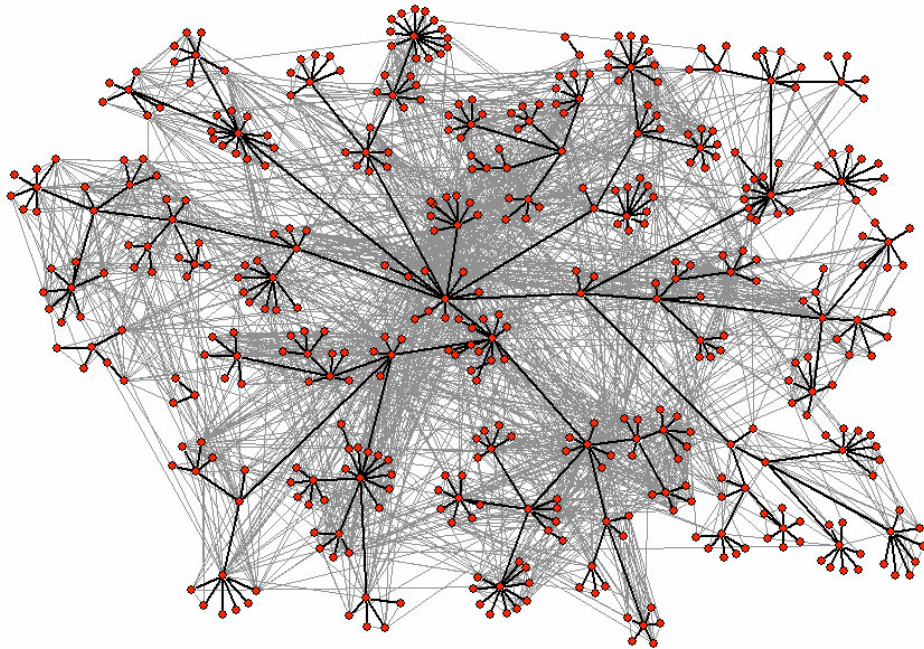


Online Social Networks and Network Economics



Who?

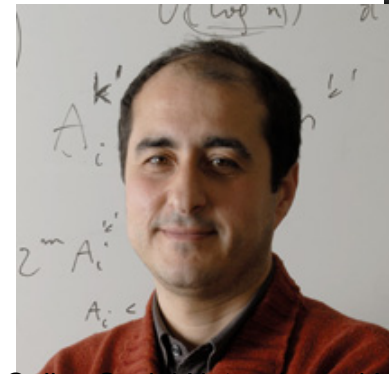
- **Dr. Aris Anagnostopoulos**



- **Prof. Guido Schäfer**



- **Prof. Stefano Leonardi**



What will We Cover?

Possible topics:

- Structure of social networks
- Models for social networks
- Epidemics and influence processes
- Mining of social networks
- Detecting communities
- Auction theory
- Computational advertising

What Do We Need?

- Some math background:
 - A bit of probability theory
 - Some basic game theory
- Some Programming
- Participation!

Exams etc.

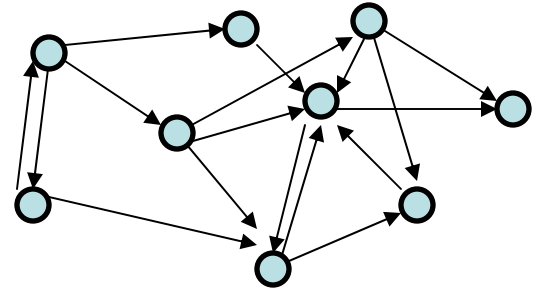
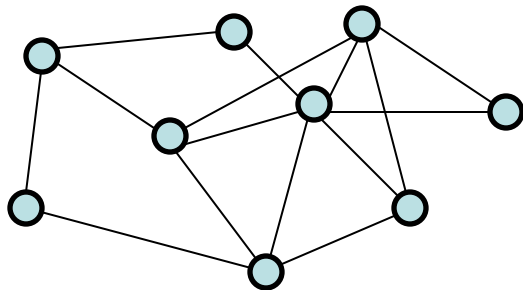
- **Class:**
 - Monday: 3.45 – 5.15
 - Wednesday: 2.00 – 3.30 (this week: 10.15 – 11.45)
- **Homeworks:**
 - 2-3 homework sets, 3 weeks for each
- **Exam:** We'll see...
 - No written class exam
- **Final project – Ideas:**
 - Programming project (e.g., write a facebook application)

Webpage – Office hours

- **Book:**
 - No book
 - Some notes for some lectures
 - Slides
- **Webpage:** www.dis.uniroma1.it/~socialnet
 - We use it for slides, notes, **announcements**
- **Office hours:**
 - Wednesday after class (Office: B118)
 - Through email

What Is a Social Network?

- **Social network:** graph that represents relationships between independent agents.

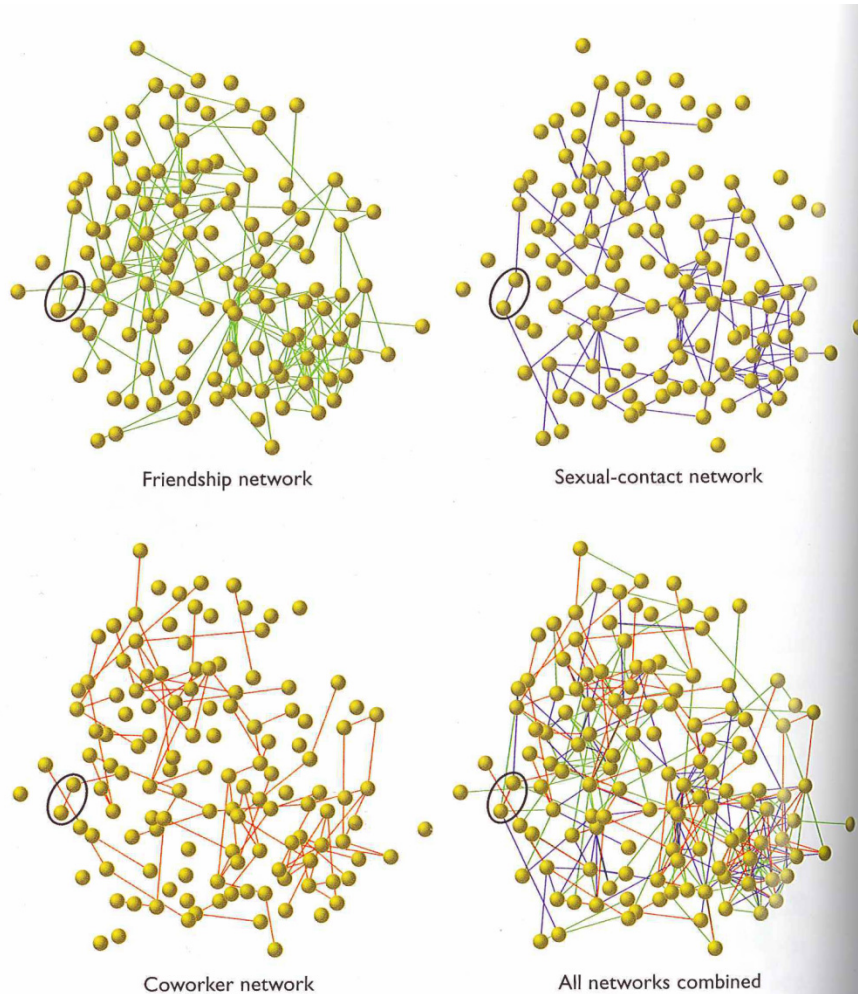


Social Networks Are Everywhere and Are Important!

Offline:

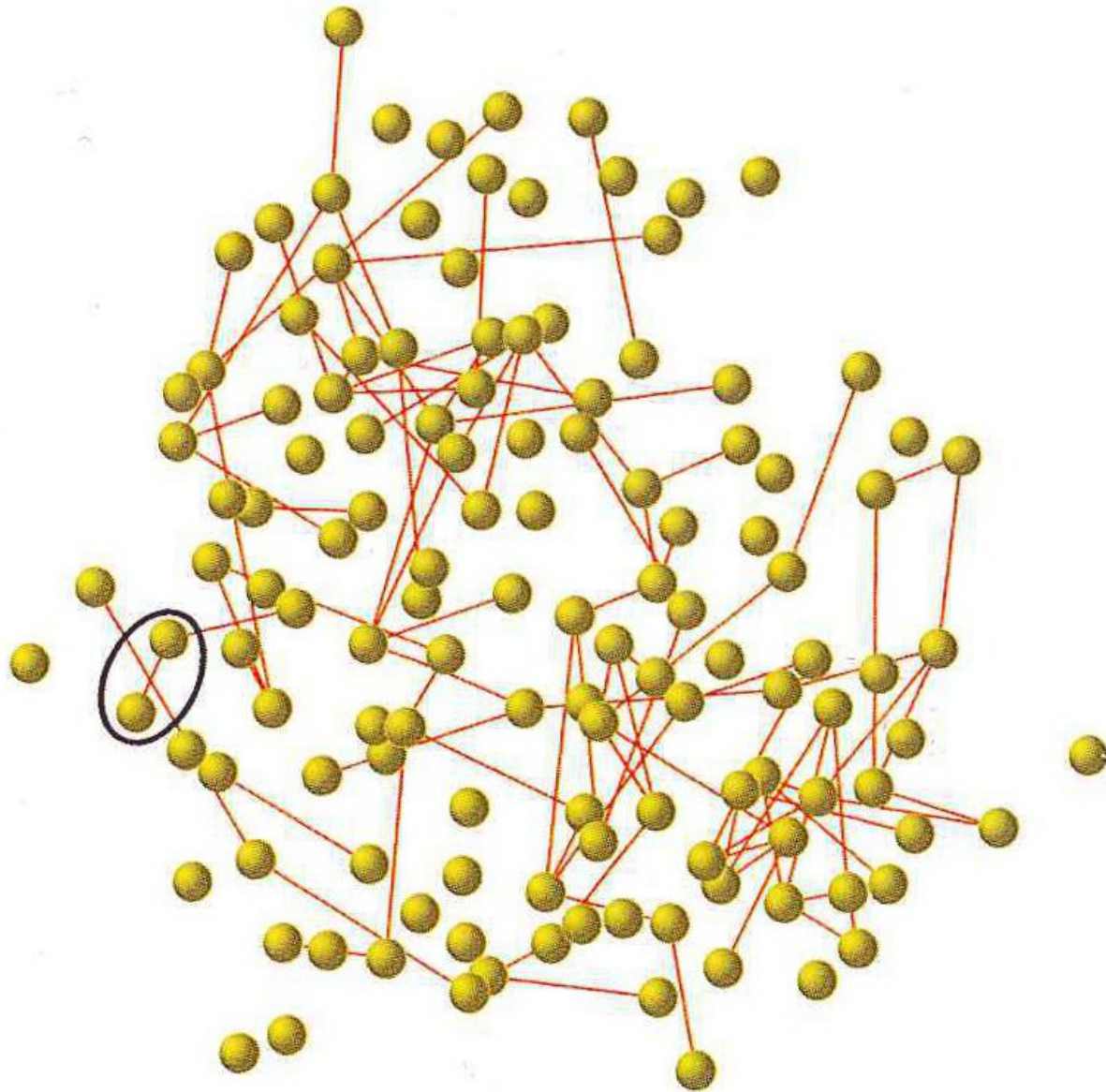
- Friendship network
 - “Show me your friend and I’ll show you who you are!”
- Professional contacts
 - Finding jobs
- Network of colleagues
 - learning new techniques
- Network of animals
 - E.g., two cows are connected if they have been in the same area
 - Mad-cow disease

