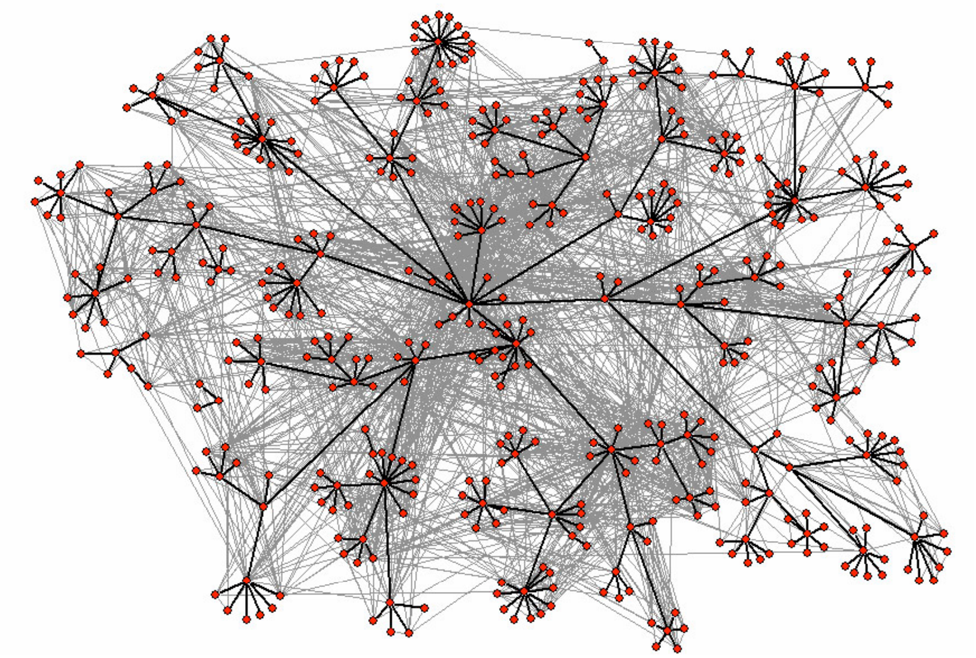
Online Social Networks and Network Economics

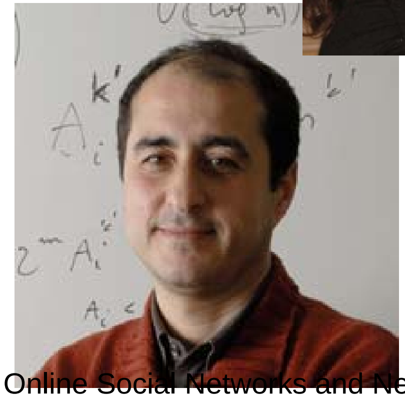


Who?

- **Dr. Aris Anagnostopoulos**

- **(Almost! Dr.) Ilaria Bordino**

- **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
- Web Mining
- Finding content in blogs
- 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.

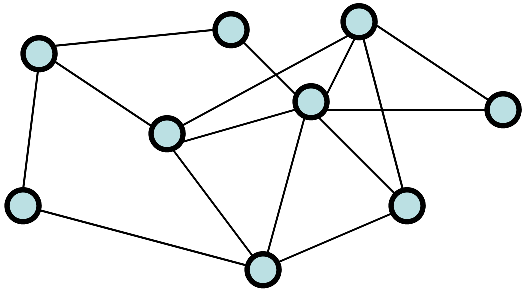
- **Class:**
  - Wednesday, 2.00 – 5.30, with a break in the middle
- **Homeworks:**
  - 2-3 homework sets, 3 weeks for each
- **Exam:** We'll see…
  - No written class exam
- **Final project – Ideas:**
  - Programming project (e.g., write a facebook application)
  - Literature review (Read some papers, summary and presentation

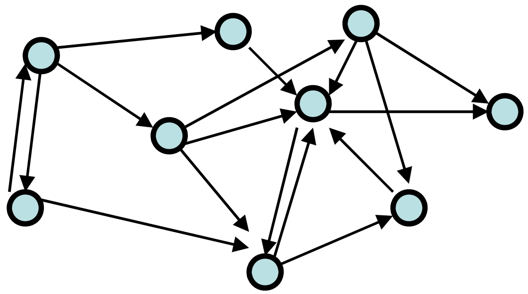
Webpage – Office hours

- **Book:**
  - No book
  - Some notes for some lectures
  - Slides
- **Webpage:** www.dis.uniroma1.it/~socialnet
  - We use it for slides, notes, announcements
- **Office hours:**
  - Monday 2pm-3pm (Aris, 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

