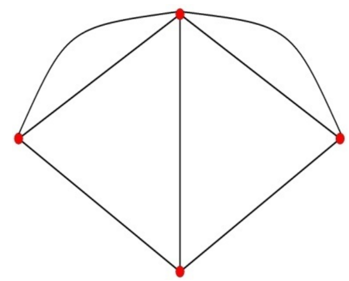
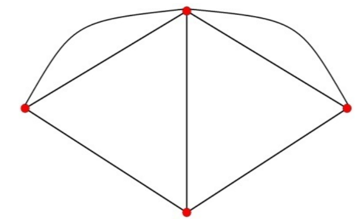
Social and Economic Networks: Models and Analysis

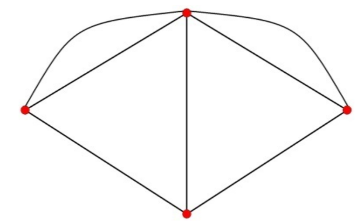


Matthew O. Jackson
Stanford University,
Santa Fe Institute, CIFAR,
www.stanford.edu\~jacksonm

1.1: Introduction

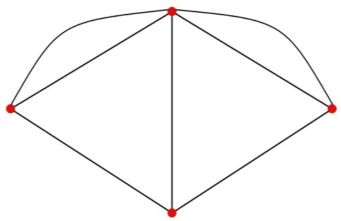


Why Study Networks?



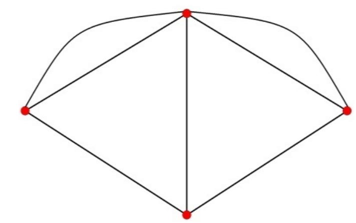
- **Many economic, political, and social interactions are shaped by the local structure of relationships**:
  - trade of goods and services, most markets are not centralized!…
  - sharing of information, favors, risk, ...
  - transmission of viruses, opinions...
  - access to info about jobs...
  - choices of behavior, education, ...
  - political alliances, trade alliances…
- **Social networks influence behavior**
  - crime, employment, human capital, voting, smoking,…
  - networks exhibit heterogeneity, but also have enough underlying structure to model
- **Pure interest in social structure**
  - understand social network structure

Primary Questions:



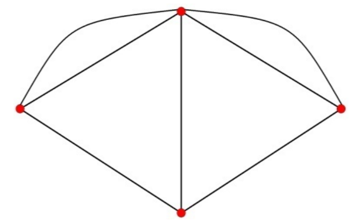
- What do we know about network structure?
- How do networks form? Do the `right' networks form?
- How do networks influence behavior? (and vice versa...)

Synthesize



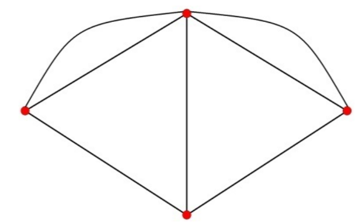
- Many literatures deal with networks
  - Sociology
  - Economics
  - Computer Science
  - Statistical Physics
  - Math (random graph)...
- What have we learned? What are important areas for future research?

Three Areas for Research



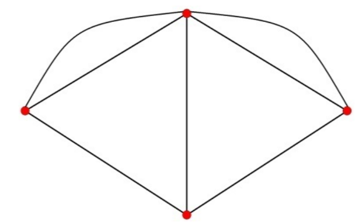
- Theory
  - network formation, dynamics, design...
  - how networks influence behavior
  - coevolution?
- Empirical and experimental work
  - observe networks, patterns, influence
  - test theory and identify regularities
- Methodology
  - how to measure and analyze networks

Central Focus



- **Models for analyzing and understanding networks:**
  - Random graph methods
  - Strategic, game theoretic techniques
  - hybrids, statistical models

Goals

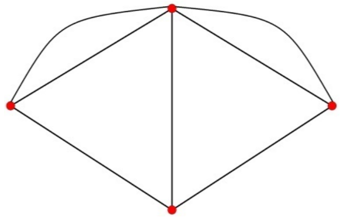


- **Presume no prior knowledge**

- **Introduce you to a variety of approaches to modeling networks (*more breadth than depth*)**

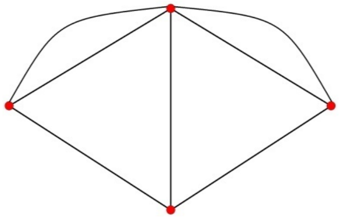
- **Give a sense of different disciplines' techniques and perspectives**

Models



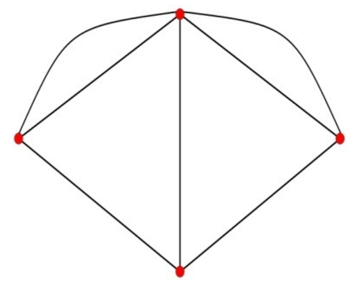
- Provide insight into why we see certain phenomena:
  - Why do social networks have short average path lengths?
- Allow for comparative statics:
  - How does component structure change with density? Important in contagion/diffusion/learning...
- Predict out of sample:
  - What will happen with a new policy (vaccine, R&D subsidy, ...)?
- Allow for statistical estimation:
  - Is there significant clustering on a local level or did it appear at random?

Outline



- Part I: Background and Fundamentals
  - Definitions and Characteristics of Networks (1,2)
  - Empirical Background (3)
- Part II: Network Formation
  - Random Network Models (4,5)
  - Strategic Network Models (6, 11)
- Part III: Networks and Behavior
  - Diffusion and Learning (7,8)
  - Games on Networks (9)

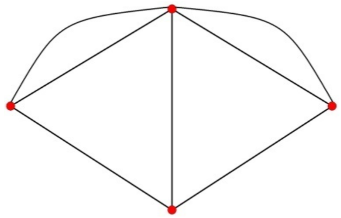
Social and Economic Networks: Models and Analysis



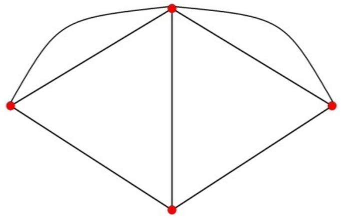
Matthew O. Jackson

Stanford University,
Santa Fe Institute, CIFAR,
www.stanford.edu\~jacksonm

1.2: Examples and Challenges

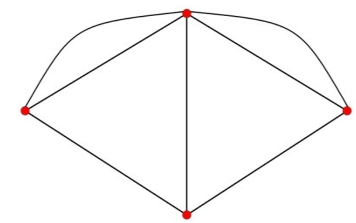


Outline

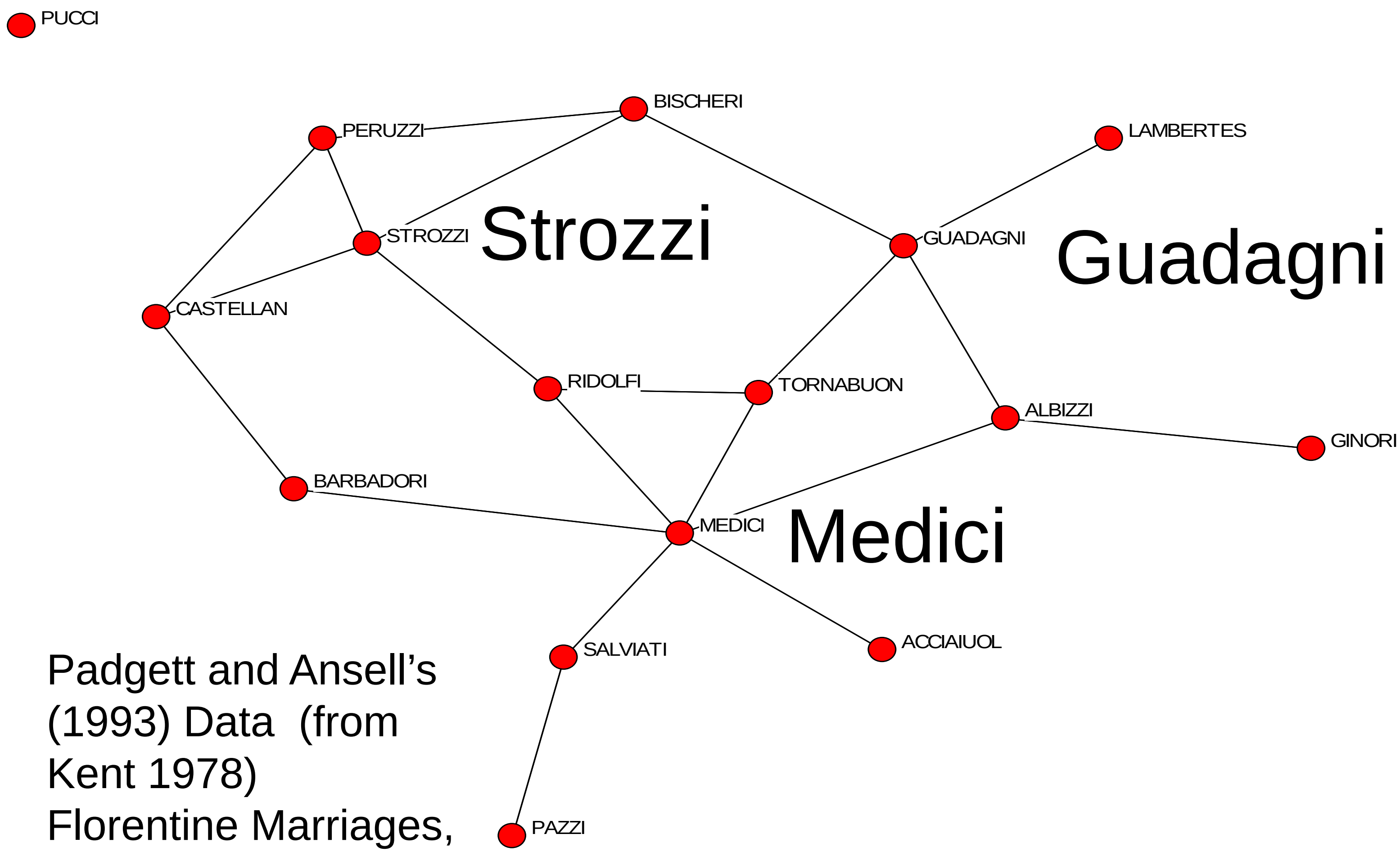


- Part I: Background and Fundamentals
  - Definitions and Characteristics of Networks (1,2)
  - Empirical Background (3)
- Part II: Network Formation
  - Random Network Models (4,5)
  - Strategic Network Models (6, 11)
- Part III: Networks and Behavior
  - Diffusion and Learning (7,8)
  - Games on Networks (9)

