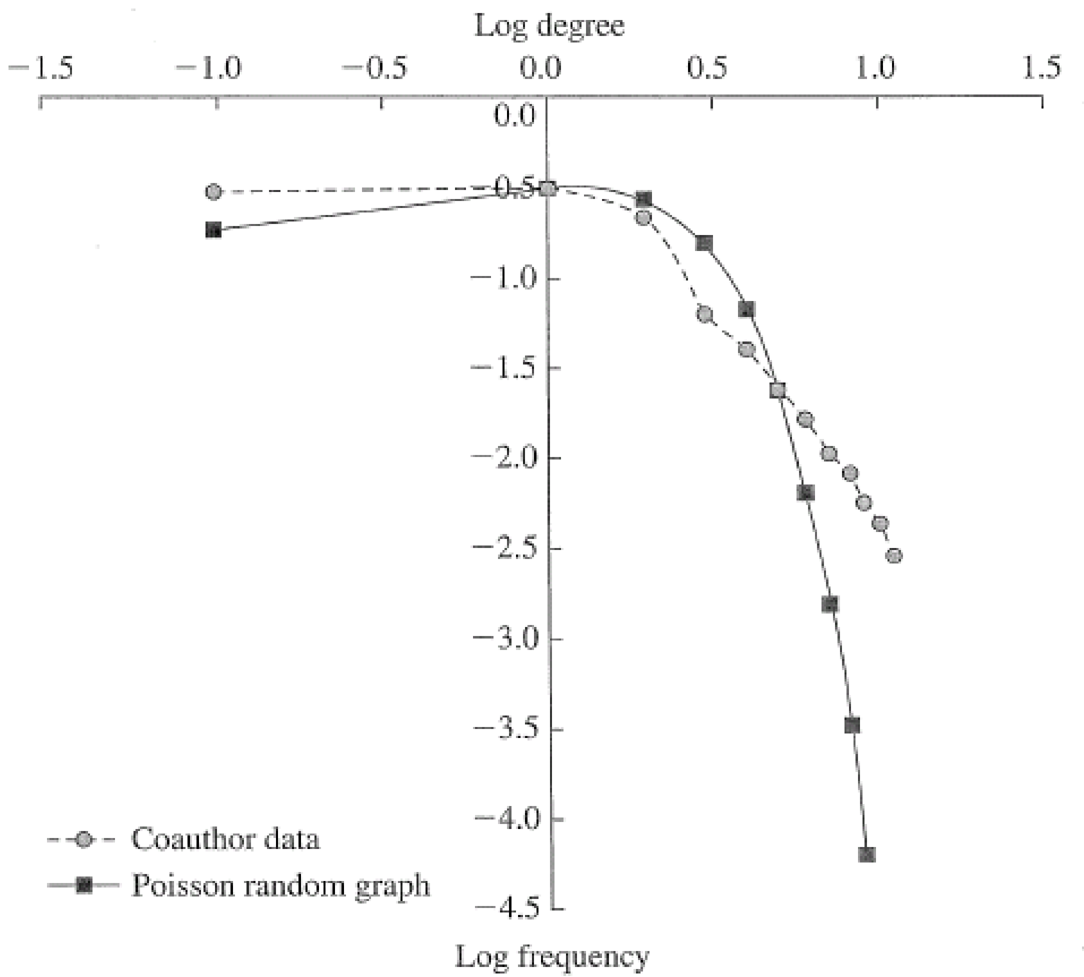


FIGURE 3.1 Comparison of the degree distributions of a coauthorship network and a Poisson random network with the same average degree.

Notre Dame and Poisson

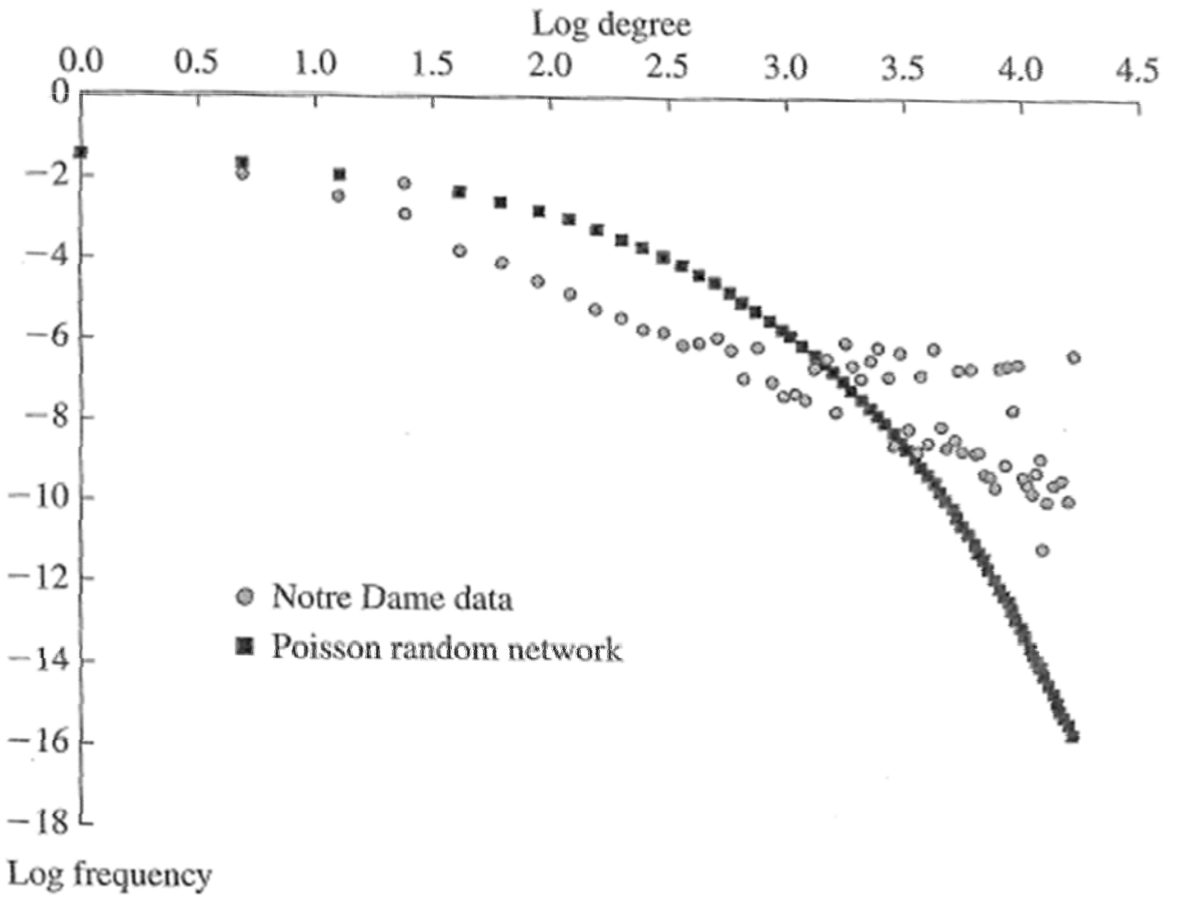


FIGURE 3.2 Distribution of in-degrees of Notre Dame web site domain from Albert, Jeong, and Barabási [9] compared to a Poisson random network.

Scale-free Distributions

- $P(d) = cd^{-\gamma},\ \ c > 0\ (\gamma \in [2,3]$ often)

Scale-free Distributions

- $P(d) = cd^{-\gamma}, \;\; c > 0 \; (\gamma \in [2,3]$ often)
- Note that $\frac{P(2)}{P(1)} = \frac{P(20)}{P(10)} = \cdots$ (hence, scale-free)
- Often called power laws

Scale-free Distributions

- $P(d) = cd^{-\gamma}, \;\; c > 0 \; (\gamma \in [2,3]$ often)
- Note that $\frac{P(2)}{P(1)} = \frac{P(20)}{P(10)} = \cdots$ (hence, scale-free)
- Often called power laws
- Notice that:

$$logP(d) = \log(c) - \gamma \log(d)$$

Scale-Free and Poisson

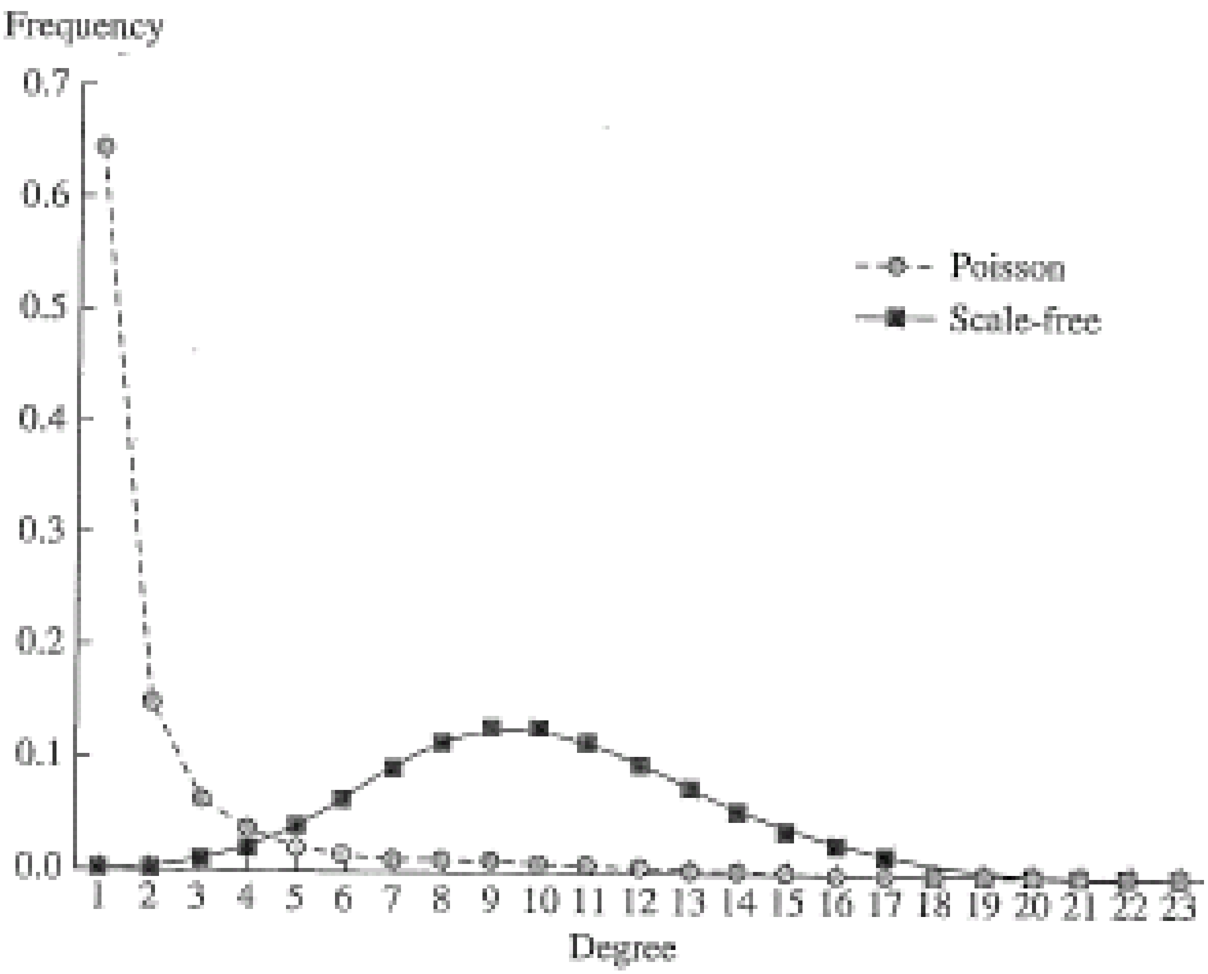


FIGURE 2.8 Comparing a scale-free distribution to a Poisson distribution.

Scale-Free and Poisson

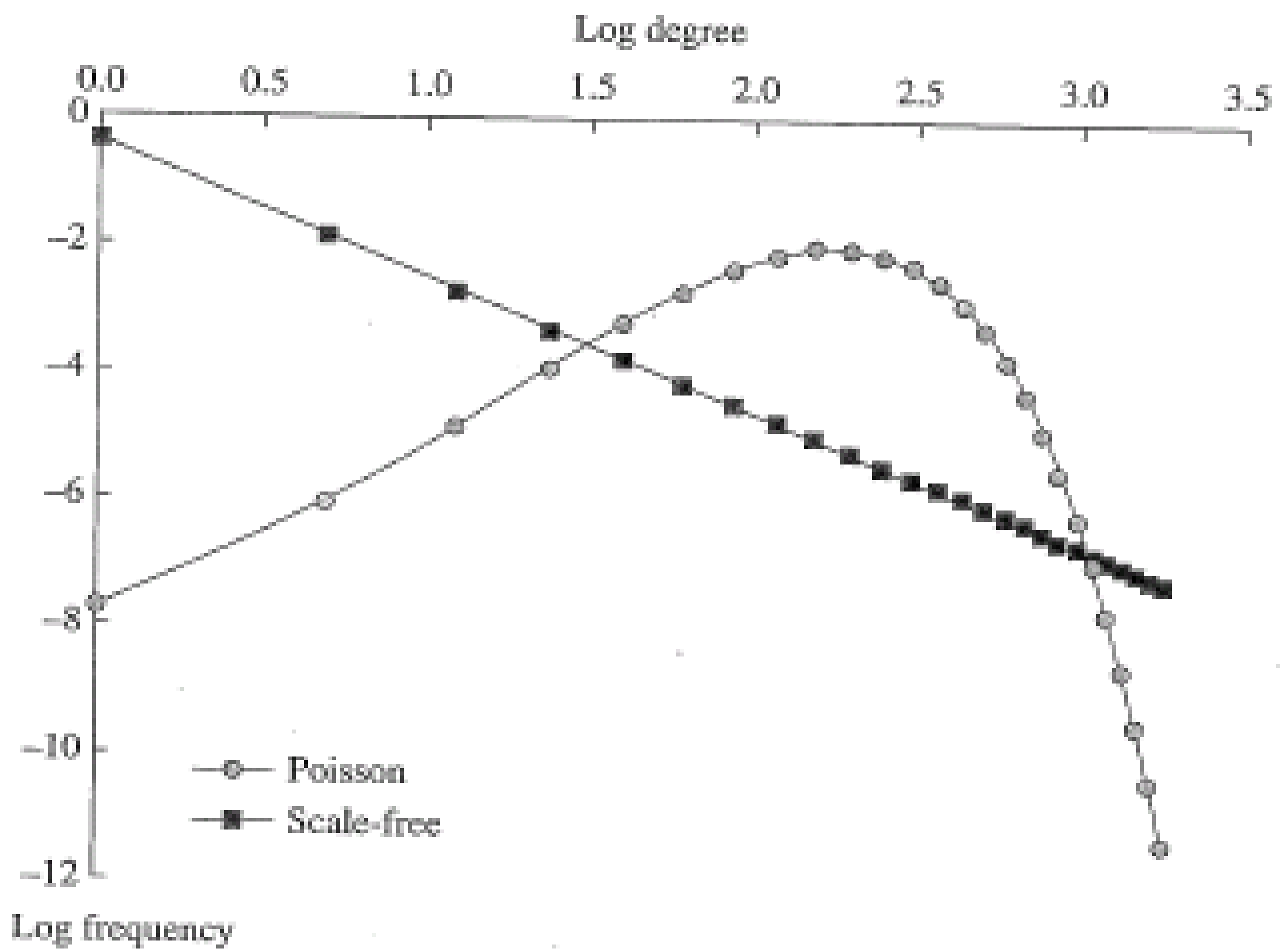
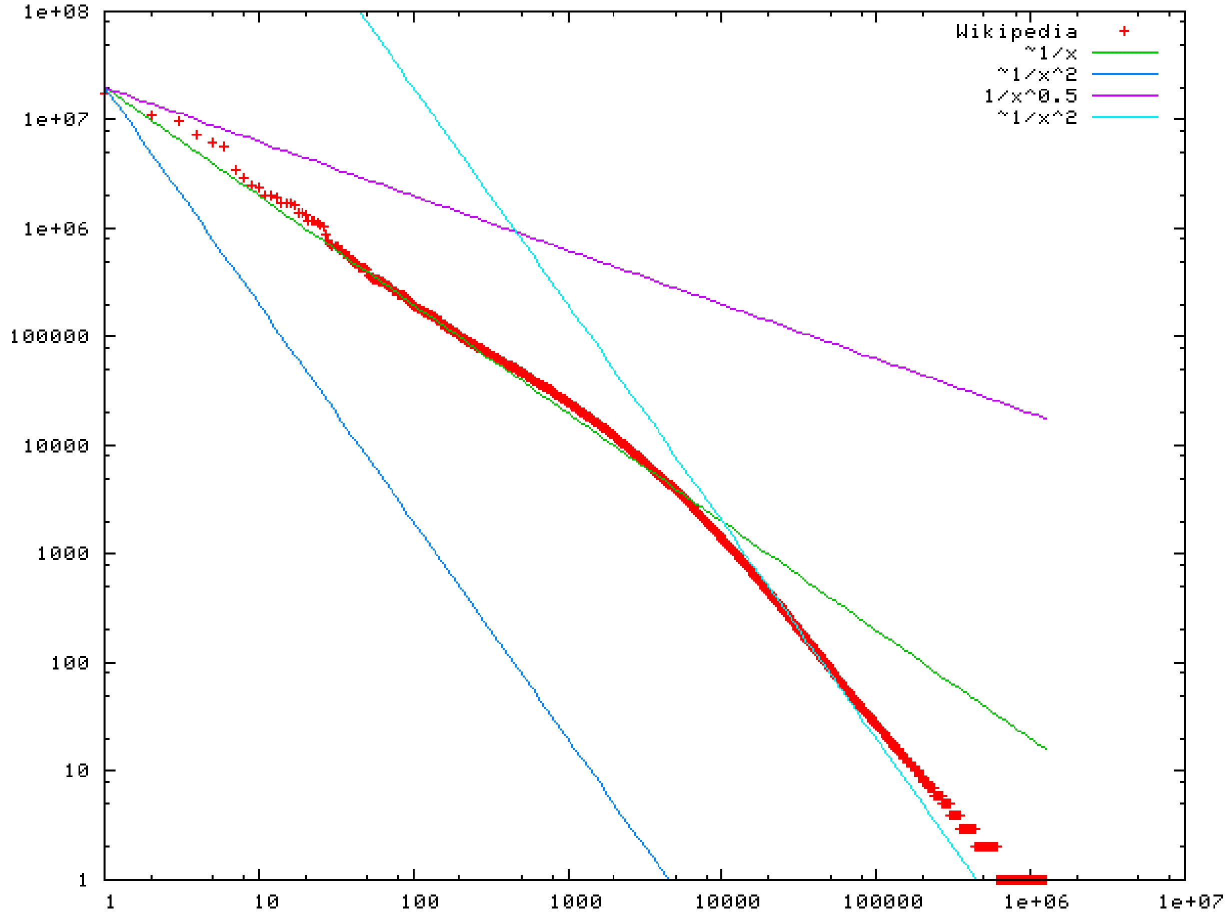


FIGURE 2.9 Comparing a scale-free distribution to a Poisson distribution: log-log plot.