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

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

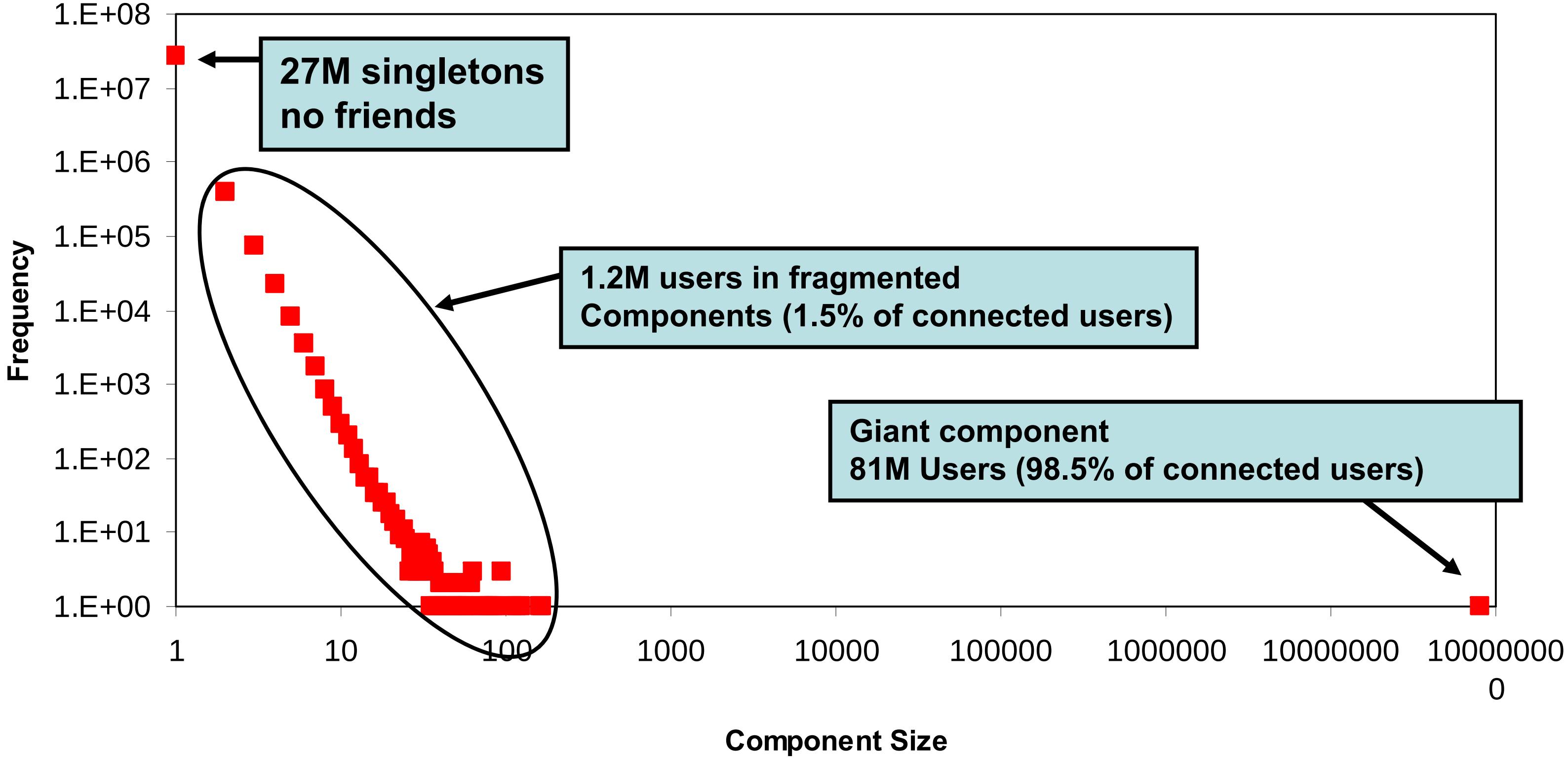
Structure of Complex Networks

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

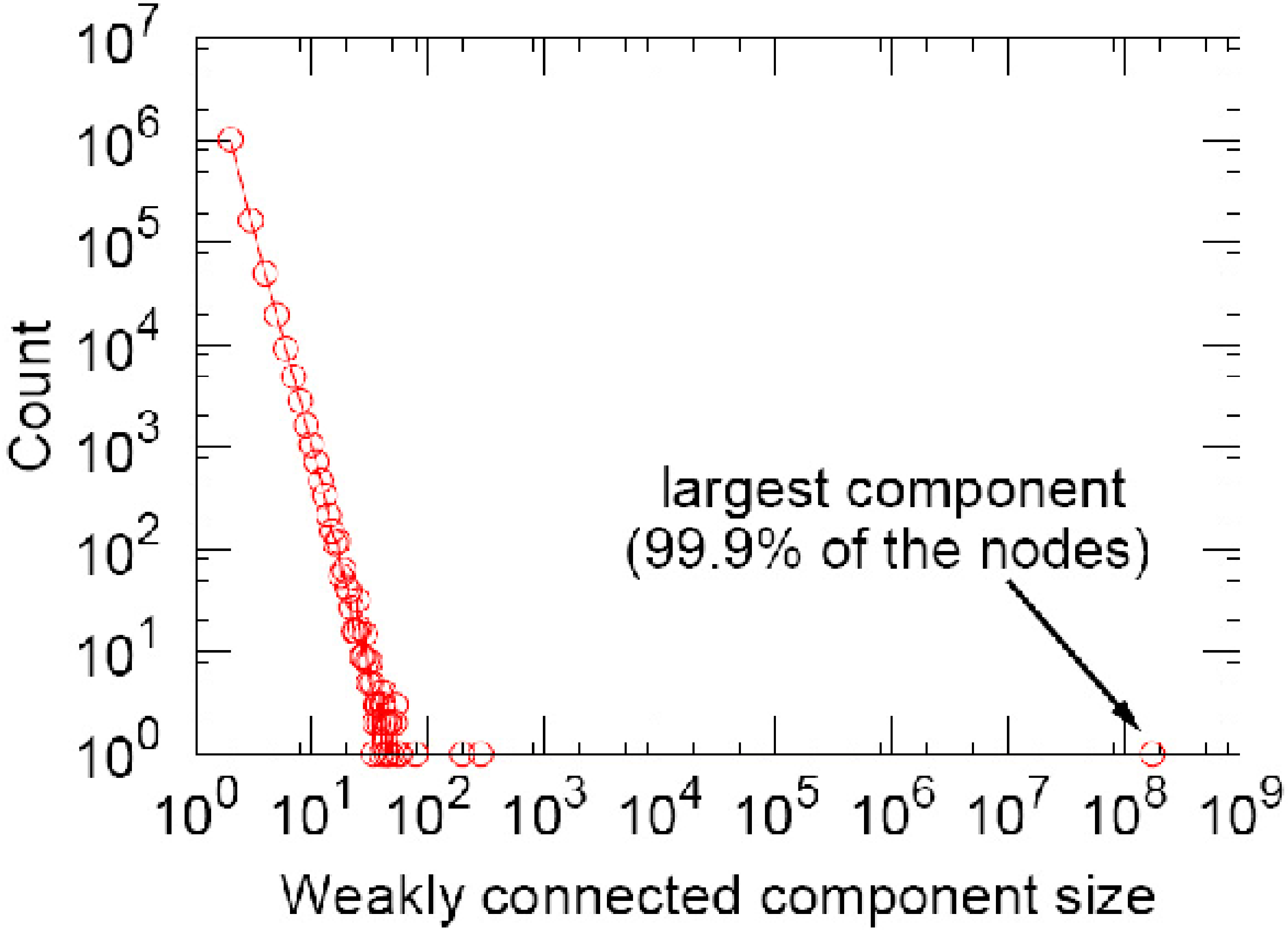
Giant Component

- 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 singletons

Yahoo! Messenger



MSN Messenger



Power-Law Degree Distributions

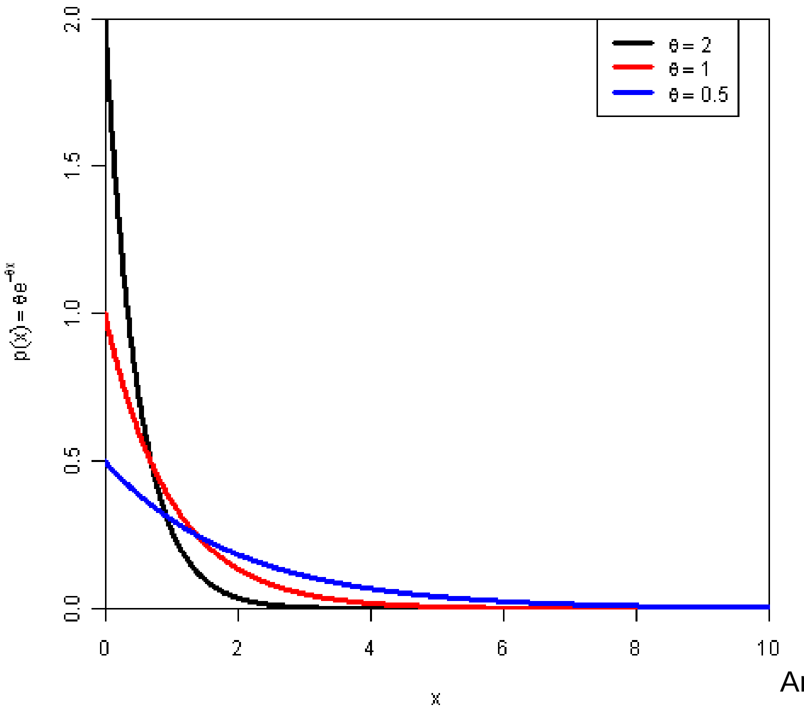
- The degree distributions of the networks follow a power-law distribution

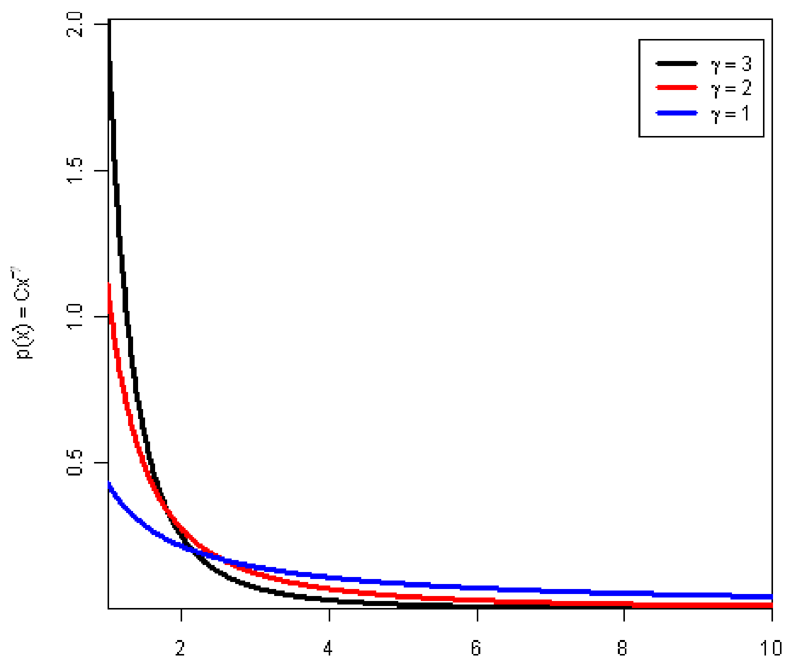
- What is power law?