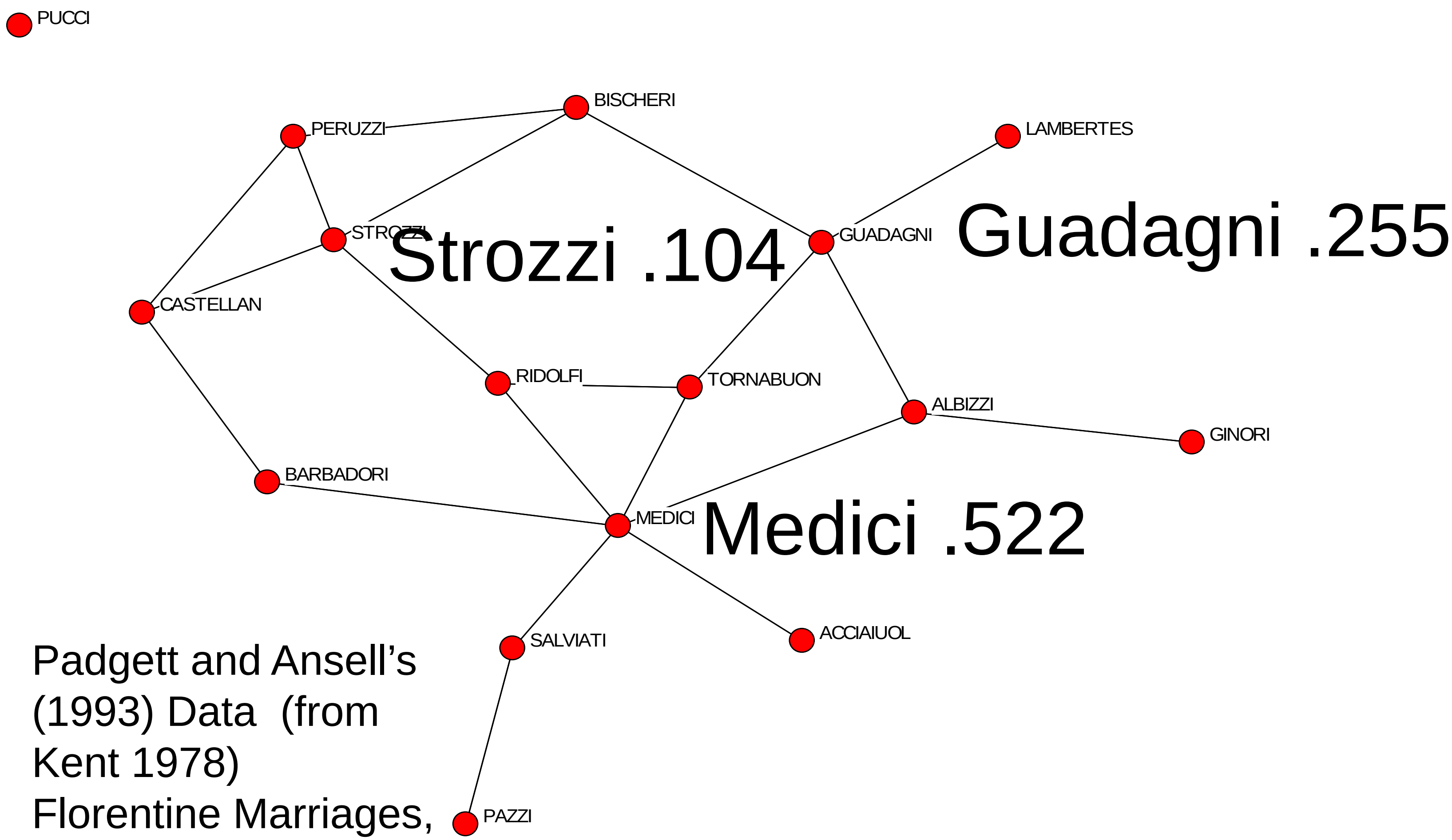
Two Examples



- Idea of data
- View of applications
- Preview some questions

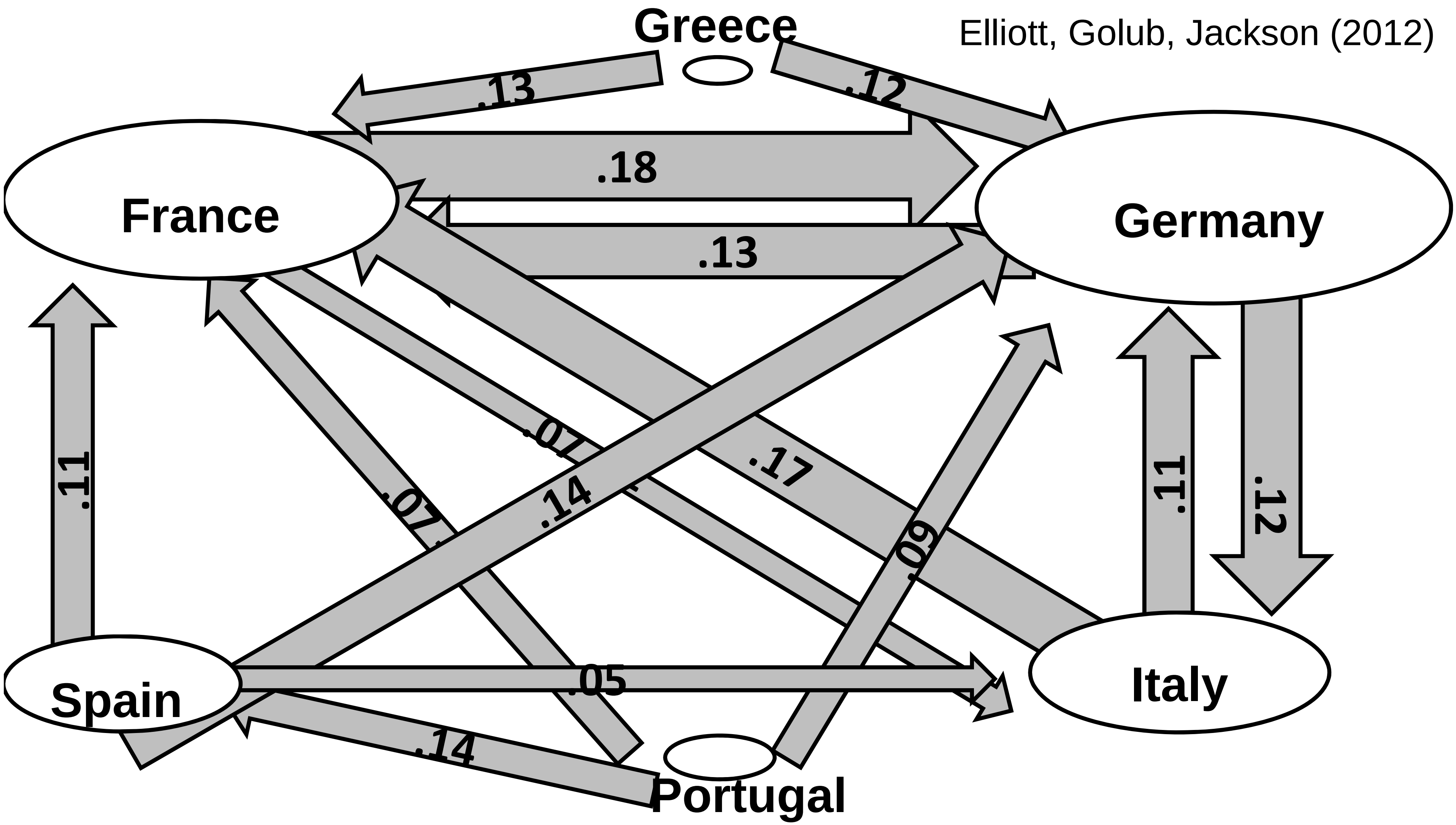


Padgett and Ansell’s (1993) Data (from Kent 1978) Florentine Marriages, 1430’s

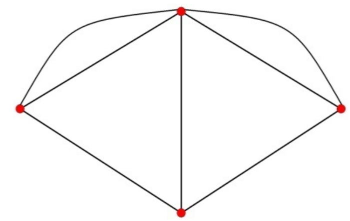


Padgett and Ansell's (1993) Data (from Kent 1978) Florentine Marriages, 1430's

Elliott, Golub, Jackson (2012)



What do We Know?

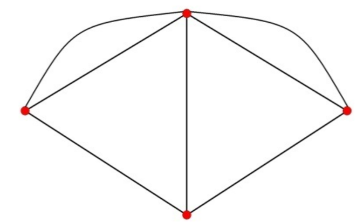


- Networks play role in many settings
  - Job contacts, crime, risk sharing, trade, politics, ...

- Network position and structure matters
  - rich sociology literature
  - Medicis not the wealthiest nor the strongest politically, but the most central

- ``Social'' Networks have special characteristics
  - small worlds, degree distributions...

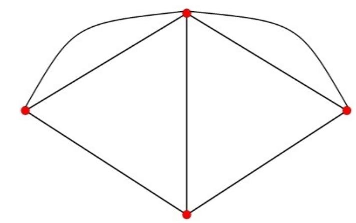
Embeddedness of Economic Activity



- Few markets are centralized, anonymous

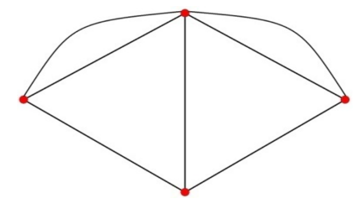
- Specific relationships matter...

Networks in Labor Markets



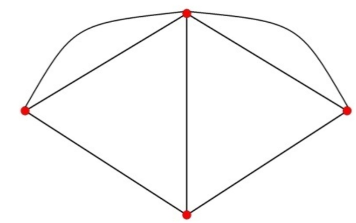
- Myers and Shultz (1951)- textile workers:
  - 62% first job from contact
  - 23% by direct application
  - 15% by agency, ads, etc.
- Rees and Shultz (1970) – Chicago market:
  - Typist 37.3%
  - Accountant 23.5%
  - Material handler 73.8%
  - Janitor 65.5%, Electrician 57.4%…
- Granovetter (1974), Ioannides and Loury (2004) ...

Other Settings



- Networks and social interactions in crime:
  - Reiss (1980, 1988) - 2/3 of criminals commit crimes with others
  - Glaeser, Sacerdote and Scheinkman (1996) - social interaction important in petty crime, among youths, and in areas with less intact households
- Networks and Markets
  - Uzzi (1996) - relation specific knowledge critical in garment industry
  - Weisbuch, Kirman, Herreiner (2000) – repeated interactions in Marseille fish markets
- Social Insurance
  - Fafchamps and Lund (2000) – risk-sharing in rural Philippines
  - De Weerdt (2002) – Tanzania, ...
- Diffusion
  - Hybrid corn adoption Ryan and Gross (1943), Griliches (1957)
  - Drug adoption Coleman, Katz, Menzel (1966)
- Sociology literature – interlocking directorates, aids transmission, language, success of immigrant groups...

The Challenge:



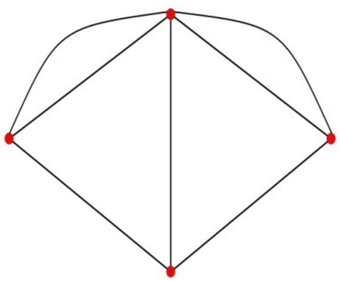
- How many networks on just 30 nodes?

- Person 1 could have 29 possible links, person 2 could have 28 not counting 1, ... total = 435

- So 435 possible links, each could either be present or not, so 2 x 2 x 2 ... 435 times = 2^{435} networks

- Atoms in the universe: between 2^{158} and 2^{246}

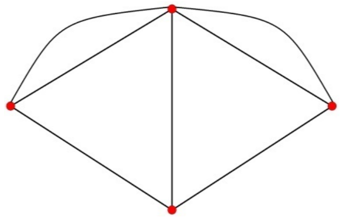
Social and Economic Networks: Models and Analysis



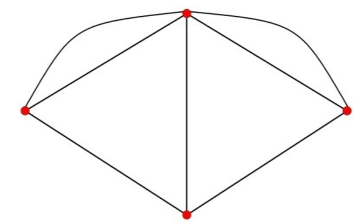
Matthew O. Jackson

Stanford University,
Santa Fe Institute, CIFAR,
www.stanford.edu\~jacksonm

1.25: Background Definitions and Notation

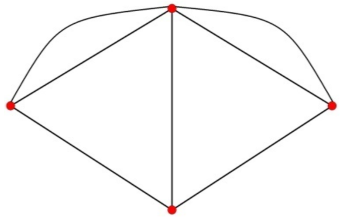


Representing Networks

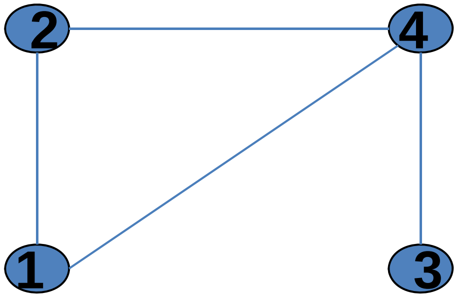


- N={1,…,n}   nodes, vertices, agents, actors, players...
- edges, links, ties: connections between nodes
  - They may have intensity (weighted)
    - How many hours do two people spend together per week?
    - How much of one country's GDP is traded with another?
  - They may just be 0 or 1 (unweighted)
    - Have two researchers written an article together?
    - Are two people ``friends'' on some social platform?
  - They may be ``undirected'' or ``directed''
    - coauthors, friends,..., relatives, spouses, ..., are mutual relationships
    - link from on web page to another, citations, following on social media..., one way

Undirected Network:



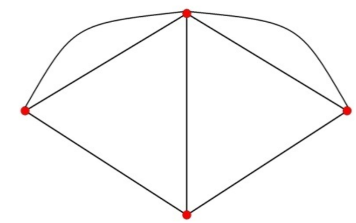
$$g = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$



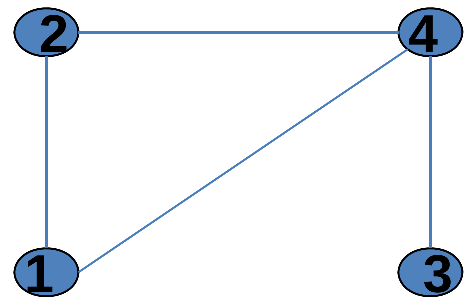
``Adjacency matrix''

g_{ij}=1 iff i & j are linked
undirected, so symmetric,

Undirected Network:



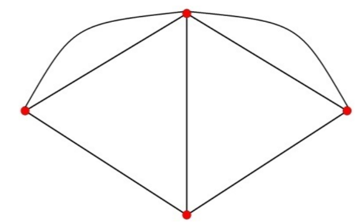
$$g = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$



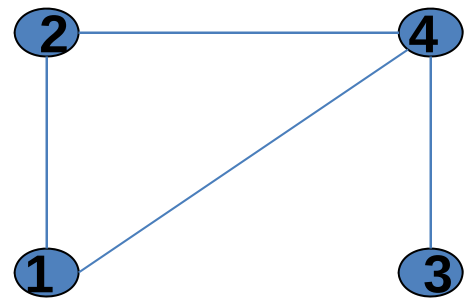
Or list the links:

$g = \{ \{1,2\}, \{1,4\}, \{2,4\}, \{3,4\} \}$

Undirected Network:



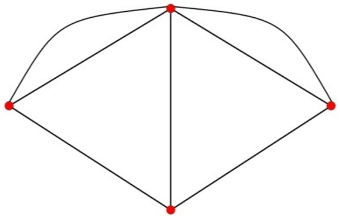
$$g = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$



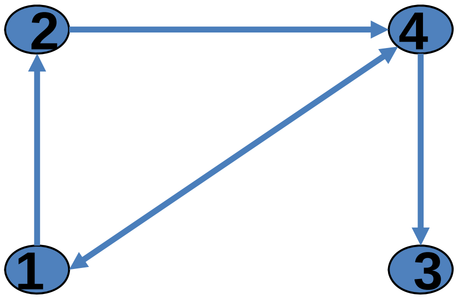
Or list the links: (why?)