Zipf's Law – Word Frequency in Wikipedia (November 27, 2006)



Zipf's Law for Cities

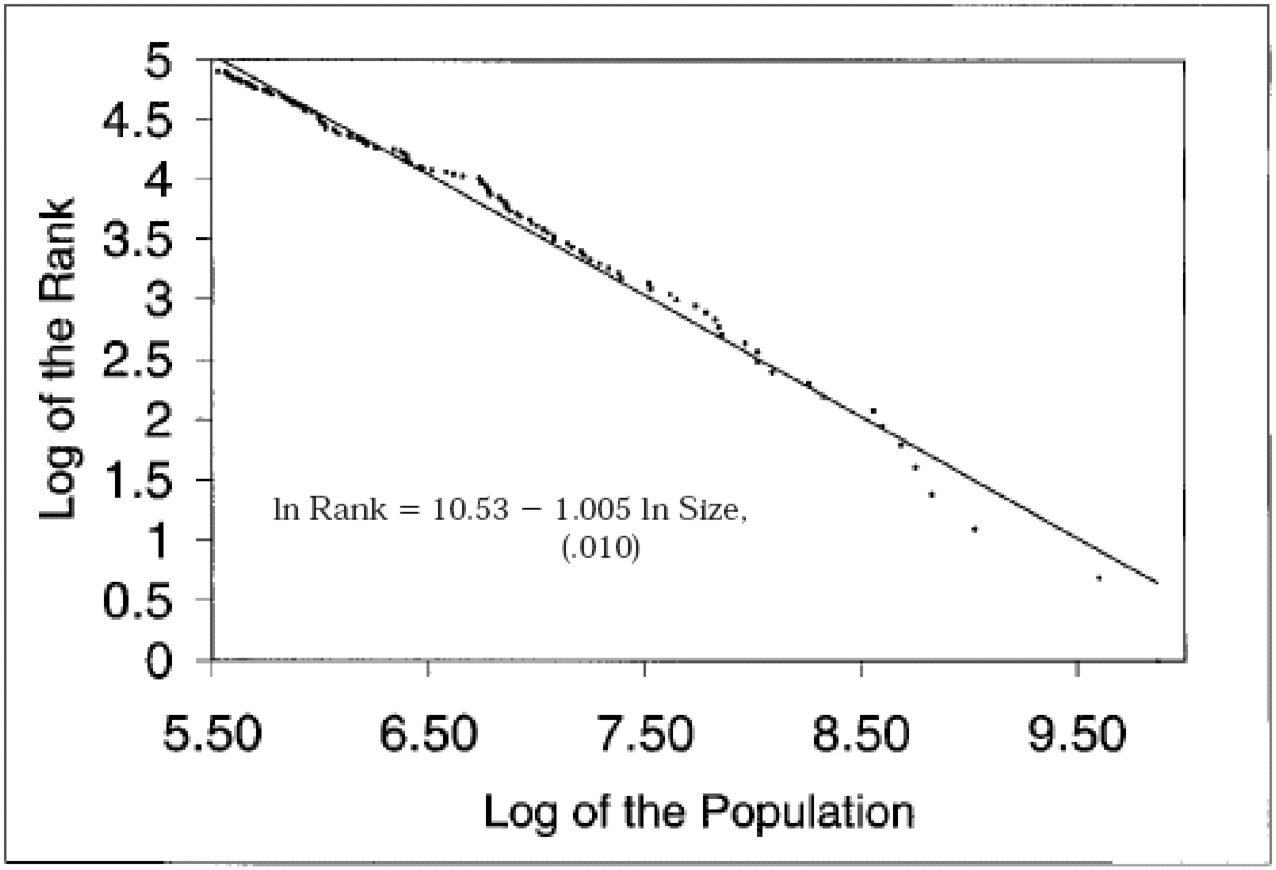
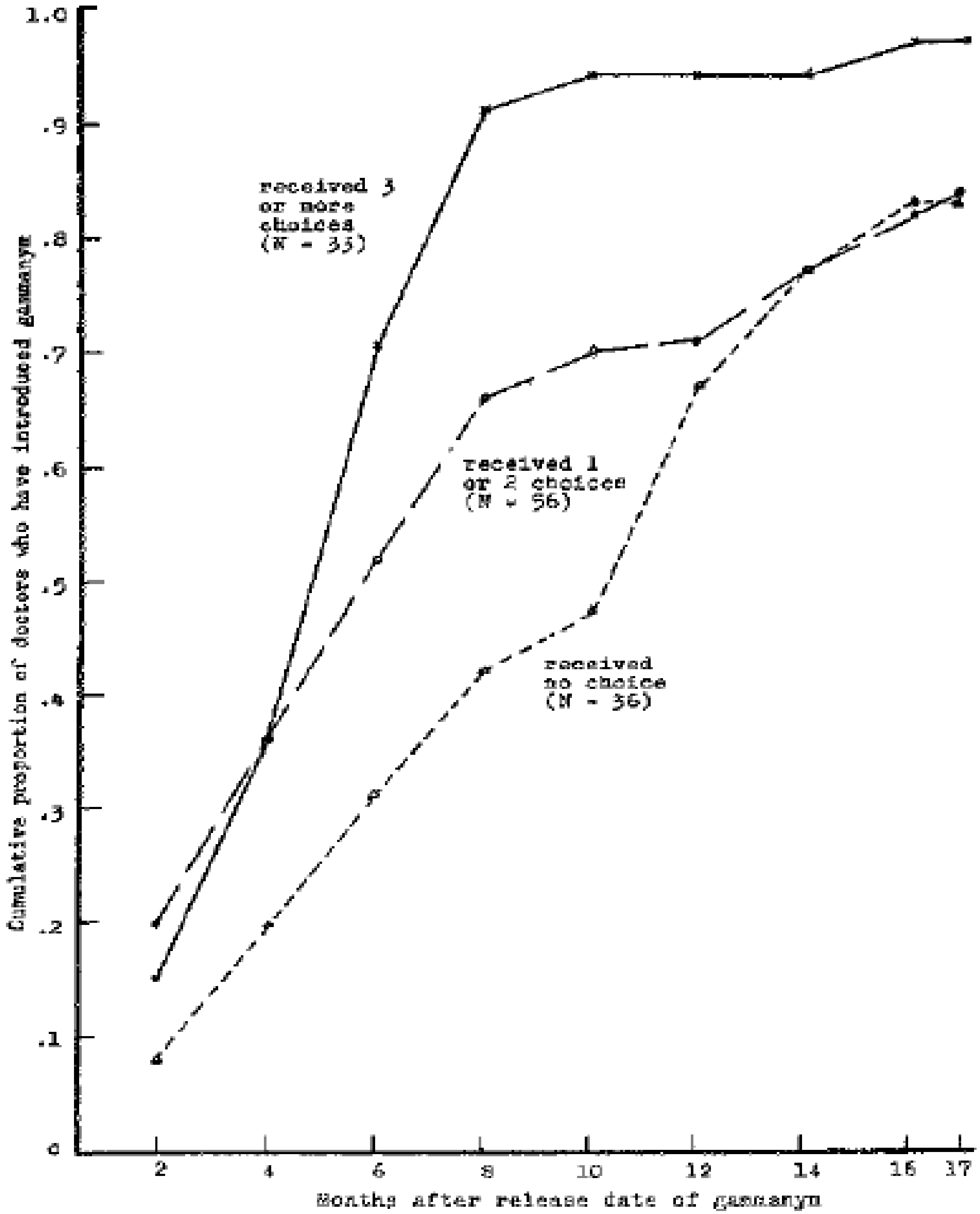


FIGURE I
Log Size versus Log Rank of the 135 largest U. S. Metropolitan Areas in 1991
Source: Statistical Abstract of the United States [1993].

Diffusion on Social Networks

- Literature precedes that of static games on social networks (though connected)
- Relevant for many applictions:
  - Epidemiology (human and technological…)
  - Learning of a language (human and technological..)
  - Product marketing
  - Transmission of information

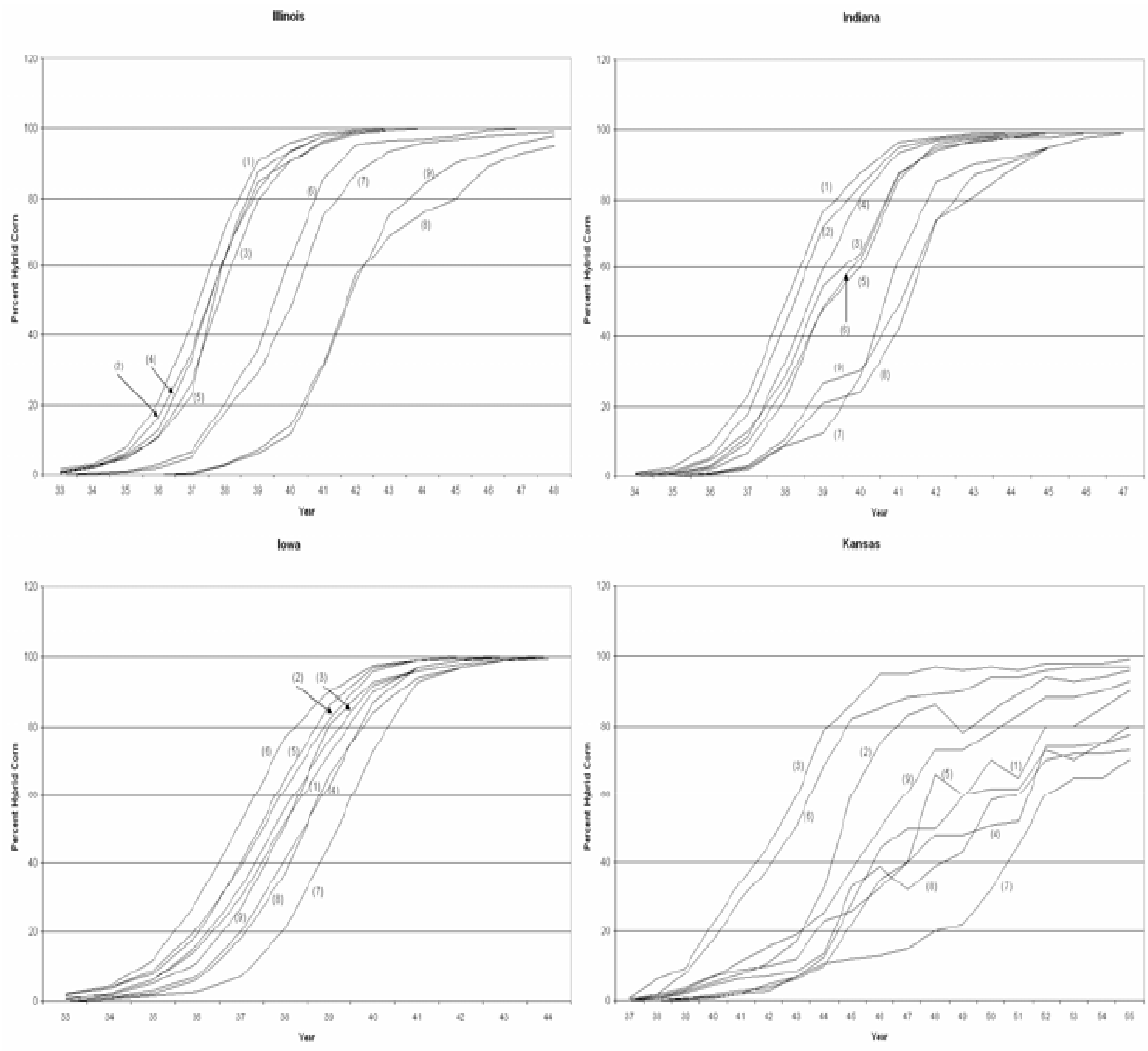


Tetracycline Adoption

(Coleman, Katz, and Menzel, 1966)