Power-Law Distribution

- Exponential distribution: $p(x) = \mu e^{-\mu x}$

- Power-law distribution: $p(x) = C \cdot x^{-\gamma}$

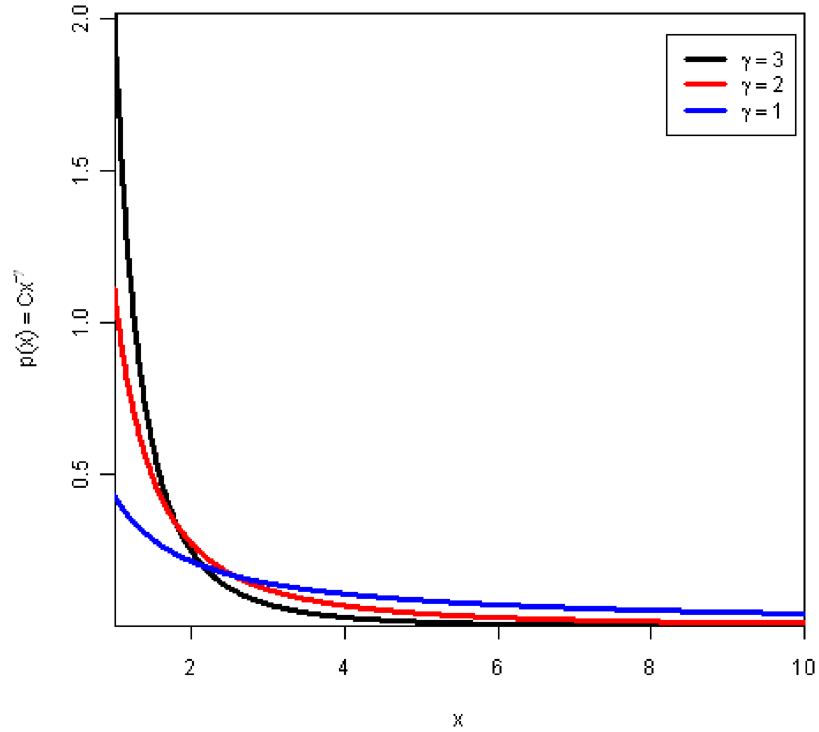


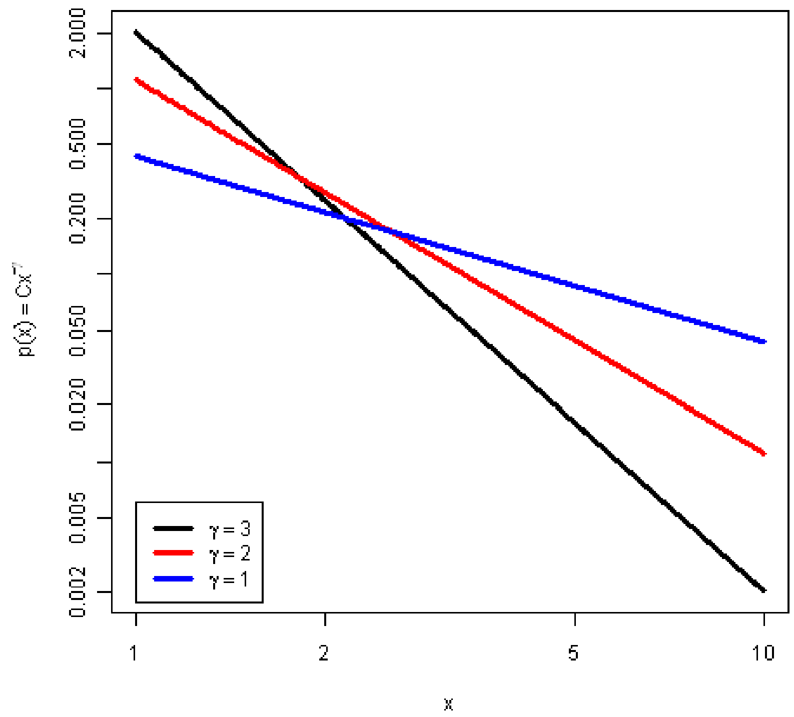


Power-Law Distribution - 2

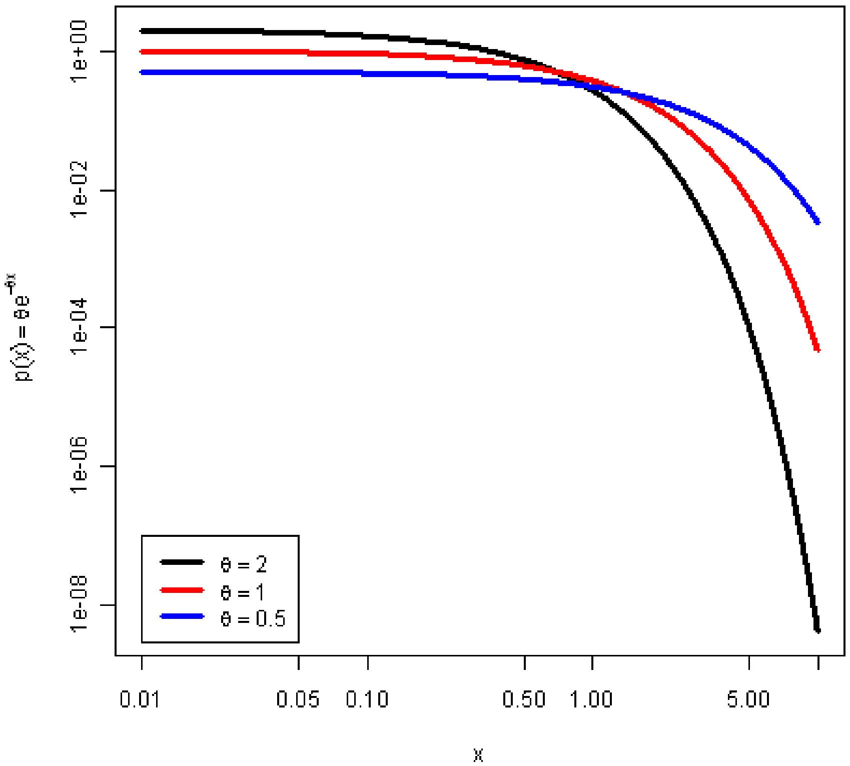
- It is a heavy-tail distribution
- **Heavy tail:** It decays slower than an exponential
- It is also called scale-free: $f(ax) = b \cdot f(x)$
- It appears in many places:
  - Degree distribution
  - Population of cities
  - Word frequencies
  - Website hits
  - Income