$g = \{ 12, 14, 24, 34 \}$

Directed Network:

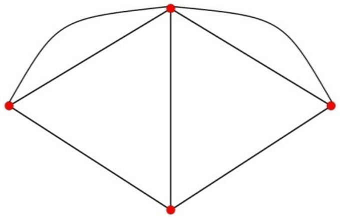


$$g = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$

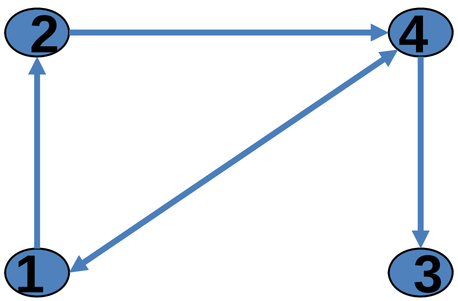


``Adjacency matrix''

Directed Network:



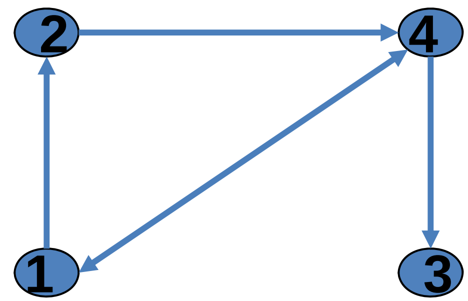
$$g = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$



$$g = \{ 12, 14, 24, 41, 43 \}$$

Directed Network:

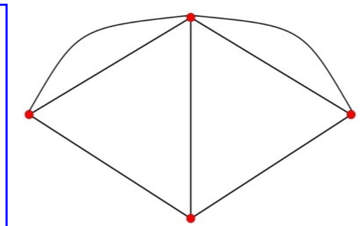
$$g = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix}$$



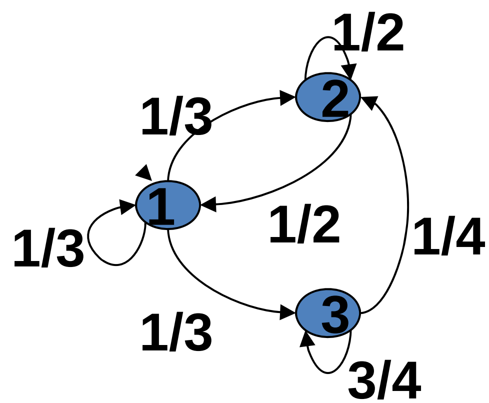
g = { 12, 14, 24, 41,43 }

order of pairs matters

Weighted Directed Network:

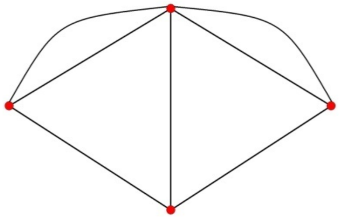


$$g = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/2 & 1/2 & 0 \\ 0 & 1/4 & 3/4 \end{pmatrix}$$

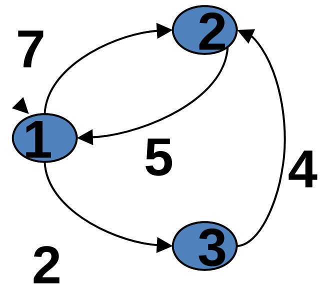


``row stochastic''

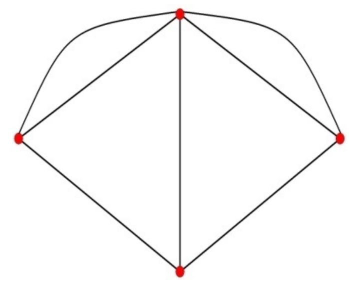
Weighted Directed Network:



$$g = \begin{pmatrix} 0 & 7 & 2 \\ 5 & 0 & 0 \\ 0 & 4 & 0 \end{pmatrix}$$



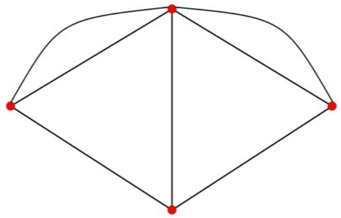
Social and Economic Networks: Models and Analysis



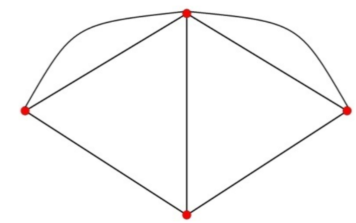
Matthew O. Jackson

Stanford University,
Santa Fe Institute, CIFAR,
www.stanford.edu\~jacksonm

1.3: Definitions and Notation

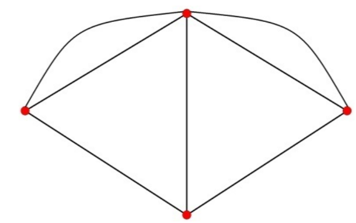


Simplifying the Complexity



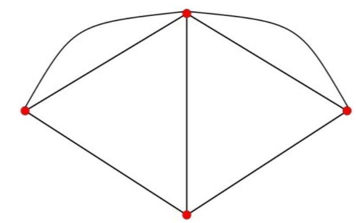
- Global patterns of networks
  - degree distributions, path lengths...
- Segregation Patterns
  - node types and homophily
- Local Patterns
  - Clustering, Transitivity, Support…
- Positions in networks
  - Neighborhoods, Centrality, Influence...

Representing Networks



- $N=\{1,\ldots,n\}$ nodes, vertices, players
- g in $\{0,1\}^{n \times n}$ adjacency matrix   (unweighted, possibly directed)
- $g_{ij} = 1$ indicates a link, tie, or edge between i and j
- *Alternative notation*: ij in g a link between i and j
- Network (N,g)

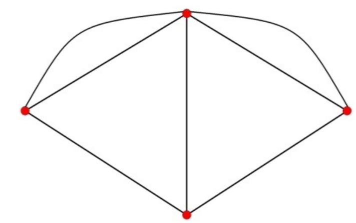
Basic Definitions



- Walk from i_1 to i_K: a sequence of nodes $(i_1,i_2,\ldots, i_K)$ and sequence of links $(i_1i_2,i_2i_3,\ldots,i_{K-1}i_K)$ such that $i_{k-1}i_k$ in g for each k

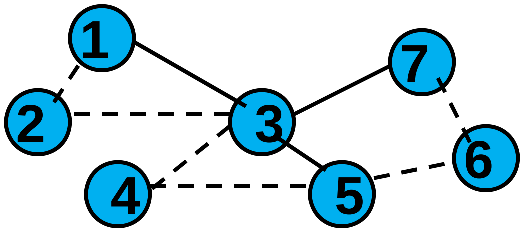
Convenient to represent it as the corresponding sequence of nodes $(i_1,i_2,\ldots, i_K)$ such that $i_{k-1}i_k$ in g for each k

Basic Definitions

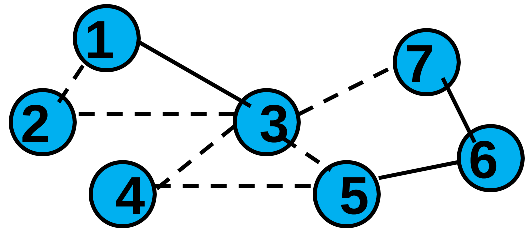


- Path: a walk $(i_1, i_2, \ldots\, i_K)$ with each node i_k distinct
- Cycle: a walk where $i_1 = i_K$
- Geodesic: a shortest path between two nodes

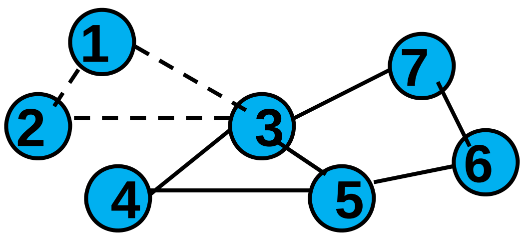
Paths, Walks, Cycles…



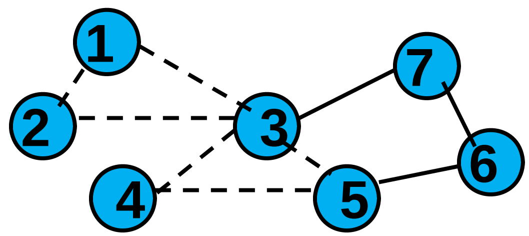
Path (and a walk) from 1 to 7:
1, 2, 3, 4, 5, 6, 7



Walk from 1 to 7 that is not a path:
1, 2, 3, 4, 5, 3, 7

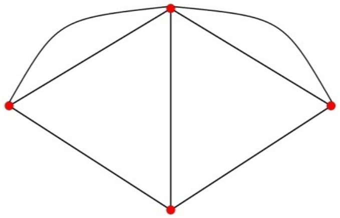


Simple Cycle (and a walk)
from 1 to 1: 1, 2, 3, 1

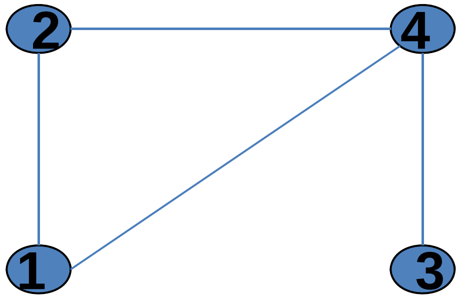


Cycle (and a walk) from 1 to 1:
1, 2, 3, 4, 5, 3, 1

Counting Walks:



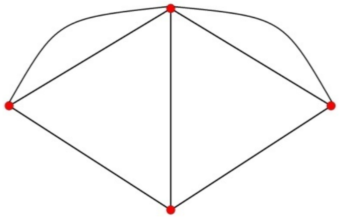
$$g = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$



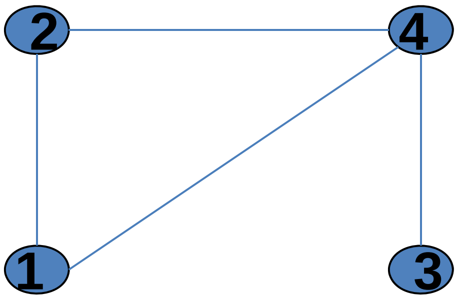
$$g^2 = \begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 3 \end{pmatrix}$$

number of walks of length 2 from i to j

Counting Walks:



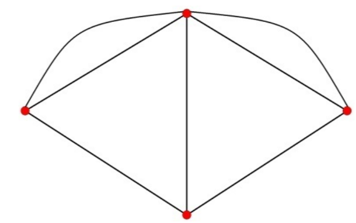
$$g = \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$



$$g^3 = \begin{pmatrix} 2 & 3 & 1 & 4 \\ 3 & 2 & 1 & 4 \\ 1 & 1 & 0 & 3 \\ 4 & 4 & 3 & 2 \end{pmatrix}$$

number of walks of length 3 from i to j

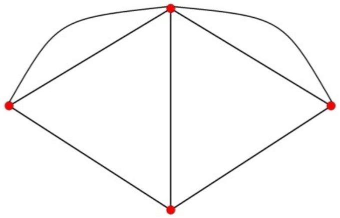
Components

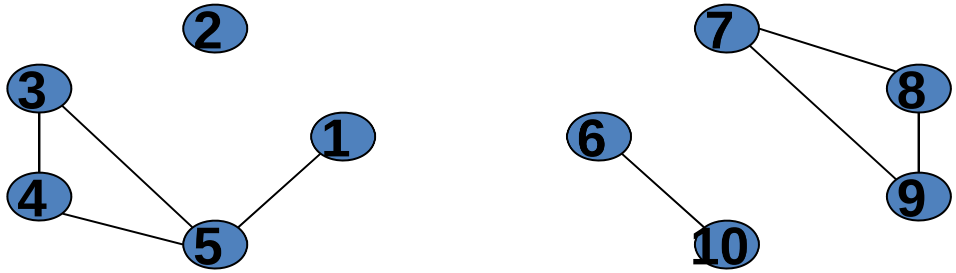


- (N,g) is connected if there is a path between every two nodes

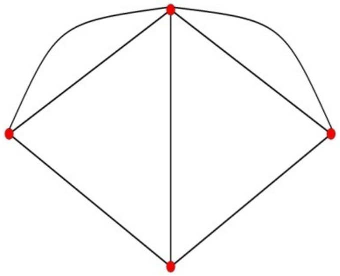
- Component: maximal connected subgraph
  - (N’,g’) is a subset of (N,g)
  - (N’,g’) is connected
  - i in N’ and ij in g implies j in N’ and ij in g’

A network with four components:





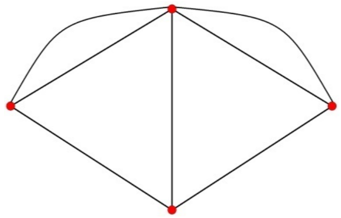
Social and Economic Networks: Models and Analysis



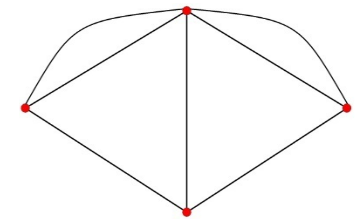
Matthew O. Jackson

Stanford University,
Santa Fe Institute, CIFAR,
www.stanford.edu\~jacksonm

1.4: Diameter



Diameter, Average Path Length



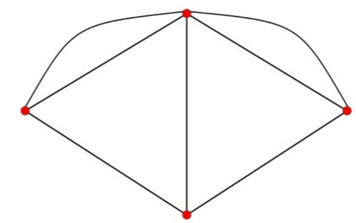
How close are nodes to each other:

- How long does it take to
  reach average node?

- How fast will information spread?...

- How does it depend on network density?

Diameter

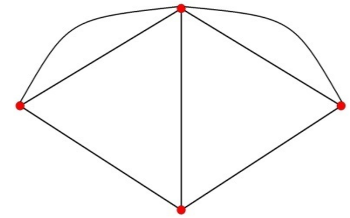


- Diameter – largest geodesic (largest shortest path)
  – if unconnected, of largest component...