Hybrid Corn, 1933-1952

(Griliches, 1957, and Young, 2006)



Main Observations

- In 1962, Everett Rogers compiles 508 diffusion studies in *Diffusion of Innovation*
- S-shaped adoption
- Different speeds of adoption for different degree agents

The Bass (1969) Model

- Ideas from Tarde (1903)
- G(t) – percentage of agents who have adopted by time t
- m – potential adopters in the population

The Bass (1969) Model

- G(t) – percentage of agents who have adopted by time t
- m – potential adopters in the population

$$G(t) = G(t-1) + p\big(m - G(t-1)\big) + q(m - G(t-1))\frac{G(t-1)}{m}$$

p – rate of innovation

q – rate of immitation

The Bass (1969) Model

- G(t) – percentage of agents who have adopted by time t

- m – potential adopters in the population

$$G(t) = G(t-1) + p\big(m - G(t-1)\big) + q(m - G(t-1))\frac{G(t-1)}{m}$$

Individuals who have
not yet adopted

The Bass (1969) Model

- G(t) – percentage of agents who have adopted by time t
- m – potential adopters in the population

$$G(t) = G(t-1) + p\big(m - G(t-1)\big) + q(m - G(t-1))\frac{G(t-1)}{m}$$

Individuals who have not yet adopted

Fraction of adopters to imitate

The Bass (1969) Model

- Continuous time version

The Bass (1969) Model

- Continuous time version

- Set m=1, g(t) rate of diffusion

$$g(t) = \big(p + qG(t)\big)\big(1 - G(t)\big)$$

The Bass (1969) Model

- Continuous time version

- Set m=1, g(t) rate of diffusion

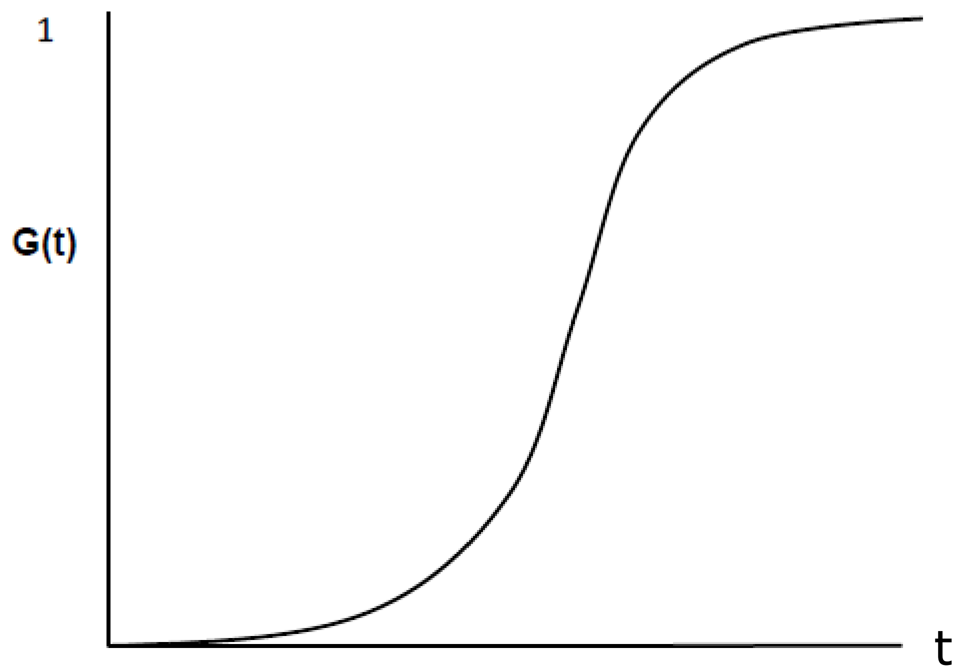
$$g(t) = \big(p + qG(t)\big)\big(1 - G(t)\big)$$

- Solve for p>0, G(0)=0:

$$G(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p} e^{-(p+q)t}}$$

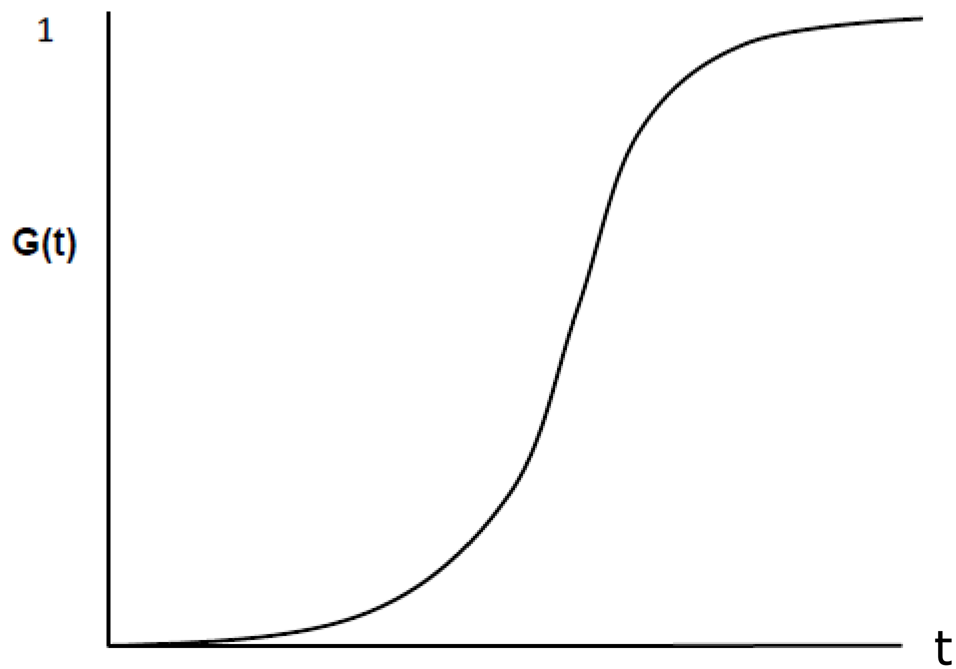
The Bass (1969) Model

- S-shaped adoption



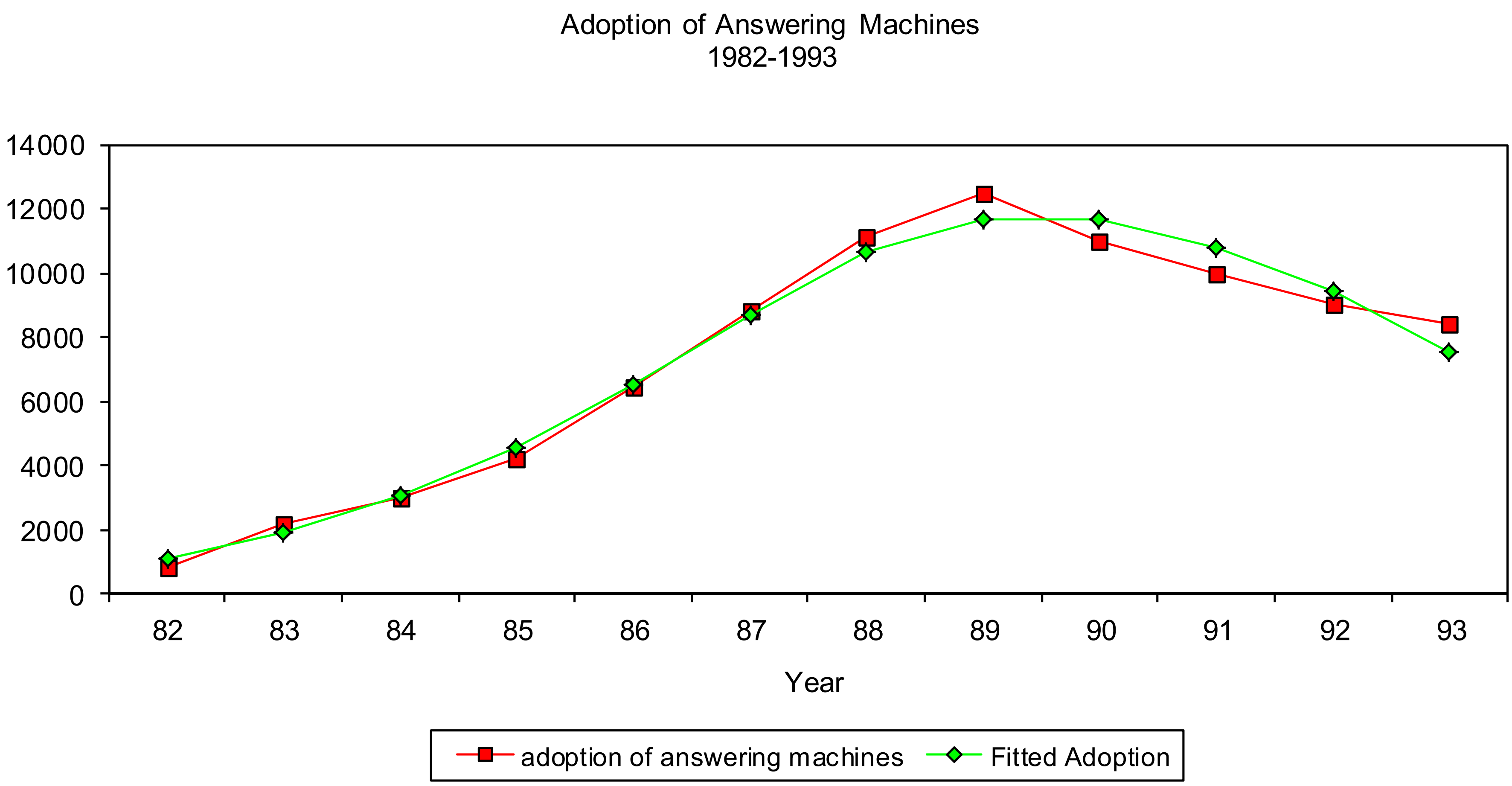
The Bass (1969) Model

- S-shaped adoption

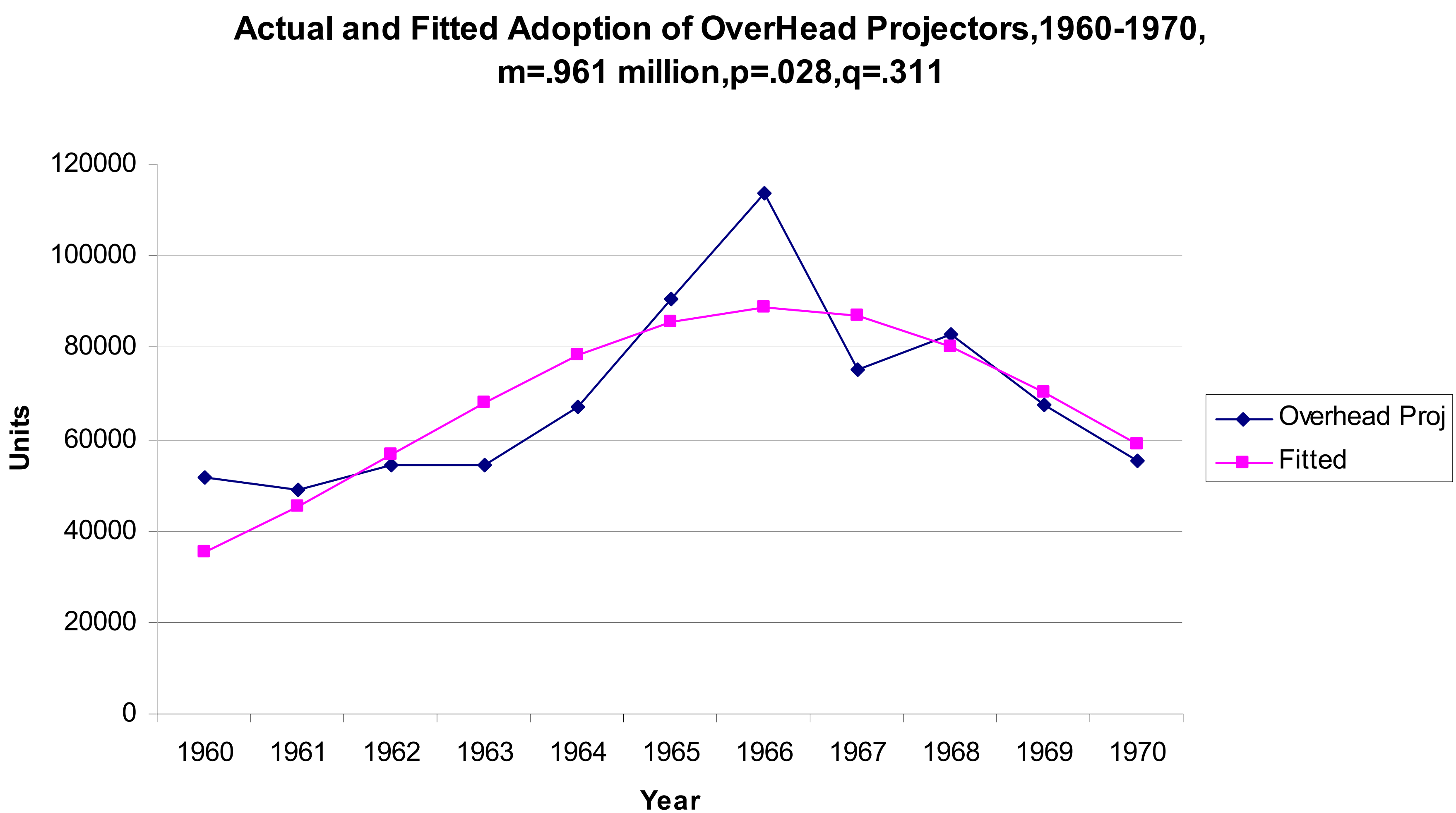


- No network effects

The Bass Model – Example 1



The Bass Model – Example 2



The Reed-Frost Model (Bailey, 1975)

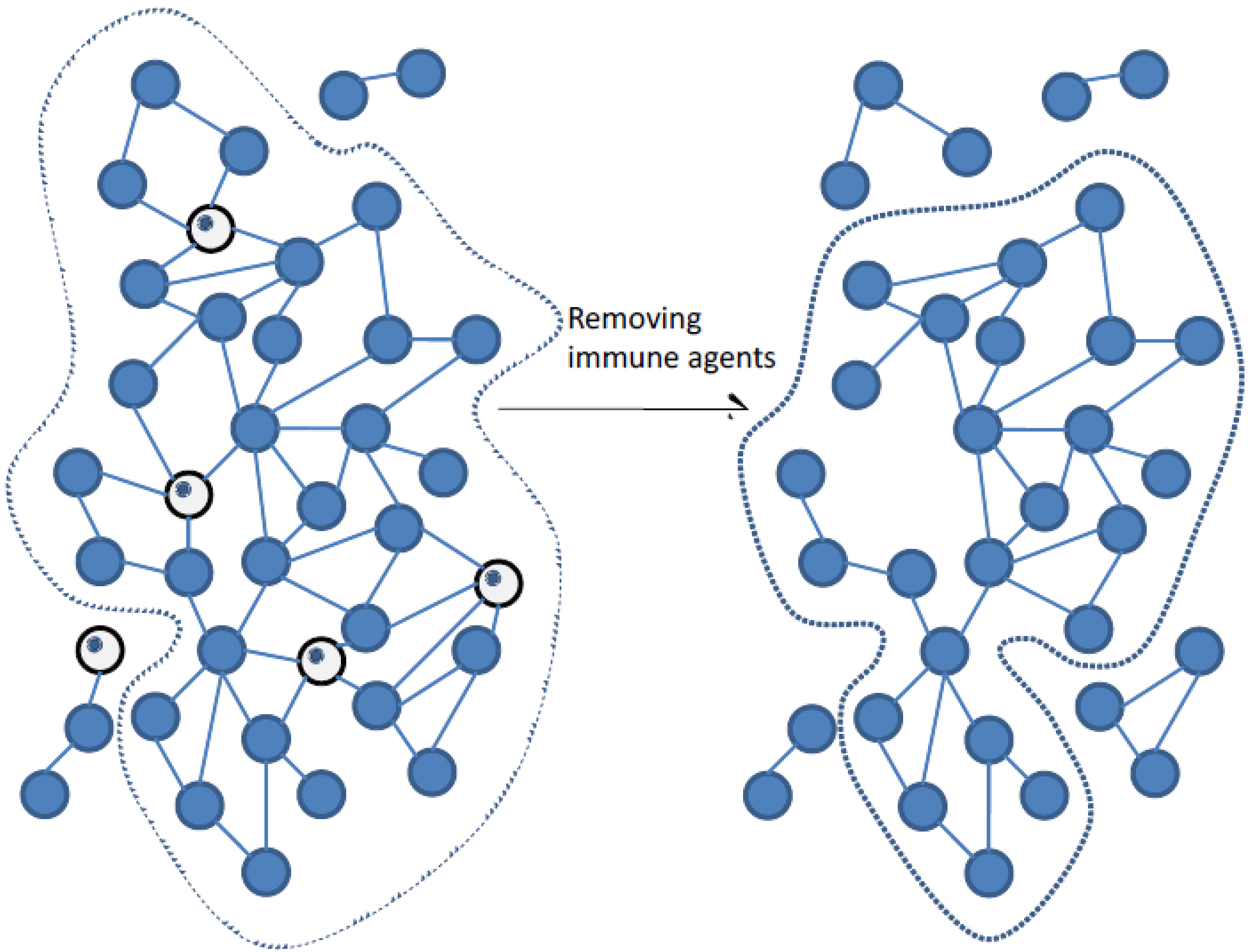
- Underlying network is an Erdos-Renyi Poisson network, with link probability p

- Each individual immune with probability π

The Reed-Frost Model (Bailey, 1975)

- Underlying network is an Erdos-Renyi Poisson network, with link probability p
- Each individual immune with probability π
- **Question:** When would a small fraction of "sick" individual contaminate a substantial fraction of society?