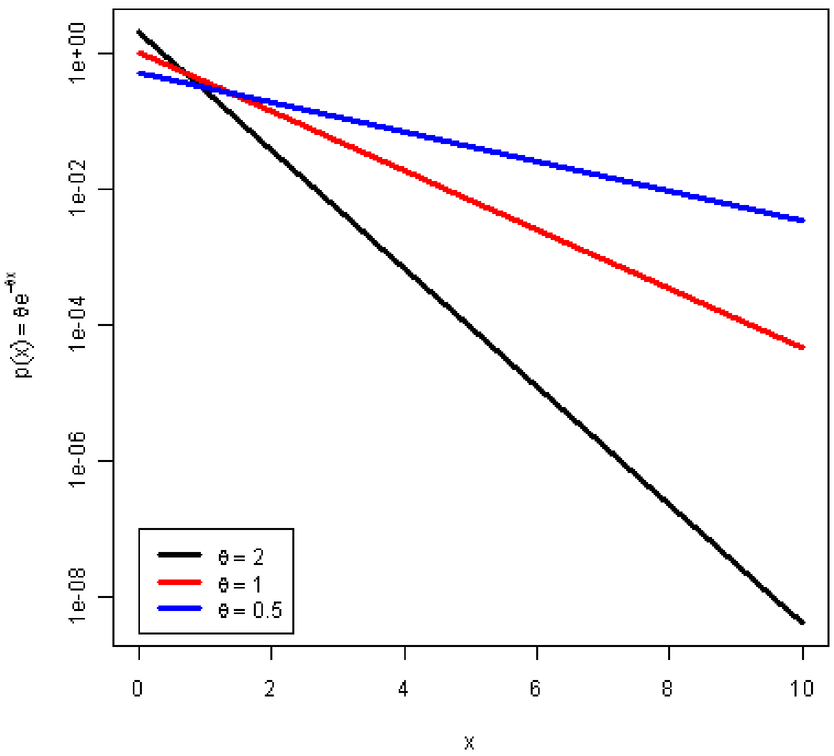
Power-Law Distribution - 3



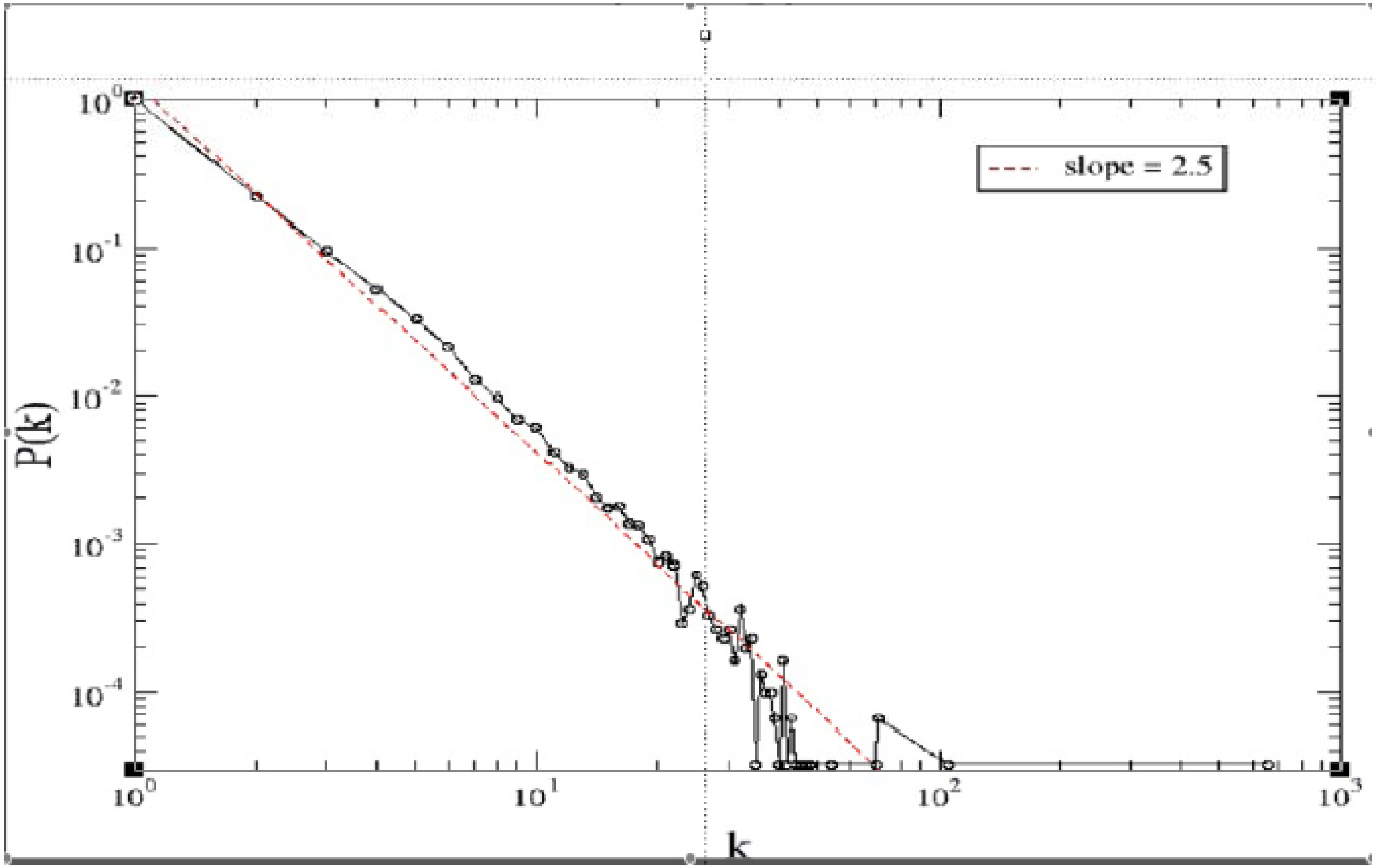


Exponential Distribution



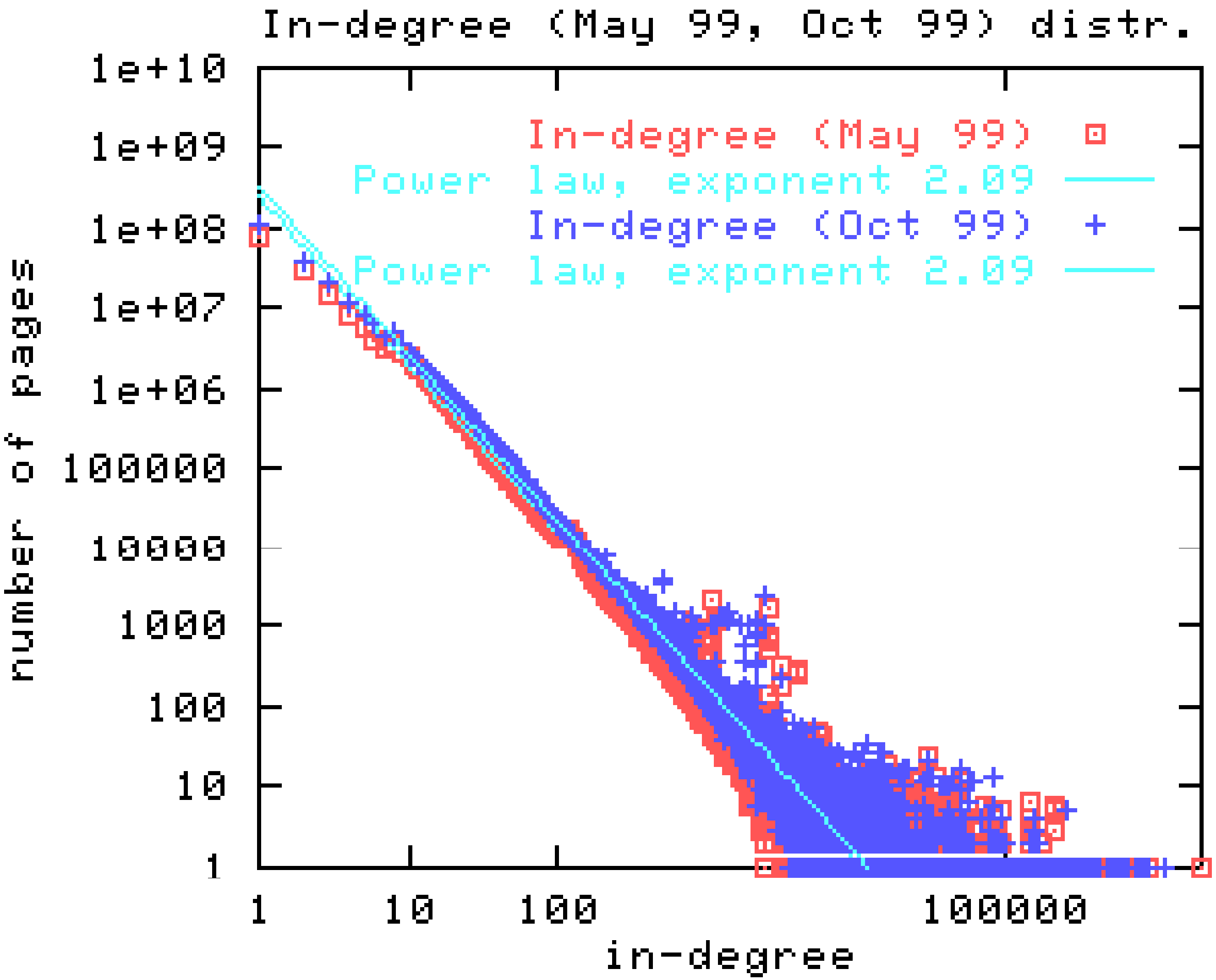


Back to Degree Distributions

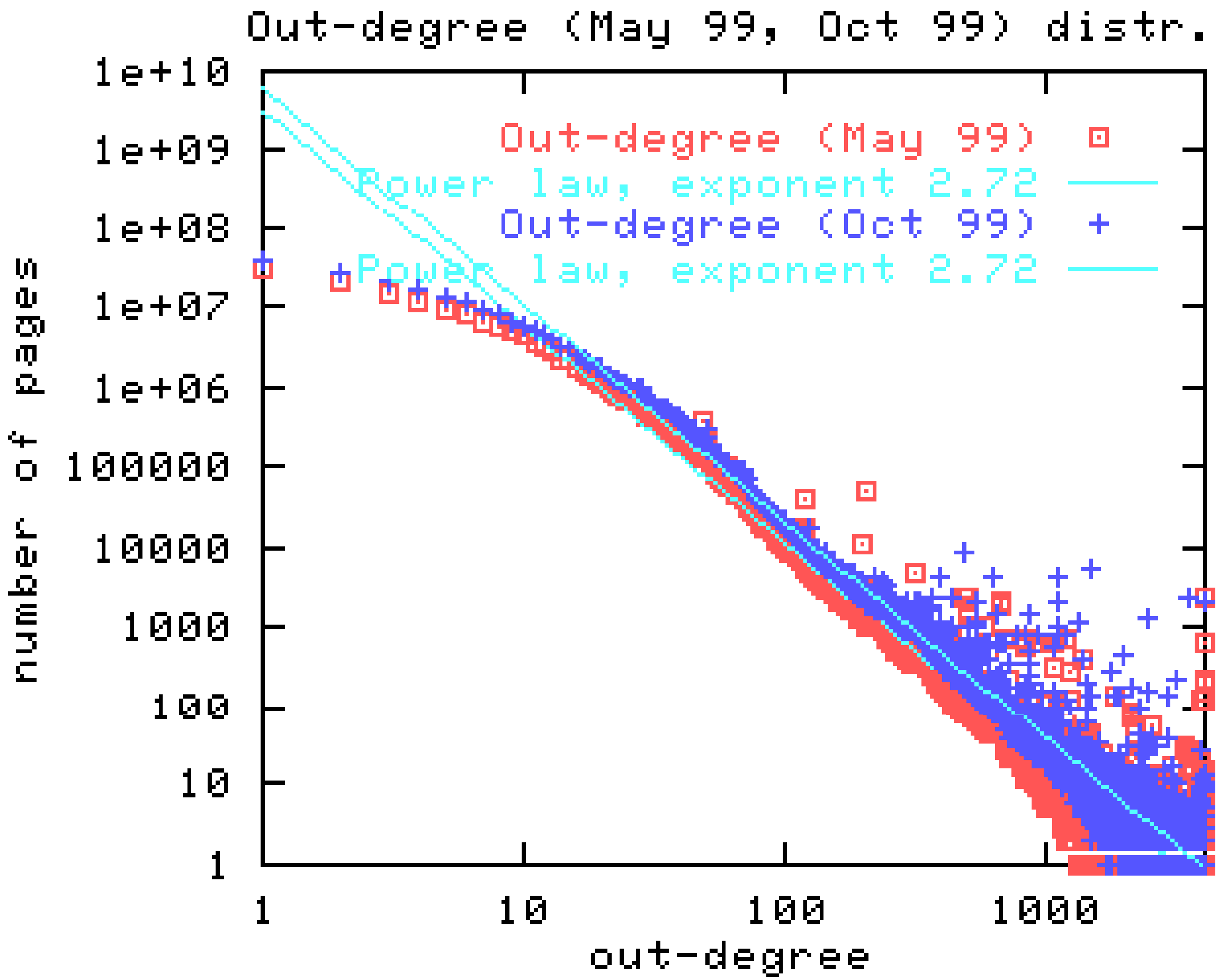


Internet Graph

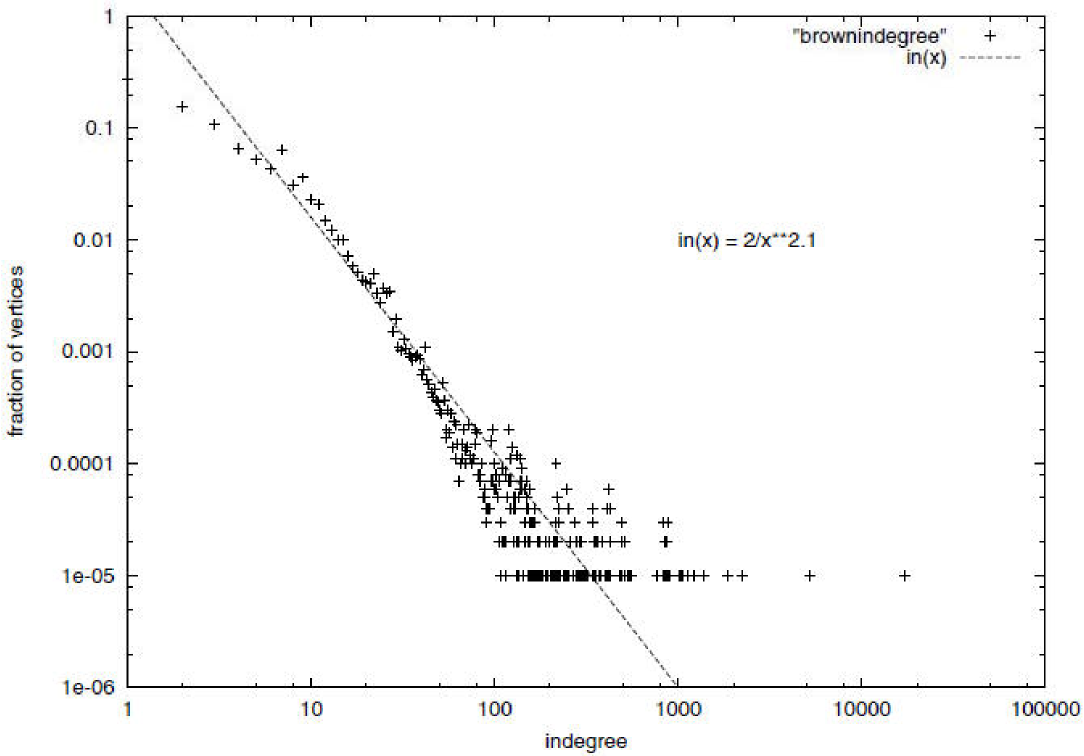
Web Graph Indegree



Web Graph Outdegree

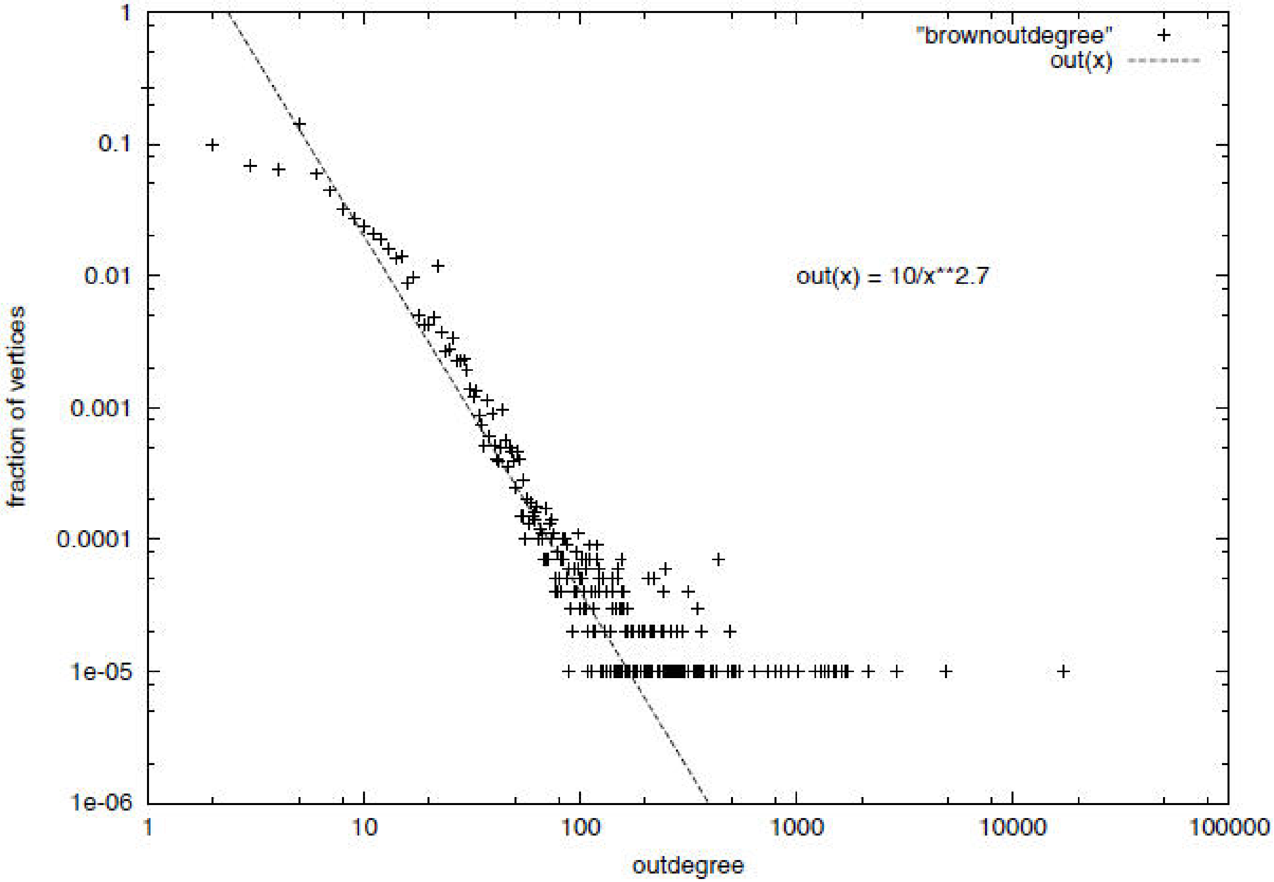


Degree Distributions



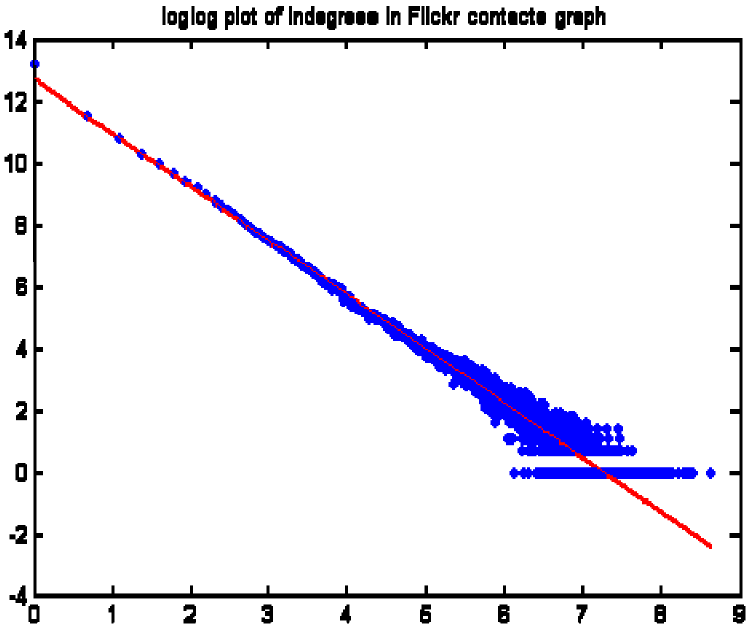
Indegree of the *.brown.edu domain