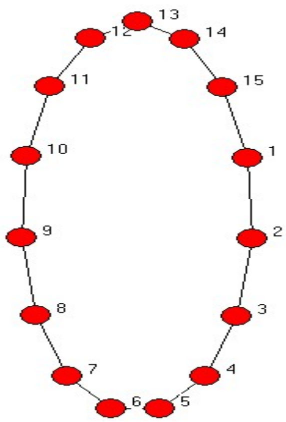
- Average path length

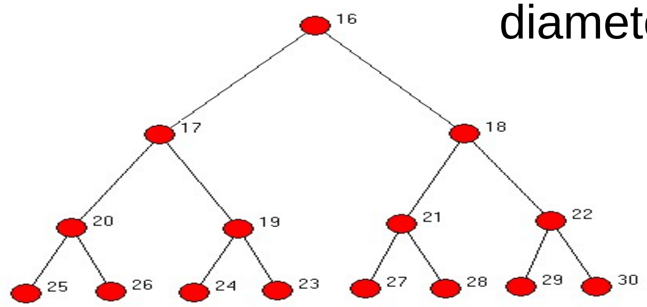
  (less prone to outliers)

Diameter:



K levels has n = 2^{K+1}-1 nodes
so, K = $\log_2(n+1)$ -1
diameter is 2K

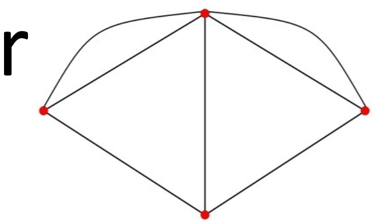




diameter is either
n/2 or (n-1)/2

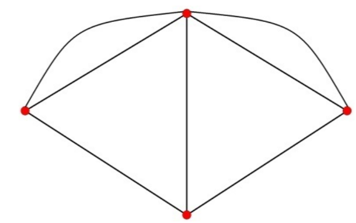
diameter is on order of
2 $\log_2(n+1)$

Small average path length and diameter



- Milgram (1967) letter experiments
  - median 5 for the 25% that made it
- Co-Authorship studies
  - Grossman (2002) Math mean 7.6, max 27,
  - Newman (2001) Physics mean 5.9, max 20
  - Goyal et al (2004) Economics mean 9.5, max 29
- WWW
  - Adamic, Pitkow (1999) – mean 3.1 (85.4% possible of 50M pages)
- Facebook
  - Backstrom et al (2012) – mean 4.74 (721 million users)

Neighborhood and Degree

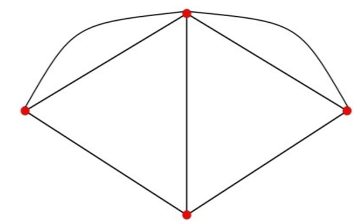


- Neighborhood: $N_i(g) = \{ j \mid ij \text{ in } g \}$

  (usual convention ii not in g)

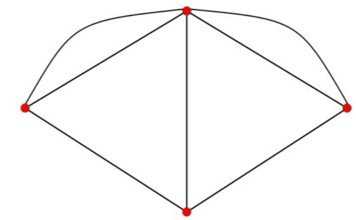
- Degree: $d_i = \# N_i(g)$

Erdos-Renyi (1959,1960) Random Graphs



- start with n nodes
- each link is formed independently with some probability p
- Serves as a benchmark ``G(n,p)''

Sequences of Networks



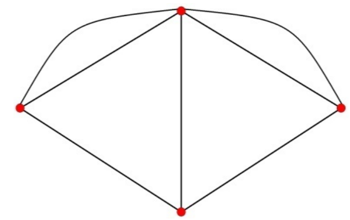
- Links are dense enough so that network is connected almost surely:

  $$d(n) \geq (1+\varepsilon)\log(n) \text{ some } \varepsilon>0$$

- $d(n)/n \to 0$:

  network is not too complete

Theorem on Network Structure



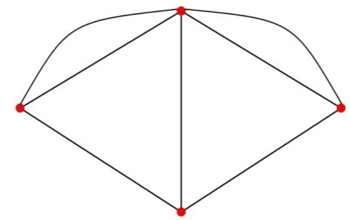
If $d(n) \geq (1+\varepsilon) \log(n)$ some $\varepsilon>0$ and $d(n)/n \to 0$

Then for large n, average path length and diameter are approximately proportional to $\log(n)/\log(d)$

(Proven for increasingly general models:

Erdos-Renyi 59 - Moon and Moser 1966, Bollobas 1981; Chung and Lu 01; Jackson 08; ...)

Theorem on Network Structure

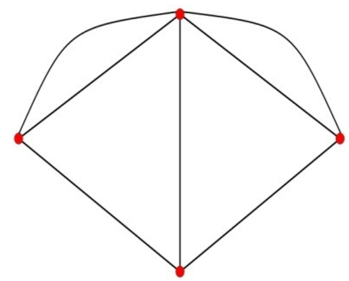


If $d(n) \geq (1+\varepsilon) \log(n)$ some $\varepsilon>0$ and $d(n)/n \to 0$

$$\frac{\textbf{AvgDist(n)}}{\textbf{log(n)/log(d(n))}} \to^{P} 1$$

same for diameter

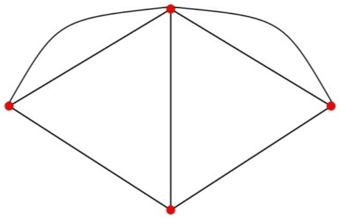
Social and Economic Networks: Models and Analysis



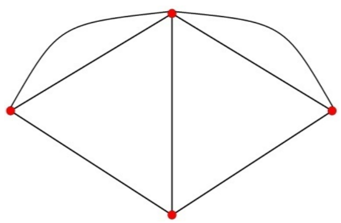
Matthew O. Jackson

Stanford University,
Santa Fe Institute, CIFAR,
www.stanford.edu\~jacksonm

1.5: Diameter and Trees

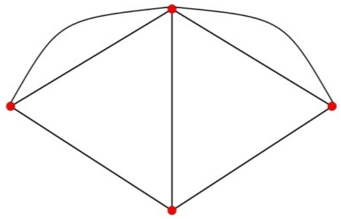


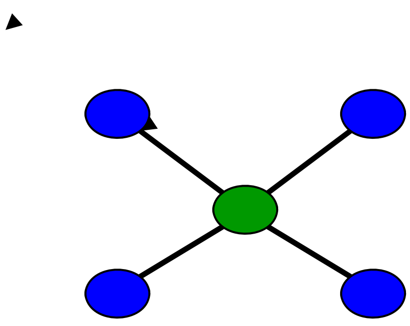
Diameter



- Bounds can be difficult – theorems are narrow, but intuition is easy
- Let’s start with an easy calculation --
- Cayley Tree: each node besides leaves has degree d

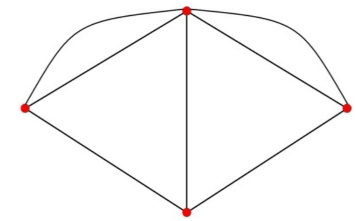
Intuition:

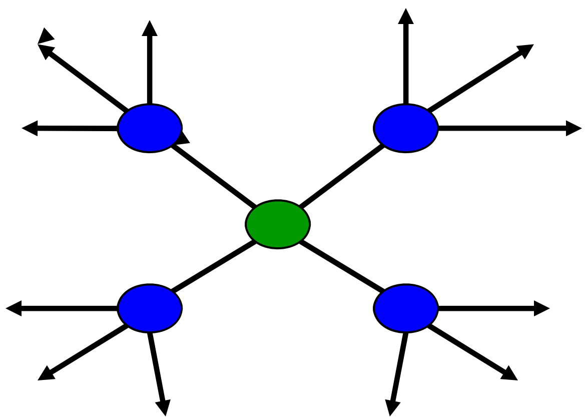




1 step: Reach d nodes,

Ideas:

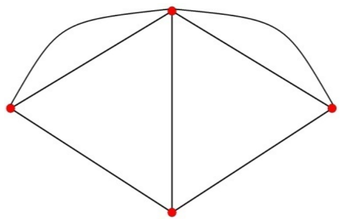


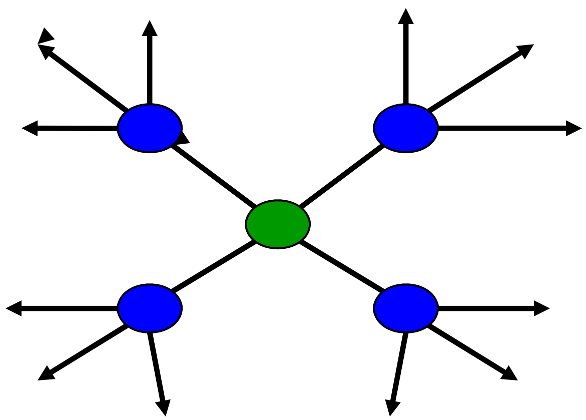


1 step: Reach d nodes,

then d(d-1),

Ideas:



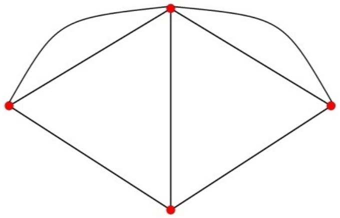


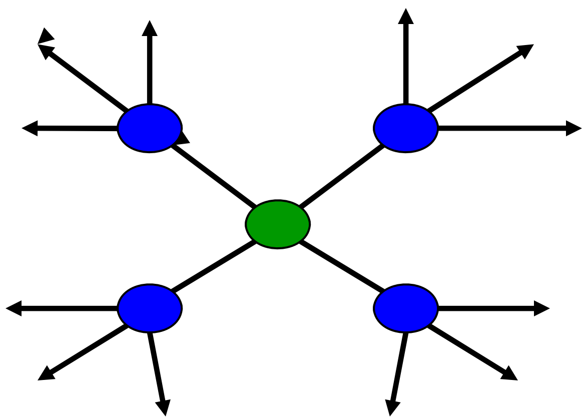
1 step: Reach d nodes,

then $d(d-1)$,

then $d(d-1)^2$,

Ideas:



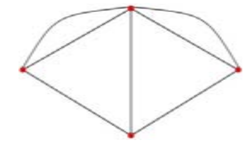


1 step: Reach d nodes,

then d(d-1),

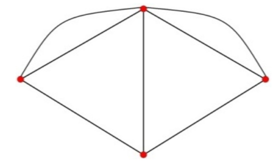
then $d(d-1)^2$, $d(d-1)^3$, ...

After ℓ steps, totals roughly d^{ℓ}



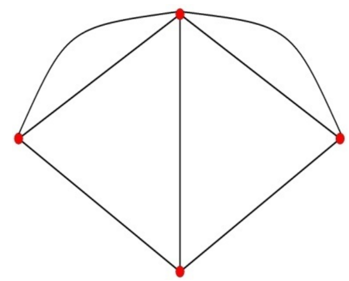
- Moving out ℓ links from root in each direction reaches $d + d(d-1) + \ldots.\ d(d-1)^{\ell-1}$ nodes
- This is $d((d-1)^{\ell} - 1)/(d-2)$ nodes: roughly $(d-1)^{\ell}$
- To reach n-1, need roughly $(d-1)^{\ell} = n$
- or ℓ on the order of $\log(n)/\log(d)$

What if not a tree, but E-R random graph?



- all have same degree – really are random
  – show that fraction of nodes that have nearly average degree is going to 1
- some links may double back
  – most nodes until the last step are still not reached!

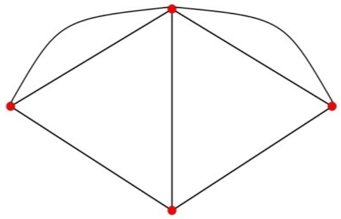
Social and Economic Networks: Models and Analysis



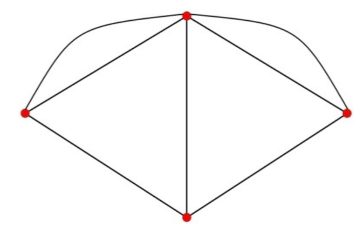
Matthew O. Jackson

Stanford University,
Santa Fe Institute, CIFAR,
www.stanford.edu\~jacksonm

1.6 : Diameters of Random Graphs



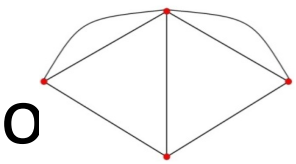
Theorem on Network Structure



If $d(n) \geq (1+\varepsilon) \log(n)$ some $\varepsilon>0$ and $d(n)/n \to 0$

$$\frac{\textbf{AvgDist(n)}}{\textbf{log(n)/log(d(n))}} \to^{P} \mathbf{1}$$

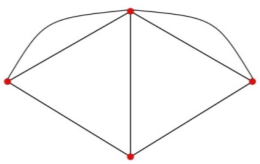
same for diameter

- Moving out ℓ links from root in each directio
  reaches $d + d(d-1) + ... \; d(d-1)^{\ell -1}$ nodes

- This is $d((d-1)^{\ell} -1)/(d-2)$ nodes or roughly $(d-1)^{\ell}$

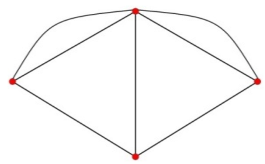
- To reach n-1, need roughly $(d-1)^{\ell} = n$ or

- ℓ on the order of $\log(n)/\log(d)$



What if not a tree, but Erdos-Renyi random graph?

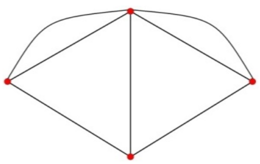
- all have same degree – really are random
  - show that fraction of nodes that have nearly average degree is going to 1
- $E[d] > (1+\varepsilon) \log(n)$



- **Chernoff Bounds:**

X is binomial variable then

$$\Pr(E[X]/3 \leq X \leq 3E[X]) \geq 1 - e^{-E[X]}$$



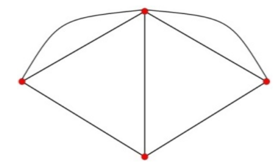
- **Chernoff Bounds:**

X is binomial variable then

$$\Pr(E[X]/3 \leq X \leq 3E[X]) \geq 1 - e^{-E[X]}$$

http://en.wikipedia.org/wiki/Chernoff_bound

- **Chernoff Bounds: Links binomial implies**

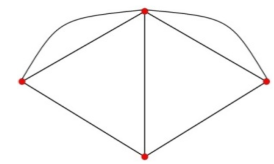


Probability that node has degree close to average:

$\Pr(\, d/3 \leq d_i \leq 3d \,) \geq 1 - e^{-d}$

$\Pr(\, d/3 \leq \text{all degrees} \leq 3d \,) \geq (1 - e^{-d})^n$

(missing steps: degrees not quite ind.)



