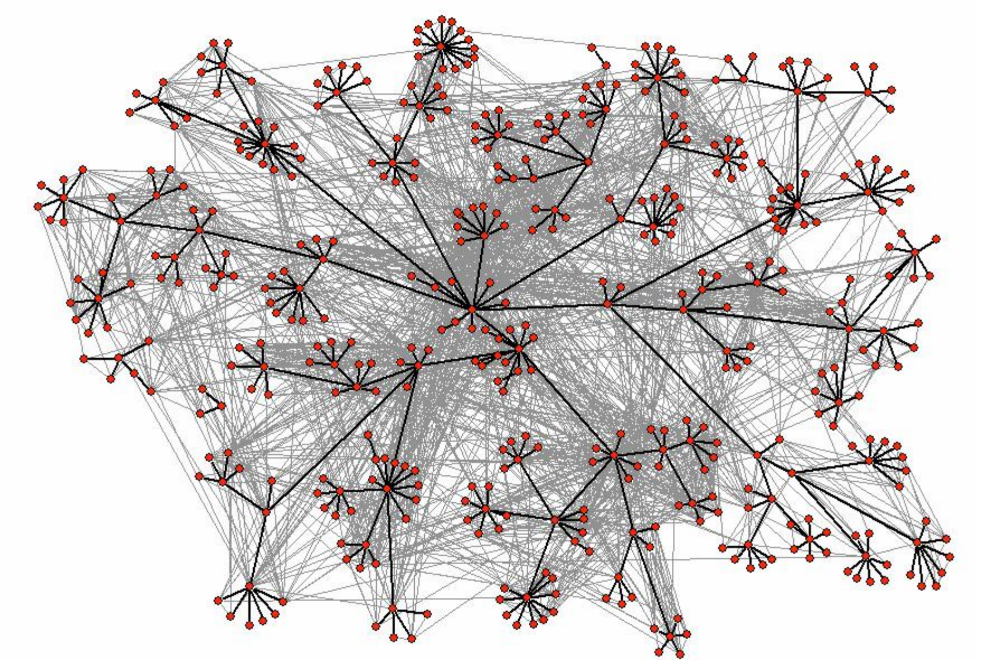
Social, Information, and Routing Networks: Models, Algorithms, and Strategic Behavior



Who?

- **Prof. Aris Anagnostopoulos**

- **Prof. Luciana S. Buriol**

- **Prof. Guido Schäfer**

What will We Cover?

Topics:

- Network properties
- Models for network formation
- Influence processes in networks
- External memory algorithms
- Streaming algorithms
- Web graph compression
- Selfish routing games
- Inefficiency of equilibria and Braess paradox
- Stackelberg routing
- Network tolls

What will We Cover?

Topics:

- Network properties
- Models for network formation
- Influence processes in networks
- External memory algorithms
- Streaming algorithms
- Web graph compression
- Selfish routing games
- Inefficiency of equilibria and Braess paradox
- Stackelberg routing
- Network tolls

What Do We Need?

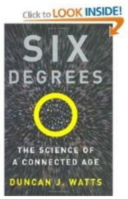
- Some math background:
  - A bit of probability theory
  - Some basic game theory
- Some background on algorithms and complexity
- Participation!

Course Schedule

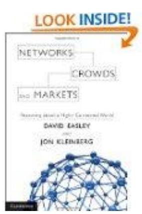
- **Days:**
  - 29/2/2012
  - 1/3/2012
  - 2/3/2012
- **Lecture Hours:**
  - 9:00 – 10:30
  - 11:00 – 12.30
  - 14:30 – 16:00
- **Open session:** 16:30 – 18:00

Resources

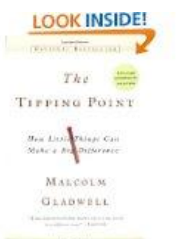
Many of the things that we cover are from papers. But some references are the following books:



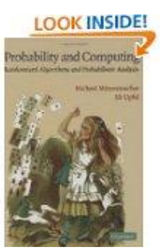
Duncan Watts: Six Degrees: The Science of a Connected Age
A nontechnical introduction for the topics we covered and more



David Easley, Jon Kleinberg: Networks, Crowds, and Markets: Reasoning About a Highly Connected World
An introductory textbook with a lot of topics on networks (social and not)
Also free at: **http://www.cs.cornell.edu/home/kleinber/networks-book/**



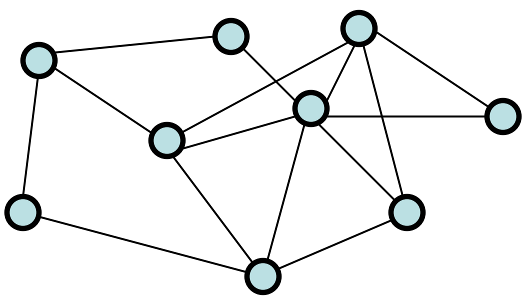
Malcolm Gladwell: The Tipping Point: How Little Things Can Make a Big Difference
About success stories about how the tipping point works; an easy and interesting read

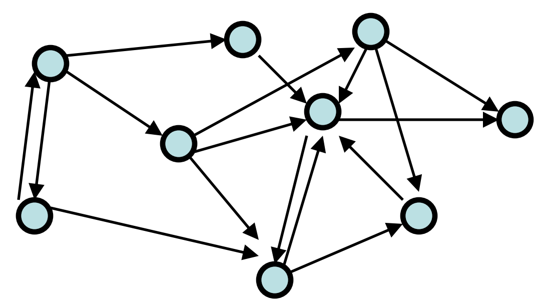


Michael Mitzenmacher and Eli Upfal: Probability and Computing: Randomized Algorithms and Probabilistic Analysis
Introduction about probabilistic techniques in computer science

What Is a Social Network?

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



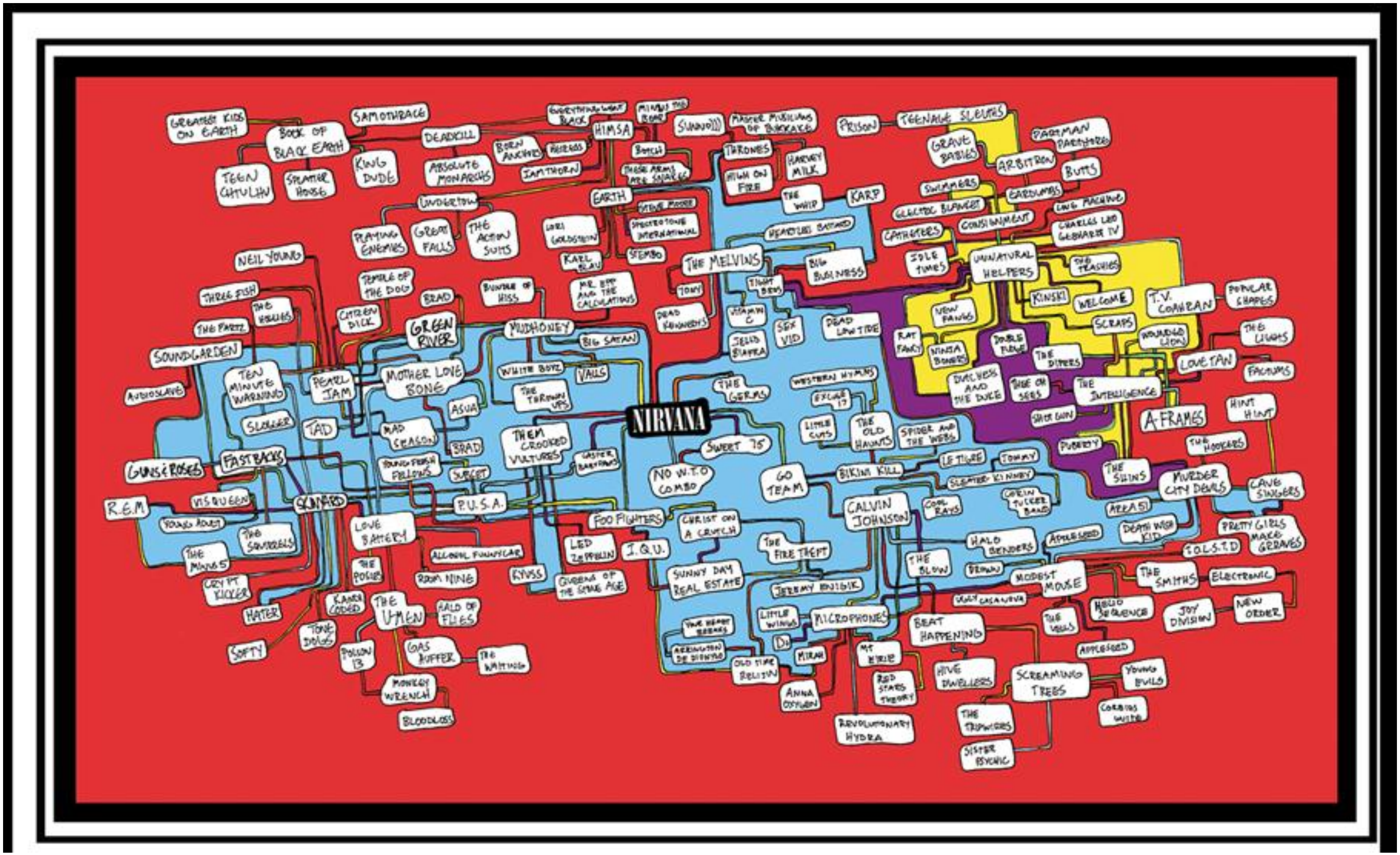


Social Networks Are Everywhere and Are Important!

Offline:

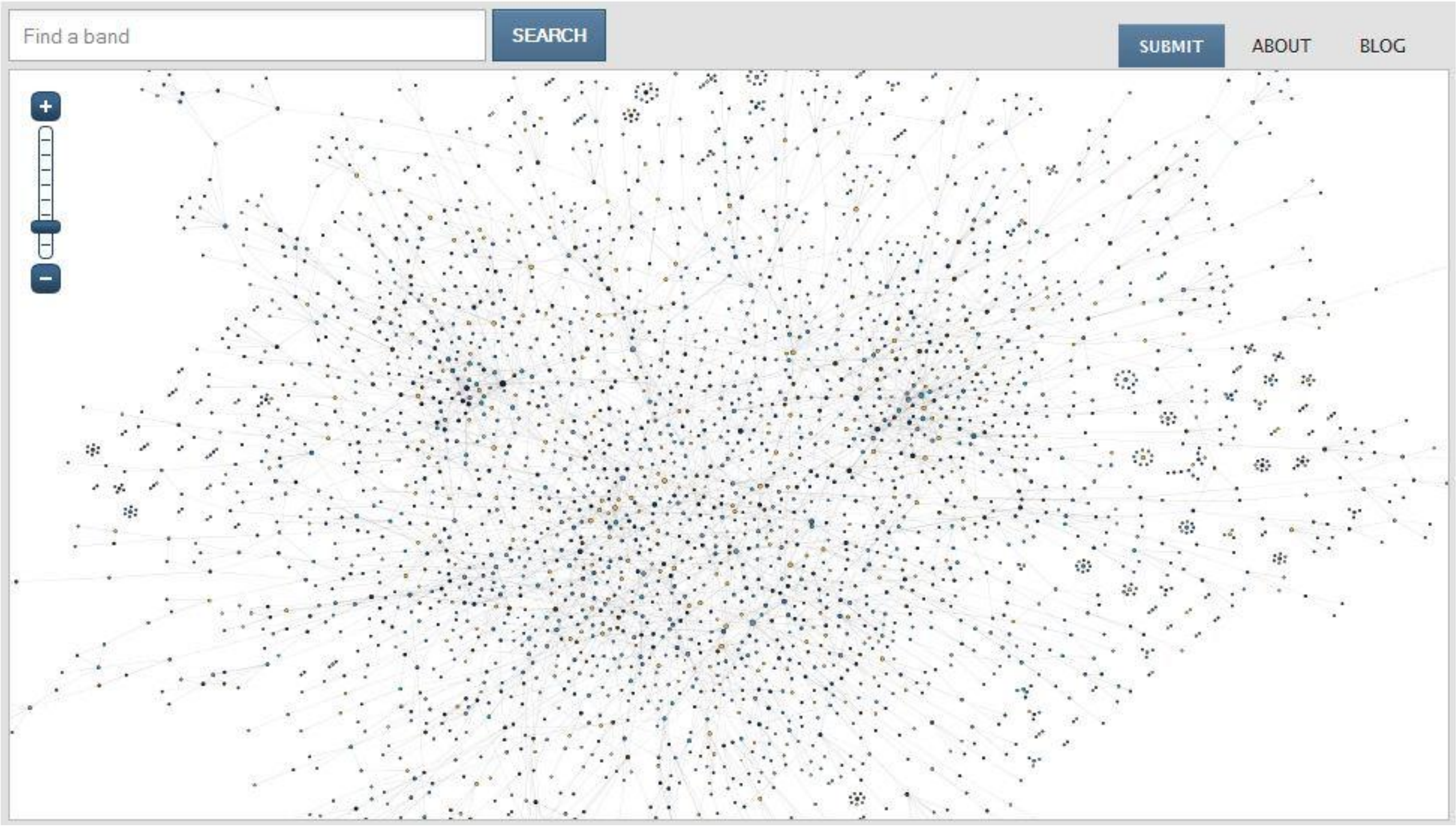
- Friendship network
  - “Show me your friend and I’ll show you who you are!”
- Professional contacts
  - Finding jobs
- Network of colleagues
  - Learning new techniques
- Network of animals
  - E.g., two cows are connected if they have been in the same area
  - Mad-cow disease