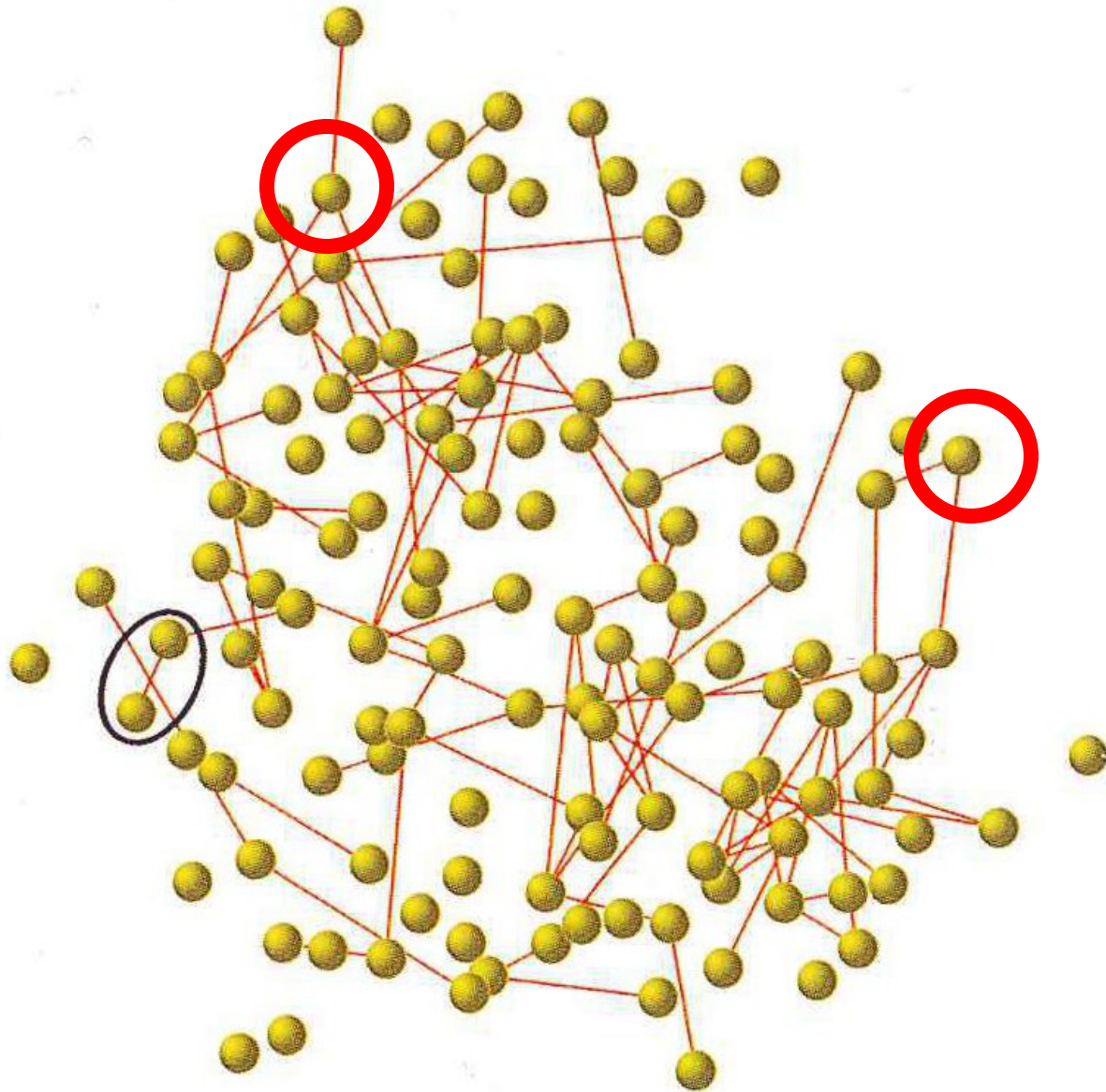
Multiple Social Networks

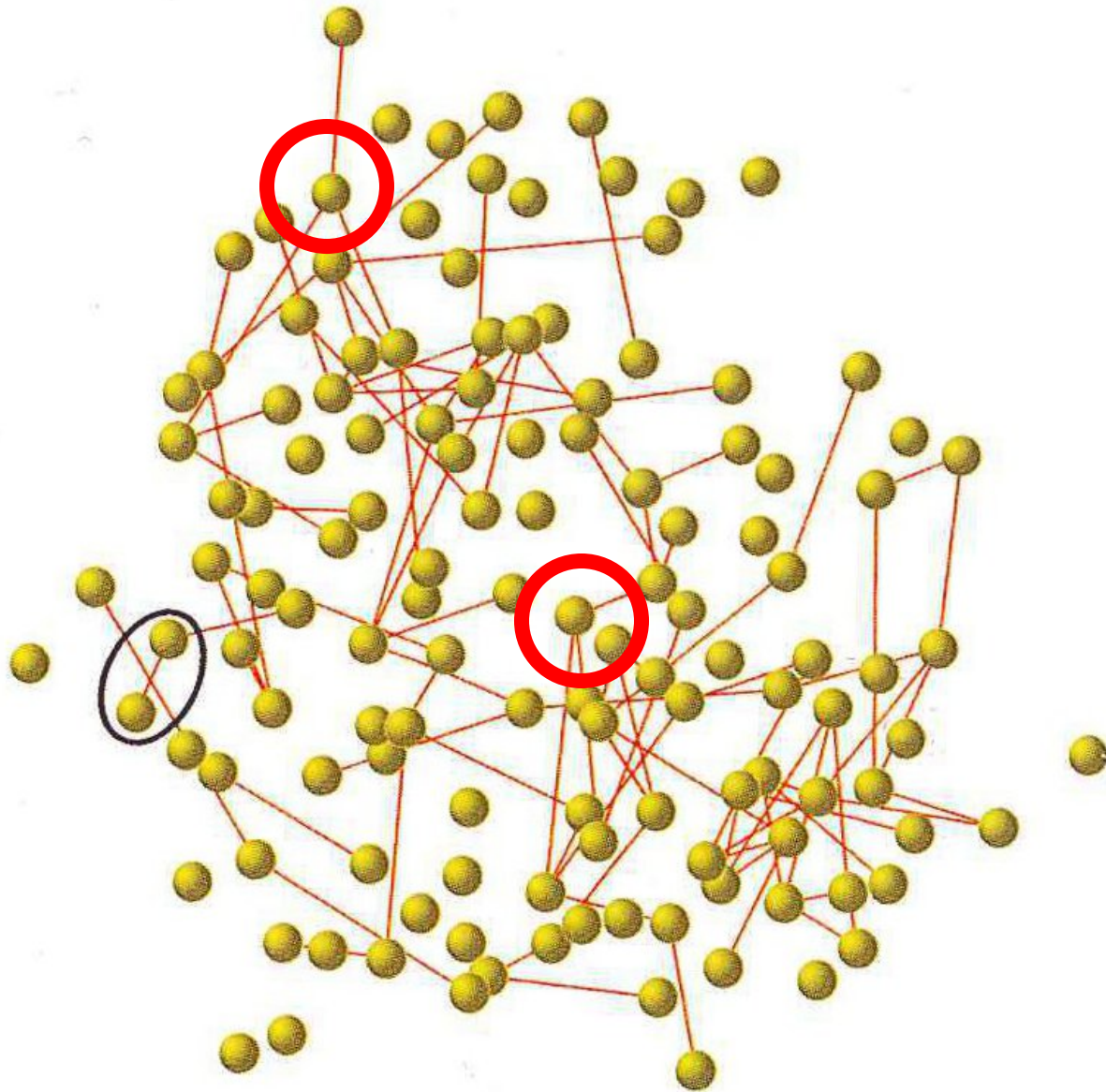


Examples:

- **Obesity:**
 - People with obese friends have higher probability to become obese
- **Smoking**
 - If your friends smoke you have higher chances to smoke
- **Happiness**
 - If your friends make you happy you become happy
- There is effect not only to friends, but to friends of friends and to friends of friends of friends







Social Networks Are Everywhere and Are Important!