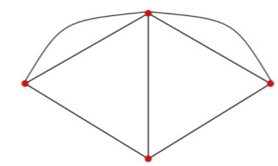
- **Chernoff Bounds:**

Pr (d/3 $\leq$ all degrees $\leq$ 3d) $\geq$ $(1 - e^{-d})^n$

- If $d > (1+\varepsilon)\log(n)$ then

Pr (d/3 $\leq$ all degrees $\leq$ 3d) $>$ $(1 - 1/n^{1+\varepsilon})^n$

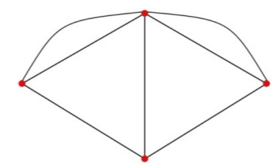
$\rightarrow \exp(-n^{-\varepsilon}) \rightarrow 1$



- **So:**

- If $d > (1+\varepsilon)\log(n)$ then

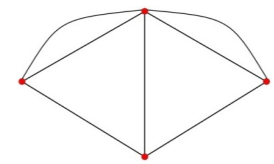
  $\Pr(\, d/3 \leq \text{all degrees} \leq 3d \,) \to 1$



- **Thus:**

- If $d > (1+\varepsilon)\log(n)$ then with prob $\to$ 1:

$$\log(n)/\log(3d) < \ell < \log(n)/\log(d/3)$$

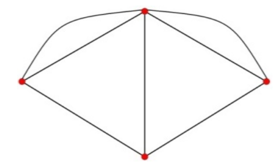


- **Avg distance and diameter:**

- Large d: log(3d) & log(d/3) tend to log(d)

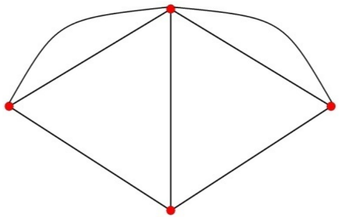
- $\log(n)/\log(3d) < \ell < \log(n)/\log(d/3)$

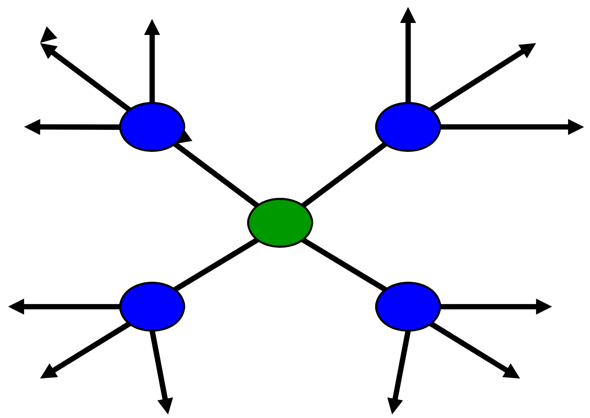
- $\log(n)/\log(d) \approx \ell$



- some links may double back
  - most nodes until the last step are still not reached, so most links still reaching new nodes!
  - After k steps reached around d^k nodes and $n - d^k$ still unreached
  - if $k < \log(n)/\log(d)$ then $n - d^k$ (much) bigger than d^k so most nodes that link to are still unreached...

Ideas:

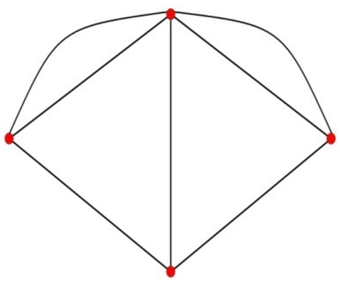




Most **at maximum distance (100, 10000, 1000000, 100000000...)**

so average distance is actually same order as diameter

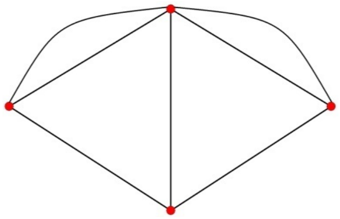
Social and Economic Networks: Models and Analysis



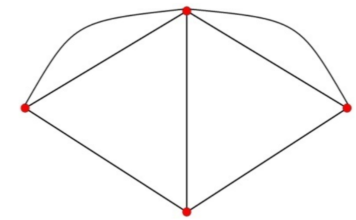
Matthew O. Jackson

Stanford University,
Santa Fe Institute, CIFAR,
www.stanford.edu\~jacksonm

1.7: Diameters in the World



Theorem on Network Structure



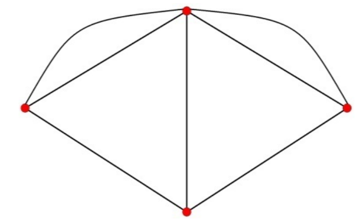
If $d(n) \geq (1+\varepsilon) \log(n)$ some $\varepsilon>0$ and $d(n)/n \rightarrow 0$

Then for large n, average path length and diameter are approximately proportional to $\log(n)/\log(d)$

(Proven for increasingly general models:

Erdos-Renyi 59 - Moon and Moser 1966, Bollobas 1981; Chung and Lu 01; Jackson 08; ...)

Small Worlds/Six Degrees of Separation

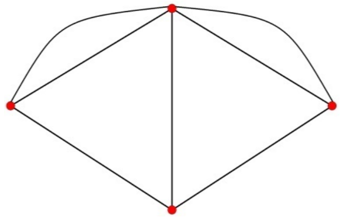


- n = 6.7 billion (world population)

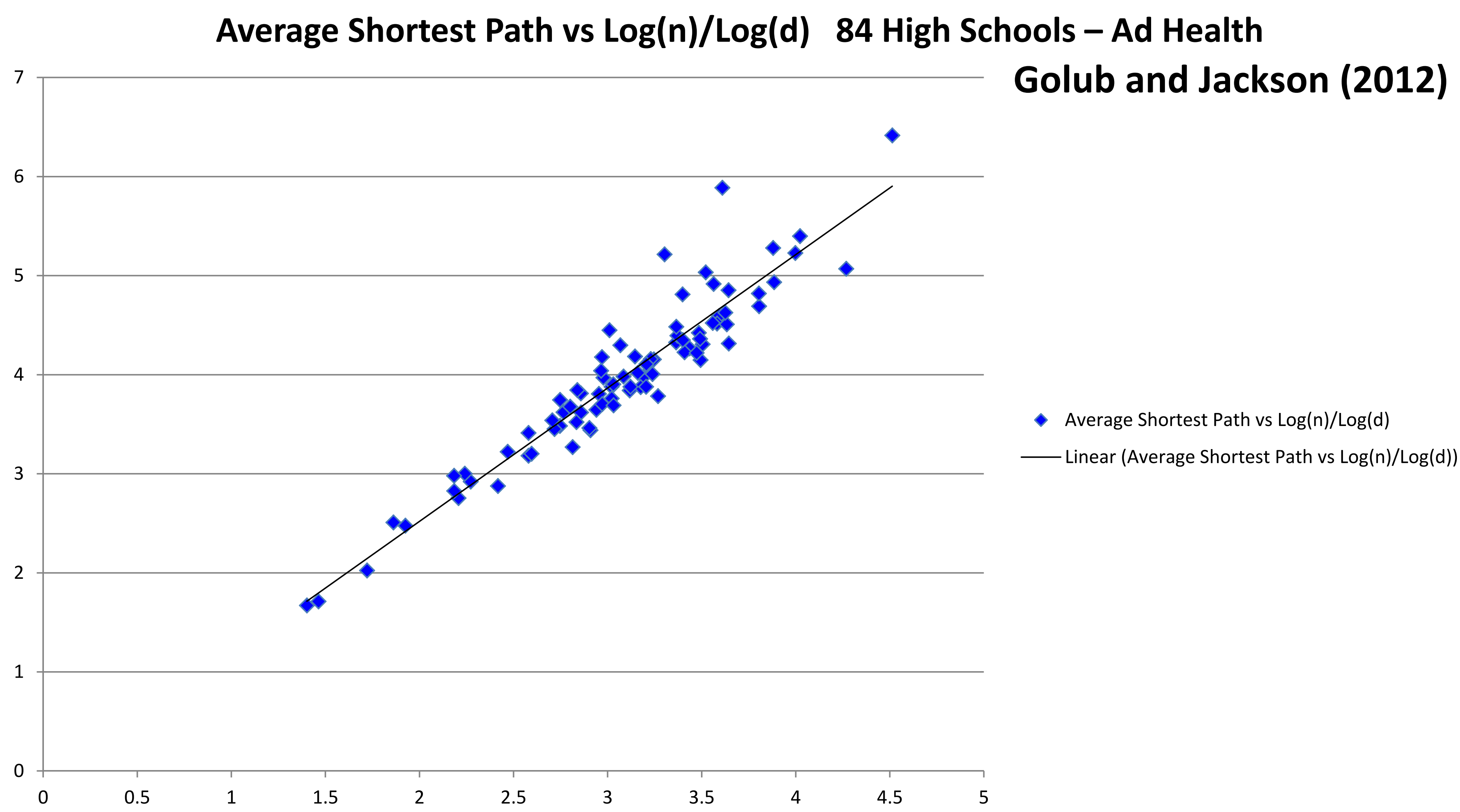
- d = 50 (friends, relatives...)

- log(n)/log(d) is about 6 !!

Examine data and diameter

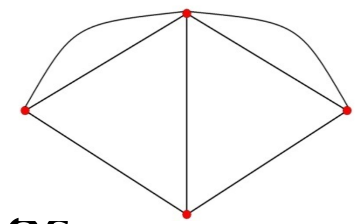


- Add Health data set
- Schools vary in average degree and homophily
- Does diameter match $\log(n)/\log(d)$?



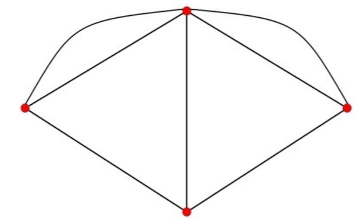
Average Shortest Path vs Log(n)/Log(d) 84 High Schools – Ad Health
Golub and Jackson (2012)
0
1
2
3
4
5
6
7
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
Average Shortest Path vs Log(n)/Log(d)
Linear (Average Shortest Path vs Log(n)/Log(d))

Erdos Numbers



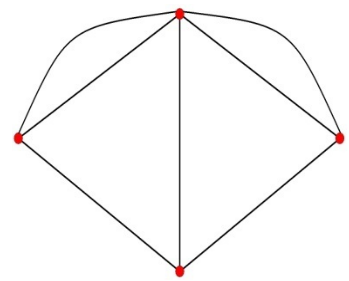
- Number of links in co-authorship network to Erdos
- Had 509 co-authors, more than 1400 papers
- 2004 auction of co-authorship with William Tozier (Erdos #=4) on E-Bay, winner paid > 1000$
- Kevin Bacon site....

Density: Average Degree



HS Friendships (CJP 09)	6.5
Romances (BMS 03)	0.8
Borrowing (BCDJ 12)	3.2
Co-authors (Newman 01, GLM 06)	
Bio	15.5
Econ	1.7
Math	3.9
Physics	9.3
Facebook (Marlow 09)	120

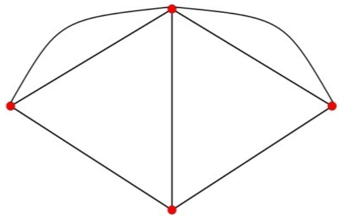
Social and Economic Networks: Models and Analysis



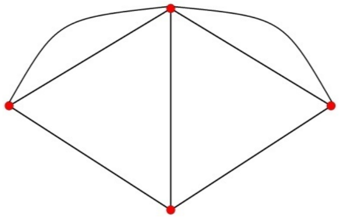
Matthew O. Jackson

Stanford University,
Santa Fe Institute, CIFAR,
www.stanford.edu\~jacksonm

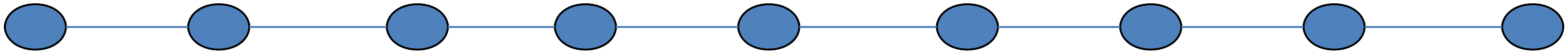
1.8: Degree Distributions

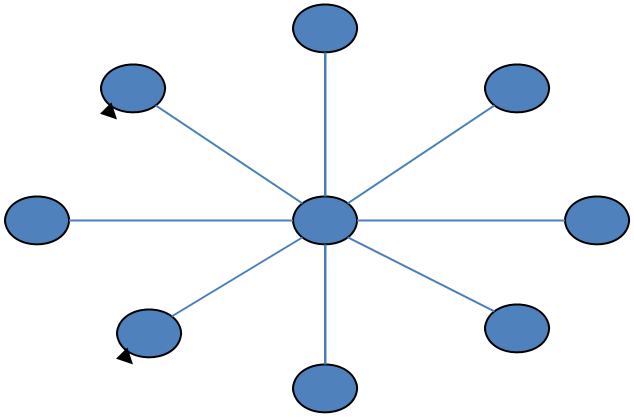


Degree Distributions

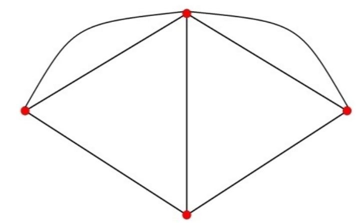


- Average degree tells only part of the story:





Degree Distribution, G(n,p):



- probability that node has d links is **binomial**

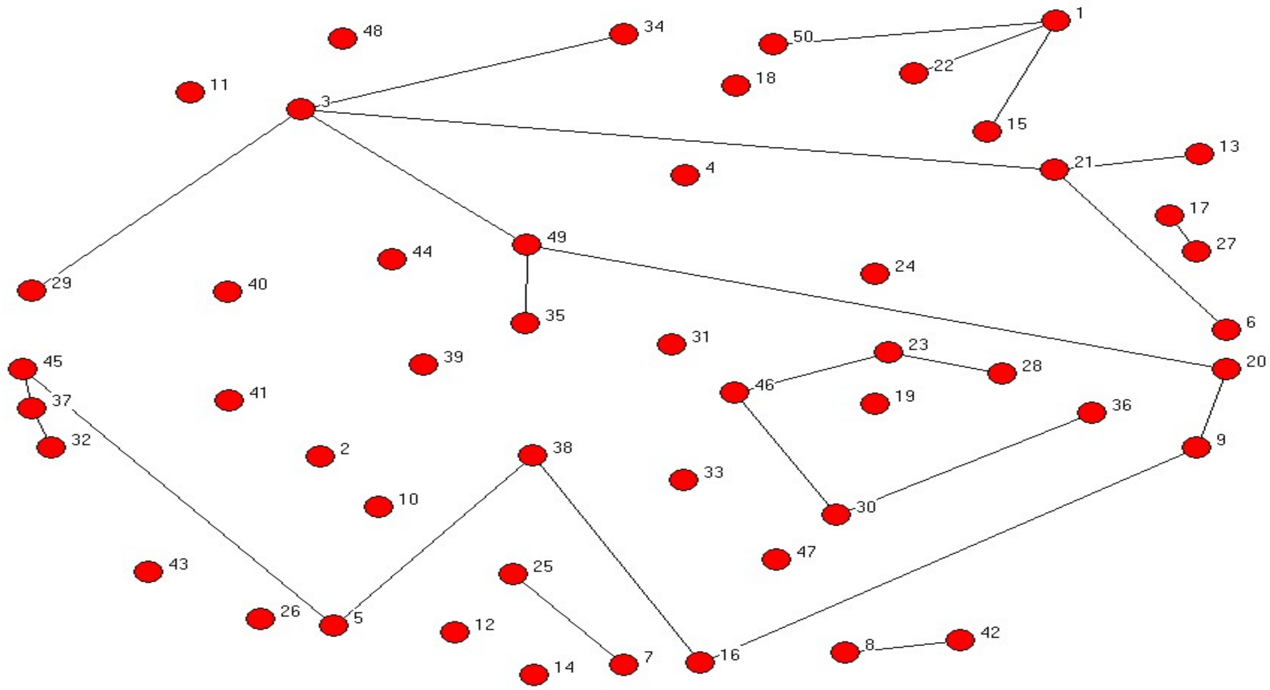
  $[\ (n-1)! \ / \ (d!(n-d-1)!)\ ]\ p^d\ (1-p)^{n-d-1}$