The Reed-Frost Model



The Reed-Frost Model

- A **component** of (N,g) is a sub-network (N',g'), such that $\emptyset \neq N' \subset N$, $g' \subset g$ such that:

  - (N',g') is connected; and
  - If $i \in N'$ and $ij \in g$, then $j \in N'$ and $ij \in g'$

Size of Large Component – Poisson

- Suppose p>1/n
- q – fraction of nodes in the largest component

Size of Large Component – Poisson

- Suppose p>1/n
- q – fraction of nodes in the largest component
- Contemplate adding a node, large n

Size of Large Component – Poisson

- Suppose $p>1/n$
- q – fraction of nodes in the largest component
- Contemplate adding a node, large n
- If it is of degree d, chance it is outside:

$$(1-q)^d$$

Size of Large Component – Poisson

- Probability of degree d is P(d):

$$1 - q = \sum_{d} P(d) * (1 - q)^{d}$$

Size of Large Component – Poisson

- Probability of degree d is P(d):

$$1 - q = \sum_d P(d) * (1 - q)^d$$

- Plugging in the Poisson:

$$q = 1 - e^{-q(n-1)p}$$

Size of Large Component – Poisson

- Probability of degree d is P(d):

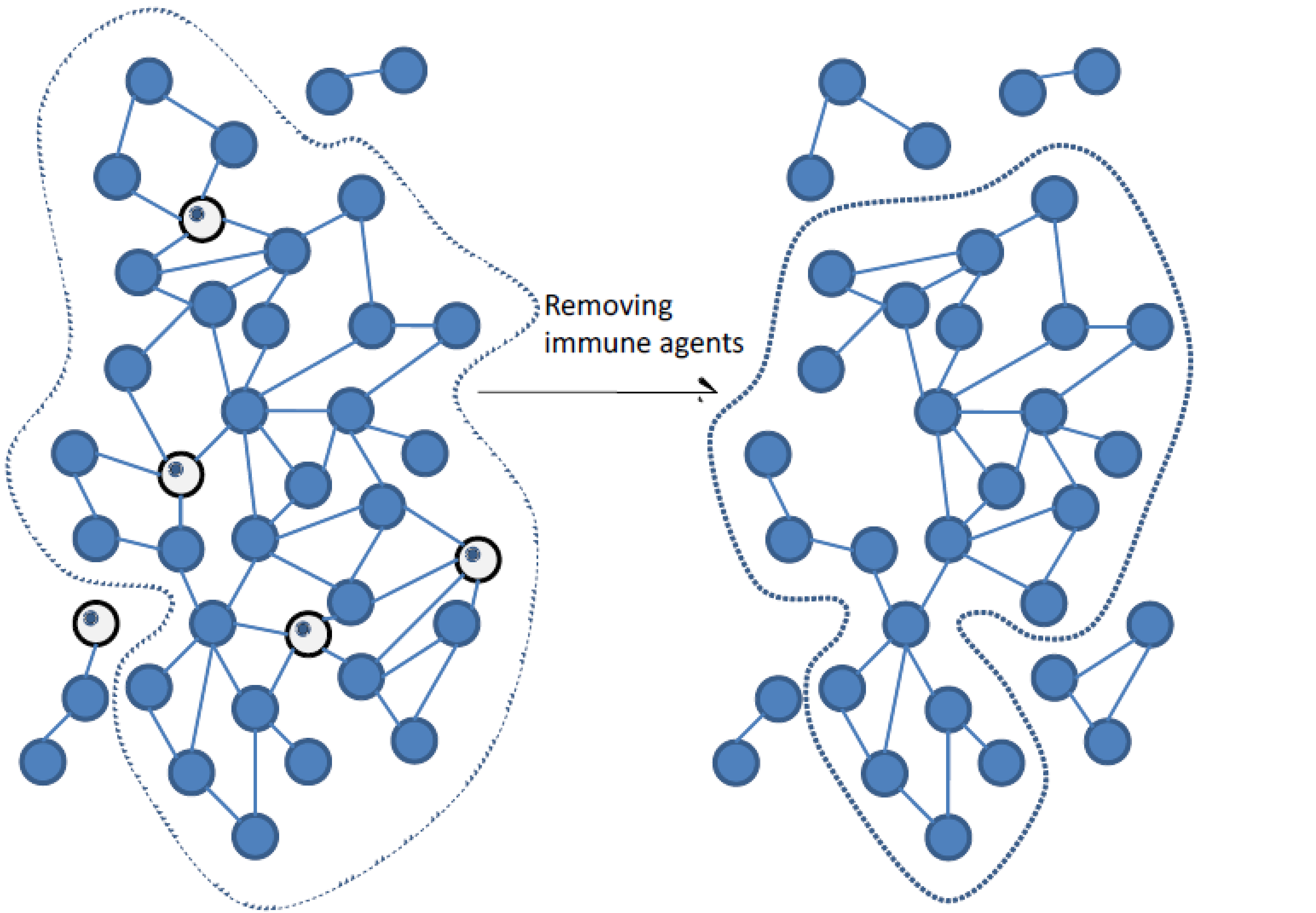
$$1 - q = \sum_{d} P(d) * (1 - q)^{d}$$

- Plugging in the Poisson:

$$q = 1 - e^{-q(n-1)p}$$

- q=0 always a solution
- When average degree > 1 (p(n-1)>1), positive q>0 solution ("phase transition" at p(n-1)=1)