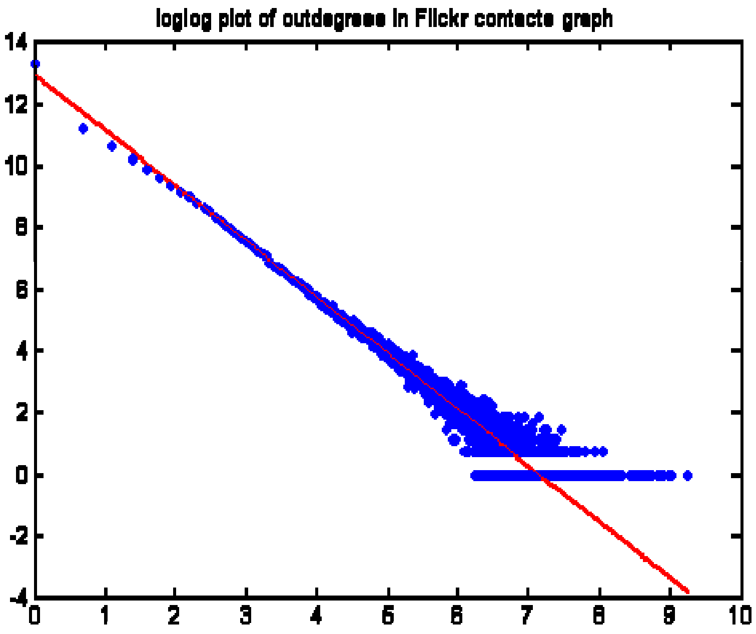
Degree Distributions

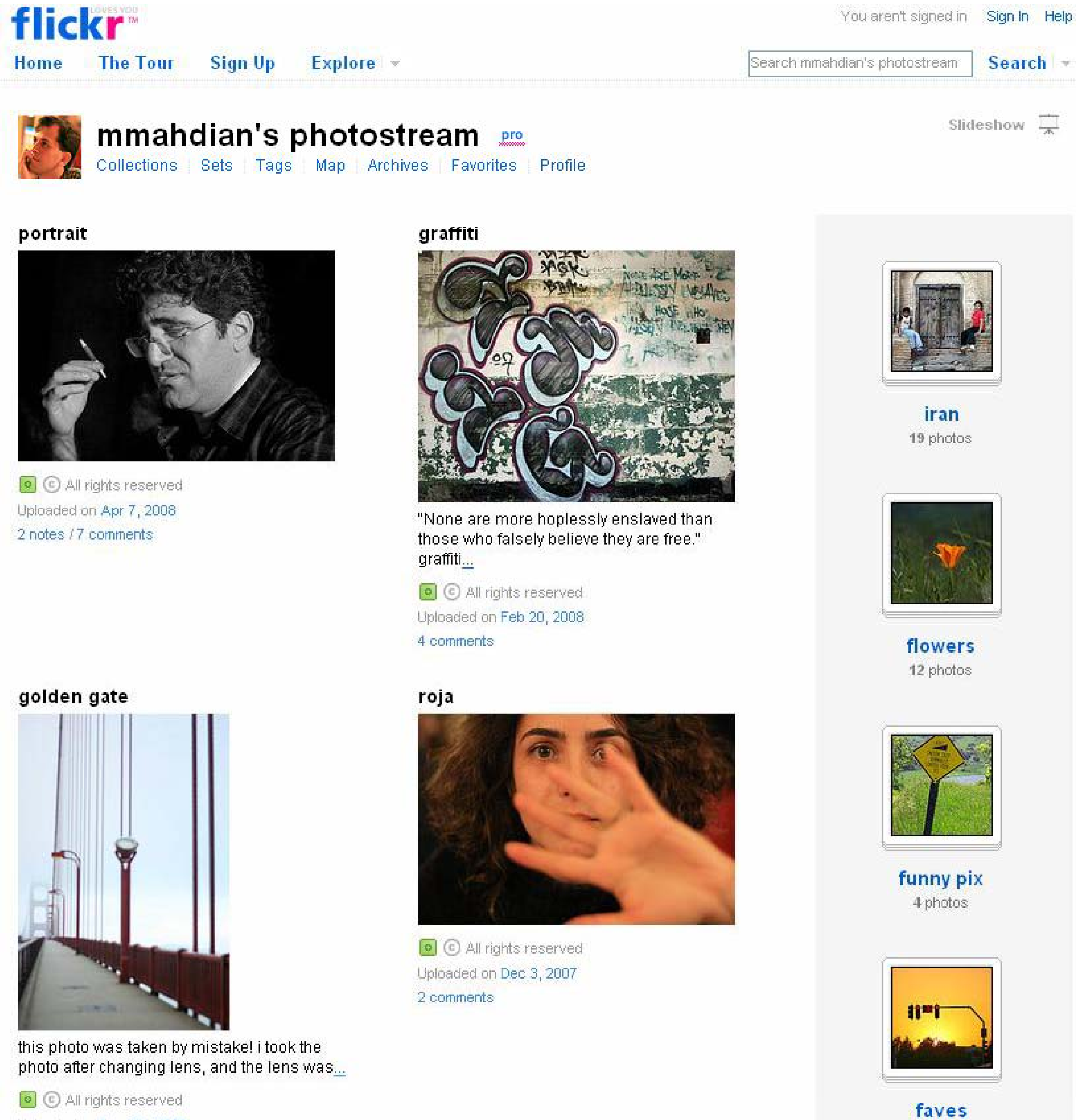


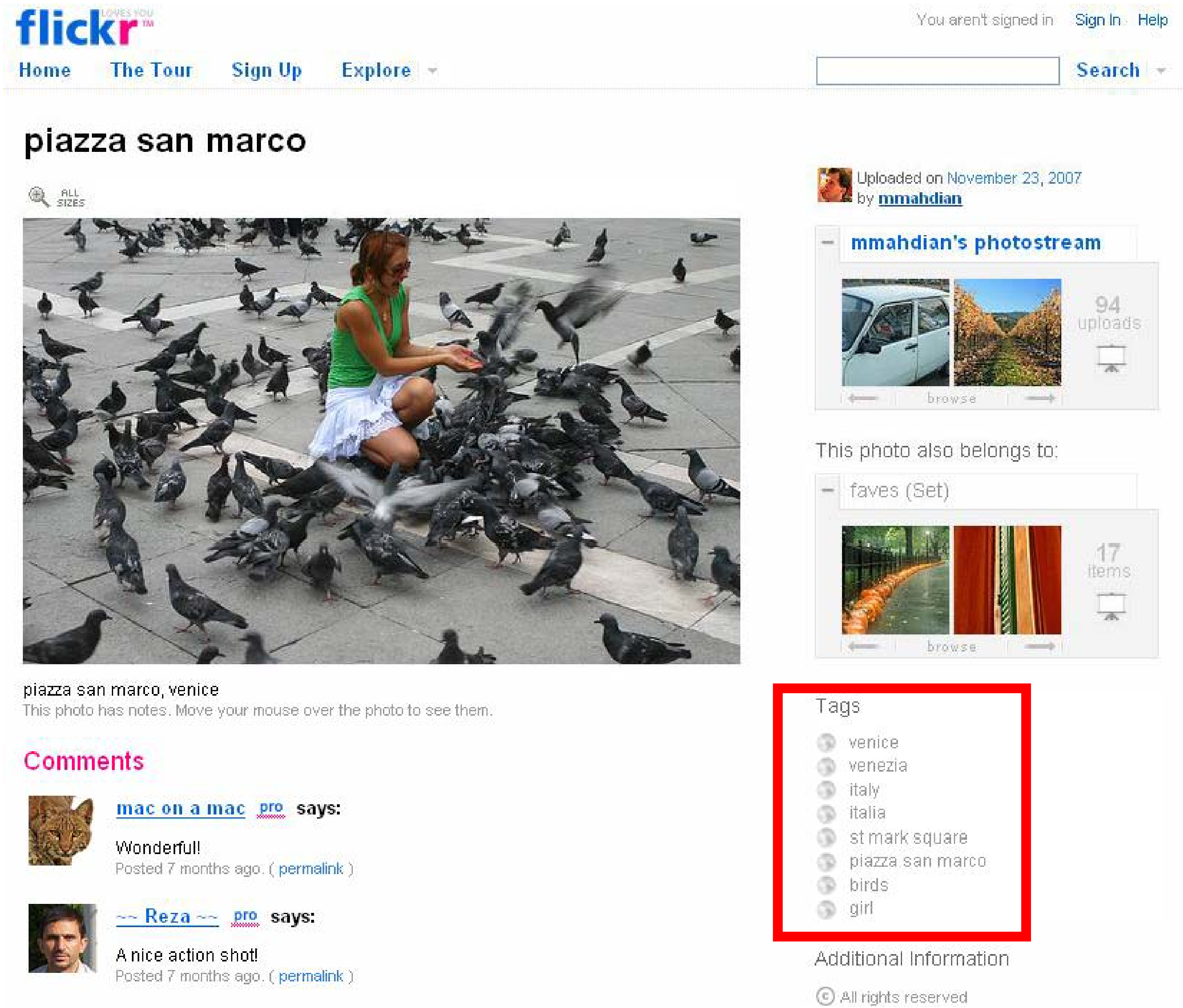
Outdegree of the *.brown.edu domain

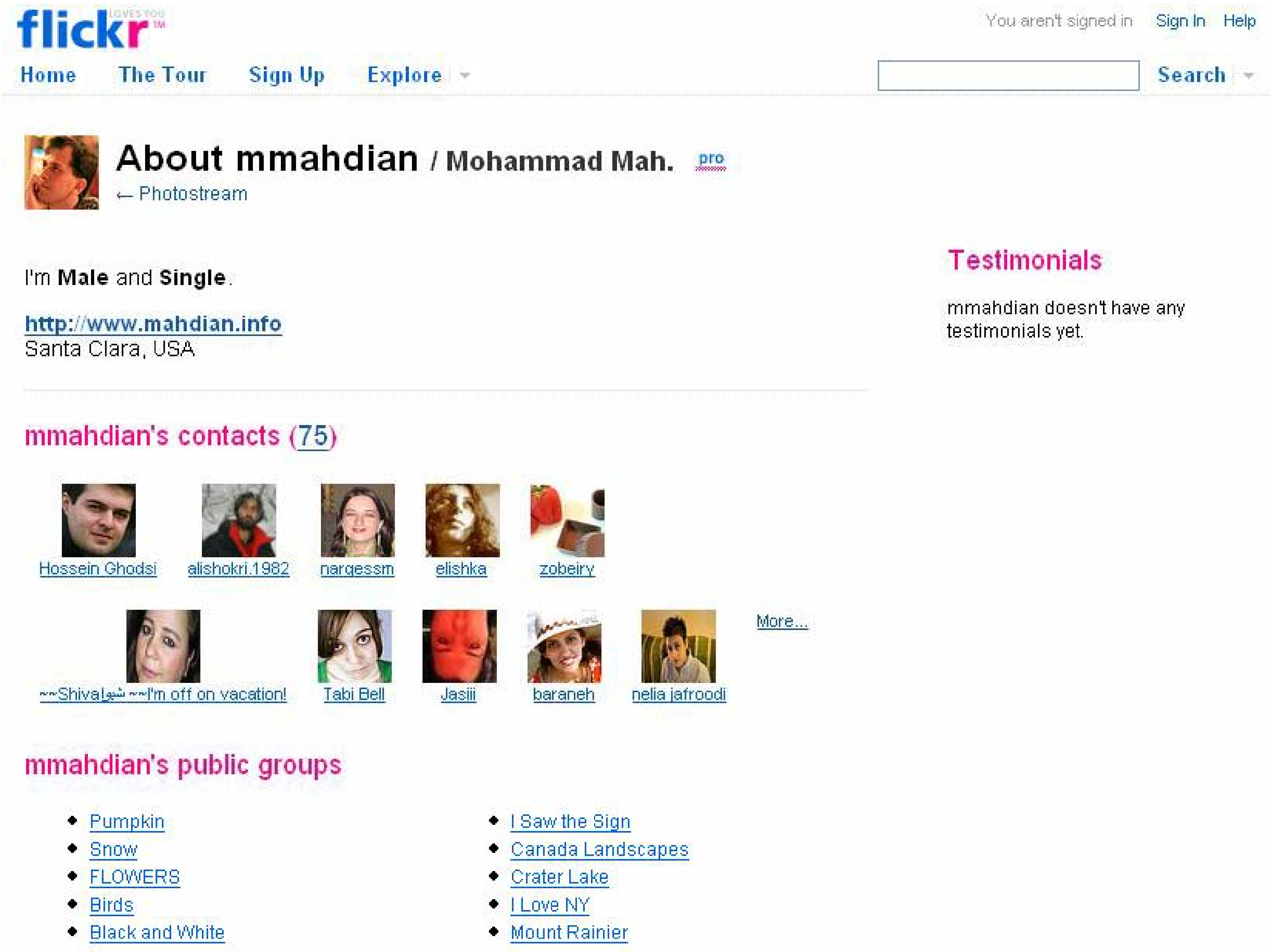
Flickr Graph, Indegrees & Outdegrees



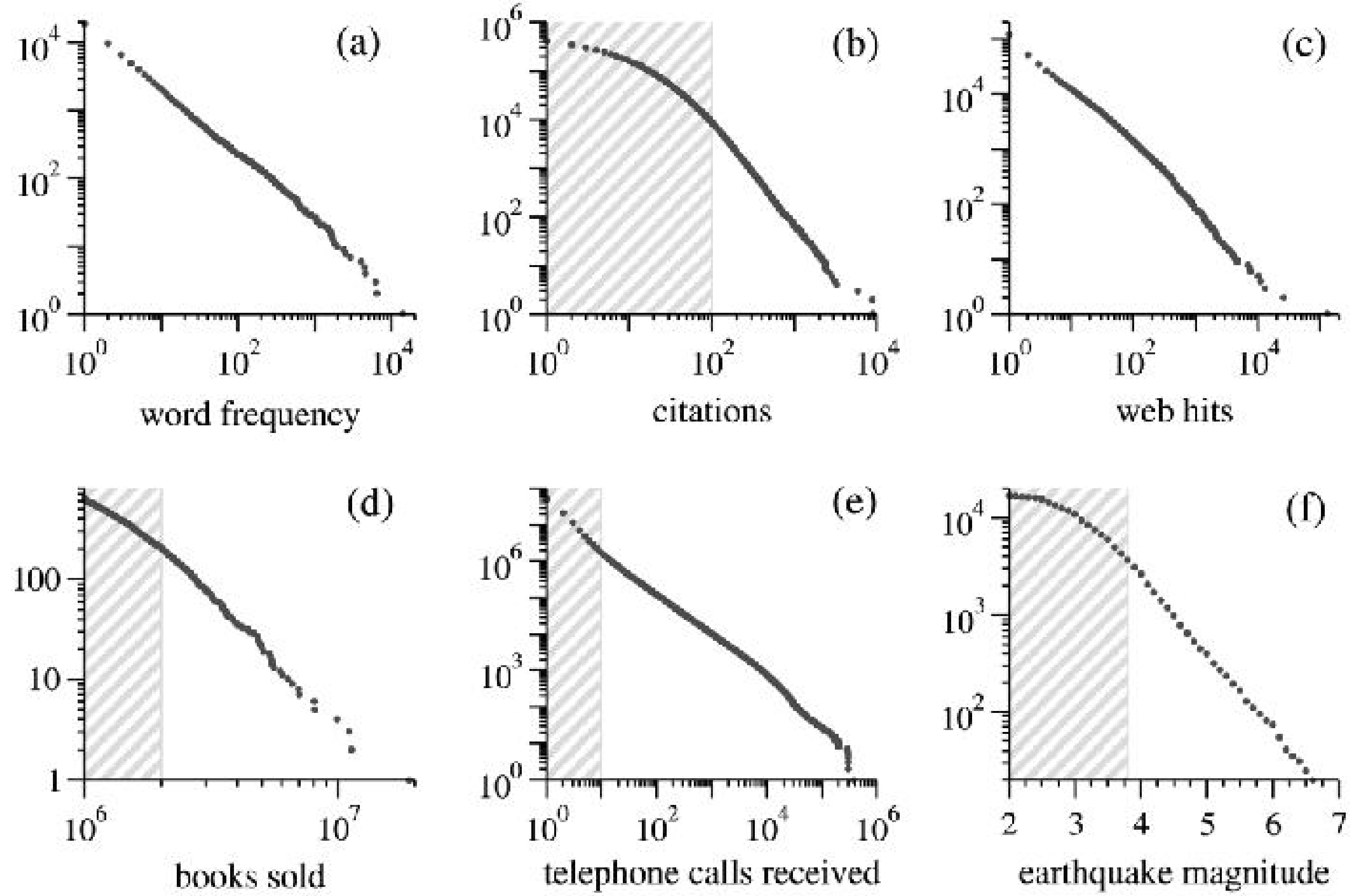



flickr
You aren't signed in
Sign In
Help
Home
The Tour
Sign Up
Explore
Search mmahdian's photostream
Search
mmahdian's photostream
pro
Slideshow
Collections
Sets
Tags
Map
Archives
Favorites
Profile
portrait
All rights reserved
Uploaded on Apr 7, 2008
2 notes / 7 comments
graffiti
"None are more hoplessly enslaved than those who falsely believe they are free." graffiti...
All rights reserved
Uploaded on Feb 20, 2008
4 comments
golden gate
this photo was taken by mistake! i took the photo after changing lens, and the lens was...
All rights reserved
roja
All rights reserved
Uploaded on Dec 3, 2007
2 comments
iran
19 photos
flowers
12 photos
funny pix
4 photos
faves