Online — Web 2.0 systems:

- Social networking systems



- Content sharing systems



- Content creation systems



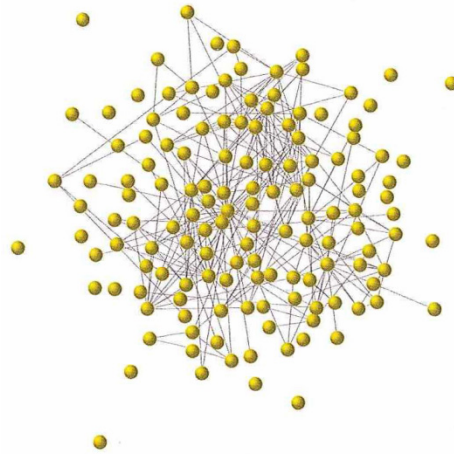
- Online games



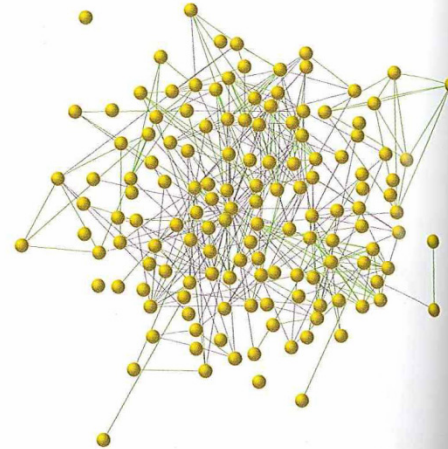
Online Revolution

- People switch more and more of their interactions from offline to online
- Pushing the # of contacts we can keep track of (Dunbar number)
- Redefining privacy
- Ideal for experiments in social sciences:
 - Ability to measure and record all activities
 - Massive data sets

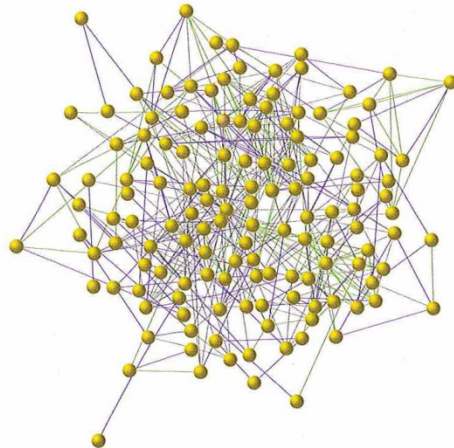
Online Revolution



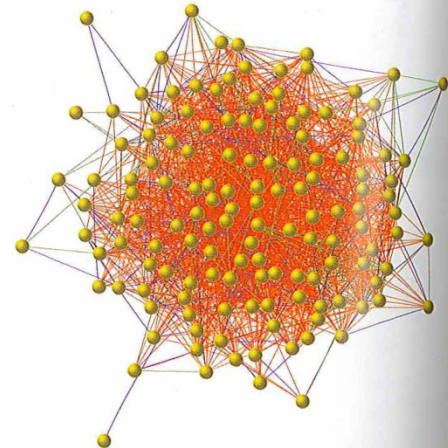
Close friends



Close friends and fellow club members



Close friends, club members, and roommates



Close friends, club members, roommates, and Facebook friends

Structure of Social Networks

- Social networks are an example of **complex networks**
- Other examples:
 - WWW, Citation graph, Biological networks, Internet, Telephone networks, Electricity grid, ...
- Studied by Mathematicians, Physicists, Computer Scientists, Sociologists, Biologists
- A lot of similar characteristics

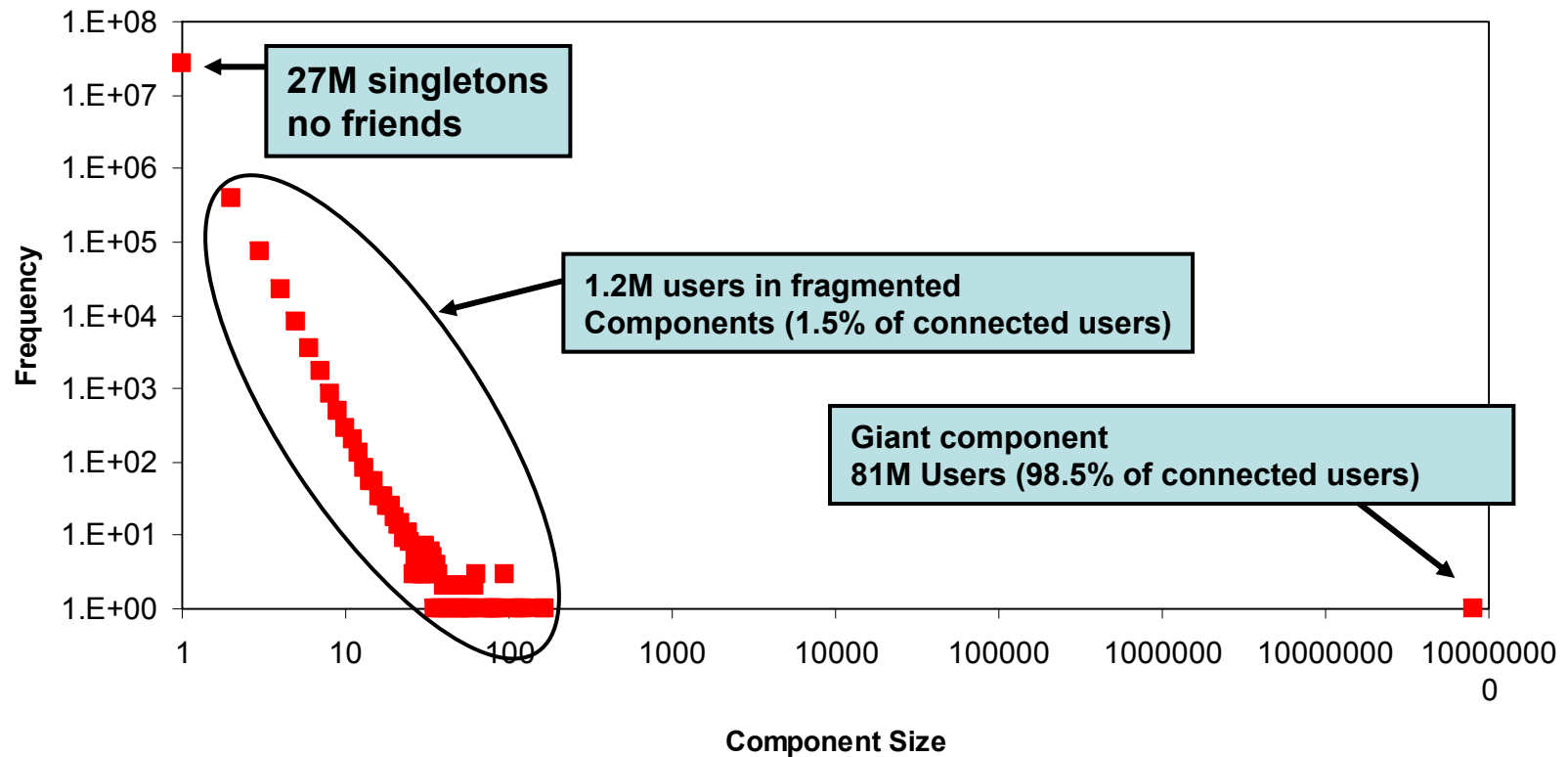
Structure of Complex Networks

1. One giant component
2. Power-law degree distributions
3. Small world
4. Globally sparse, locally dense

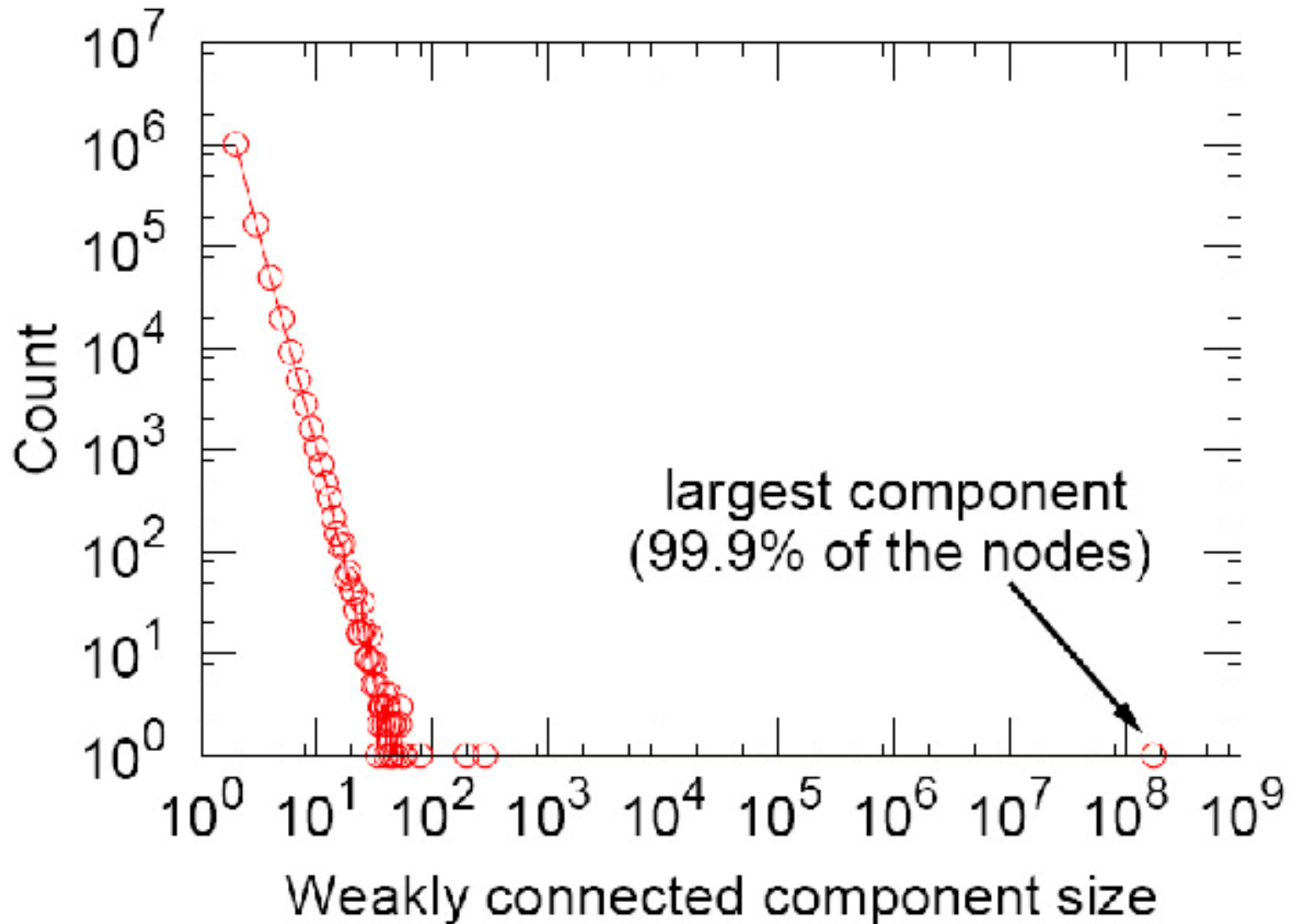
Giant Component

- There is a large connected component containing the vast majority of the nodes
- The second smallest is much much smaller
- There are a lot of disconnected nodes