Nirvana



NIRVANA
GREATEST KIDS ON EARTH
BOOK OF BLACK EARTH
SAMOTHRACE
KING DUDE
TEEN CHTULHU
SPLATTER HOUSE
DEADKILL
ABSOLUTE MONARCHS
HIMSA
UNDERTOW
PLAYING ENEMIES
GREAT FALLS
THE ACTION SUITS
EARTH
STEVE MOORE
LORI GOLDSTEIN
KARL BLAU
STEMBO
THE MELVINS
BIG BUSINESS
THRONES
HARVEY MILK
HIGH ON FIRE
THE WHIP
KARP
PRISON
TEENAGE SLEUTHS
GRAVE BABIES
ARBITRON
BUTTS
SWIMMERS
ELECTRIC BLANKET
CONSIGNMENT
CATHETERS
IDLE TIMES
UNNATURAL HELPERS
KINSKI
WELCOME
SCRAPS
T.V. COAHRAN
POPULAR SHAPES
THE LIGHTS
LOVE TAN
FACTUMS
HINT HINT
A-FRAMES
THE INTELLIGENCE
THE HOOKERS
MURDER CITY DEVILS
CAVE SINGERS
THE SHINS
PUBERTY
DEATH WISH KID
PRETTY GIRLS MAKE GRAVES
T.O.L.S.T.D
THE SMITHS
ELECTRONIC
NEW ORDER
JOY DIVISION
MODEST MOUSE
APPLESEED
YOUNG EVILS
SCREAMING TREES
THE TRIPWIRES
SISTER PSYCHIC
HIVE DWELLERS
BEAT HAPPENING
REVOLUTIONARY HYDRA
RED STARS THEORY
MICROPHONES
ANNA OXYGEN
MIRAH
LITTLE WINGS
JEREMY ENIGK
SUNNY DAY REAL ESTATE
THE FIRE THEFT
CHRIST ON A CRUTCH
FOO FIGHTERS
I.Q.U.
LED ZEPPELIN
QUEENS OF THE STONE AGE
KYUSS
P.U.S.A.
SWEET 75
NO W.T.O COMBO
GO TEAM
CALVIN JOHNSON
BIKINI KILL
LE TIGRE
SLEATER-KINNEY
COOL RAYS
HALO BENDERS
THE BLOW
THE GERMS
WESTERN HYMNS
EXCUSE 17
THE OLD HAUNTS
SPIDER AND THE WEBS
DEAD LOW TIDE
SEX VID
JELLO BIAFRA
DEAD KENNEDYS
MR. EPP AND THE CALCULATIONS
BUNDLE OF HISS
MUDHONEY
BIG SATAN
VALIS
THE THROWN UPS
GREEN RIVER
BRAD
CITIZEN DICK
TEMPLE OF THE DOG
NEIL YOUNG
THREE FISH
THE FARTZ
SOUNDGARDEN
AUDIOSLAVE
TEN MINUTE WARNING
PEARL JAM
MOTHER LOVE BONE
MAD SEASON
TAD
FASTBACKS
GUNS & ROSES
R.E.M
VISQUEEN
SKIN YARD
THE MINUS 5
THE SQUIRRELS
HATER
SOFTY
TONE DOGS
THE U-MEN
LOVE BATTERY
ALCOHOL FUNNYCAR
ROOM NINE
THE POSIES
HALO OF FLIES
GAS HUFFER
MONKEY WRENCH
BLOODLOSS
THEM CROOKED VULTURES
YOUNG FRESH FELLOWS

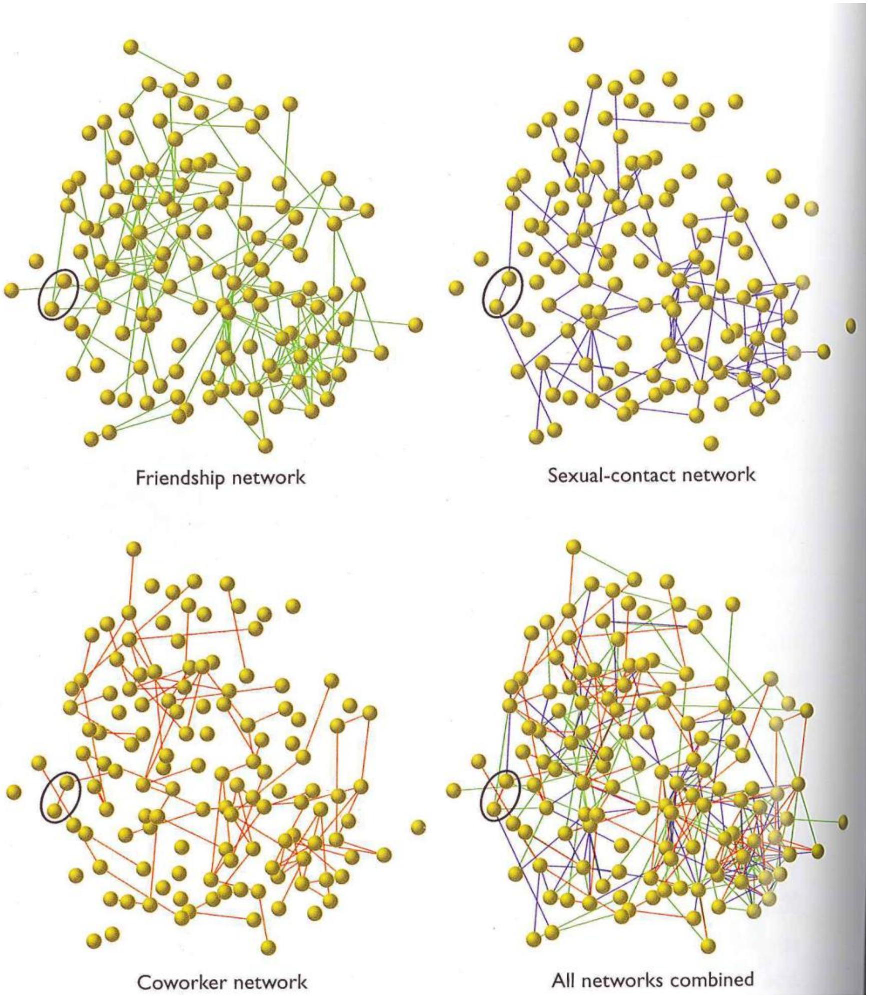
http://www.seattlebandmap.com



Seattle

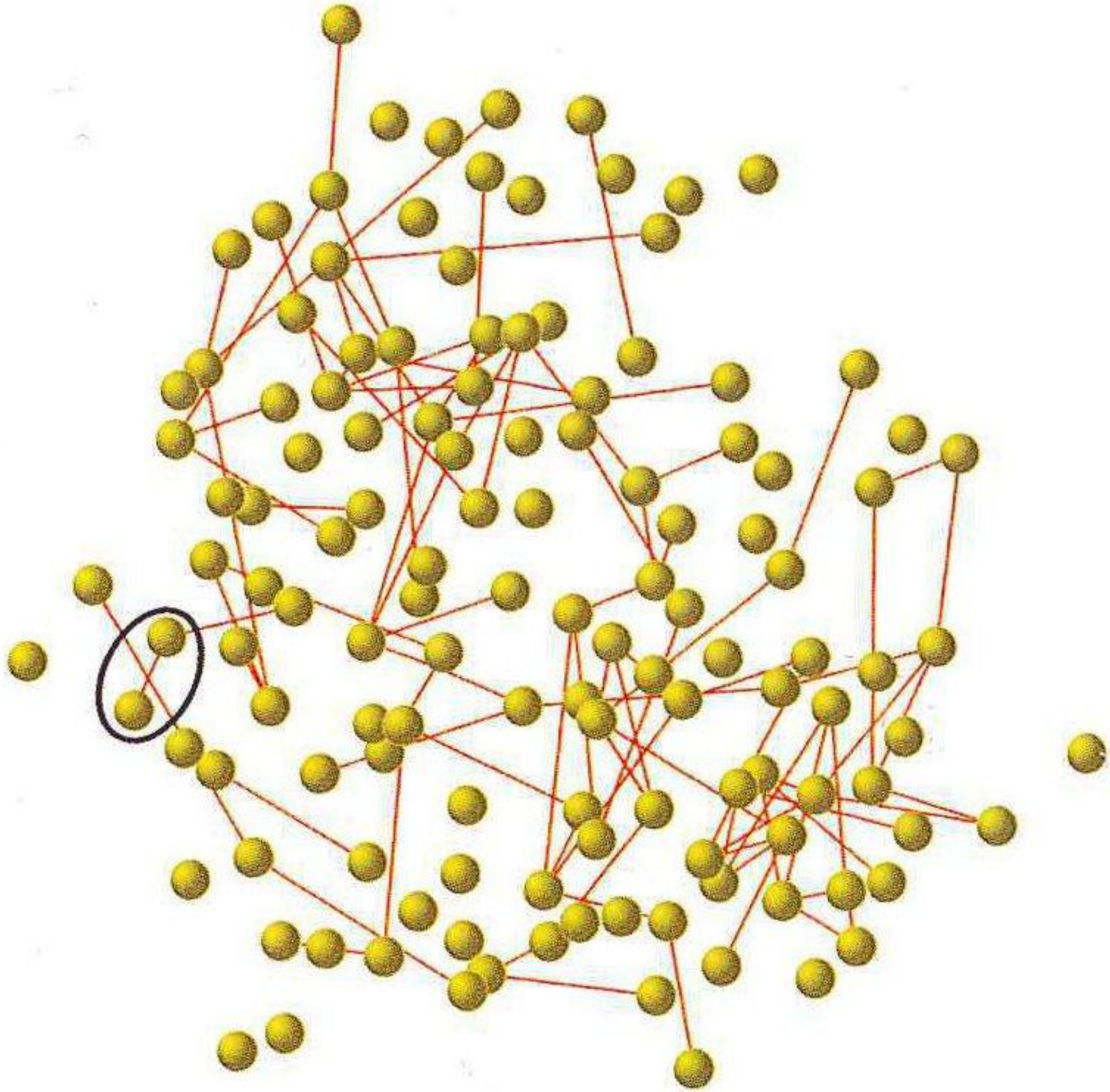

CANADA
WASHINGTON
MONTANA
NORTH DAKOTA
MINNESOTA
MAINE
OREGON
IDAHO
WYOMING
SOUTH DAKOTA
WISCONSIN
MICHIGAN
NEW YORK
NEVADA
UTAH
NEBRASKA
IOWA
ILLINOIS
INDIANA
OHIO
PENNSYLVANIA
CALIFORNIA
COLORADO
KANSAS
MISSOURI
WEST VIRGINIA
VIRGINIA
KENTUCKY
TENNESSEE
NORTH CAROLINA
ARIZONA
NEW MEXICO
OKLAHOMA
ARKANSAS
SOUTH CAROLINA
MISSISSIPPI
ALABAMA
GEORGIA
TEXAS
LOUISIANA
FLORIDA
PACIFIC OCEAN
ATLANTIC OCEAN
HAWAII
ALASKA
CANADA
MEXICO
GULF OF MEXICO
THE BAHAMAS
CUBA