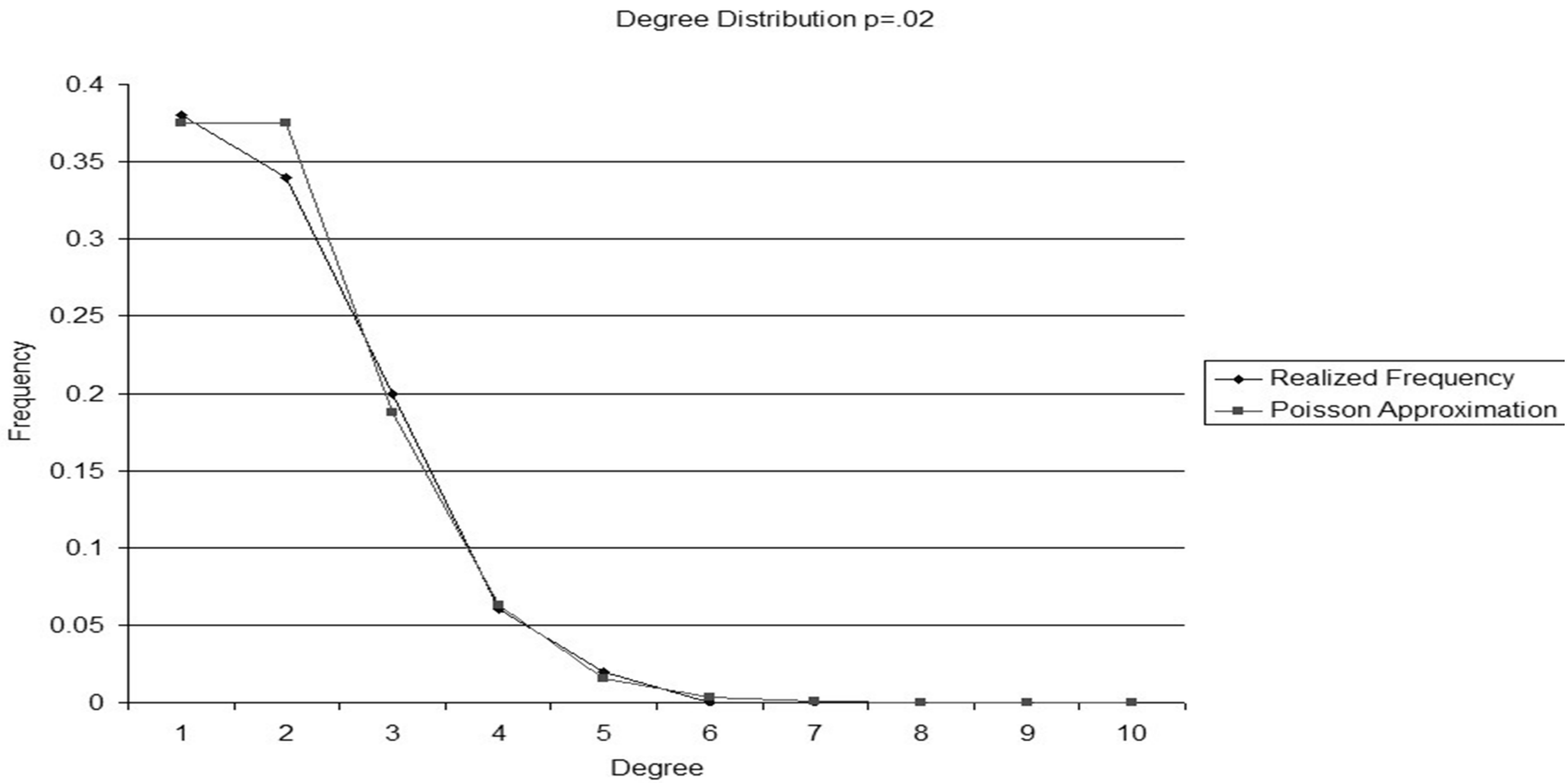
- Large n, small p, this is approximately a **Poisson** distribution:

  $[\ (n-1)^d \ / \ d!\ ]\ \ p^d\ e^{-(n-1)p}$

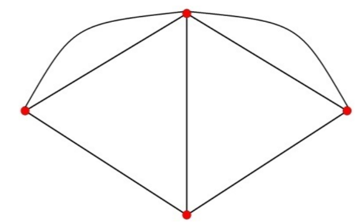
- hence name ``**Poisson random graphs**''

Random network
p=.02, 50 nodes




Degree Distribution p=.02
Frequency
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
1
2
3
4
5
6
7
8
9
10
Degree
Realized Frequency
Poisson Approximation

Note

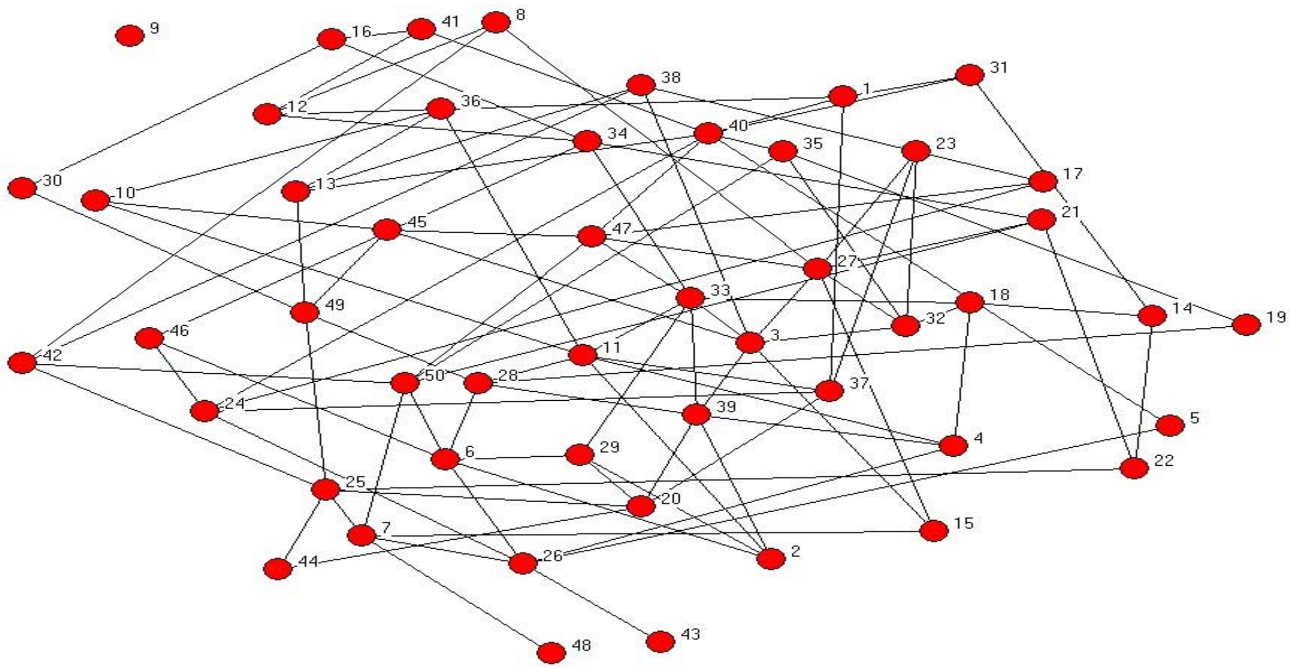


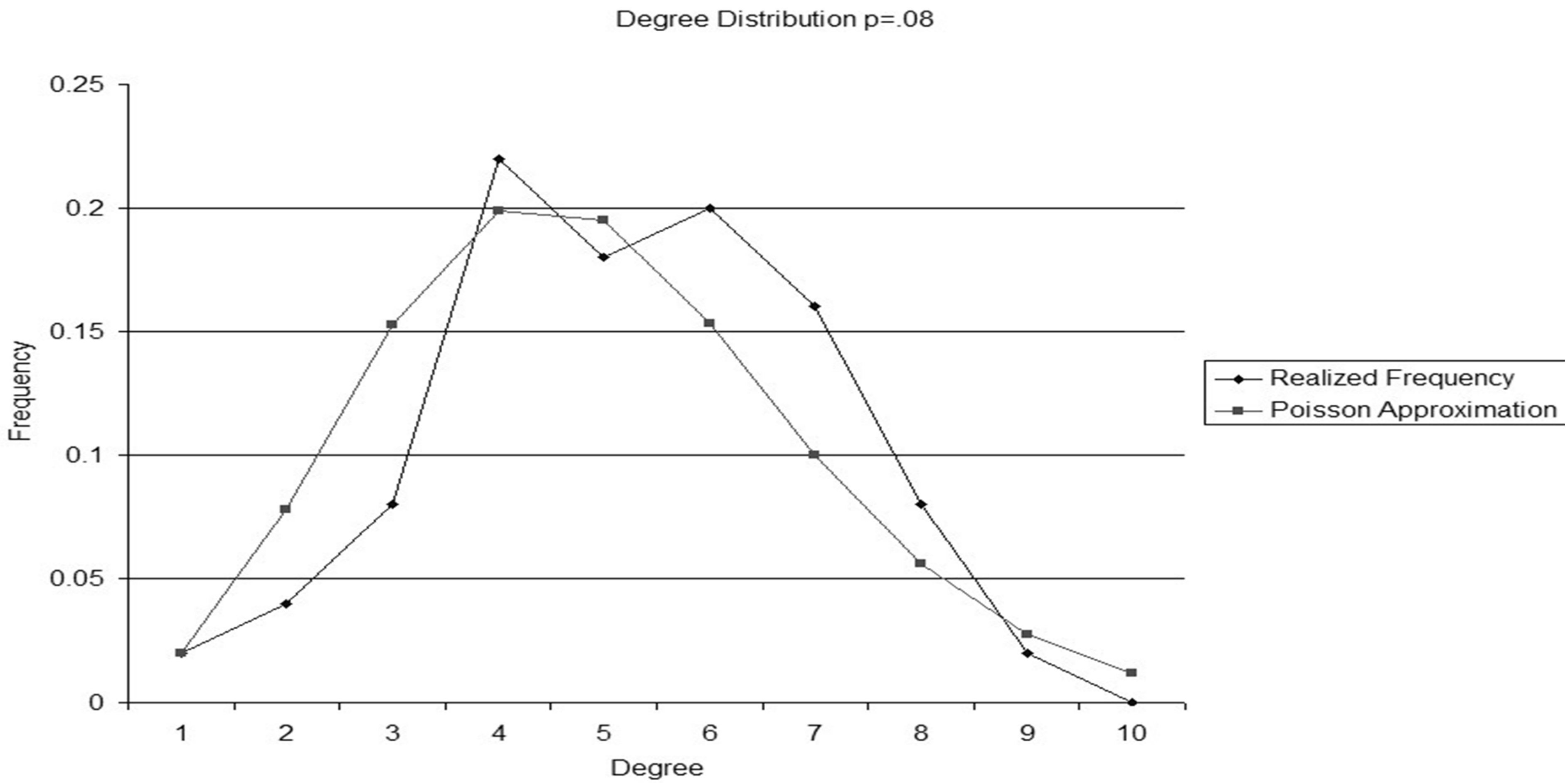
- many isolated nodes

- several components

- no component has more than a small fraction of the nodes, just starting to see one large one emerge

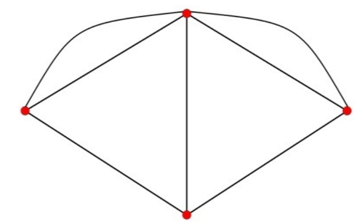
Random Network
p=.08, 50 nodes





Degree Distribution p=.08
Frequency
Degree
0.25
0.2
0.15
0.1
0.05
0
1
2
3
4
5
6
7
8
9
10
Realized Frequency
Poisson Approximation

Distribution of links per node: Fat tails (Price 1965)