Yahoo! Messenger



MSN Messenger



Power-Law Degree Distributions

- The degree distributions of the networks follow a power-law distribution
- What is power law?

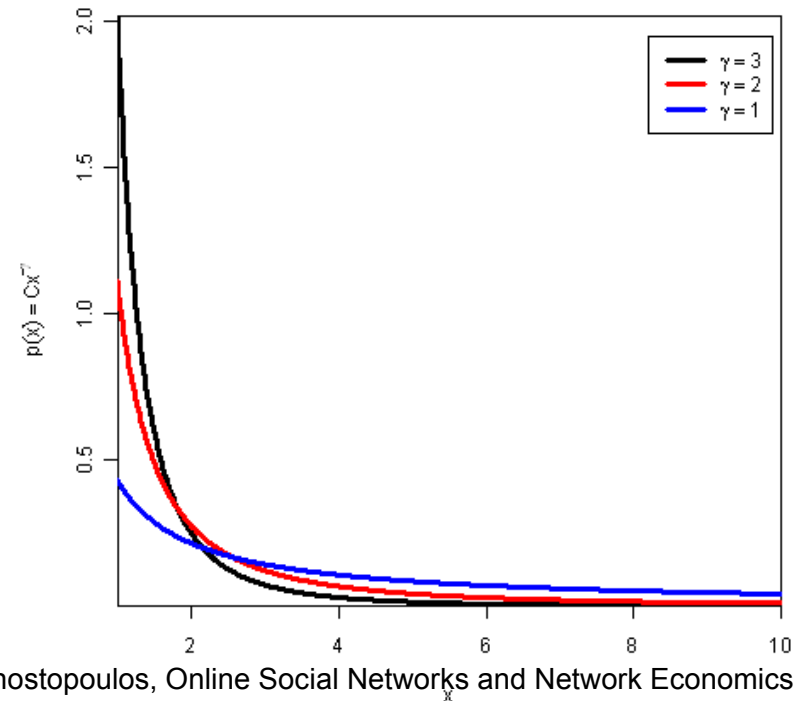
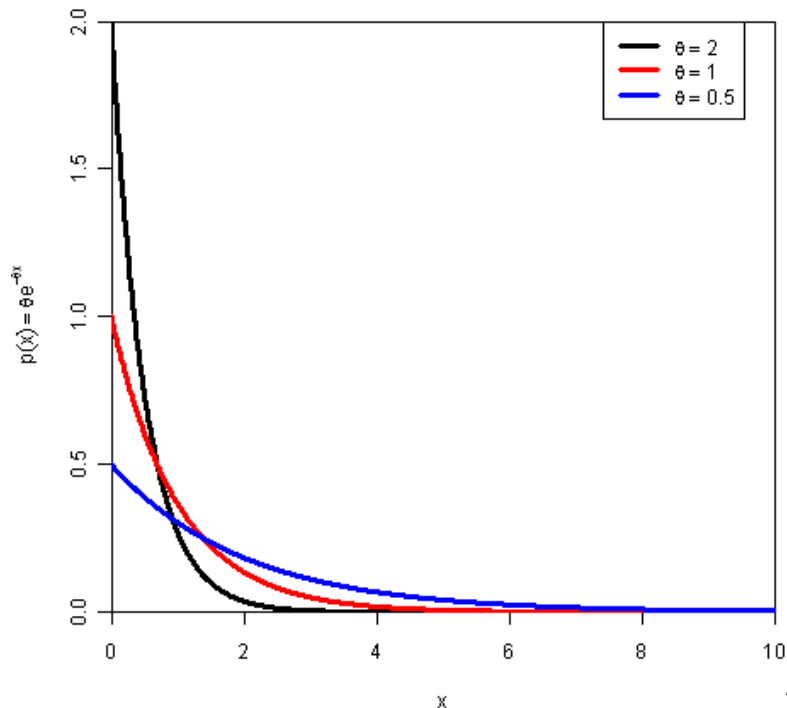
Power-Law Distribution

- Exponential distribution:

$$p(x) = \theta e^{-\theta x}$$

- Power-law distribution:

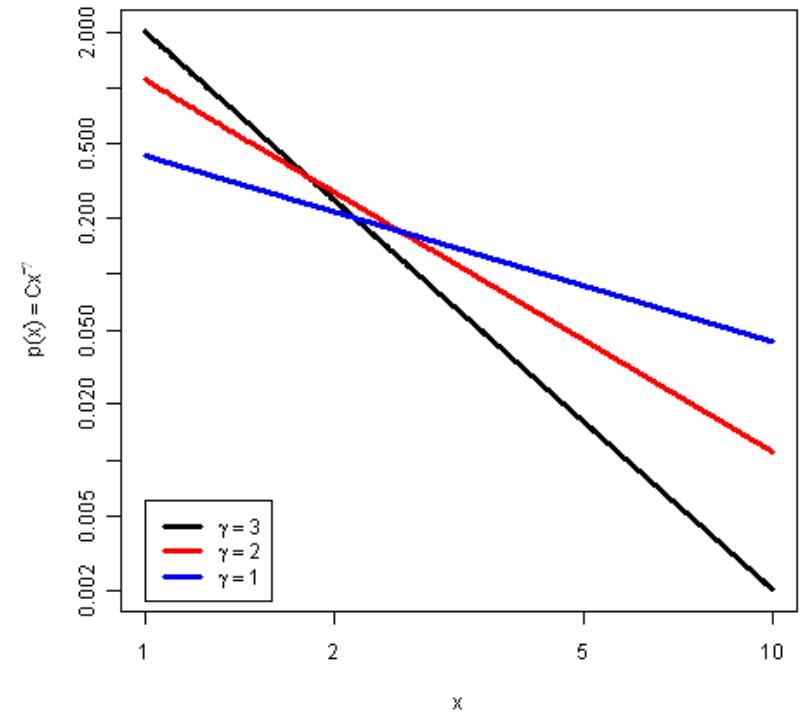
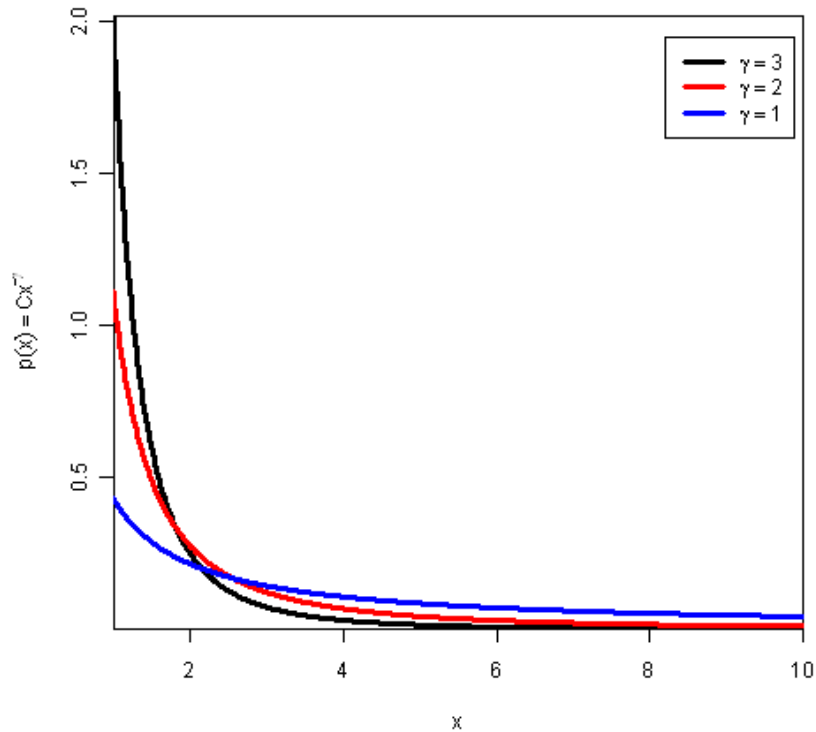
$$p(x) = C \cdot x^{-\gamma}$$