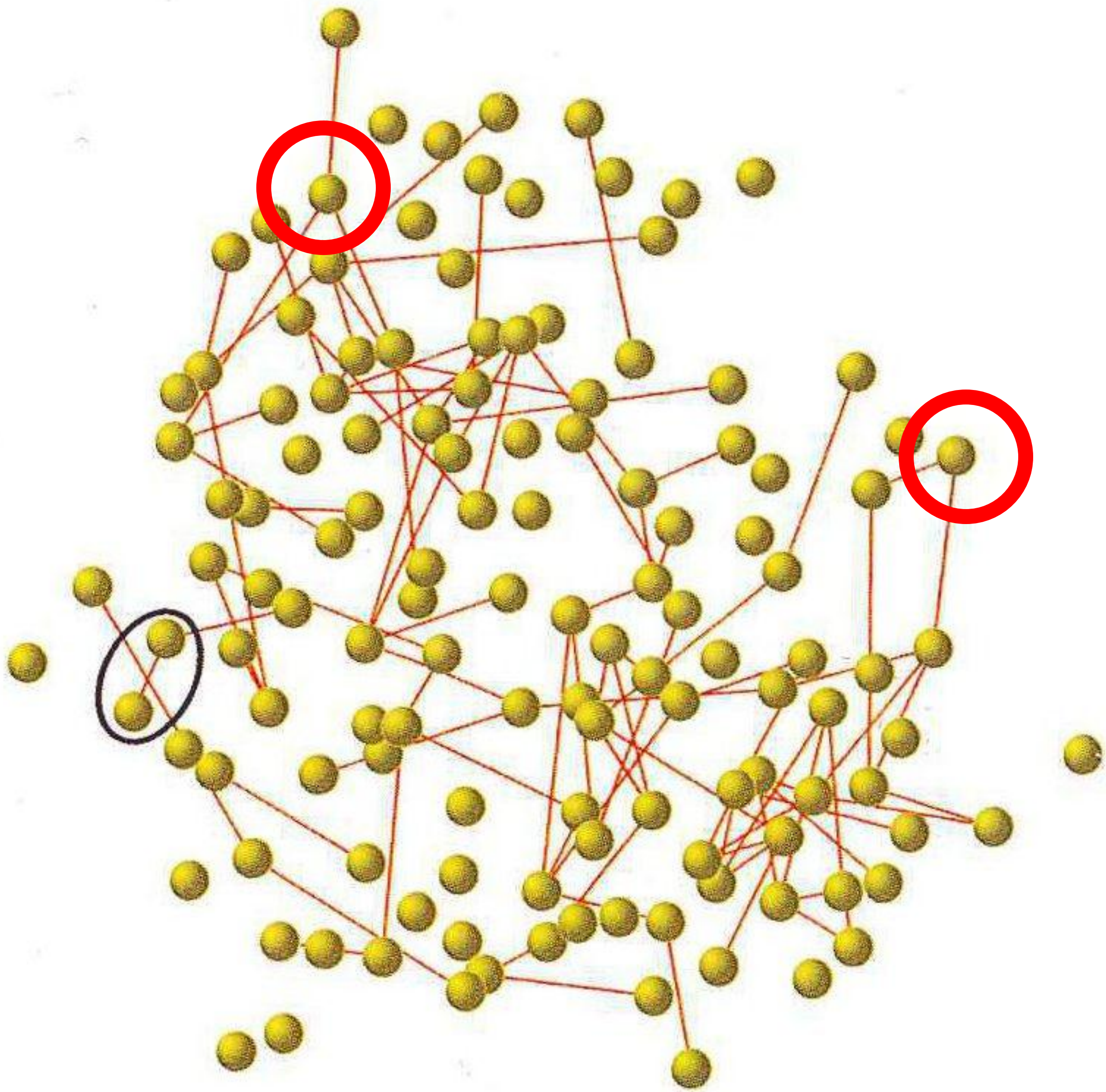
Multiple Social Networks

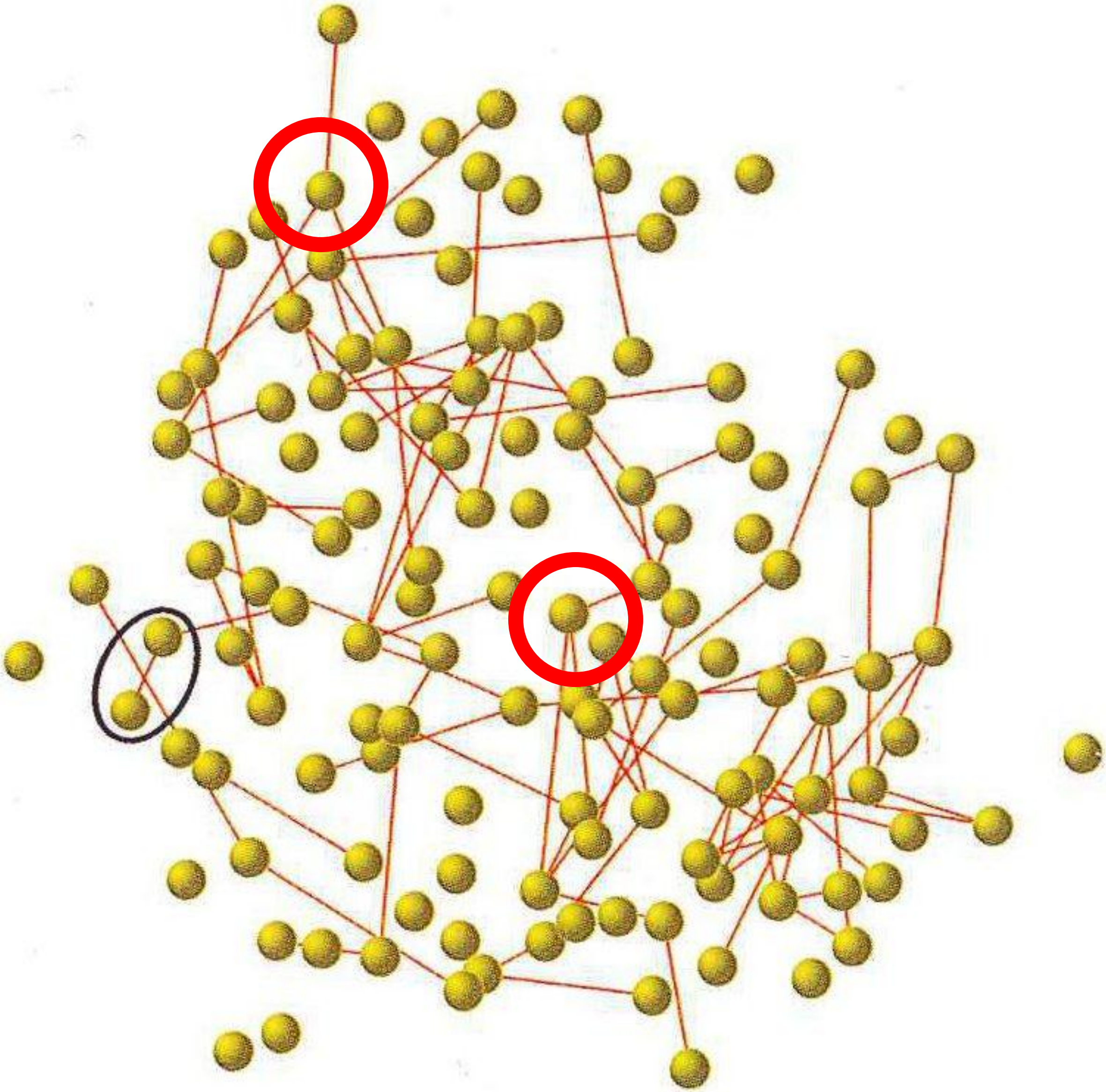


Examples:

- **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:

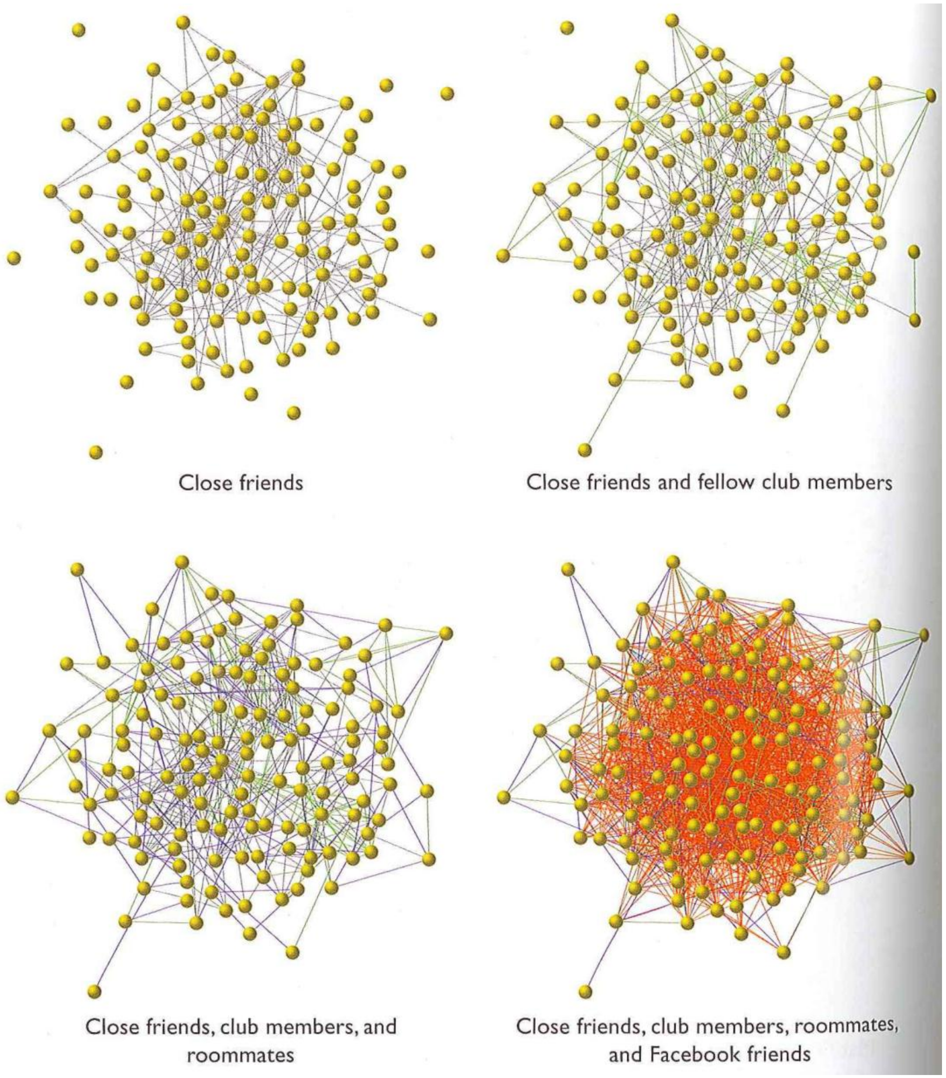
- Social networking systems
- Content sharing systems
- Content creation systems
- Online games