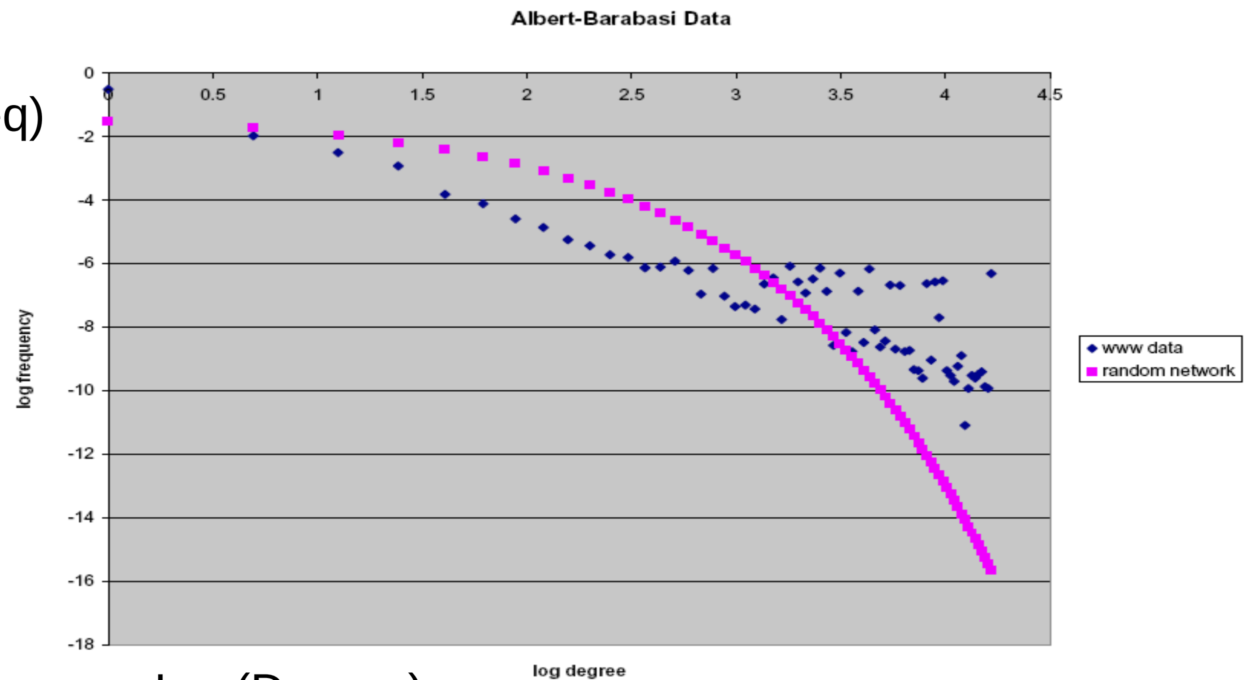


- More high and low degree nodes than predicted at random
  - Citation Networks - too many with 0 citations, too many with high numbers of citations to have citations drawn at random
  - ``Fat tails'' compared to random network

- Related to other settings (wealth, city size, word usage...): Pareto (1896), Yule (1925), Zipf (1949), Simon (1955),

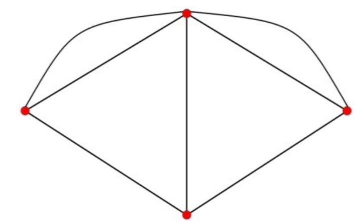
Degree – ND www Albert, Jeong, Barabasi (1999)

log(freq)



Log(Degree)

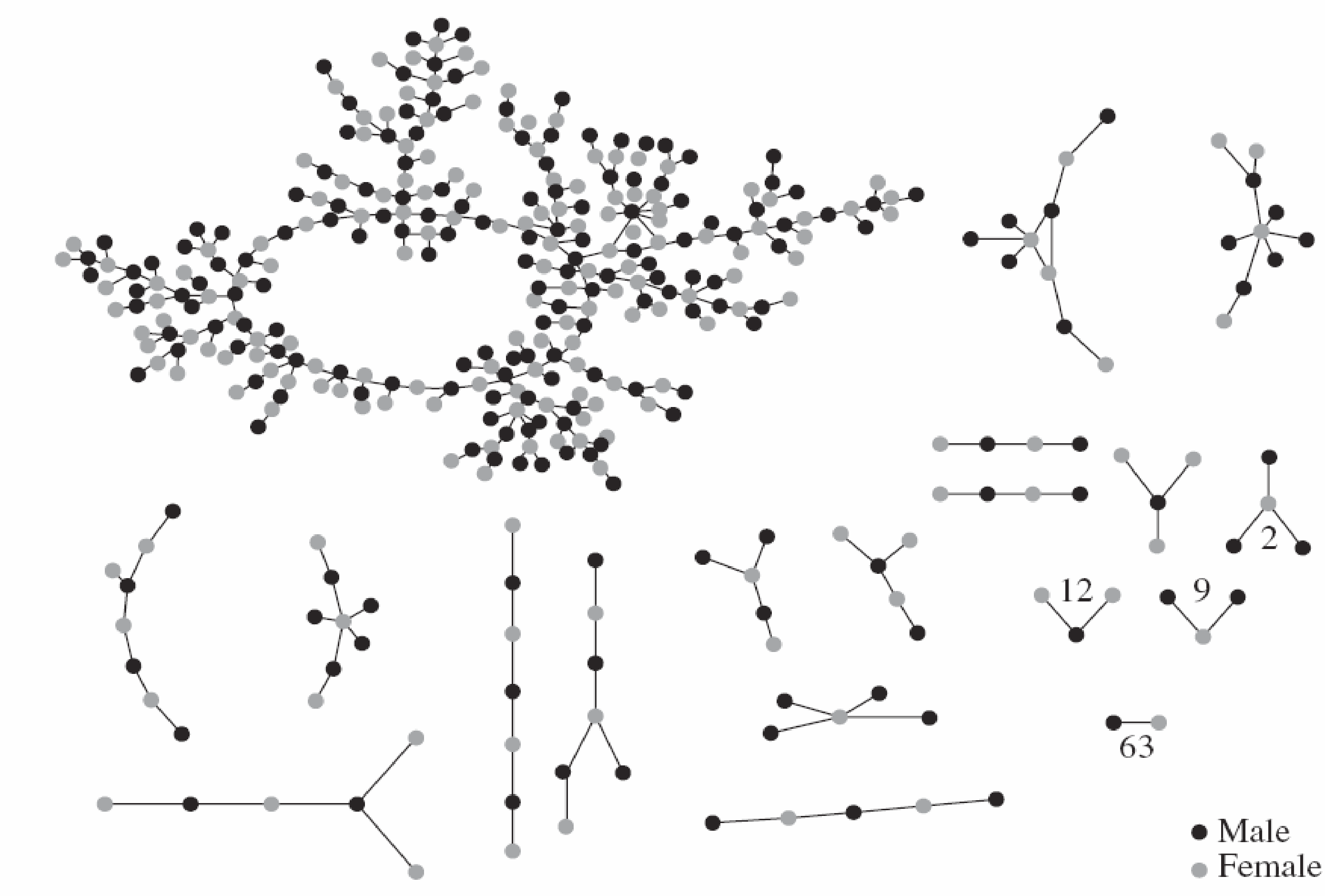
Scale Free Distributions



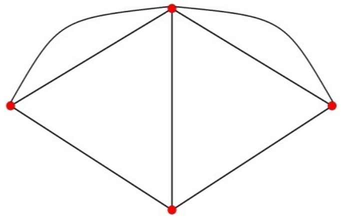
- $P(d) = c\, d^{-a}$

- $\log(P(d)) = \log (c) - a \log(d)$

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

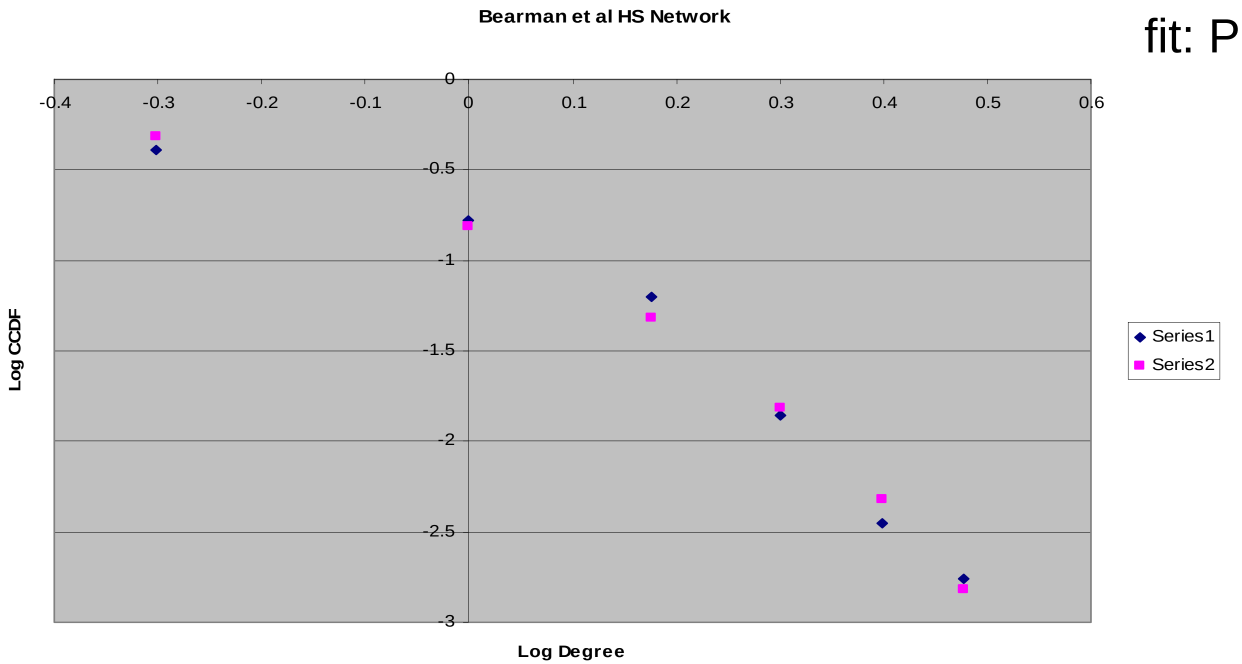


Romance Network

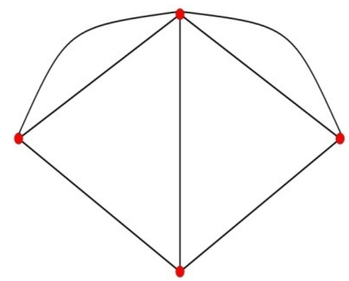


fit: Uniform at Random .99

fit: Power .84



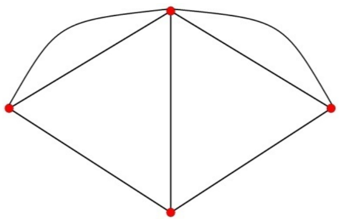
Social and Economic Networks: Models and Analysis



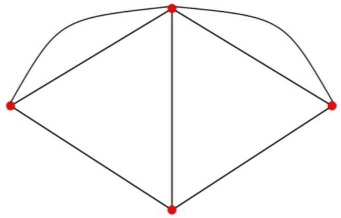
Matthew O. Jackson

Stanford University,
Santa Fe Institute, CIFAR,
www.stanford.edu\~jacksonm

1.9: Clustering



Clustering

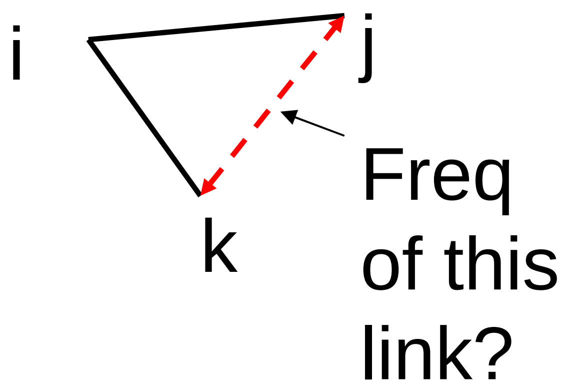


- What fraction of my friends are friends of each other?

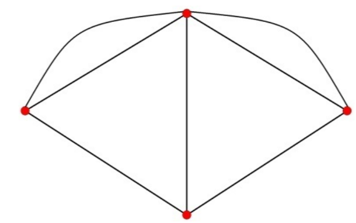
- $Cl_i(g) = \#\{ kj \text{ in } g \mid k, j \text{ in } N_i(g)\} \;/\; \#\{ kj \mid k, j \text{ in } N_i(g)\}$

- Average clustering:

$$Cl^{avg}(g) = \sum_i Cl_i(g) / n$$



Clustering



- What fraction of my friends are friends?

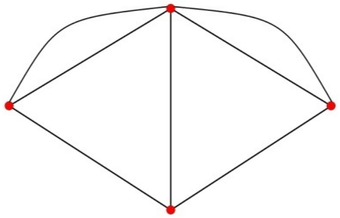
- $Cl_i(g) = \#\{ kj \text{ in } g \mid k, j \text{ in } N_i(g)\} \;/\; \#\{ kj \mid k, j \text{ in } N_i(g)\}$

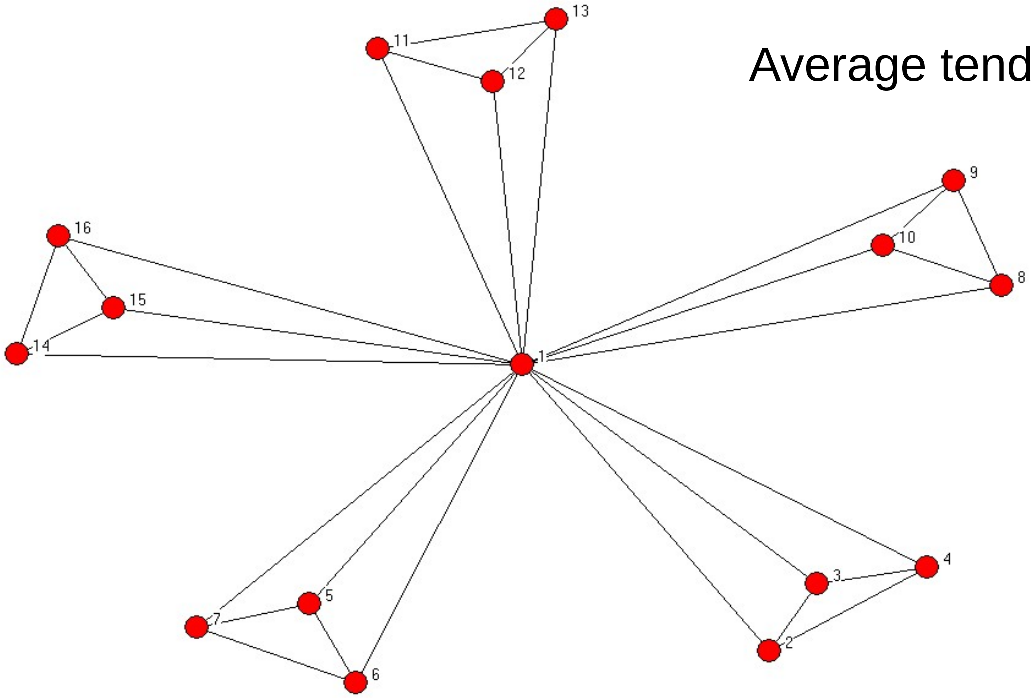
- Average clustering: $Cl^{avg}(g) = \sum_i Cl_i(g) / n$