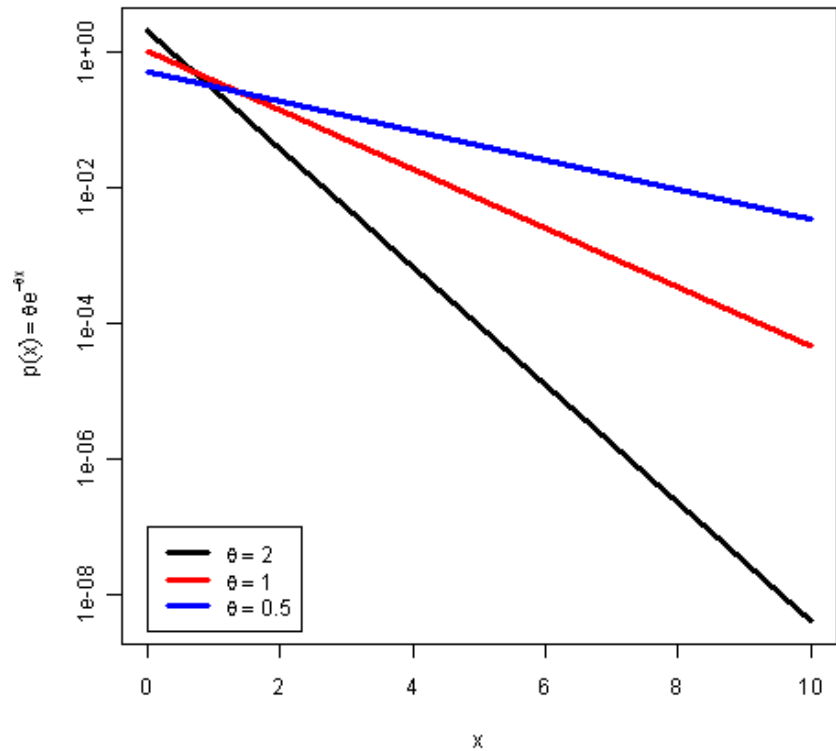
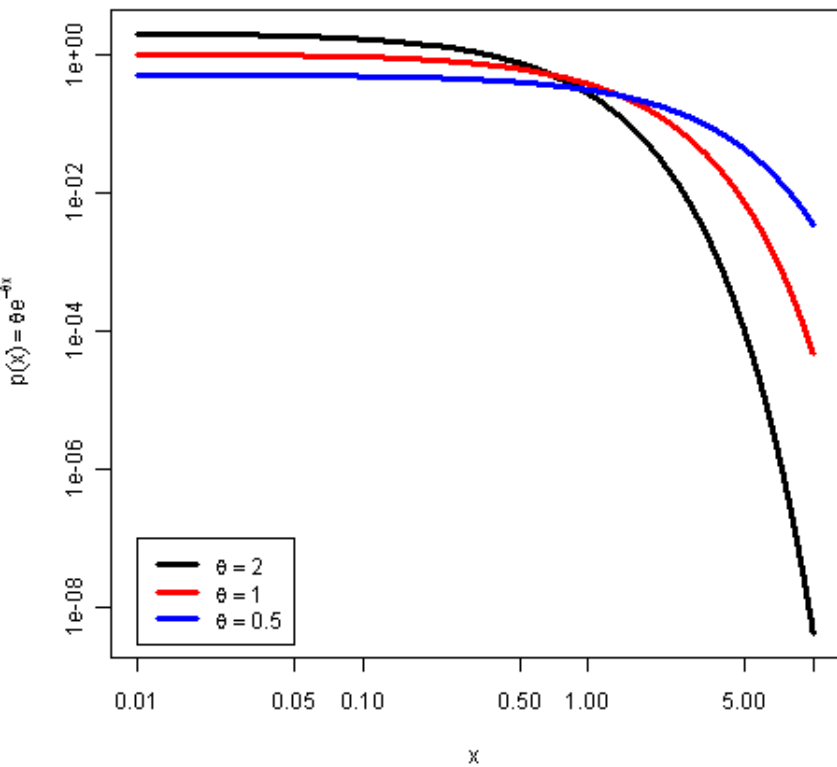
Power-Law Distribution - 2

- It is a heavy-tail distribution
- **Heavy tail:** It decays slower than an exponential
- It is also called scale-free: $f(ax) = b \cdot f(x)$
- It appears in many places:
 - Degree distribution
 - Population of cities
 - Word frequencies
 - Website hits
 - Income