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



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

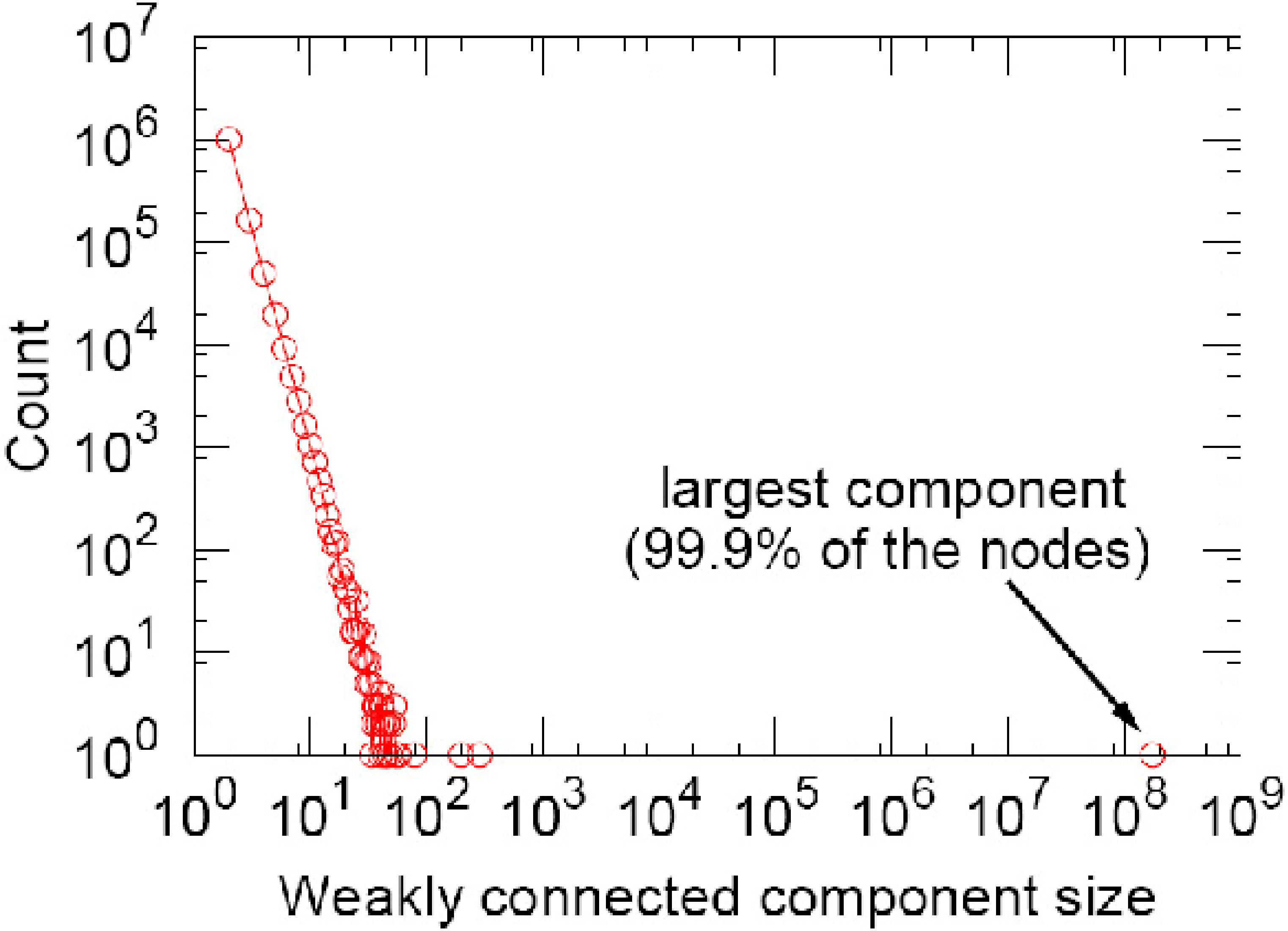
Structure of Complex Networks

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

MSN Messenger



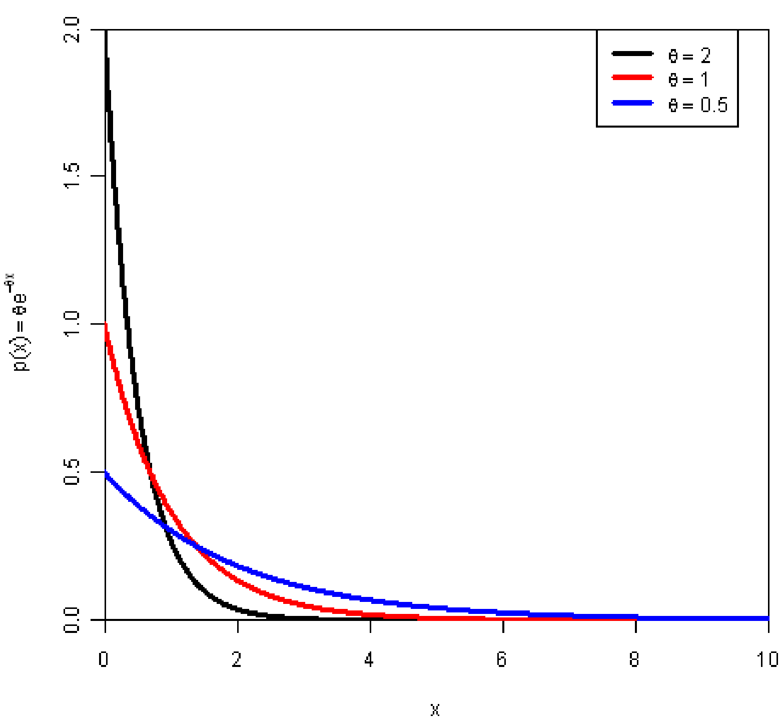
Power-Law Degree Distributions

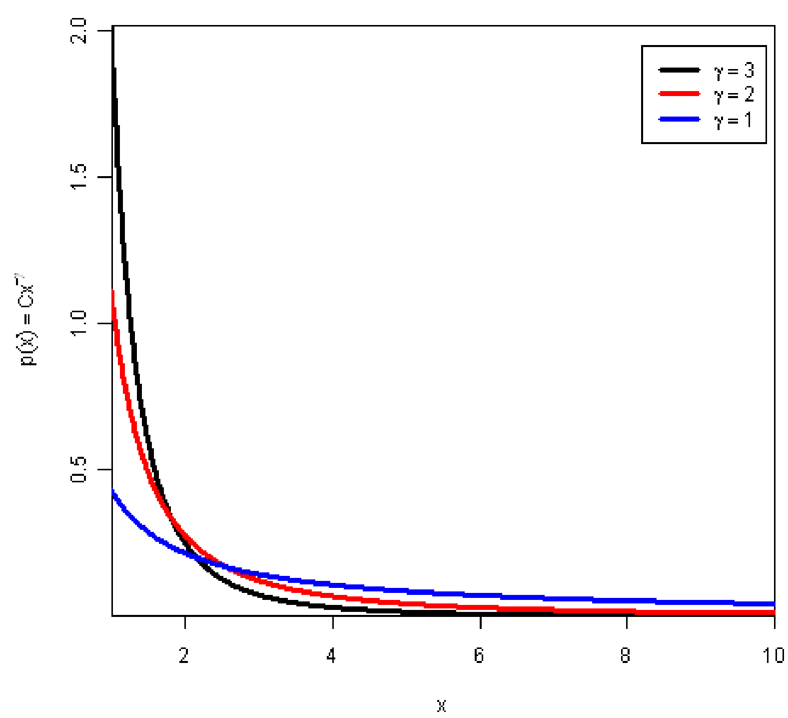
- The degree distributions of the networks follow a power-law distribution

- What is power law?

Power-Law Distribution

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





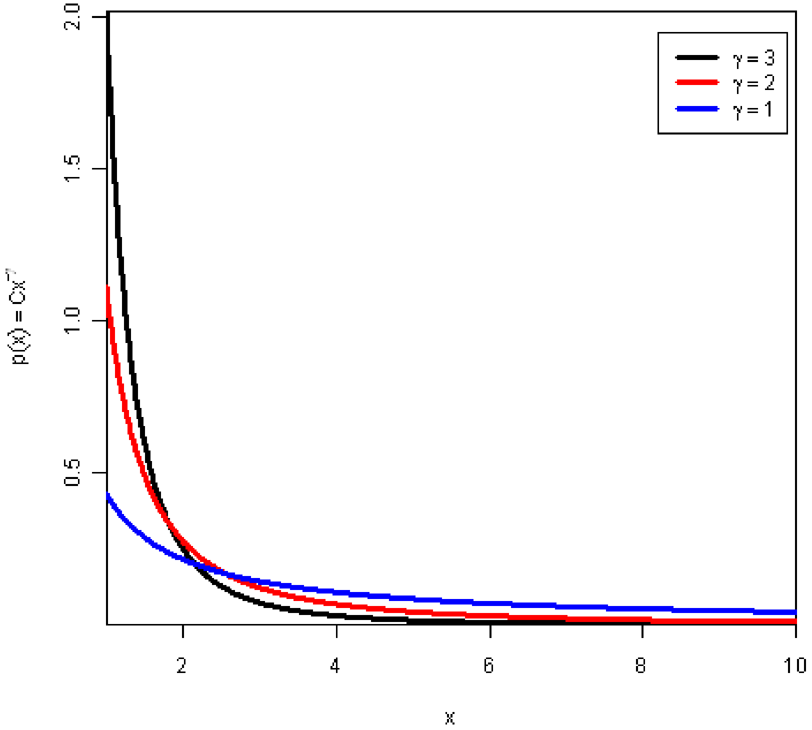
Power-Law Distribution - 2

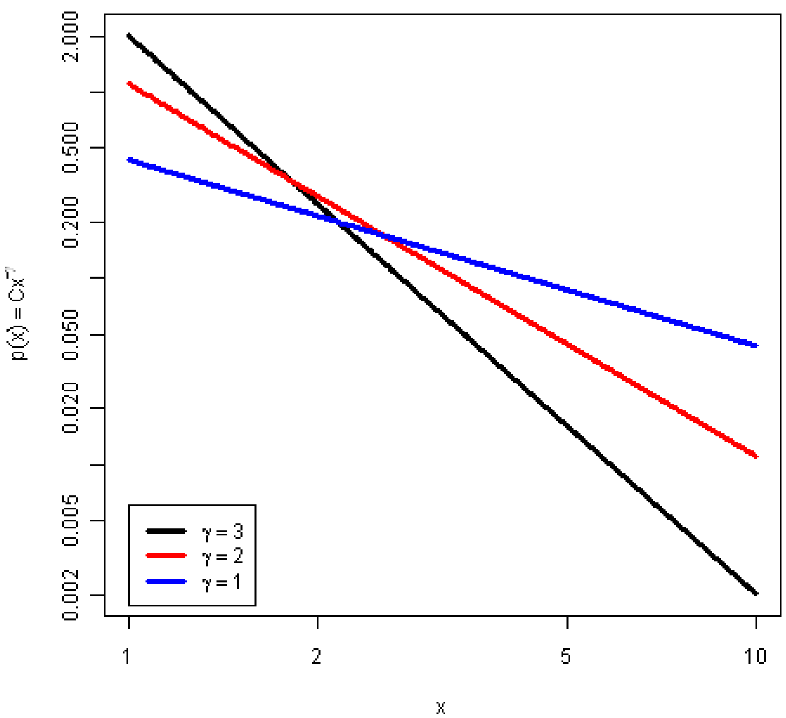
- It is a heavy-tail distribution
- **Heavy tail:** It decays slower than an exponential

- It is also called scale-free: $f(ax) = b \cdot f(x)$

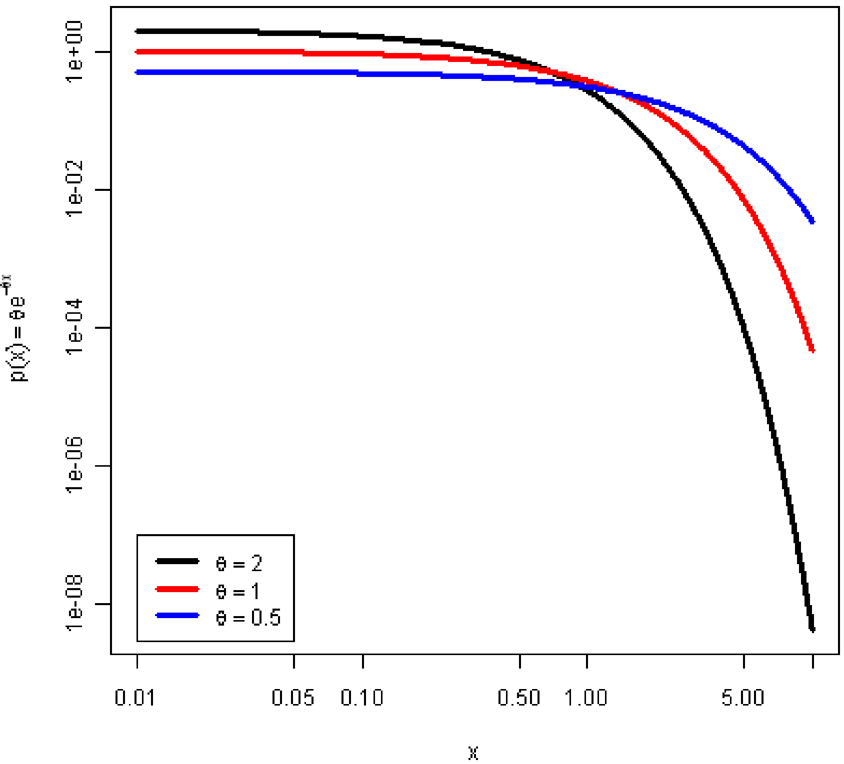
- It appears in many places:
  - Degree distribution
  - Population of cities
  - Word frequencies
  - Website hits
  - Income