The Reed-Frost Model

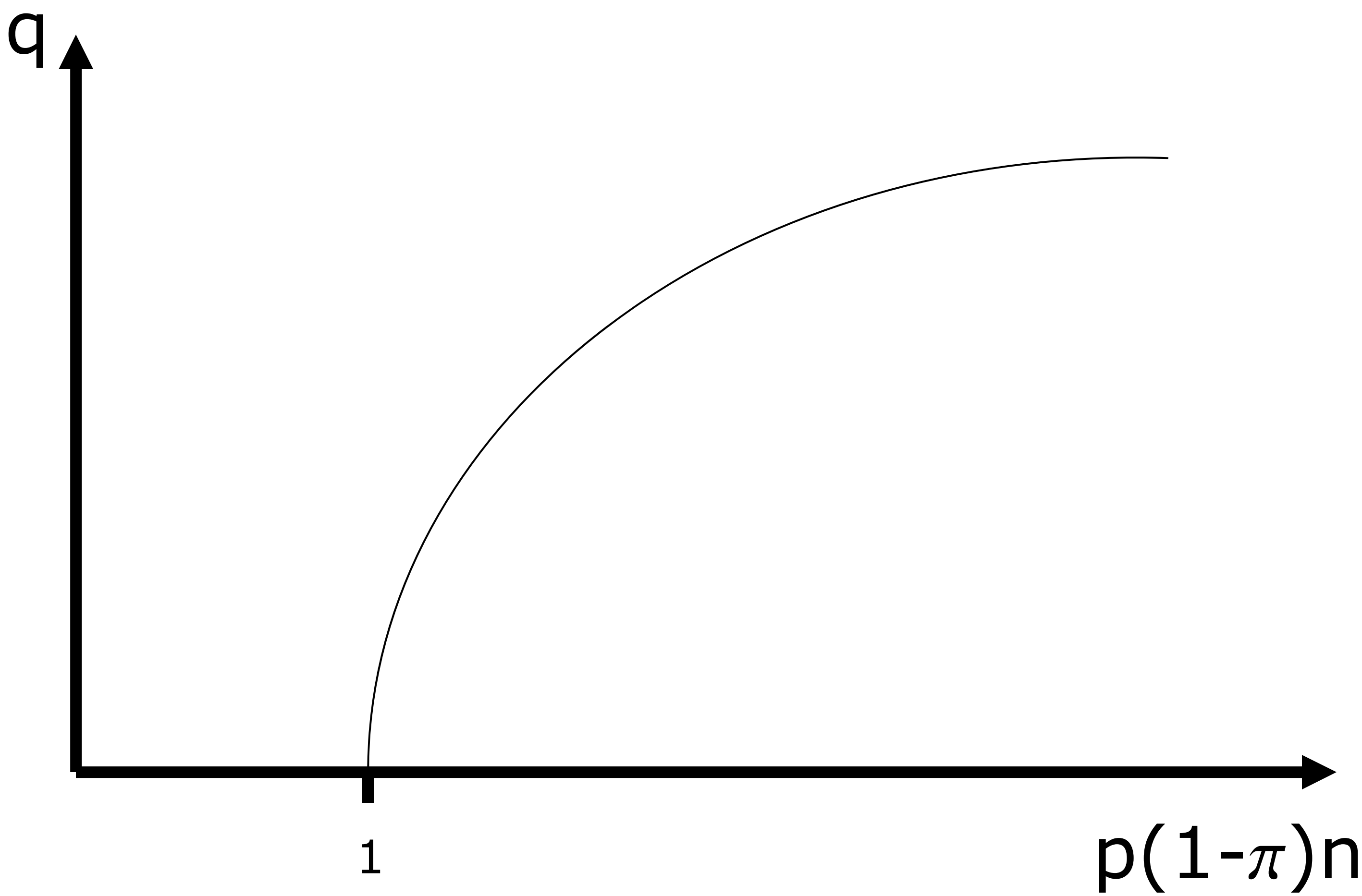


Back to Reed-Frost

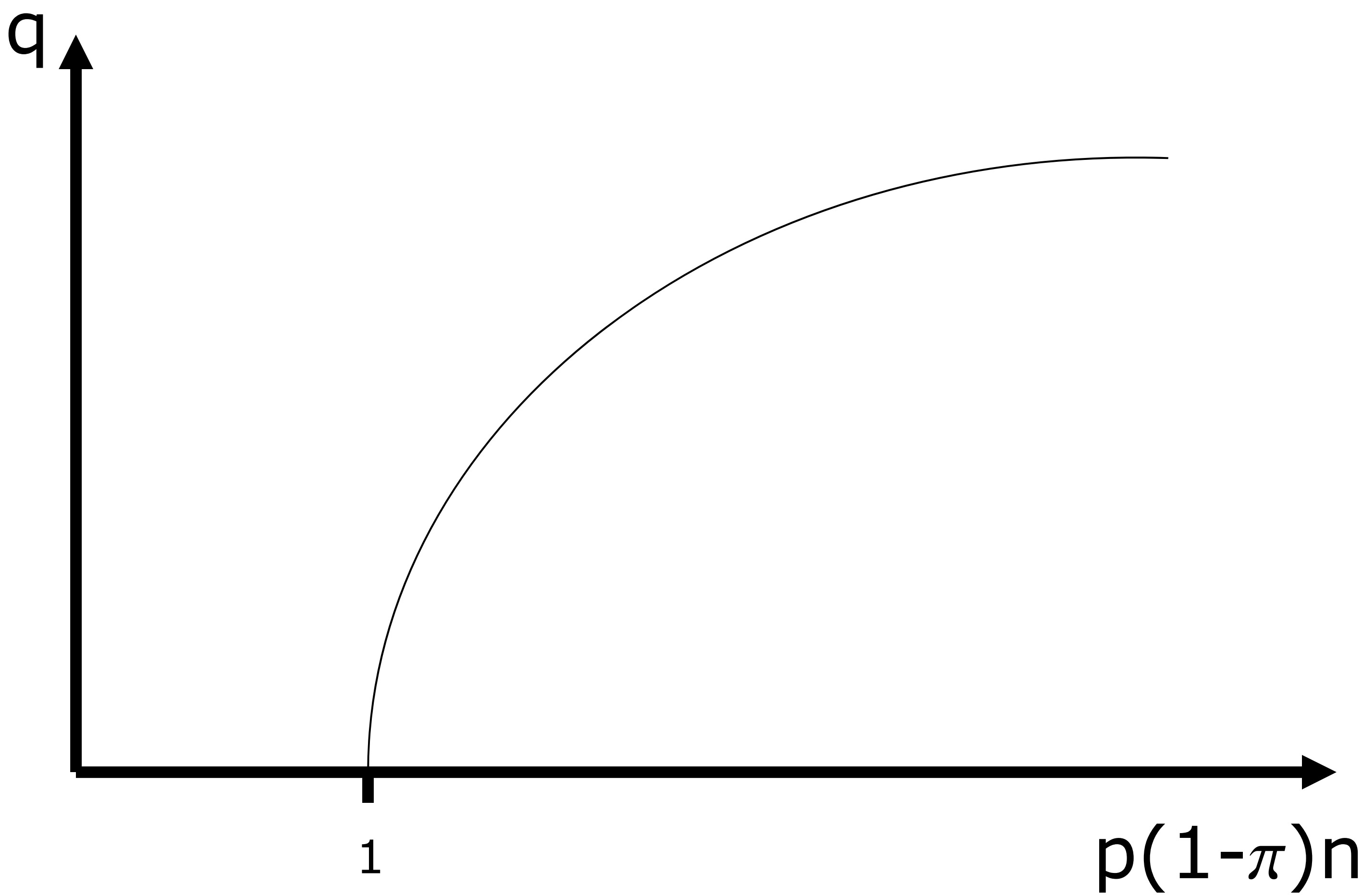
- $(1-\pi)n$ relevant nodes
- If $p(1-\pi)n<1$, no giant component and small fraction infected will die out
- If $p(1-\pi)n>1$, small infection may spread to the giant component:

$$q = 1 - e^{-q(1-\pi)np}$$

The Reed-Frost Model



The Reed-Frost Model



- No strategies, no dynamics… Which is next!

Questions:

- How do choices to invest in education, learn a language, etc., depend on social network structure and location within a network?

Questions:

- How do choices to invest in education, learn a language, etc., depend on social network structure and location within a network?
  - How does network structure impact behavior and welfare? Complexity of calculating equilibria?
  - How does relative location in a network impact behavior and welfare?

Questions:

- How do choices to invest in education, learn a language, etc., depend on social network structure and location within a network?
  - How does network structure impact behavior and welfare? Complexity of calculating equilibria?
  - How does relative location in a network impact behavior and welfare?
- How does behavior propagate through network (important for marketing, epidemiology, etc.)?

Example - Experimentation

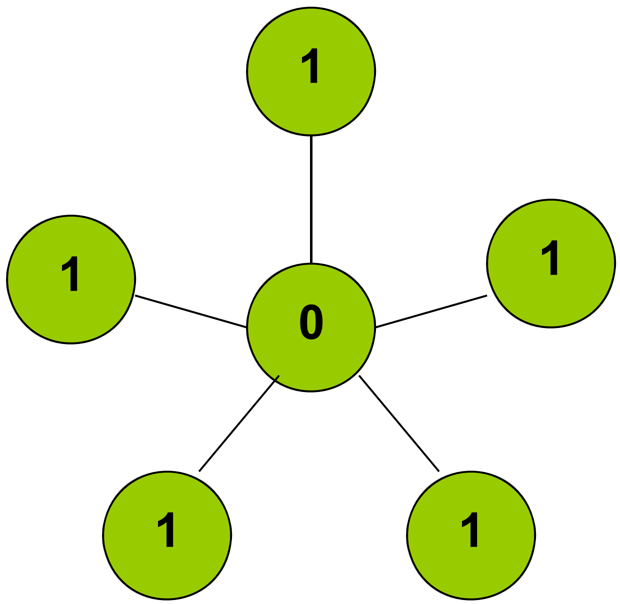
- Suppose you gain 1 if anyone experiments, 0 otherwise, but experimentation is costly (grains, software, etc.)

Example - Experimentation

- Suppose you gain 1 if anyone experiments, 0 otherwise, but experimentation is costly (grains, software, etc.) EXPERIMENTATION – 1

  NO EXPERIMENTATION - 0

- Knowing the network structure

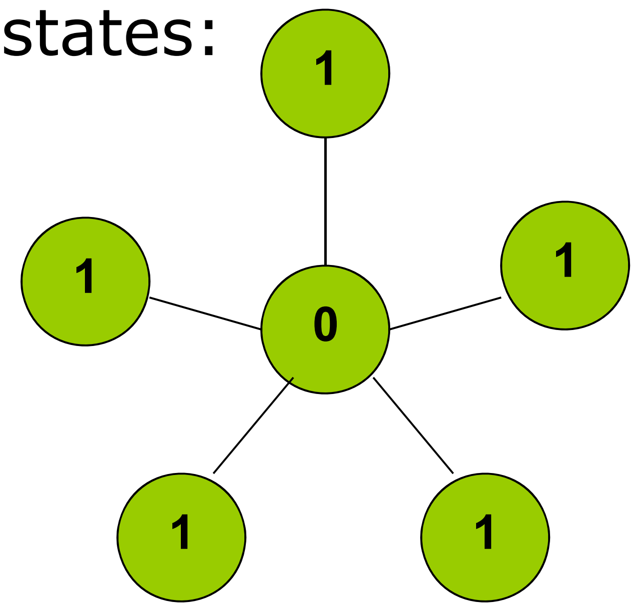


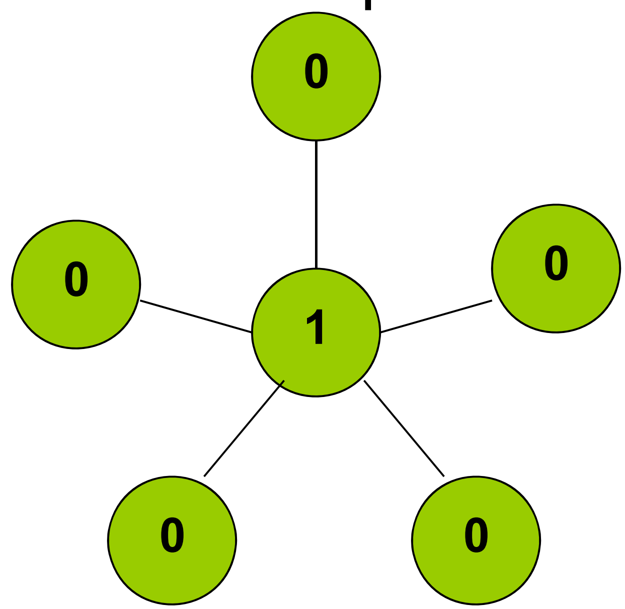
Example - Experimentation

- Suppose you gain 1 if anyone experiments, 0 otherwise, but experimentation is costly (grains, software, etc.) EXPERIMENTATION – 1

  NO EXPERIMENTATION - 0

- Knowing the network structure – multiple stable states:





Example – Experimentation (2)

Not knowing the structure

Example – Experimentation (2)

Not knowing the structure

- Probability p of a link between any two agents (Poisson..).

Example – Experimentation (2)

Not knowing the structure

- Probability p of a link between any two agents.
- Symmetry

Example – Experimentation (2)

Not knowing the structure

- Probability p of a link between any two agents.
- Symmetry
- Probability that a neighbor experiments independent of own degree (number of neighbors)
  - → Higher degree less willing to choose 1
  - → Threshold equilibrium: low degrees experiment, high degrees do not.

Example – Experimentation (2)

Not knowing the structure

- Probability p of a link between any two agents.
- Symmetry
- Probability that a neighbor experiments independent of own degree (number of neighbors)
  - → Higher degree less willing to choose 1
  - → Threshold equilibrium: low degrees experiment, high degrees do not.
- Strong dependence on p
  - p=0→ all choose 1,
  - p=1→ only one chooses 1.

General Messages

- **Information Matters**

General Messages

- Information Matters
- **Location Matters**
  - Monotonicity with respect to degrees
    - Regarding behavior (complementarities…)
    - Regarding expected benefits (externalities…)

General Messages

- Information Matters
- Location Matters
  - Monotonicity with respect to degrees
    - Regarding behavior (complementarities…)
    - Regarding expected benefits (externalities…)
- **Network Structure Matters**
  - Adding links affects behavior monotonically (complementarities…)
  - Increasing heterogeneity has regular impacts.

Challenge

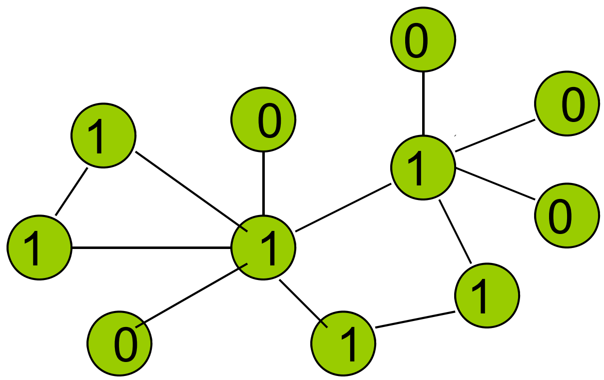
- Complexity of networks
- Tractable way to study behavior outside of simple (regular structures)?

Focus on key characteristics:

- Degree Distribution
  - Degree of node = number of neighbors
- How connected is the network?
  - average degree, FOSD shifts.
- How are links distributed across agents?
  - variance, skewness, etc.