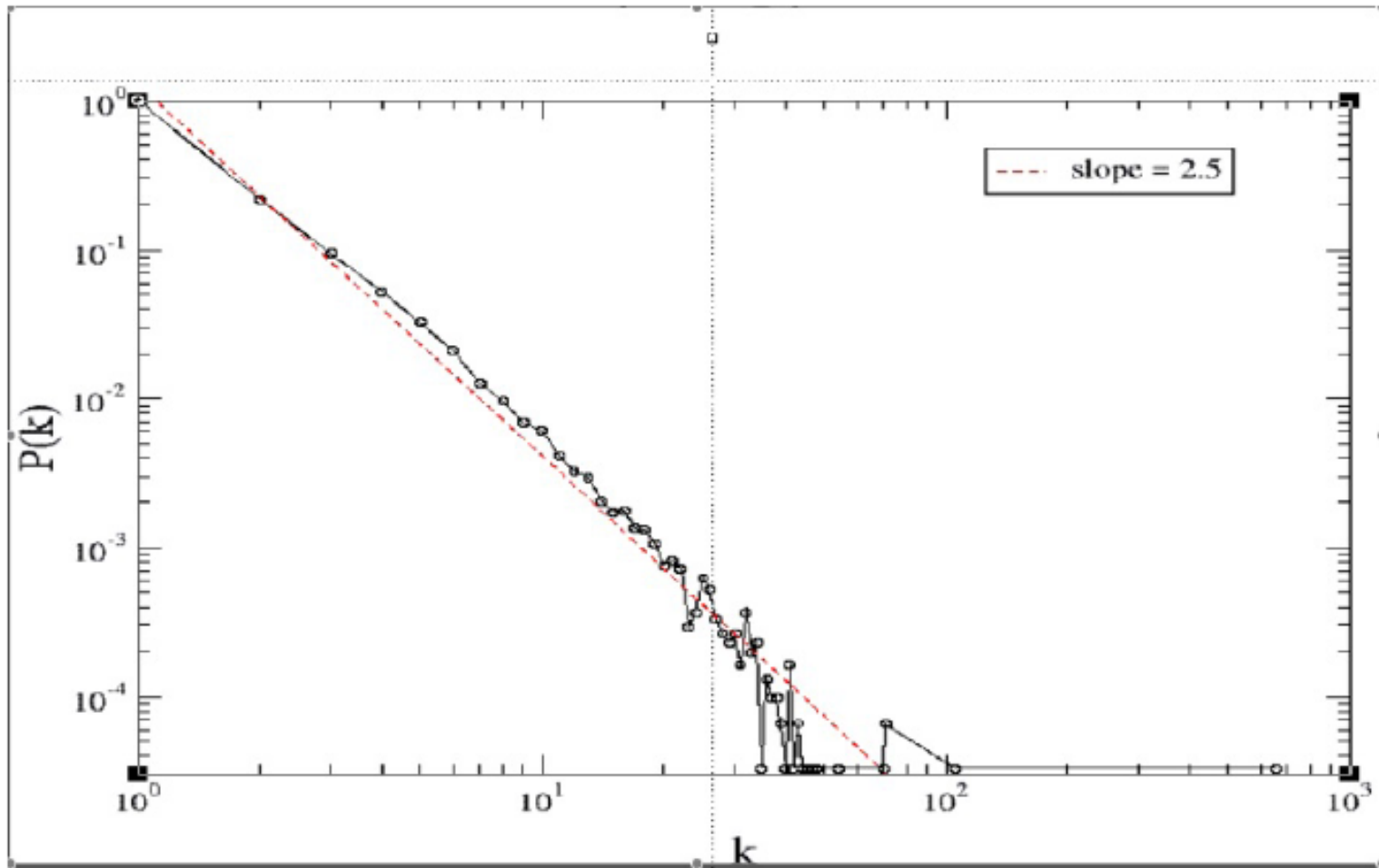
Power-Law Distribution - 3



Exponential Distribution



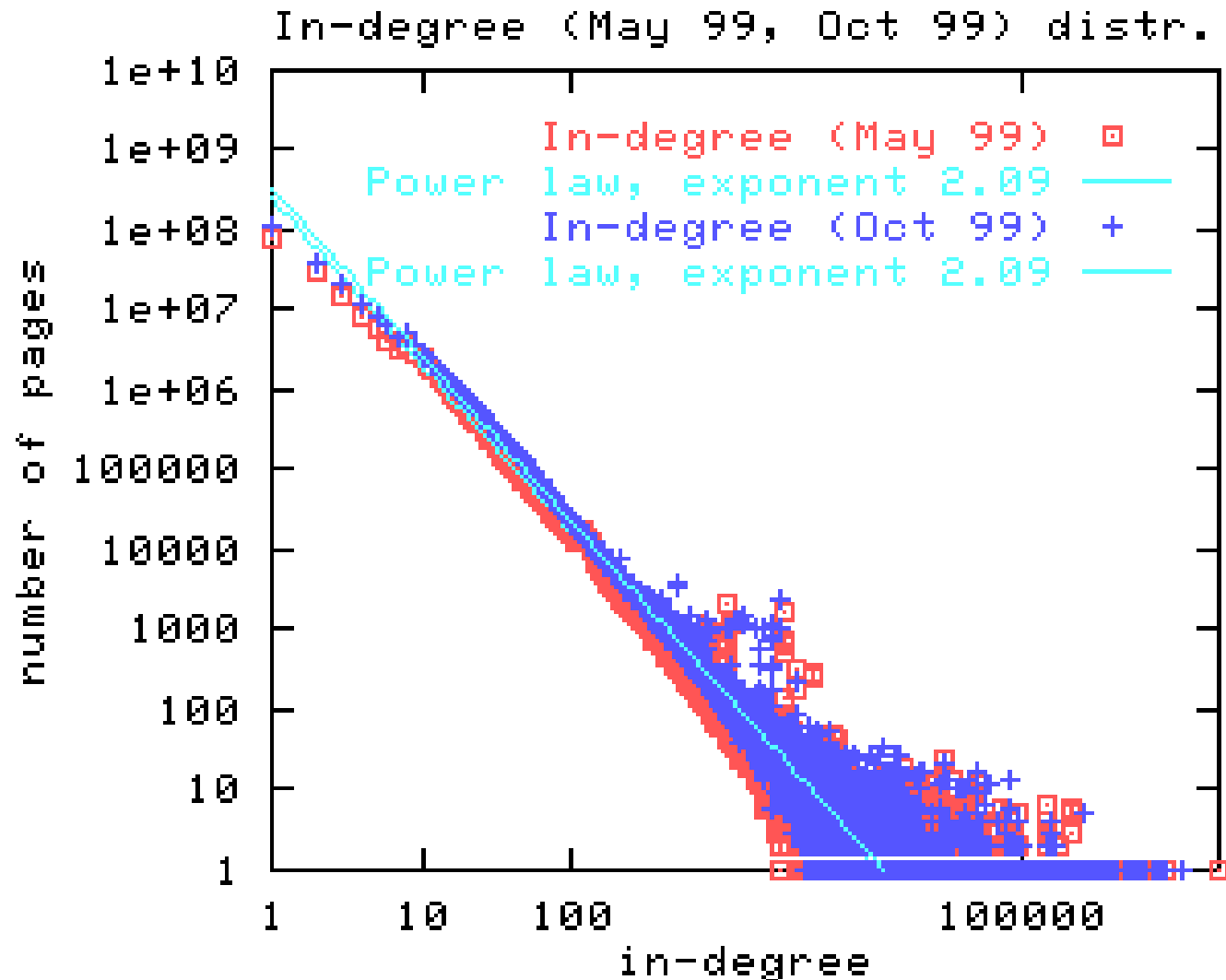
Back to Degree Distributions



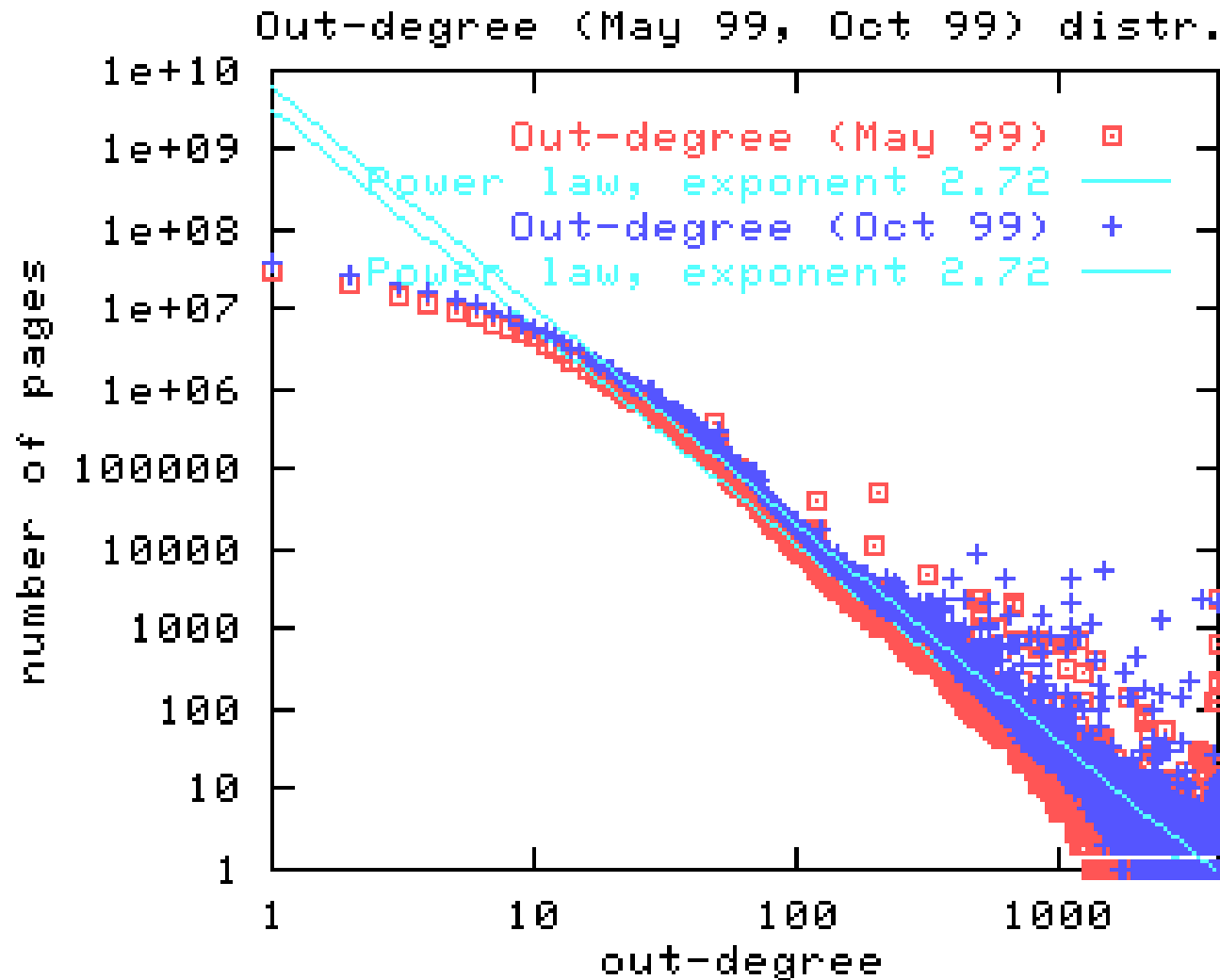
Internet Graph

Aris Anagnostopoulos, Online Social Networks and Network Economics