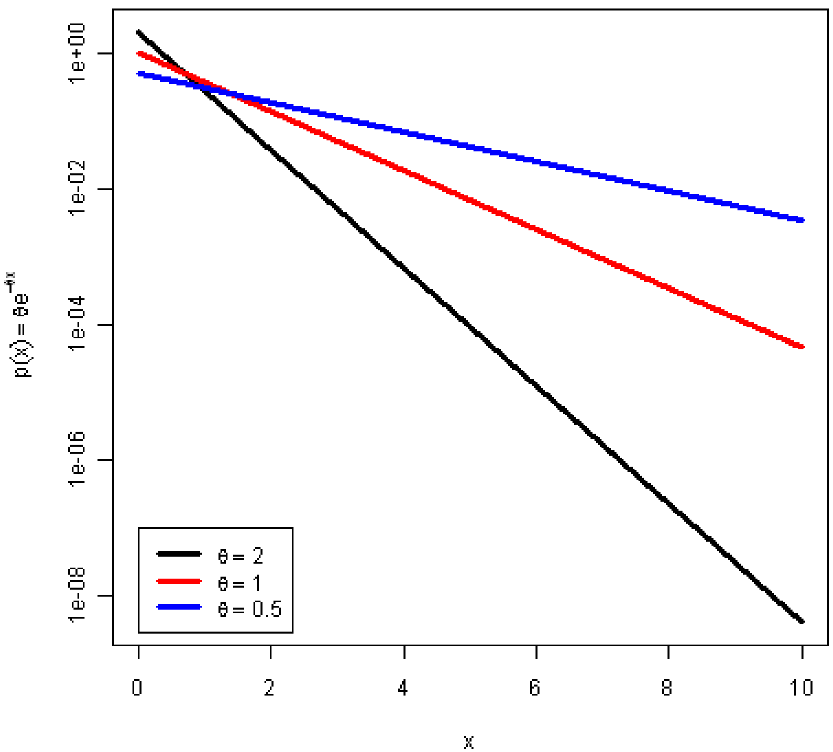
Power-Law Distribution - 3



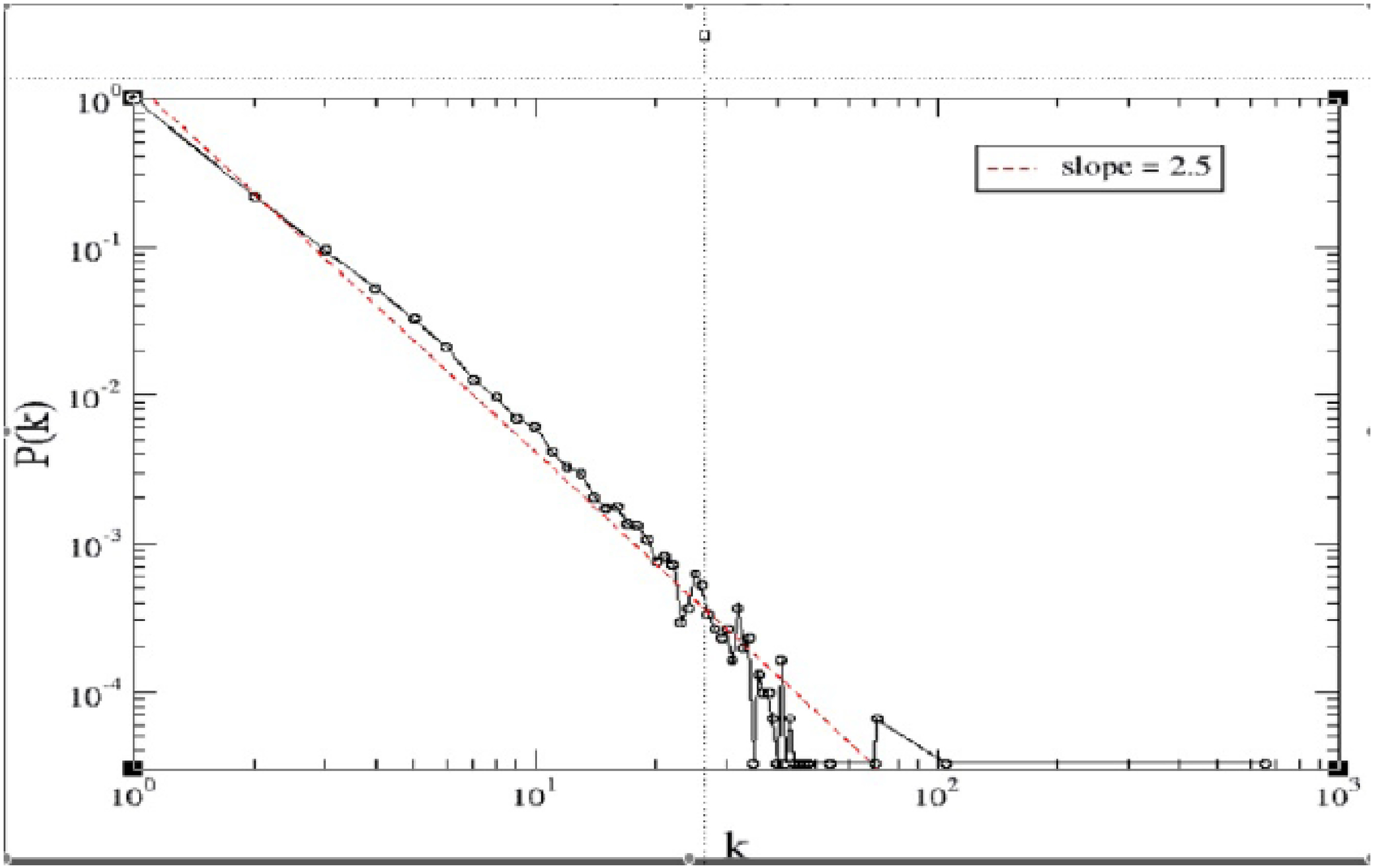


Exponential Distribution



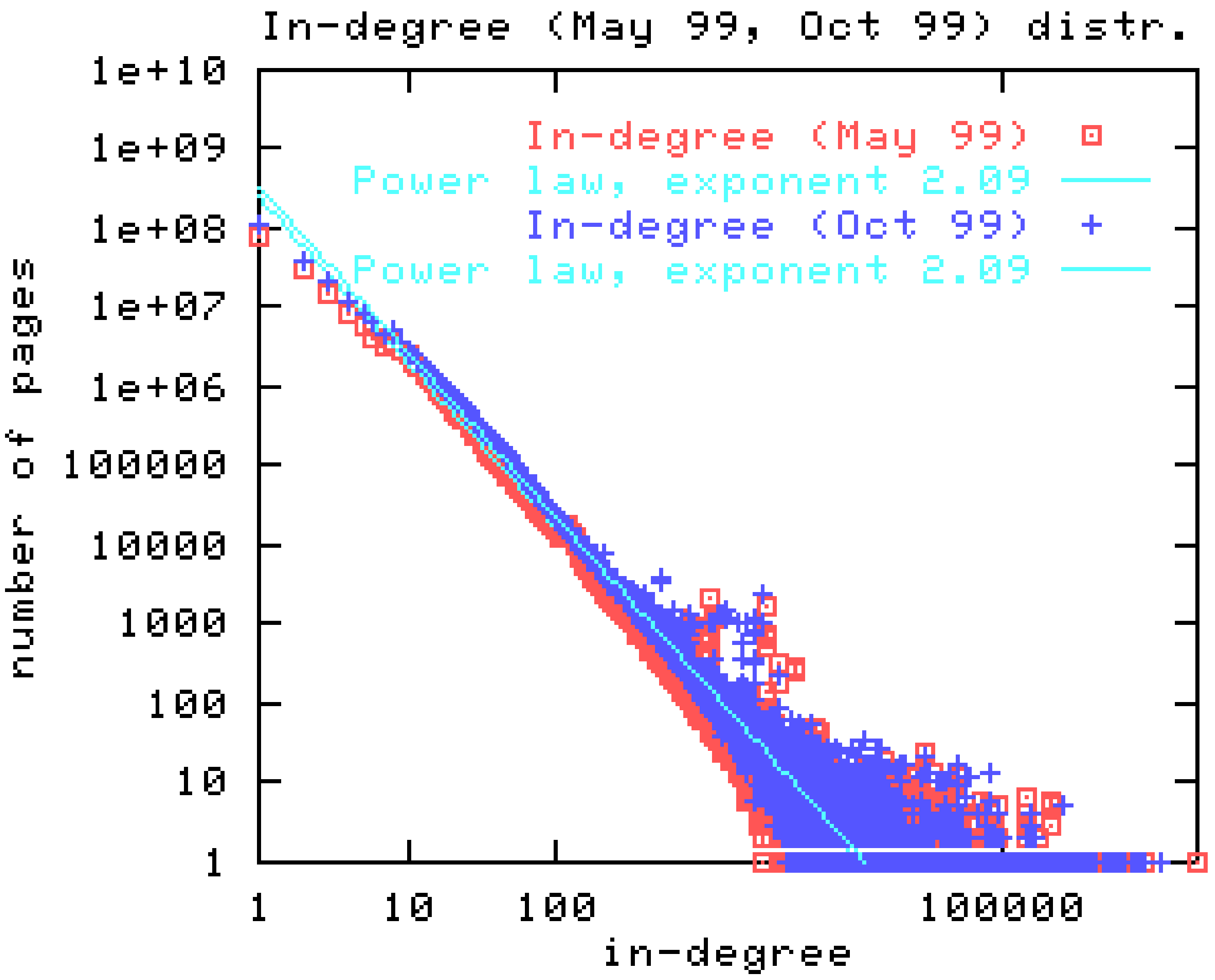


Back to Degree Distributions

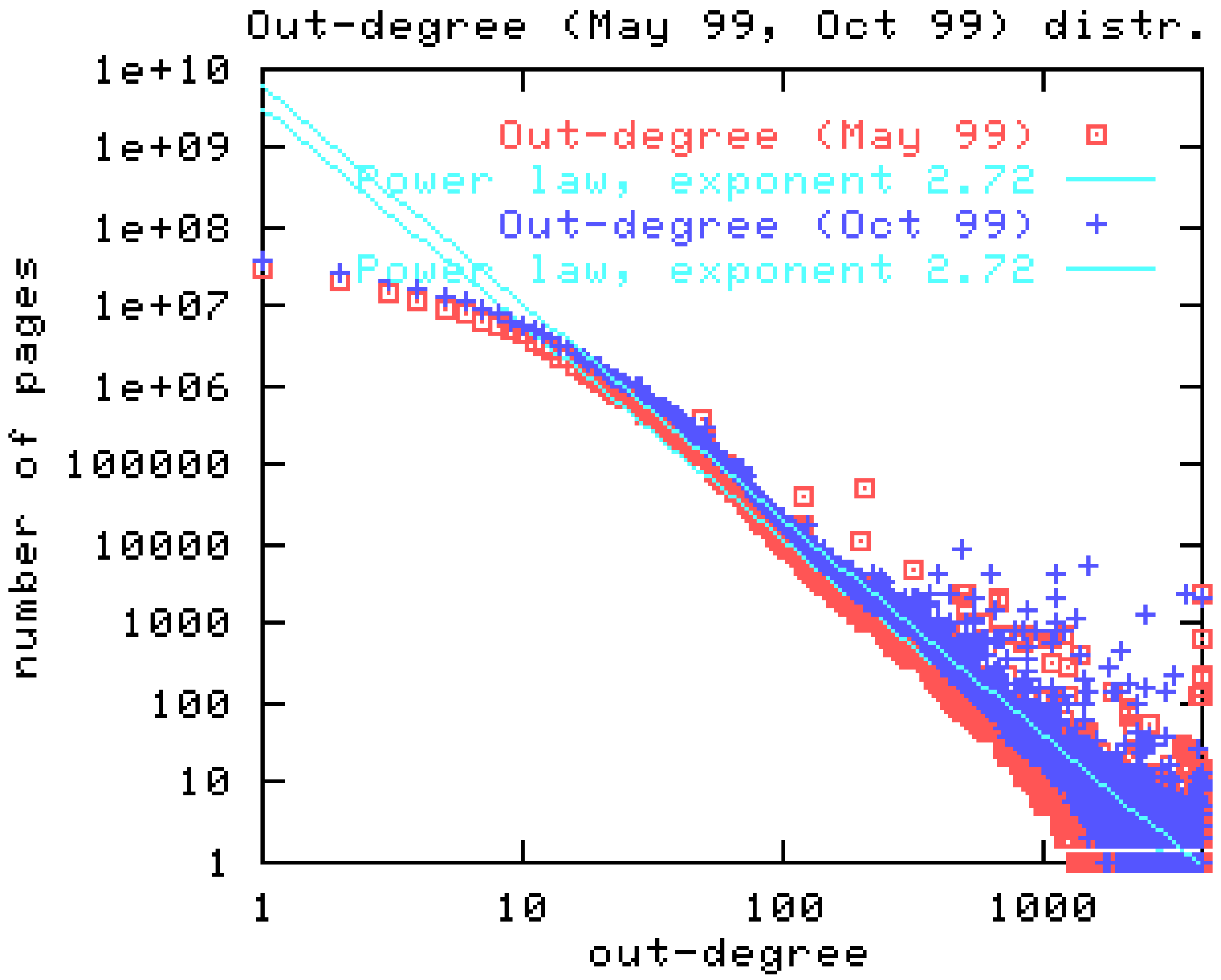


Internet Graph

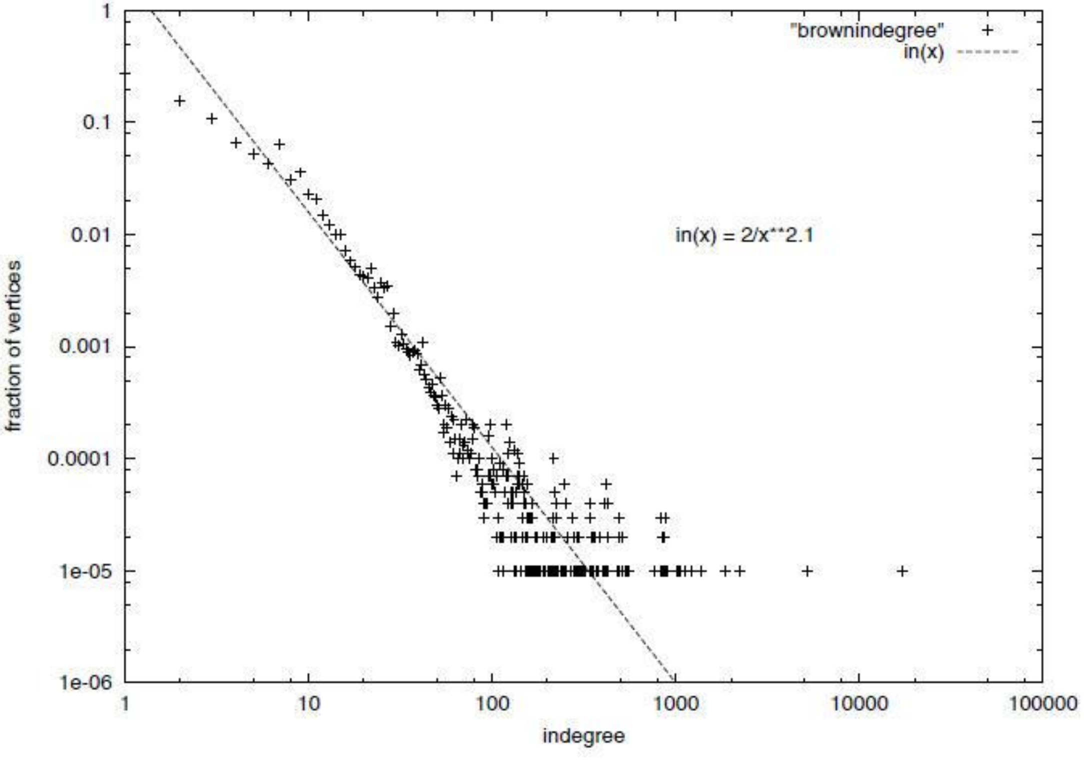
Web Graph Indegree



Web Graph Outdegree

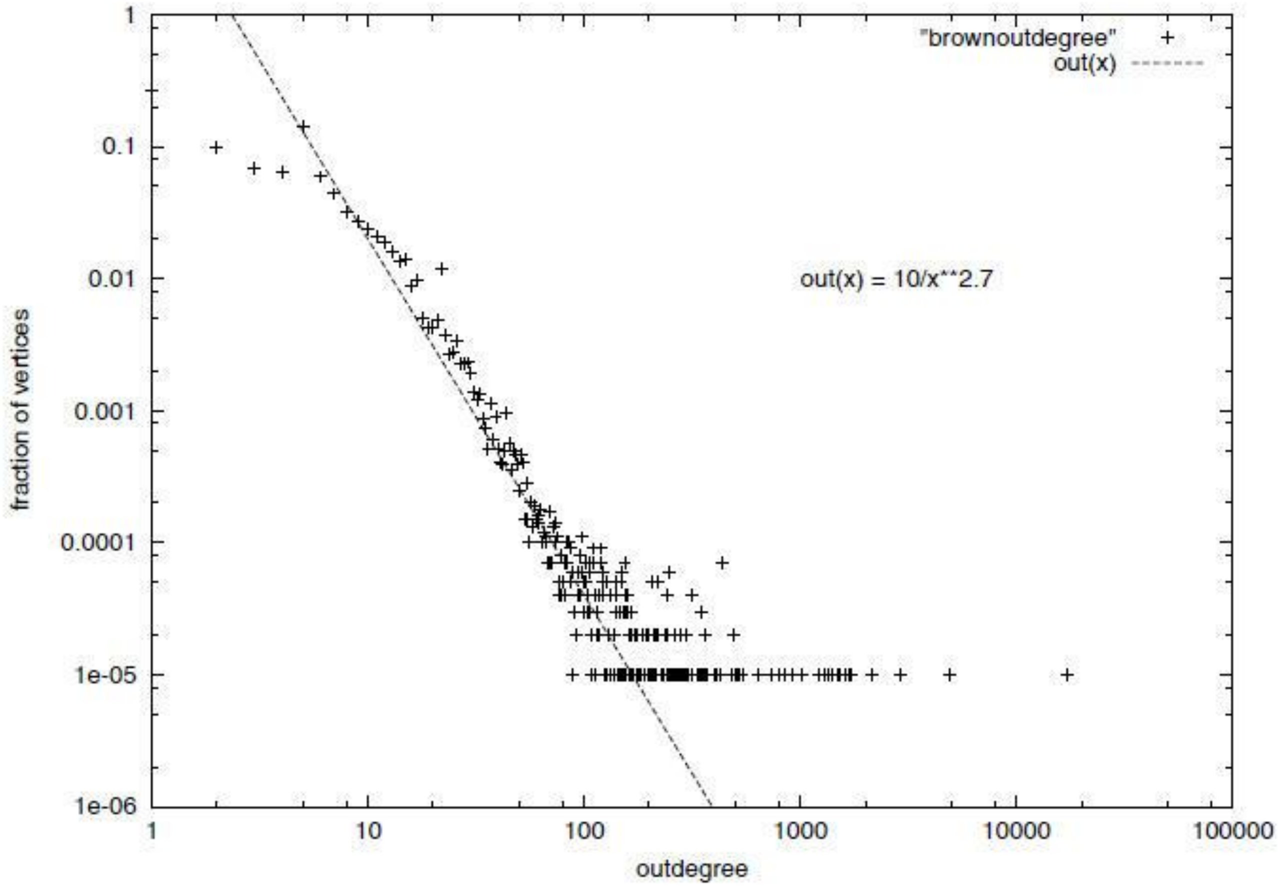


Degree Distributions



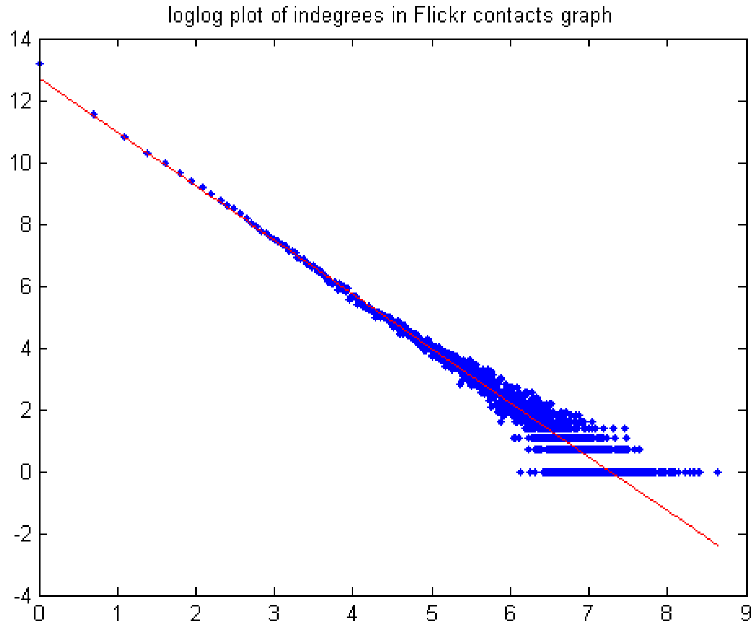
Indegree of the *.brown.edu domain

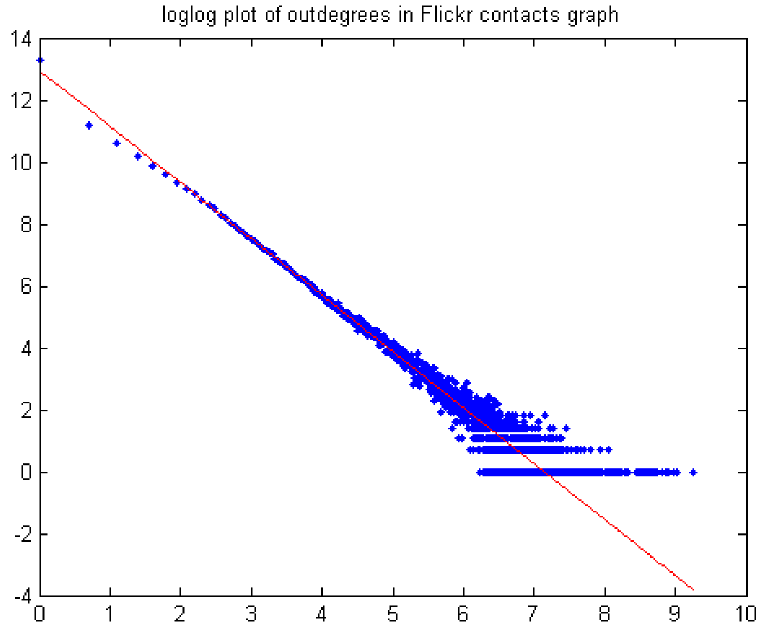
Degree Distributions



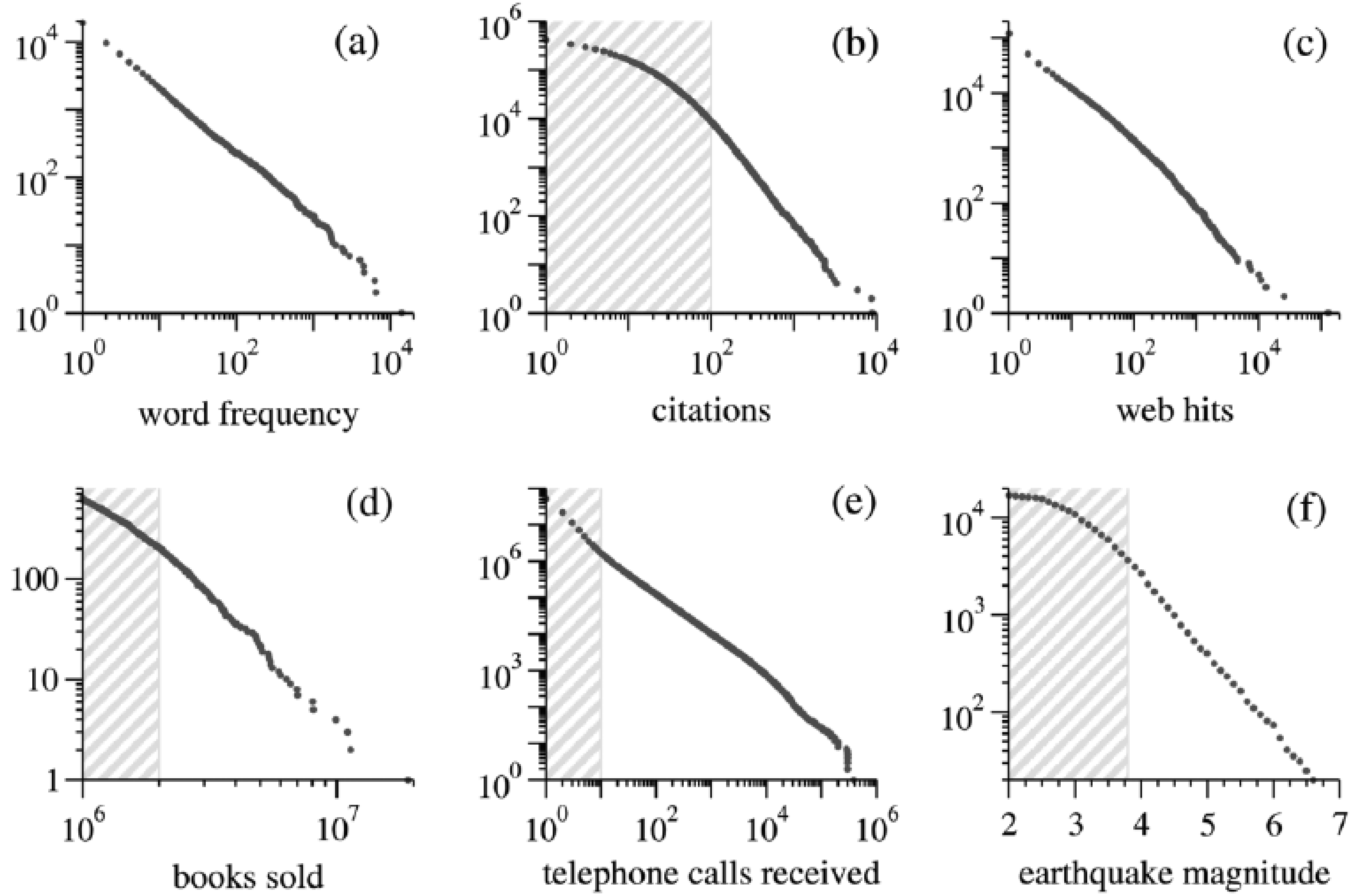
Outdegree of the *.brown.edu domain

Flickr Graph, Indegrees & Outdegrees

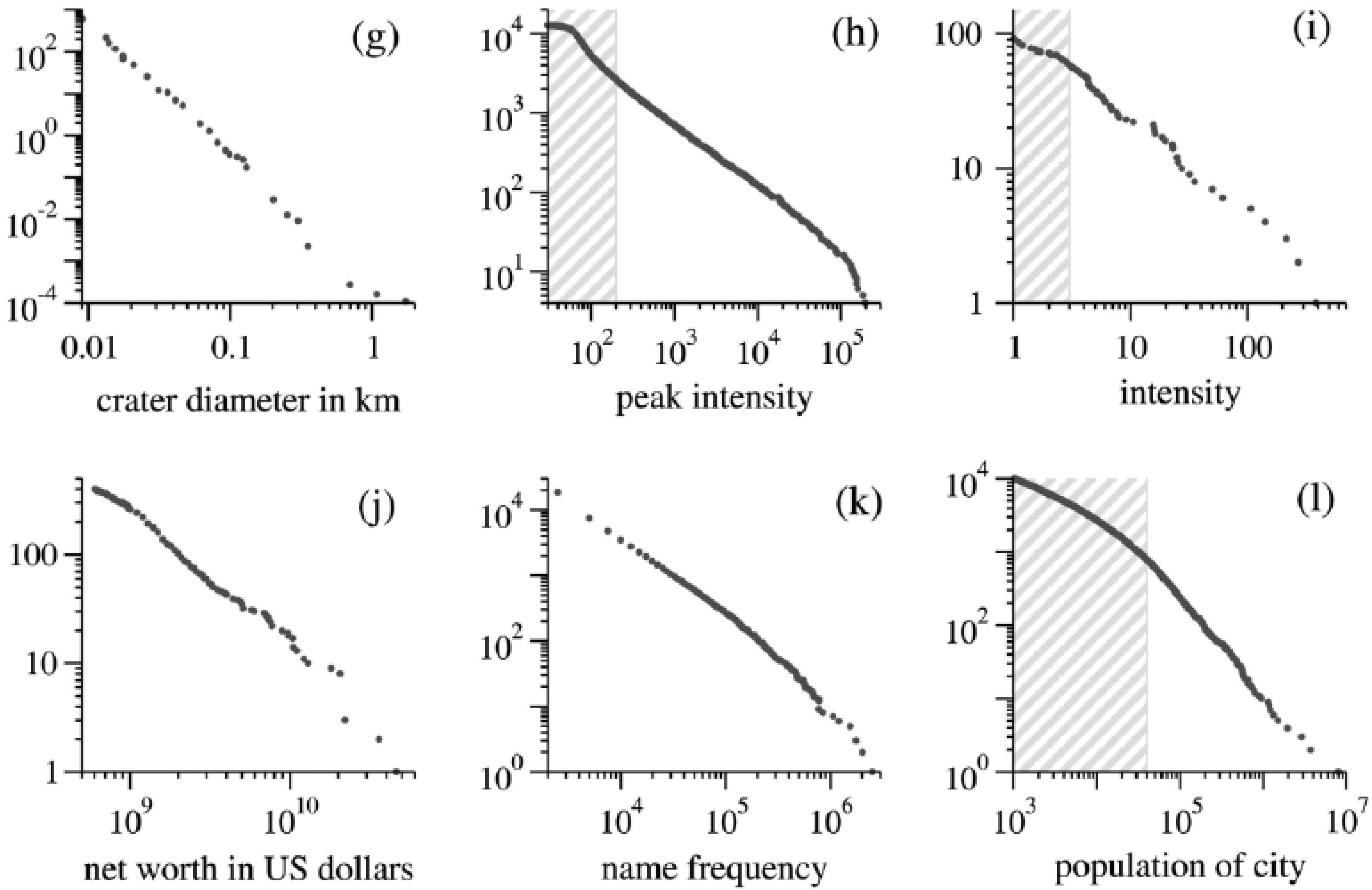




Power Laws Everywhere