- Overall clustering:

$Cl(g) = \sum_i \#\{ kj \text{ in } g \mid k, j \text{ in } N_i(g)\} \;/\; \sum_i \#\{ kj \mid k, j \text{ in } N_i(g)\}$

Differences in Clustering

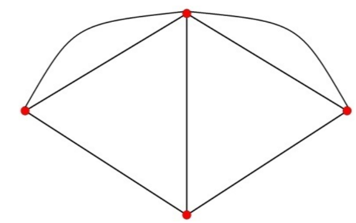




Average tends to 1

Overall tends to 0

Clustering in a Poisson Random Network



- Average and Overall clustering tend to 0, if max degree is bounded and network becomes large:

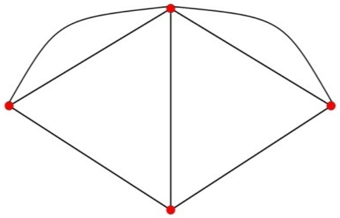
$Cl(g) = \sum_i \#\{ kj \text{ in } g \mid k, j \text{ in } N_i(g)\} \;/\; \sum_i \#\{ kj \mid k, j \text{ in } N_i(g)\}$

is simply p

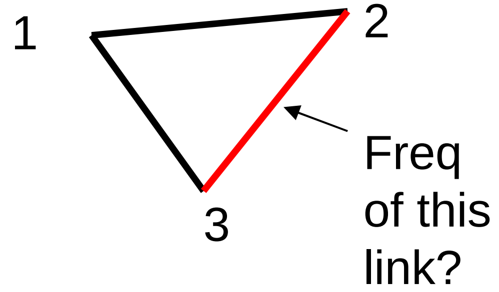
- If degree is bounded, then p (n-1) is bounded

- So p goes to 0 as n grows

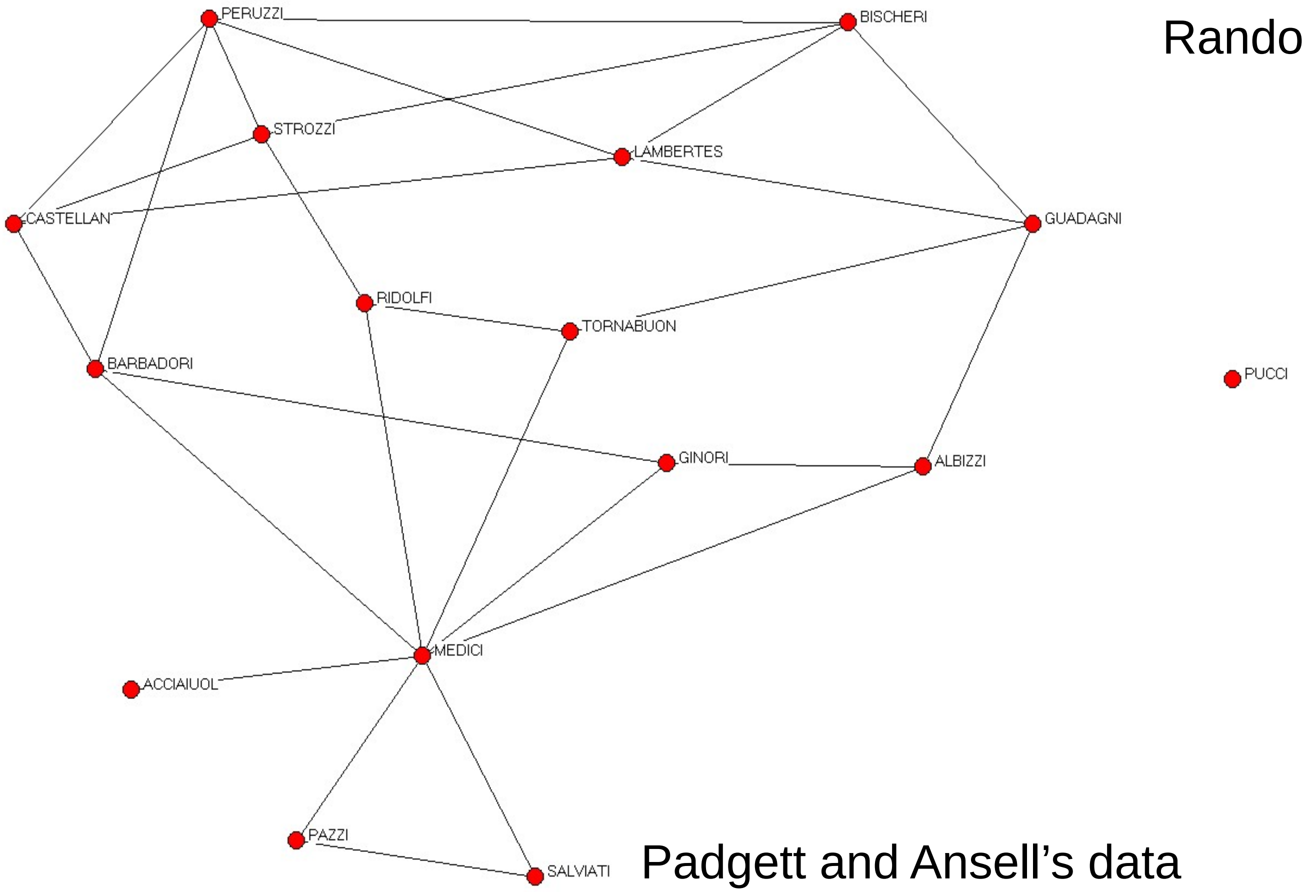
High? Clustering Coefficients -



- Prison friendships
  - **.31** (MacRae 60) vs .0134
- co-authorships
  - **.15** math (Grossman 02) vs .00002,
  - **.09** biology (Newman 01) vs .00001,
  - **.19** econ (Goyal et al 06) vs .00002,
- www
  - .11 for web links (Adamic 99) vs .0002

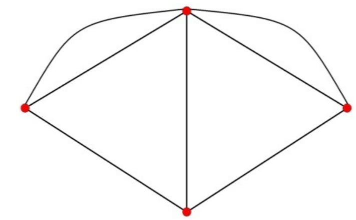


Clustering: .46
Random: .29



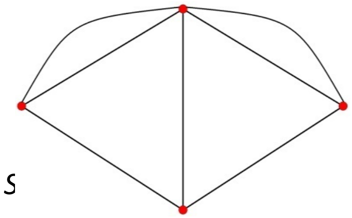
Padgett and Ansell's data
1430's Florentine marriages
and business dealings

Week 1 Wrap



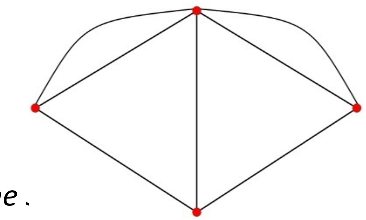
- Many relationships are ``networked'' and understanding network structure can help understand behavior and outcomes
- Networks are complex, but can be partly described by some characteristics
  - degree distributions
  - clustering
  - diameter ...
- Tree-like structures are generated by random links lead to short paths
- Many observed social networks are more clustered than would arise at random

Week 1: References In Order Mentioned



- Jackson, M.O. (2008) Social and Economic Networks, Princeton University Press
- Padgett, J.F., and C.K. Ansell (1993) “Robust Action and the Rise of the Medici, 1400–1434,” *American Journal of S*
- Kent, D. (1978) *The Rise of the Medici: Faction in Florence 1426–1434*, Oxford: Oxford University Press.
- Elliott, M.L., B. Golub and M.O. Jackson (2012) ``Financial Networks and Contagion,’’ SSRN working paper 2175056
- Myers C.A., and G.P. Shultz (1951) *The Dynamics of a Labor Market*, New York: Prentice-Hall.
- Rees AJ, Shultz GP. 1970. Workers in an Urban Labor Market. Chicago: Univ. Chicago Press
- Granovetter M. 1973. The strength of weak ties. Am. J. Sociol. 78: 1360 – 80
- Ioannides YM, Datcher-Loury L. 2004. Job information networks, neighborhood effects and inequality. J. Econ. Lit. 424: 1056 – 93
- Reiss AJ. 1980. Understanding changes in crime rates. In Indicators of Crime and Criminal Justice: Quantitative Studies. Washington, DC: Bur. Justice Stat.
- Reiss AJ. 1988. Co-offending and criminal careers. In Crime and Justice: A Review of Research, Vol. 10, ed. M Tonry. Chicago: Univ. Chicago Press
- Glaeser E, Sacerdote B, Scheinkman J. 1996. Crime and social interactions. Q. J Econ. 111: 507 – 48
- Uzzi B. 1996. The sources and consequences of embeddedness for the economic performance of organizations: the network effect. Am. Sociol. Rev. 61: 674 98
- Weisbuch, G., A. Kirman, and D. Herreiner (2000) “Market Organization,” *Economica* 110:411–436.
- Fafchamps M, Lund S. 2003. Risk-sharing networks in rural Philippines. J. Dev. Econ. 71: 261 – 87
- De Weerdt J. 2004. Risk-sharing and endogenous network formation. In Insurance Against Poverty, ed. S Dercon. Oxford: Oxford Univ. Press
- Ryan, B., and N.C. Gross (1943) “The Diffusion of Hybrid Seed Corn in Two Iowa Communities,” *Rural Sociology* 8:15–24.
- Griliches, Z. (1957) “Hybrid Corn: An Exploration in the Economics of Technological Change,” *Econometrica* 25(4):501–522.
- Coleman, J.S., E. Katz, and H. Menzel (1966) *Medical Innovation: A Diffusion Study,* Indianapolis, Ind.: Bobbs-Merrill.

Week 1: References Cont'd



- Milgram, S. (1967) "The Small-World Problem," *Psychology Today* 2:60–67.
- Grossman, J.W. (2002) "The Evolution of the Mathematical Research Collaboration Graph," in *Proceedings of the* *Conference on Combinatorics* (Congressus Numerantium, Vol. 158).
- Newman, M.E.J. (2001) Scientific collaboration networks. I. Network construction and fundamental results Phys. Rev. E 64, 016131
- Goyal, S., M. van der Leij, and J.-L. Moraga-Gonz´alez (2006) "Economics: An Emerging Small World," *Journal of Political Economy* 114(2):403–412.
- Adamic, L.A. (1999) The SmallWorldWeb, in Proceedings of the ECDL. Lecture Notes in Computer Science 1696, Berlin: Springer-Verlag.
- Backstrom, L., P. Boldiy, M. Rosay, .J. Ugander S. Vignay (2012) ``Four Degrees of Separation'' arXiv 1111.4570v3
- Erd¨os, P., and A. R´enyi (1959) "On Random Graphs," *Publicationes Mathematicae Debrecen* 6:290–297.
- ---(1960) "On the Evolution of Random Graphs," *Publication of the Mathematical Institute of the Hungarian Academy of Sciences* 5:17–61.
- --- (1961) "On the Strength of Connectedness of a Random Graph," *ActaMathamatica Academy of Sciences of Hungarica* 12:261–267.
- JW Moon, L Moser (1966) ``Almost all (0, 1) matrices are primitive,'' Studia Sci. Math. Hungar,
- B Bollobás (1981) ``Diameter of Random Graphs,'' - Trans. Am. Math. Soc., 1981 - ams.org
- Chung, F., and L. Lu (2002) The Average Distances in Random Graphs with Given Expected Degrees, Proceedings of the National Academy of Sciences 99:15879 15882.
- Jackson, M.O. (2008b) Average Distance, Diameter, and Clustering in Social Networks with Homophily, arXiv:0810.2603v1, in the Proceedings of the Workshop in Internet and Network Economics (WINE 2008), Lecture Notes in Computer Science, Edited by C. Papadimitriou and S. Zhang, Springer Verlag, Berlin Heidelberg.
- Golub, B. and M.O.Jackson (2012) ``Network Structure and the Speed of Learning: Measuring Homophily Based on its Consequences,'' *Annals of Economics and Statistics* 107/108 July-Dec., 2012.

Week 1: References Cont'd

- Currarini, S., M.O. Jackson, and P. Pin (2010) "Identifying the roles of race-based choice and chance in high school friendship network formation," in the Proceedings of the National Academy of Sciences, 107(11): 4857 – 4861
- Bearman, P., J. Moody, and K. Stovel (2004) "Chains of Affection: The Structure of Adolescent Romantic and Sexual Networks," Chicago, University of Chicago, manuscript.
- Banerjee, A., A.G.Chandrasekhar, E. Duflo, M.O. Jackson (2012) ``Diffusion of Microfinance,'' *Science*, 26 July 2013: Vol. 341 no. 6144, DOI: 10.1126/science.1236498.
- Marlow, C. (2009). Maintained Relationships on Facebook. mimeo
- Price DJS. 1965. Networks of scientific papers. Science 149: 510 15
- Pareto, V. (1896) *Cours d'Economie Politique,* Geneva: Droz
- Yule, G. (1925) "A Mathematical Theory of Evolution Based on the Conclusions of Dr. J.C. Willis," *F.R.S. Philosophical Transactions of the Royal Society of London B* 213:21–87.
- Zipf, G. (1949) *Human Behavior and the Principle of Least Effort*, Cambridge, Mass.: Addison-Wesley.
- Simon, H. (1955) "On a Class of Skew Distribution Functions," *Biometrika* 42(3–4):425–440.
- Albert, R., H. Jeong, and A.L. Barab´asi (1999) "Diameter of theWorldWideWeb," *Nature* 401:130–131.
- MacRae, J. (1960) "Direct Factor Analysis of Sociometric Data," *Sociometry* 23:360–371.