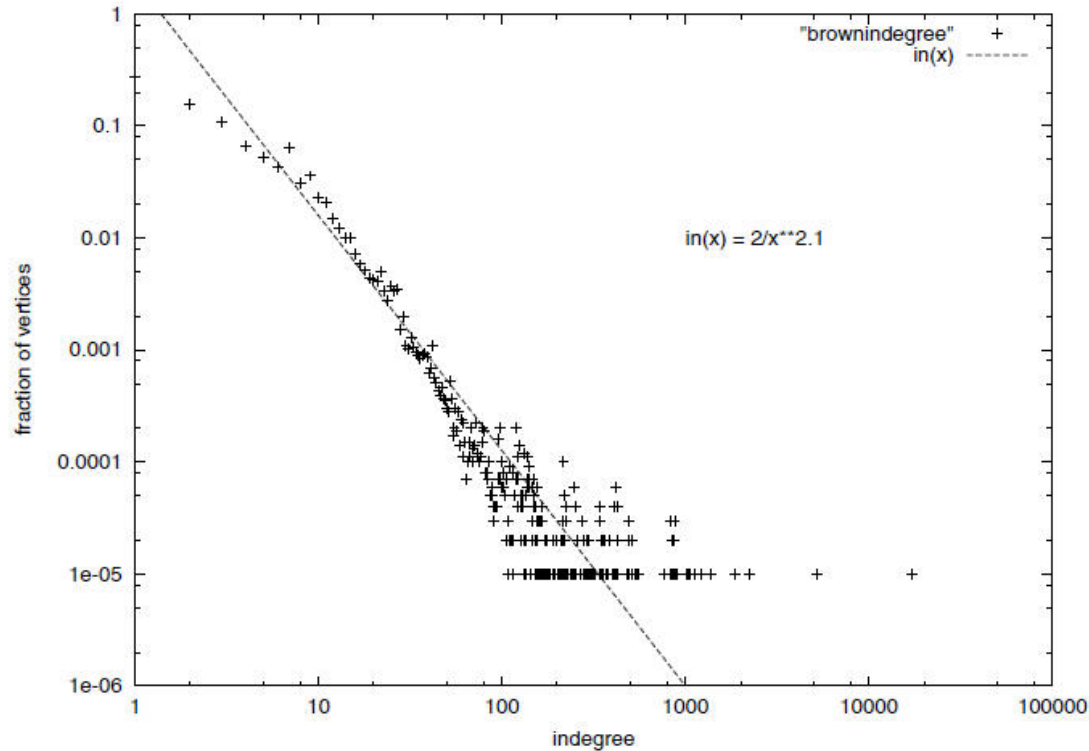
Web Graph Indegree



Web Graph Outdegree

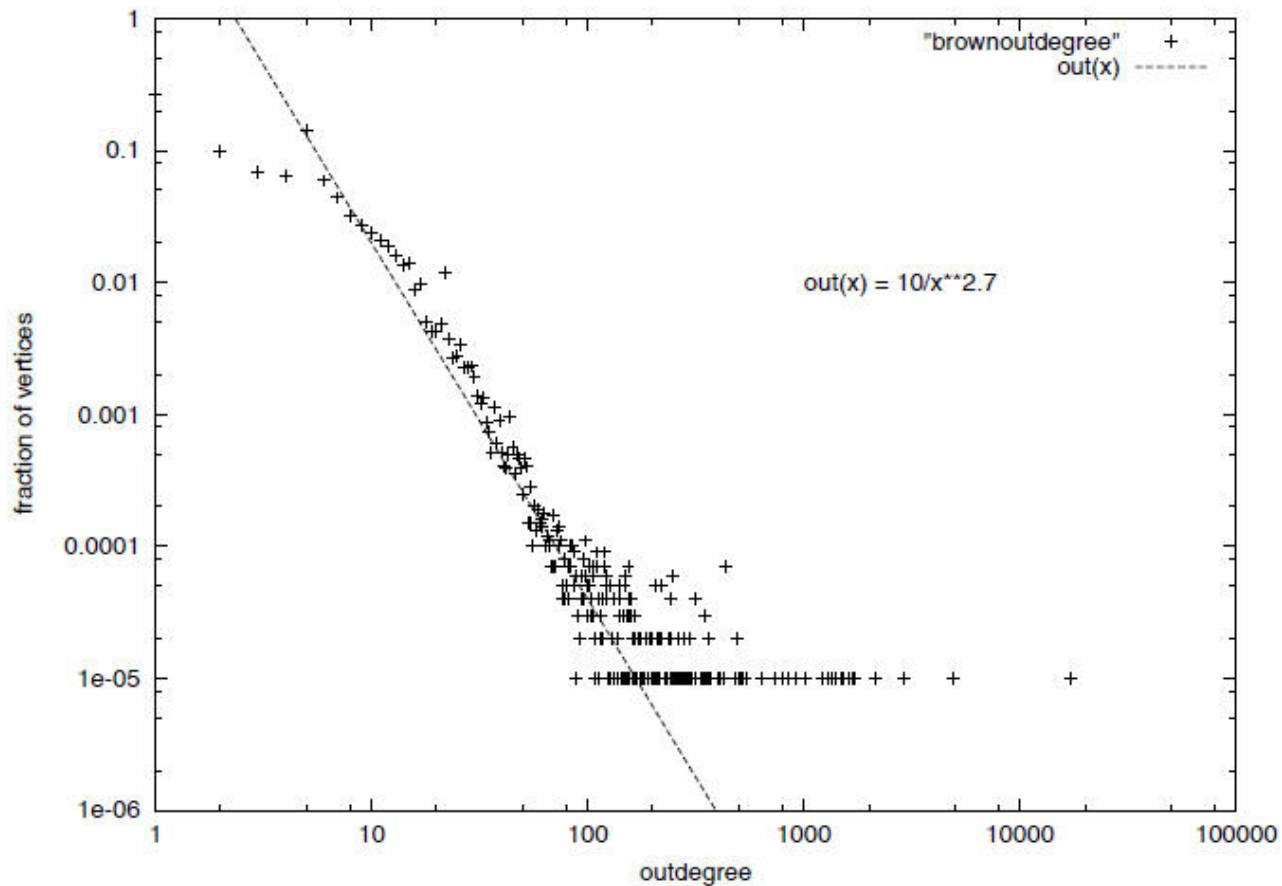


Degree Distributions



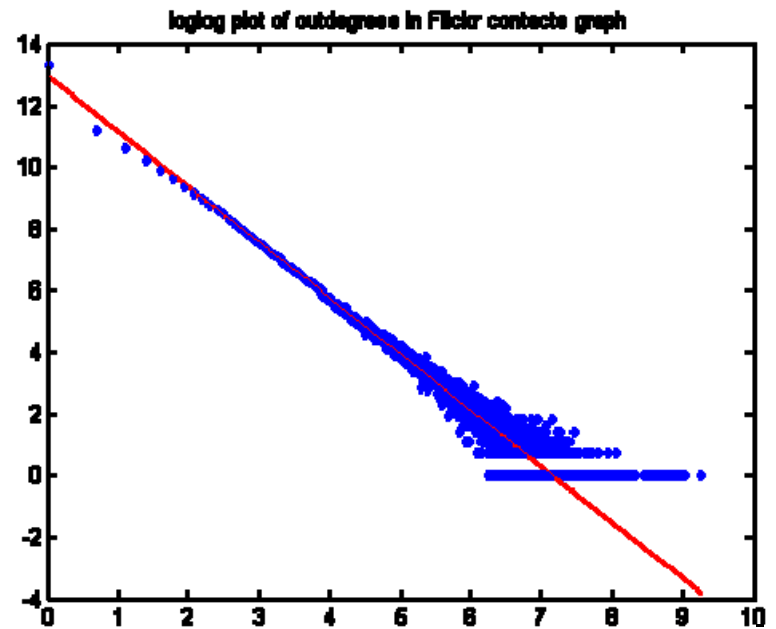
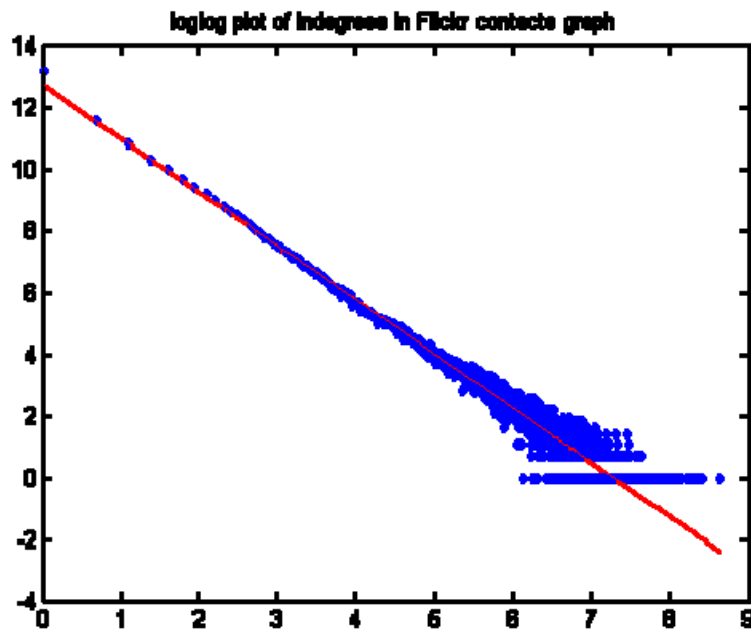
Indegree of the *.brown.edu domain