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.

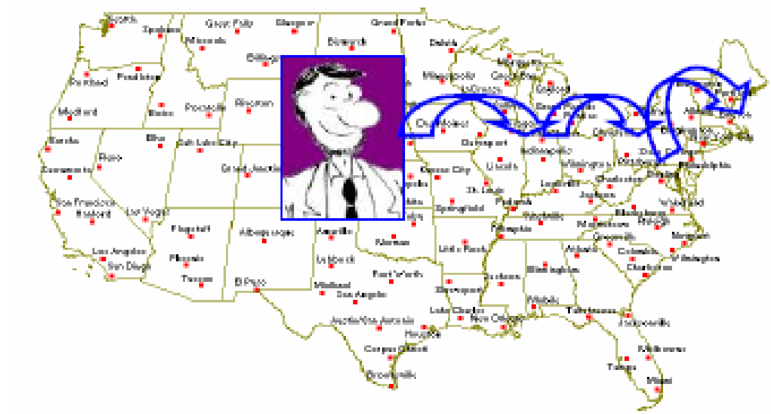
Small World

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

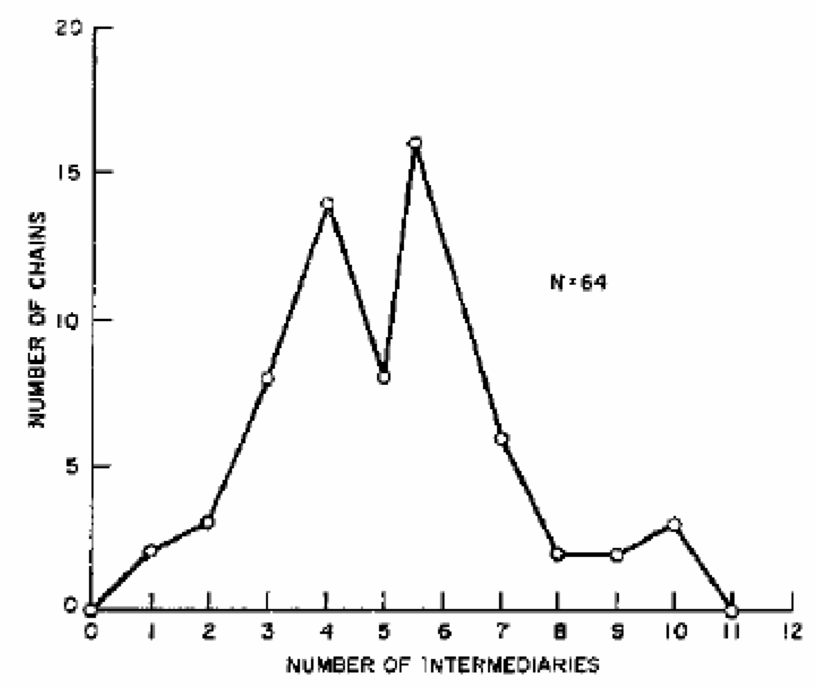
Rules for participants: Local routing

- 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

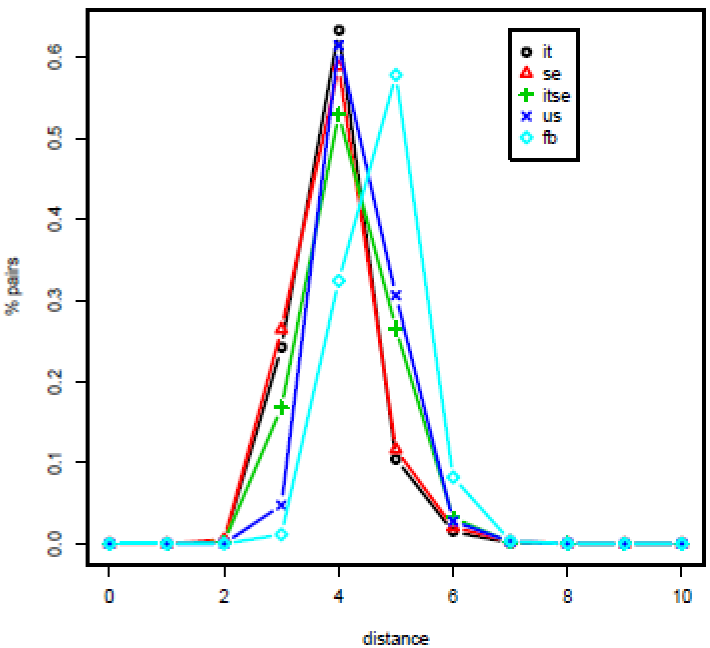
- 29% of the letters reached the destination
- Average path length, 5.2
- Bimodality is not accident: target reached through
  - hometown, 6.1
  - business contacts, 4.6
- Role of geography
  - Boston, 4.4
  - Nebraska, 5.5
- Role of occupation
  - random, 5.7
  - stockholders, 5.4

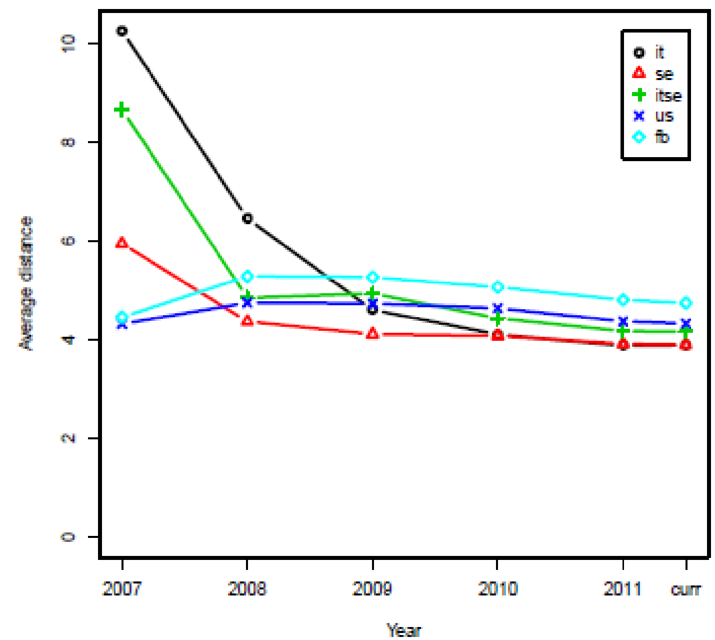


Small World – Facebook study

In January 2012 researchers from Facebook and University of Milan published results on the Facebook network

- Active users on May 2011
- n = 721M, m = 69 B
- Average distance = 4.74





Small World – Facebook study

FOLHA.com

27 DE FEVEREIRO DE 2012 - 11:09 SP 24ºC RIO 27ºC
VEJA O TEMPO EM MAIS CIDADES

BUSCAR

22/11/2011 - 18h40

No Facebook, 6 graus de separação viram 4,74

JOHN MARKOFF E SOMINI SENGUPTA
DO "NEW YORK TIMES"

Recomendar 448 +1 6

O mundo é ainda menor do que você imaginava.

Acrescentando um novo capítulo à pesquisa que fez com que o termo "seis graus de separação" se tornasse corrente, cientistas do Facebook e da Universidade de Milão reportaram na segunda-feira (21) que o número de conhecidos que separam duas pessoas quaisquer no planeta não era de seis, mas de 4,74.