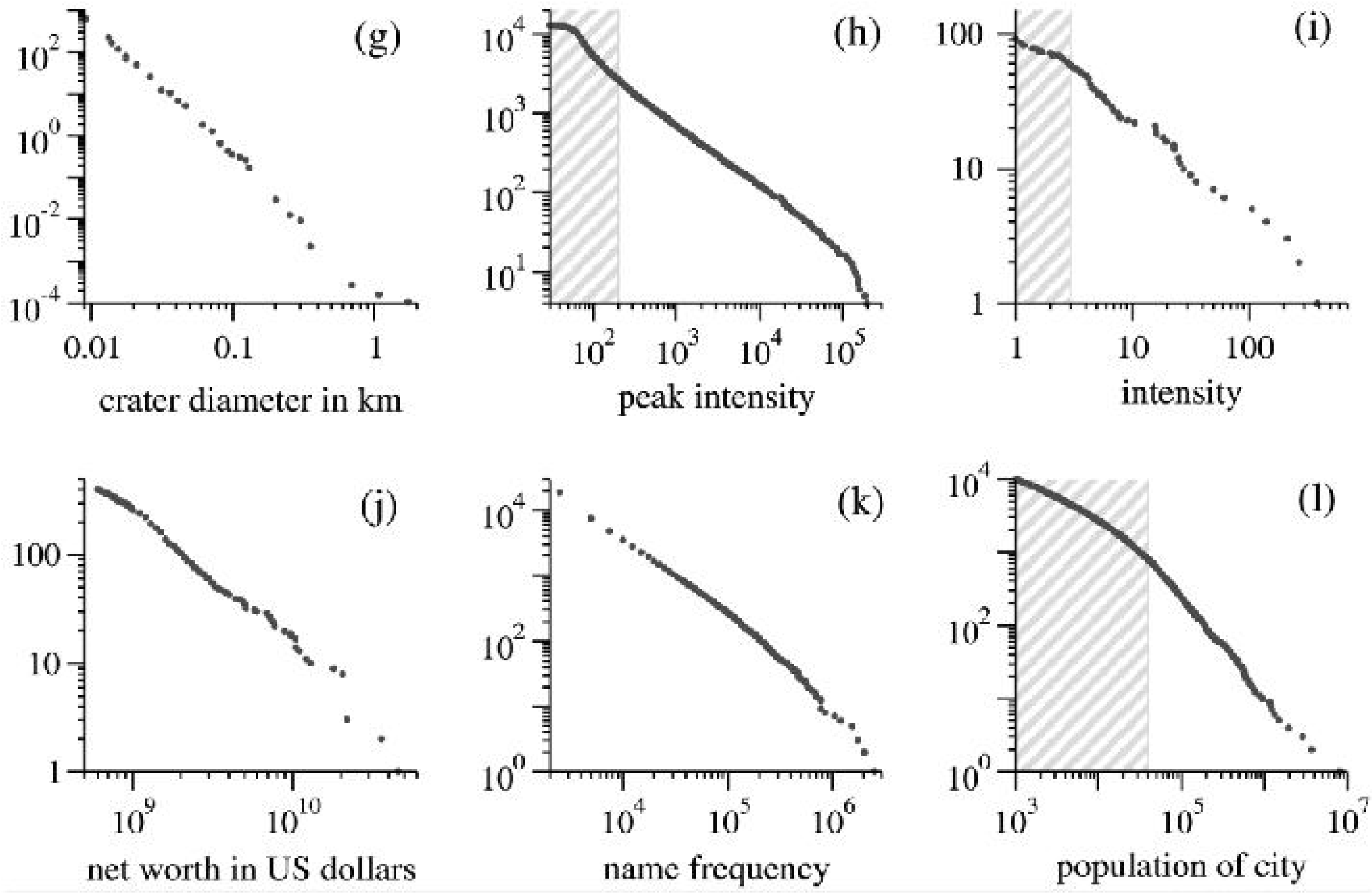

flickr
You aren't signed in Sign In Help
Home The Tour Sign Up Explore
Search
piazza san marco
ALL SIZES
Uploaded on November 23, 2007
by mmahdian
mmahdian's photostream
94 uploads
browse
This photo also belongs to:
faves (Set)
17 items
browse
piazza san marco, venice
This photo has notes. Move your mouse over the photo to see them.
Comments
mac on a mac pro says:
Wonderful!
Posted 7 months ago. (permalink)
~~ Reza ~~ pro says:
A nice action shot!
Posted 7 months ago. (permalink)
Tags
venice
venezia
italy
italia
st mark square
piazza san marco
birds
girl
Additional Information
All rights reserved


flickr LOVES YOU
You aren't signed in
Sign In
Help
Home
The Tour
Sign Up
Explore
Search
About mmahdian / Mohammad Mah.
pro
← Photostream
I'm Male and Single.
http://www.mahdian.info
Santa Clara, USA
Testimonials
mmahdian doesn't have any testimonials yet.
mmahdian's contacts (75)
Hossein Ghodsi
alishokri.1982
nargessm
elishka
zobeiry
~~Shivaشیوا~~I'm off on vacation!
Tabi Bell
Jasiii
baraneh
nelia jafroodi
More...
mmahdian's public groups
Pumpkin
Snow
FLOWERS
Birds
Black and White
I Saw the Sign
Canada Landscapes
Crater Lake
I Love NY
Mount Rainier

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.

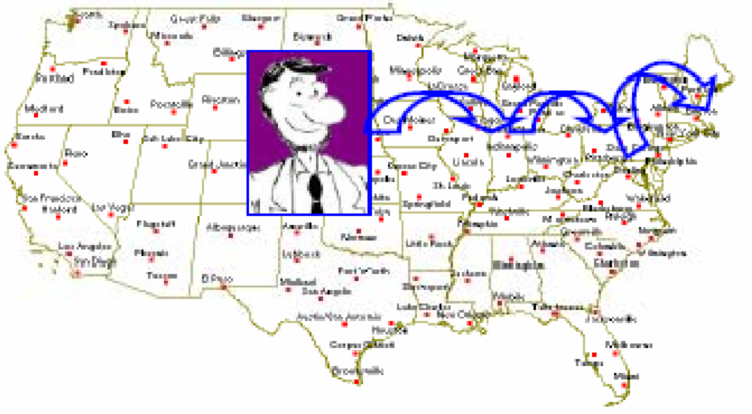
Small World

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

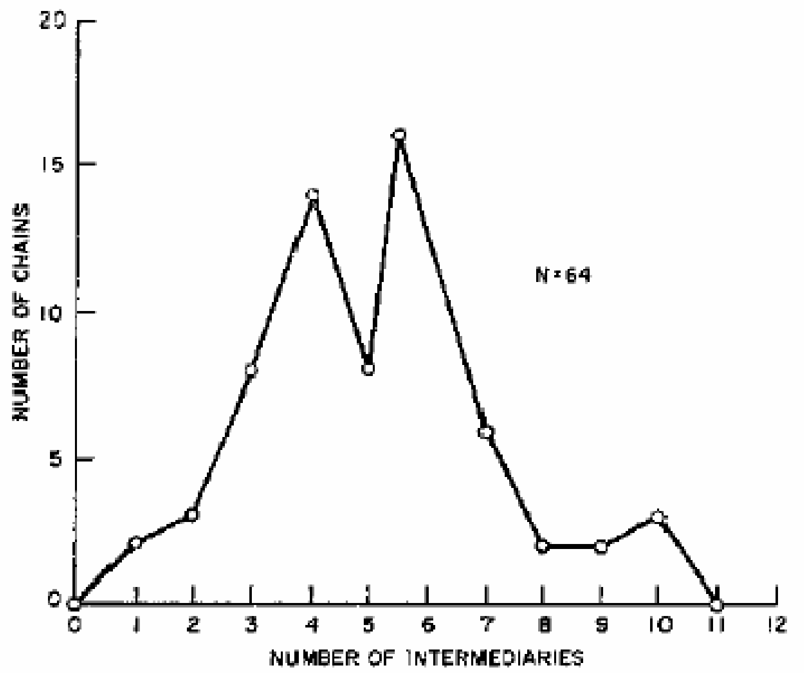
Rules for participants: Local routing

- 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."