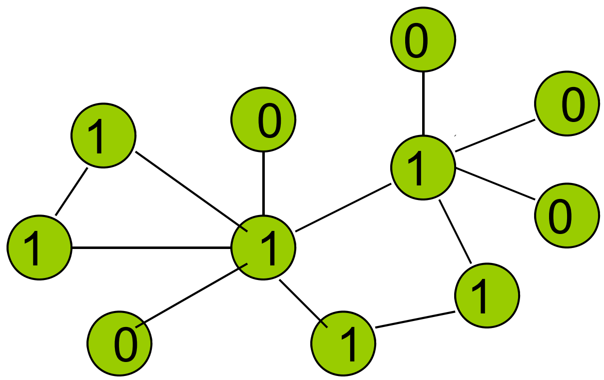
What we analyze:

- A network describes who neighbors are, whose actions a player cares about:



What we analyze:

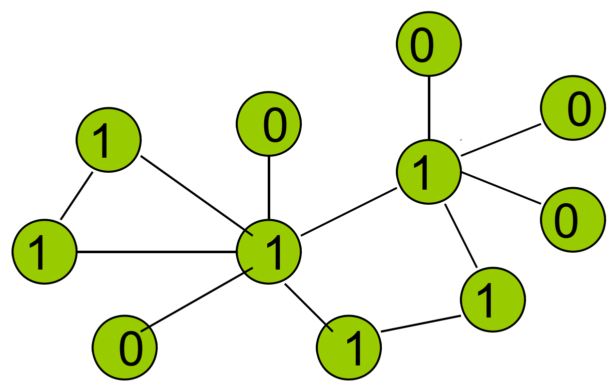
- A network describes who neighbors are, whose actions a player cares about:



- Players choose actions (today: in {0,1})

What we analyze:

- A network describes who neighbors are, whose actions a player cares about:



- Players choose actions (today: in {0,1})
- Examine
  - **equilibria**
  - **how play diffuses** through the network

Games on Networks

- g is network (in $\{0,1\}^{nxn}$):

$$\boldsymbol{g}_{ij} = \begin{cases} 1 & \boldsymbol{i} \neq \boldsymbol{j} \text{ connected} \\ 0 & \text{otherwise} \end{cases}$$

Games on Networks

- g is network (in $\{0,1\}^{n x n}$):

$$g_{ij} = \begin{cases} 1 & i \neq j \text{ connected} \\ 0 & \text{otherwise} \end{cases}$$

- $N_i(g)$ i's neighborhood,

$$N_i(g) = \{ j \mid g_{ij} = 1 \}$$

Games on Networks

- g is network (in $\{0,1\}^{nxn}$):

$$g_{ij} = \begin{cases} 1 & i \neq j \text{ connected} \\ 0 & \text{otherwise} \end{cases}$$

- $N_i(g)$ i's neighborhood,

$$N_i(g) = \{ j | g_{ij} = 1 \}$$

- $d_i(g) = |N_i(g)|$ i's degree

Games on Networks

- g is network (in $\{0,1\}^{nxn}$):

$$g_{ij} = \begin{cases} 1 & i \neq j \text{ connected} \\ 0 & \text{otherwise} \end{cases}$$

- $N_i(g)$ i's neighborhood,

$$N_i(g) = \{ j \mid g_{ij} = 1 \}$$

- $d_i(g)=|N_i(g)|$ i's degree

- Each player chooses an action in {0,1}

Payoff Structure: (today) Complements

- Payoffs depend only on the number of neighbors choosing 0 or 1.

Payoff Structure: (today) Complements

- Payoffs depend only on the number of neighbors choosing 0 or 1.

- normalize payoff of all neighbors choosing 0 to 0

Payoff Structure: (today) Complements

- Payoffs depend only on the number of neighbors choosing 0 or 1.

- normalize payoff of all neighbors choosing 0 to 0

- $v(d,x) - c_i$ payoff from choosing 1 if degree is d and a fraction x of neighbors choose 1
  - Increasing in x

Payoff Structure: (today) Complements

- Payoffs depend only on the number of neighbors choosing 0 or 1.
- normalize payoff of all neighbors choosing 0 to 0
- $v(d,x) - c_i$ payoff from choosing 1 if degree is d and a fraction x of neighbors choose 1
  - Increasing in x
- c_i distributed according to H

Examples (payoff: v(d,x)-c)

- **Average Action**: v(d,x)=v(d)x= x
  (classic coordination games, choice of technology)

- **Total Number**: v(d,x)=v(d)x=dx
  (learn a new language, need partners to use new good or technology, need to hear about it to learn)

- **Critical Mass**: v(d,x)=0 for x up to some M/d and v(d,x)=1 above M/d
  (uprising, voting, …)

- **Decreasing**: v(d,x) declining in d
  (information aggregation, lower degree correlated with leaning towards adoption)

Information (covered networks, payoffs)

- Incomplete information
  - know only own degree and assume others' types are governed by degree distribution
  - presume no correlation in degree
  - Bayesian equilibrium – as function of degree

Information (covered networks, payoffs)

- Incomplete information
  - know only own degree and assume others' types are governed by degree distribution
  - presume no correlation in degree
  - Bayesian equilibrium – as function of degree
- Complete information
  - "know g" (or at least know actions in neighborhood)
  - Nash equilibrium

Information (covered networks, payoffs)

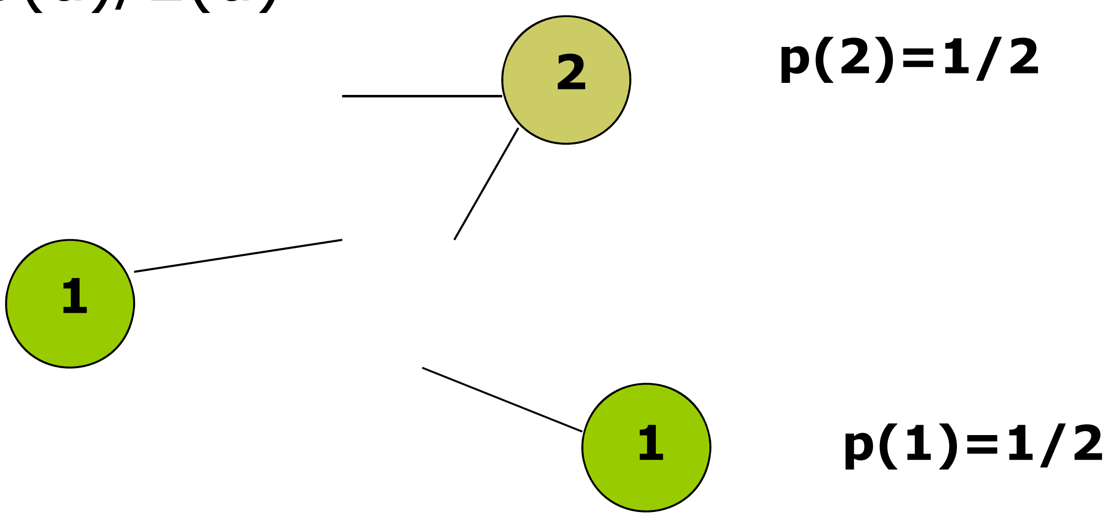
- Incomplete information
  - know only own degree and assume others' types are governed by degree distribution
  - presume no correlation in degree
  - Bayesian equilibrium – as function of degree
- Complete information
  - "know g" (or at least know actions in neighborhood)
  - Nash equilibrium
- Intermediate...

(today) Incomplete information case:

- g drawn from some set of networks G such that:
  - degrees of neighbors are independent
  - Probability of any node having degree d is p(d)
  - probability of given neighbor having degree d is $P(d)=dp(d)/E(d)$

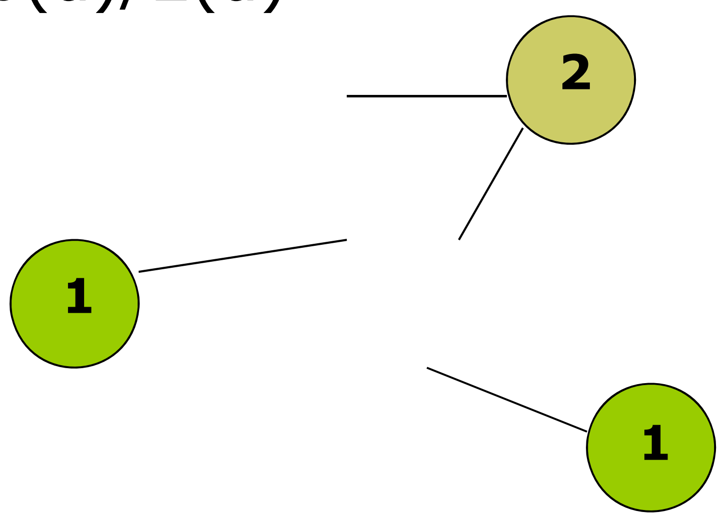
(today) Incomplete information case:

- g drawn from some set of networks G such that:
  - degrees of neighbors are independent
  - Probability of any node having degree d is p(d)
  - probability of given neighbor having degree d is P(d)=dp(d)/E(d)



(today) Incomplete information case:

- g drawn from some set of networks G such that (assuming large population):
  - degrees of neighbors are independent
  - Probability of any node having degree d is p(d)
  - probability of given neighbor having degree d is P(d)=dp(d)/E(d)



Probability of hitting **2** is twice as high as that of hitting **1** → **P(2)=2/3.**

(today) Incomplete information case:

- g drawn from some set of networks G such that:
  - degrees of neighbors are independent
  - Probability of any node having degree d is p(d)
  - probability of given neighbor having degree d is P(d)=dp(d)/E(d)

- type of i is ($d_i(g)$, c_i); space of types T_i

(today) Incomplete information case:

- g drawn from some set of networks G such that:
  - degrees of neighbors are independent
  - Probability of any node having degree d is p(d)
  - probability of given neighbor having degree d is $P(d)=dp(d)/E(d)$
- type of i is ($d_i(g)$, c_i); space of types T_i
- strategy: $\sigma_i: T_i \rightarrow \Delta(X)$

Equilibrium as a fixed point:

- H(v(d,x)) is the percent of degree d types adopting action 1 if x is fraction of random neighbors adopting.
- Equilibrium corresponds to a fixed point:

$$\mathbf{x = \varphi(x) = \Sigma\, P(d)\, H(v(d,x))}$$
$$\mathbf{= \Sigma\, d\, p(d)\, H(v(d,x)) \,/\, E[d]}$$