Degree Distributions

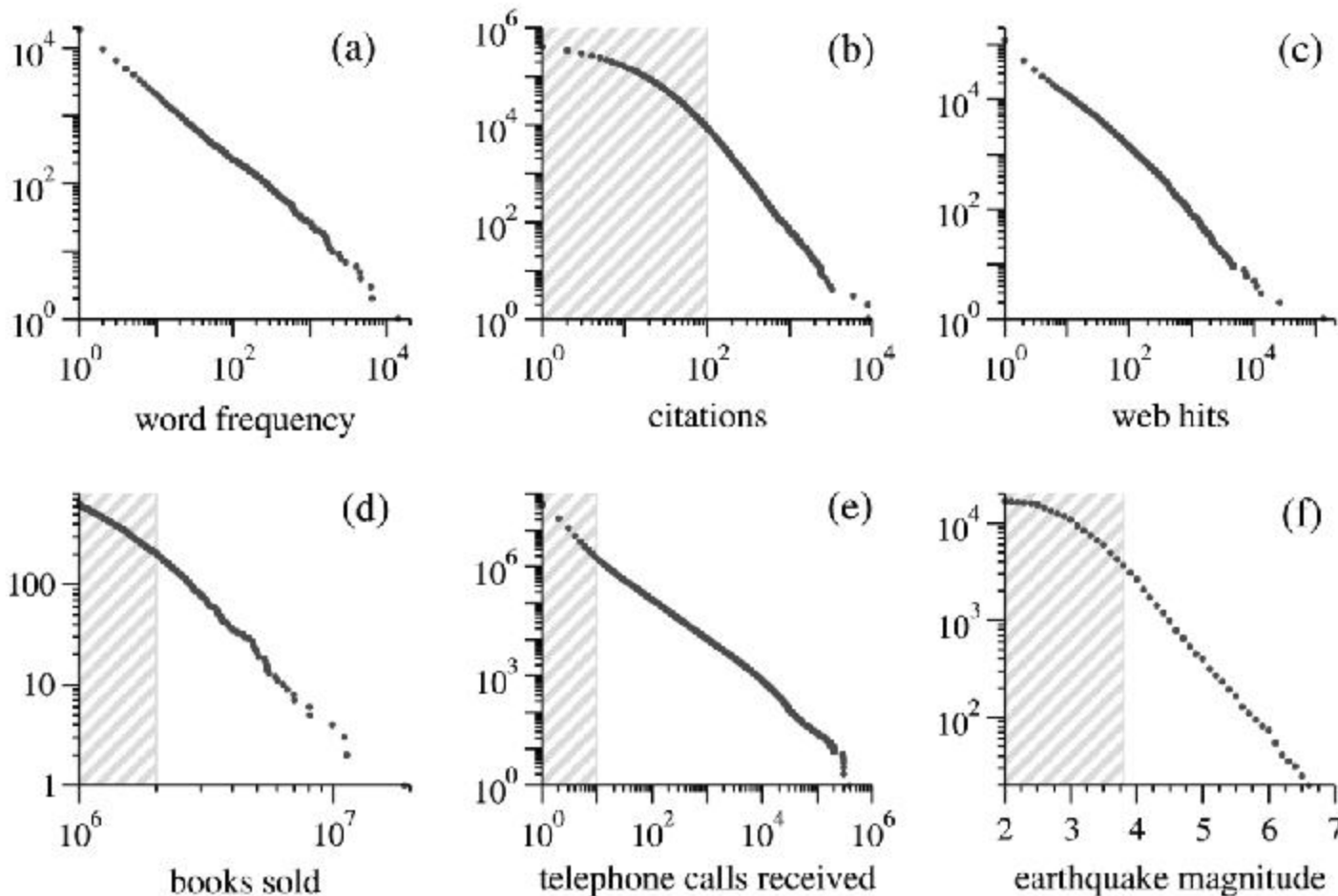


Outdegree of the *.brown.edu domain

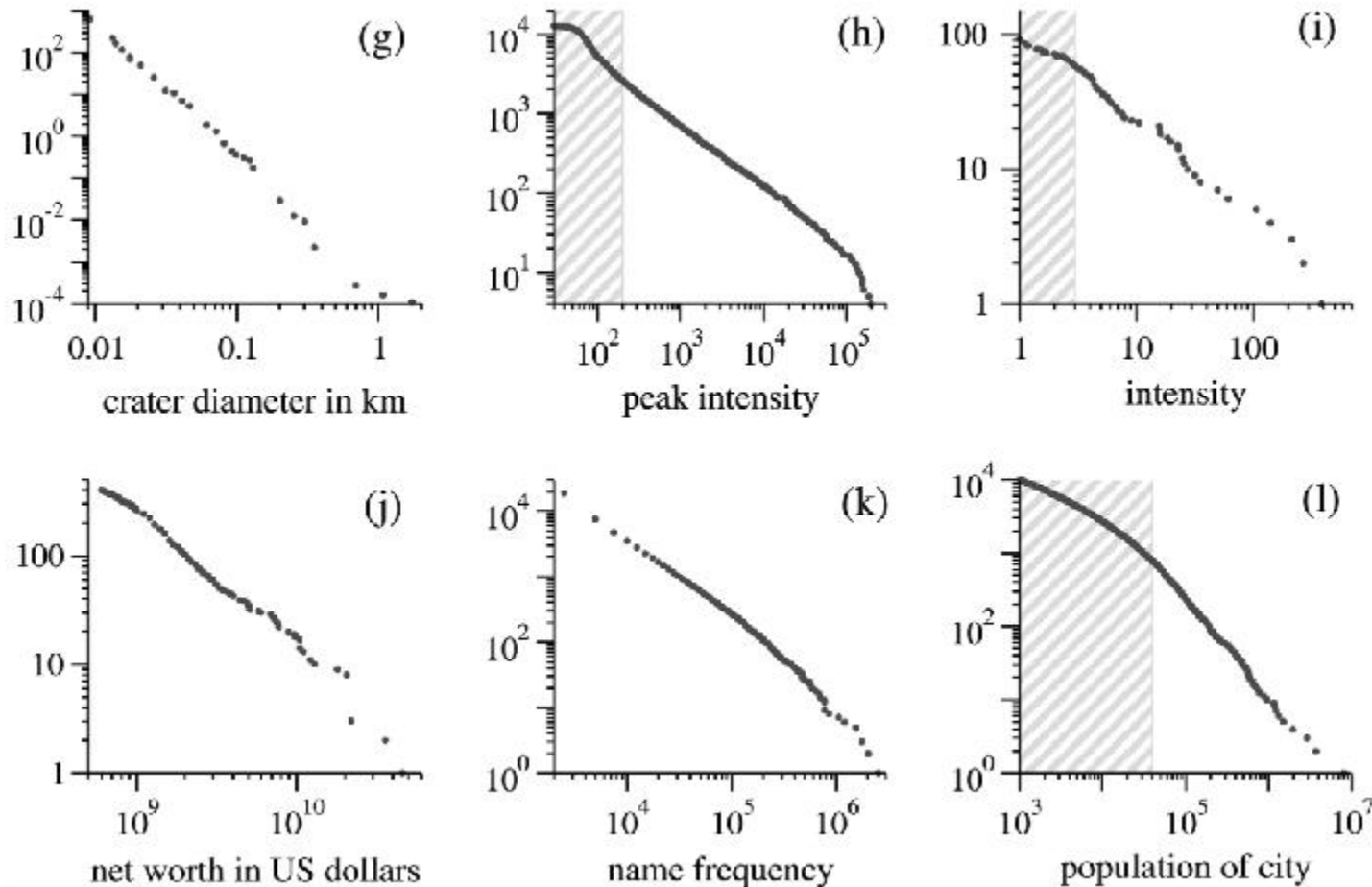
Flickr Graph, Indegrees & Outdegrees



Power Laws Everywhere



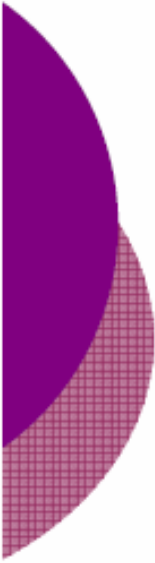
Power Laws Everywhere - 2



Power Laws Everywhere - 3

Figure 4. Cumulative distributions or 'rank/frequency plots' of twelve quantities reputed to follow power laws. The distributions were computed as described in Appendix A. Data in the shaded regions were excluded from the calculations of the exponents in table 1. Source references for the data are given in the text. (a) Numbers of occurrences of words in the novel *Moby Dick* by Hermann Melville. (b) Numbers of citations to scientific papers published in 1981, from time of publication until June 1997. (c) Numbers of hits on web sites by 60000 users of the America Online Internet service for the day of 1 December 1997. (d) Numbers of copies of bestselling books sold in the US between 1895 and 1965. (e) Number of calls received by AT&T telephone customers in the US for a single day. (f) Magnitude of earthquakes in California between January 1910 and May 1992. Magnitude is proportional to the logarithm of the maximum amplitude of the earthquake, and hence the distribution obeys a power law even though the horizontal axis is linear. (g) Diameter of craters on the moon. Vertical axis is measured per square kilometre. (h) Peak gamma-ray intensity of solar flares in counts per second, measured from Earth orbit between February 1980 and November 1989. (i) Intensity of wars from 1816 to 1980, measured as battle deaths per 10000 of the population of the participating countries. (j) Aggregate net worth in dollars of the richest individuals in the US in October 2003. (k) Frequency of occurrence of family names in the US in the year 1990. (l) Populations of US cities in the year 2000.

Small World



Small world problem

- What is the probability that two random people will know each other
 - directly
 - through a path of acquaintances
 - through a short path of acquaintances
- Social networks are
 - tightly woven
 - individuals far in physical/social space linked to each other
- How to study this?

Small World

Milgram experiment, 1967



Target person in Boston, sources in Nebraska

Letter must be passed according to: "If you do not know the target person on a personal basis, do not try to contact him directly. Instead, mail this folder to a personal acquaintance who is more likely than you to know the target person."

5% of letters made it to the destination

Two random people were connected on average by a path of six acquaintances: six degrees of separation

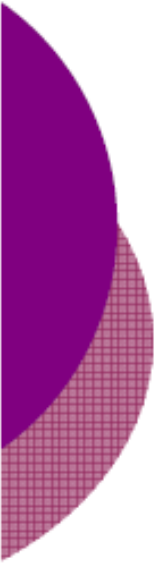
Small World



Travers-Milgram experiment, 1969

- More detailed and scientific study
- Arbitrary target
 - Lives in Sharon, MA
 - Works in Boston, MA
 - Stockbroker
- Three sets of sources
 - ~100 random people in Boston
 - ~100 random people in Nebraska
 - ~100 random blue-chip stockholders in Nebraska

Small World



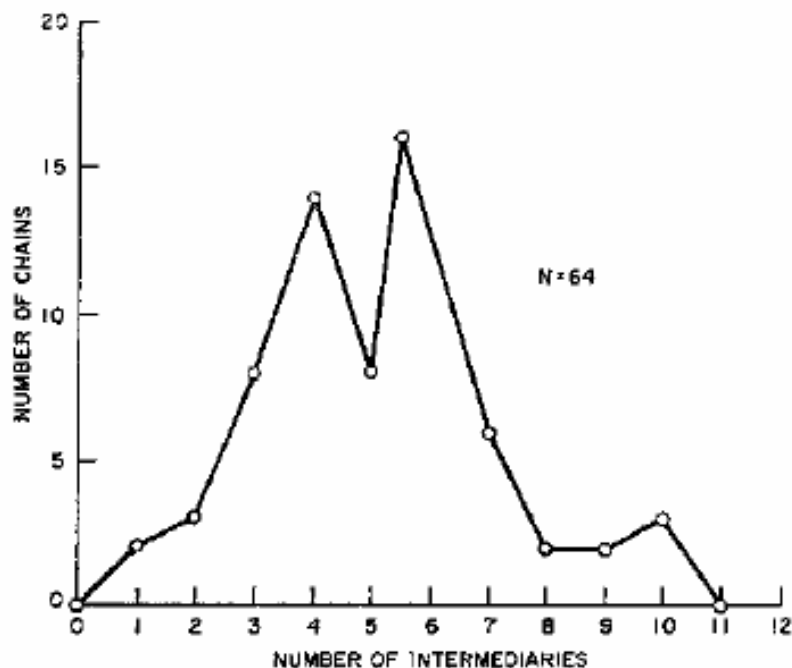
Rules for participants: Local routing

- Description of the study
- Name of the target person, address, occupation, place of employment, college/year of graduation, military service, wife's name and hometown
- “If you know the target person on a personal basis,, mail this folder directly to him (her). Do this only if you have previously met the target person and know each other on a first name basis. If you do not know the target person on a personal basis, do not try to contact him directly. Instead, mail this folder to a personal acquaintance who is more likely than you to know the target person.”

Small World

Experimental findings

- 29% of the letters reached the destination
- Average path length, 5.2
- Bimodality is not accident: target reached through
 - hometown, 6.1
 - business contacts, 4.6
- Role of geography
 - Boston, 4.4
 - Nebraska, 5.5
- Role of occupation
 - random, 5.7
 - stockholders, 5.4



Globally Sparse, Locally Dense

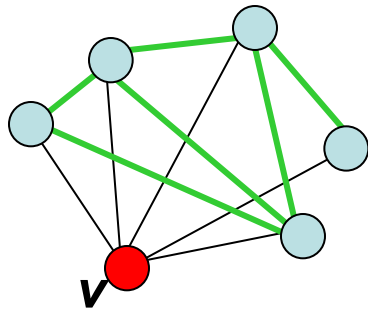
- Social networks are sparse, i.e. small number of edges (think of facebook)
- They are locally dense: many of my friends are friends with each other

Can we measure that?

Clustering Coefficient

How many of your friends are friends?

Clustering Coefficient C_v of user v measures the density of its neighborhood.



$$C_v = \frac{6}{\binom{5}{2}} = \frac{6}{10}$$

$$C_v = \frac{|\{e_{uw} : u, w \text{ neighbors of } v\}|}{\binom{d_v}{2}}$$

$C_v = 1$ if all friends also linked to each-other

$C_v = 0$ if no friends linked to each-other

For the entire graph:

$$C(G) = \frac{1}{n} \sum_{v \in V} C_v$$

Globally Sparse, Locally Dense

- For Yahoo! Messenger:

$$\frac{|E|}{\binom{n}{2}} = 4 \cdot 10^{-8}$$

$$C(G) = 0.16$$

- One explanation: **Communities**
 - We will talk about them later