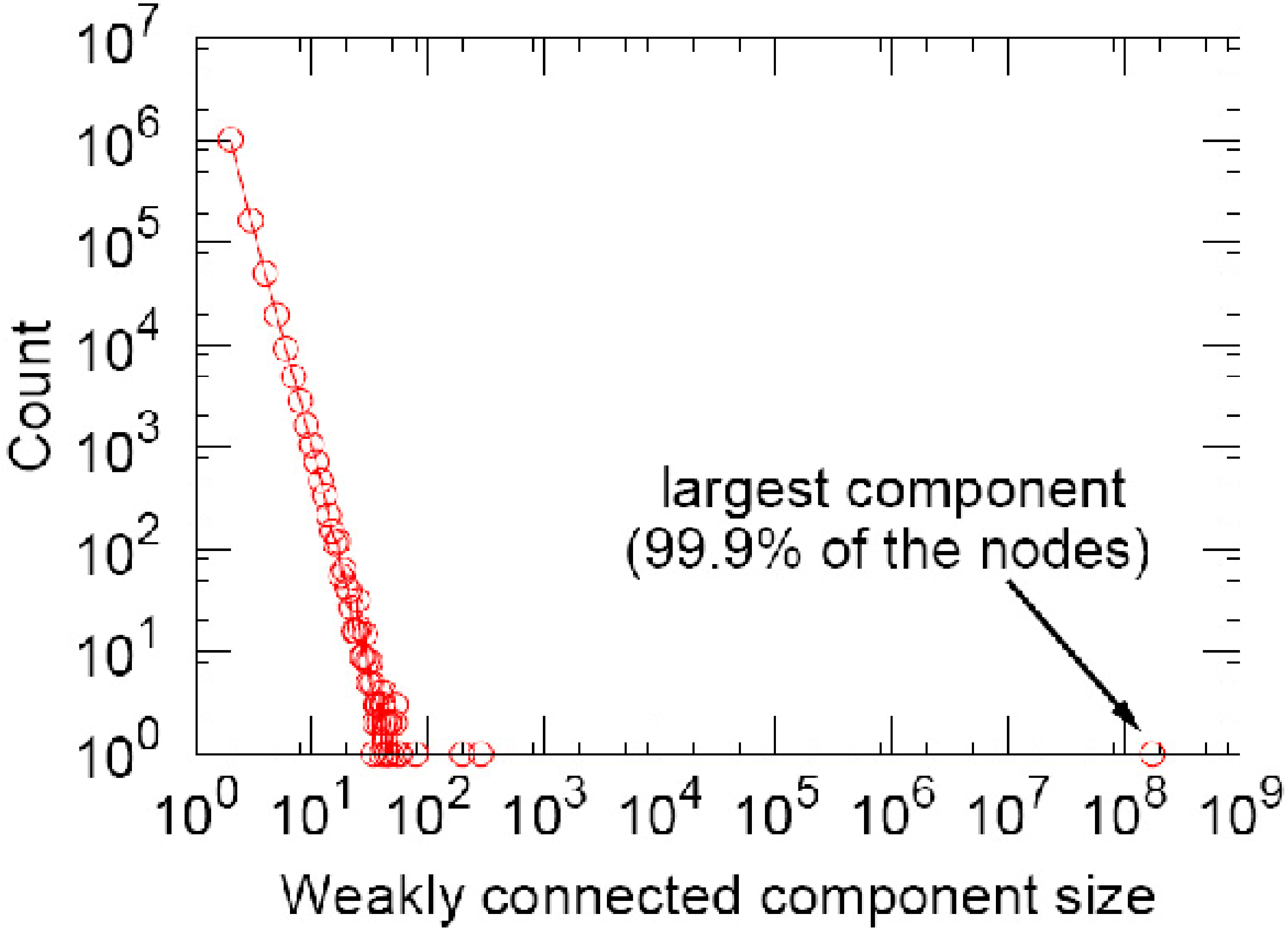
Small World

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



MSN Messenger



Globally Sparse, Locally Dense

- Social networks are sparse, i.e. small number of edges

  (think of facebook)

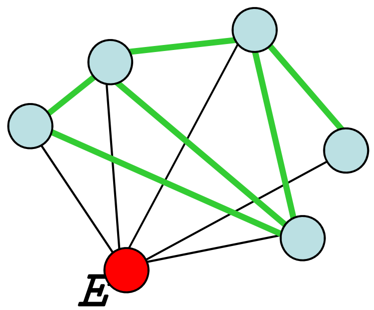
- They are locally dense: many of my friends are friends with each other

Can we measure that?

Clustering Coefficient

How many of your friends are friends?

Clustering Coefficient $\}_{E}$ of user E measures the density of its neighborhood.



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

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

$\}_{E}$ = 1 if all friends also linked to each-other

$\}_{E}$ = 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! Messanger:

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

$$C(G) = 0.16$$

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

Models for Social Networks

We saw some properties of Social Networks

Can we develop models that generate these properties?

Why develop models?

- Understand the process of network formation
- Use them for prediction
- Use them for simulations

Erdős-Rényi Random Graph Model

- It’s been out since the 50’s
- The simplest possible
- Not a good model for social networks
- Will give us intuition

Erdős-Rényi Random Graph Model

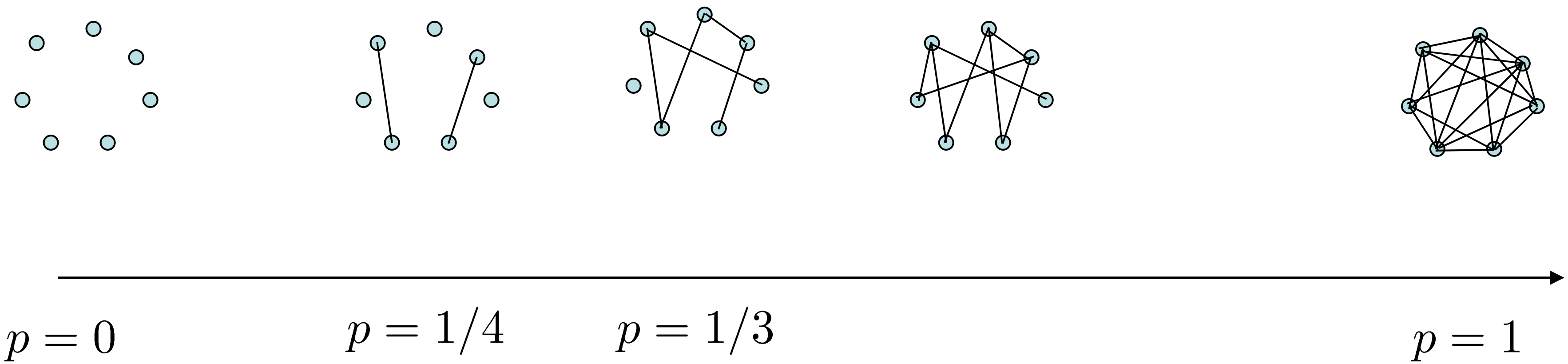
- It is also called $G_{n,p}$
- A graph has n vertices
- So it has at most $\binom{n}{2} = \frac{n(n-1)}{2}$ edges
- Each edge exists w.p. p, independently of the other
- p is in general a function of n

We want to study properties of the graph for large n

Erdős-Rényi Random Graph Model

- What is the expected number of edges?
- What is the expected degree?
- How many possible graphs can be constructed?
- Assume that a graph has m edges. What is the probability it will be created by the $G_{n,p}$ model?

Graph Properties



Graph Properties

- We say that a property holds with high probability (whp) if the probability goes to 1 as $n \rightarrow$ –

- We can also say that it holds for almost all graphs.

Example: A graph is connected whp. if

$$p = \frac{c \ln n}{n}, \quad c > 1$$

We study properties as p goes from 0 to 1.

Monotone Graph Properties

A property is called monotone if adding edges does not destroy it.

Example: Connectivity, small diameter, Hamiltonian cycle

Theorem: If a monotone property holds for $G_{n,p}$ whp then it also holds for $G_{n,p'}$ whp. if $p' > p$.

Phase Transition

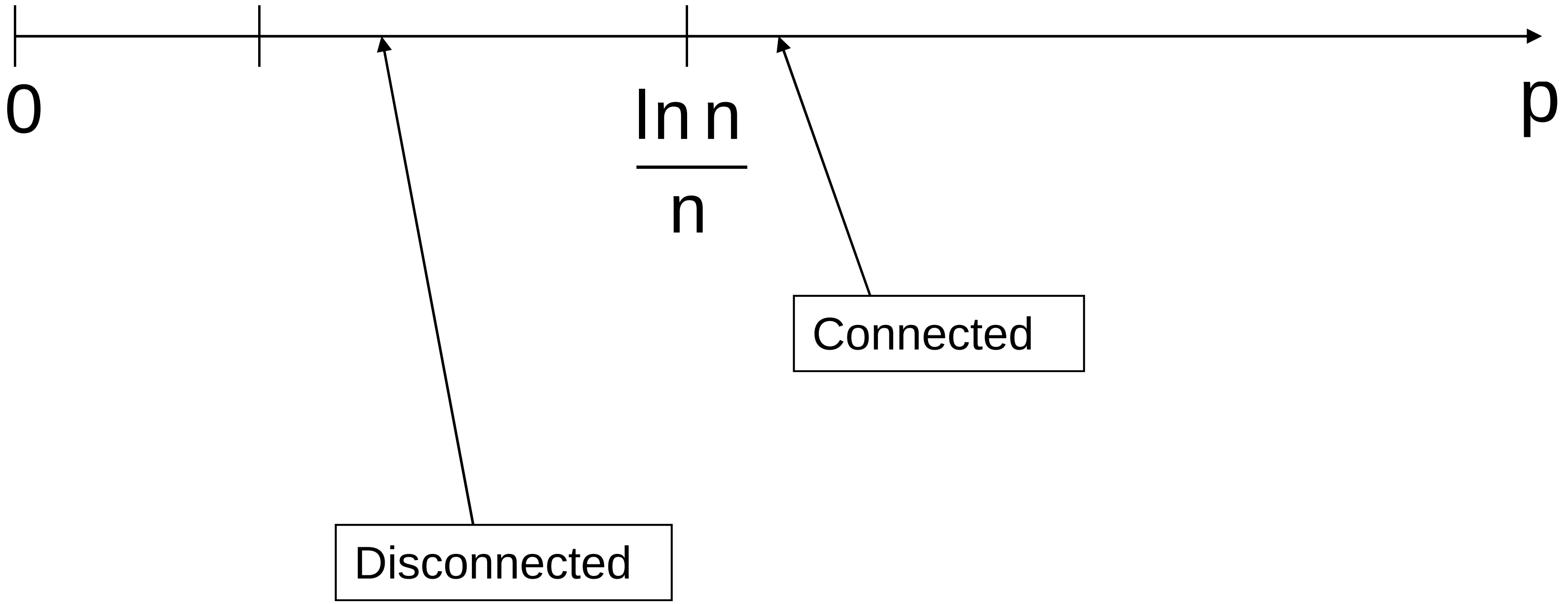
For many properties we have a **threshold**: p^*

- If $p < p^*$ the property holds for almost no graphs.
- If $p > p^*$ the property suddenly holds for almost all graphs.

We say that we have a **phase transition** at p^*

Connectivity

We study the connectivity of the graph



Degree Distribution

Degree distribution is **Binomial**:

$$\mathbf{Pr}(d_v = k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$