A constatação original quanto aos "seis graus", publicada em 1967 pelo psicólogo Stanley Milgram, foi estabelecida por meio de um estudo entre 296 voluntários, convidados a enviar uma mensagem por cartão postal, retransmitida por amigos e amigos de amigos, a uma determinada pessoa em um subúrbio de Boston.

A nova pesquisa utiliza um universo um pouquinho maior de participantes: os 721 milhões de usuários do Facebook, equivalentes a mais de um décimo da população mundial. Os resultados foram postados no site do Facebook na noite de domingo.

A experiência levou um mês. Os pesquisadores utilizaram um conjunto de algoritmos desenvolvido pela Universidade de Milão a fim de calcular a distância média entre duas

Globally Sparse, Locally Dense

- Social networks are sparse, i.e. small number of edges (think of facebook)

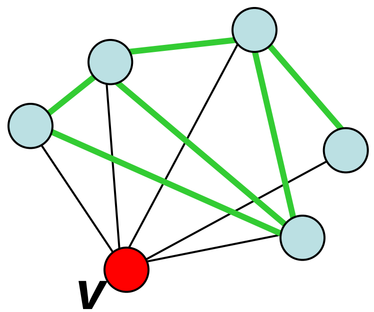
- They are locally dense: many of my friends are friends with each other

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 Facebook:

$$\frac{|E|}{\binom{n}{2}} = \frac{69 \cdot 10^9}{\binom{721 \cdot 10^6}{2}} = 2.7 \cdot 10^{-7}$$

- For Yahoo! Messanger:

$$\frac{|E|}{\binom{n}{2}} = 4 \cdot 10^{-8} \qquad C(G) = 0.16$$

- One explanation: Communities

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

Modeling Approaches

Two main types of mathematical models:

Probabilistic
- There is a random process that creates the networks
- Allow fitting the model to the data to estimate parameters
- Can be used to make predictions
- Answer **how**

This lecture

Game Theoretic
- Agents (Users) are rational players in a "game"
- Answer **why**

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 for the small world

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 others
- *p* is in general a function of *n*

We want to study properties of the graph for large *n*

Simple Graph Questions

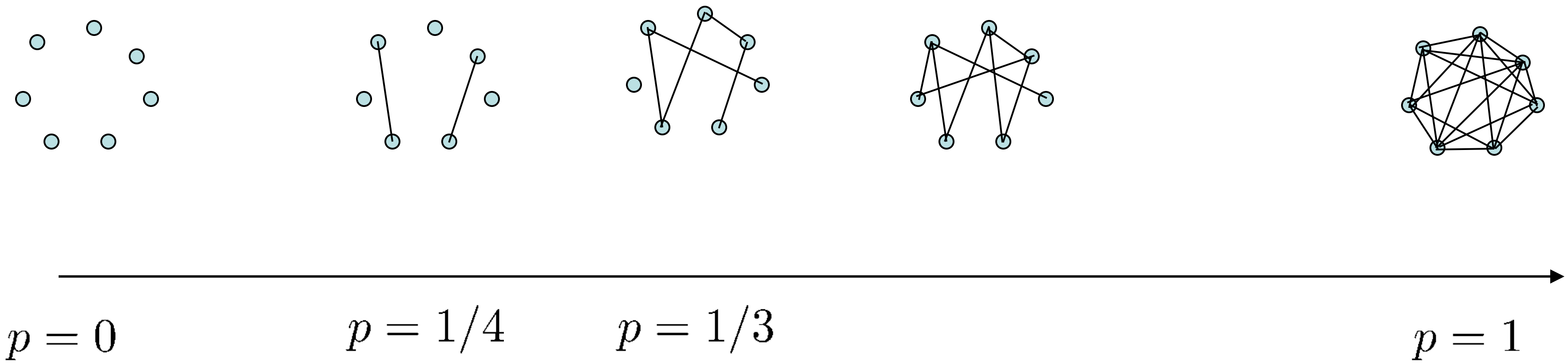
- What is the expected number of edges?

- 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 $n \to \infty$

- 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

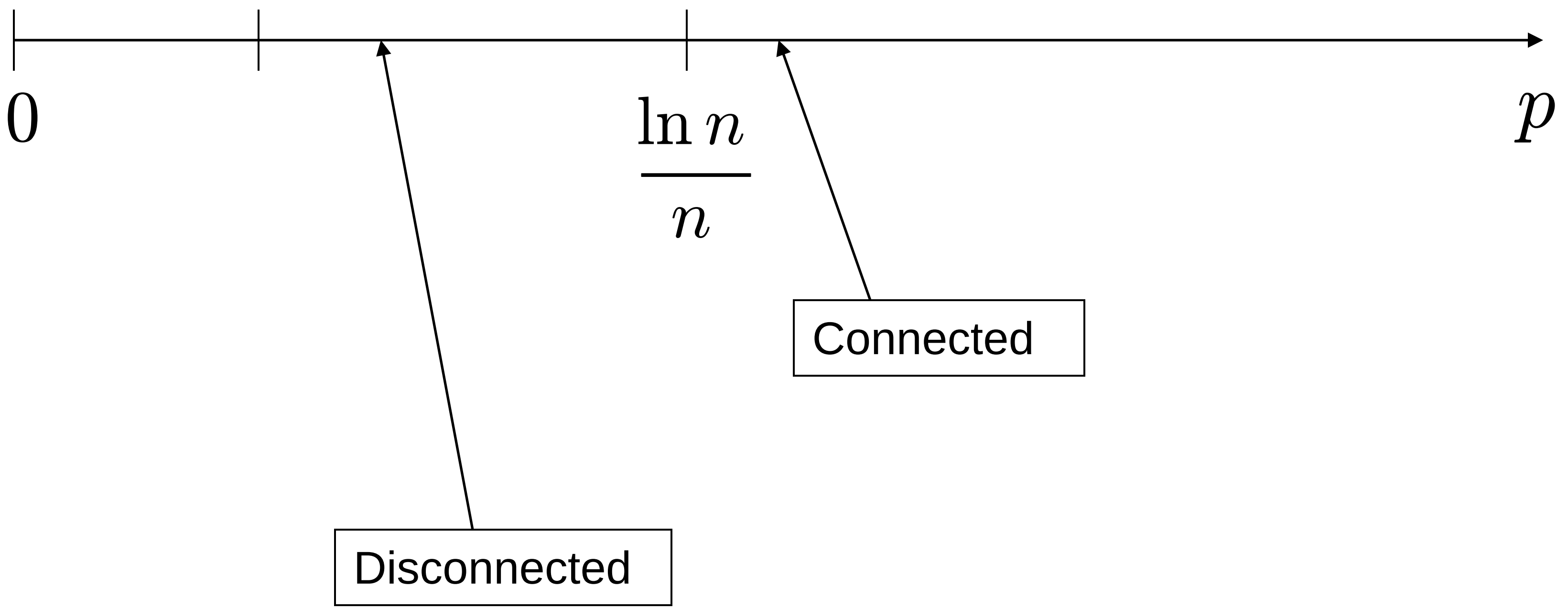
For many properties we have a **threshold**: p^*

- 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

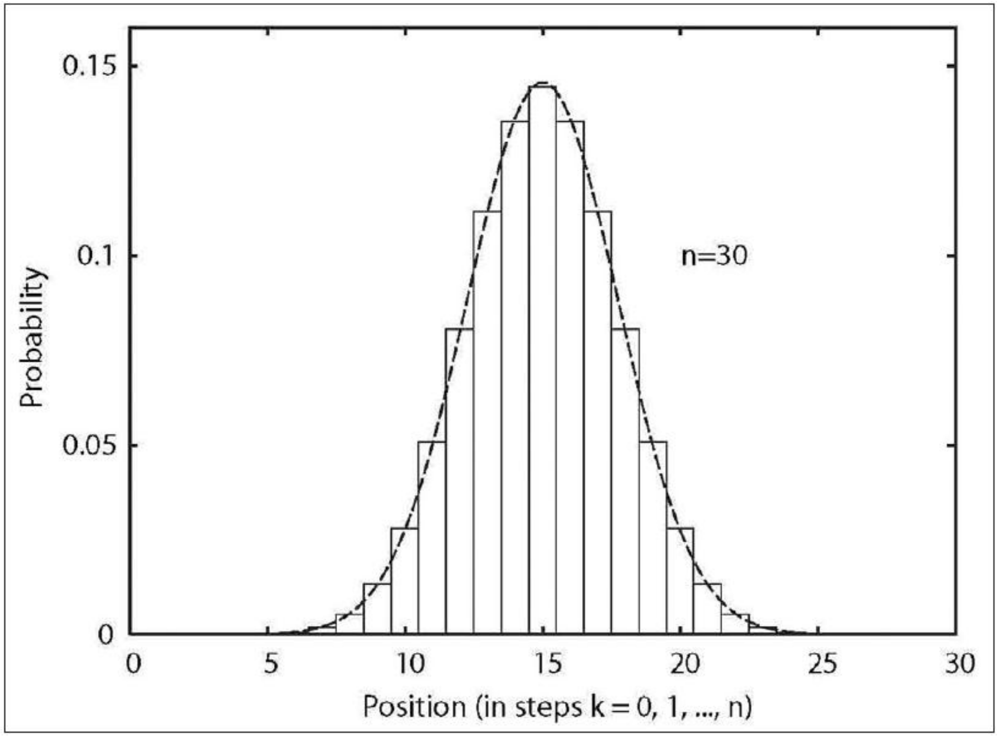
We study the connectivity of the graph



Degree Distribution

Degree distribution is **Binomial**:

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



Random Graphs for Social Network Modeling

What are the properties of social networks that random graphs create?

What are the nonrealistic properties?

Random Graphs for Social Network Modeling

What are the properties of social networks that random graphs create?

- Small worlds

What are the nonrealistic properties?

Random Graphs for Social Network Modeling

What are the properties of social networks that random graphs create?

- Small worlds

What are the nonrealistic properties?

- In real networks nodes are not connected to random nodes, we have communities

Random Graphs for Social Network Modeling

What are the properties of social networks that random graphs create?

- Small worlds

What are the nonrealistic properties?

- In real networks nodes are not connected to random nodes, we have communities
- Degree distribution is power-law, not binomial

Random Graphs for Social Network Modeling

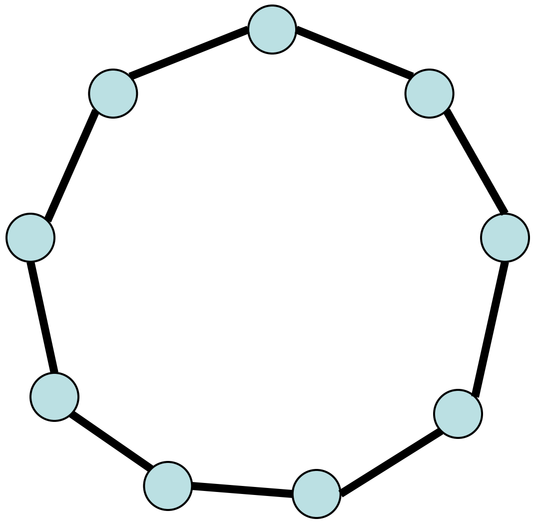
In real networks we have links between people “in our neighborhood” and some “random” links

This brings us to the Watts-Strogatz model

Watts-Strogatz Model

Parameters: *n*, *k*, *p*

- *n* nodes are arranged on a ring
- Each node connects to k of its closest neighbors
- Each link is independently "rewired" to be random with probability *p*



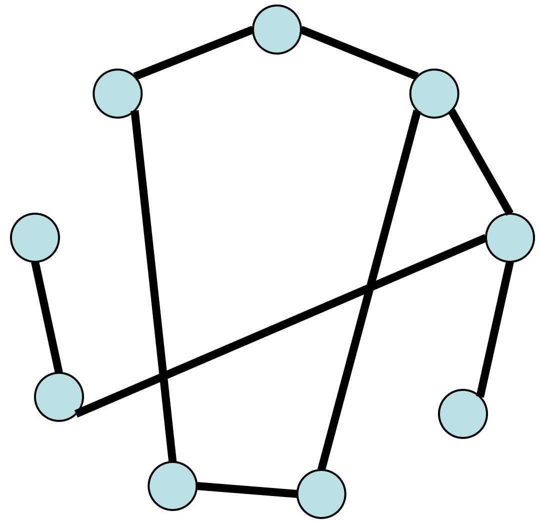
- *p* controls randomness

$n = 9$, $k = 2$, $p = 1/3$

Watts-Strogatz Model

Parameters: *n*, *k*, *p*

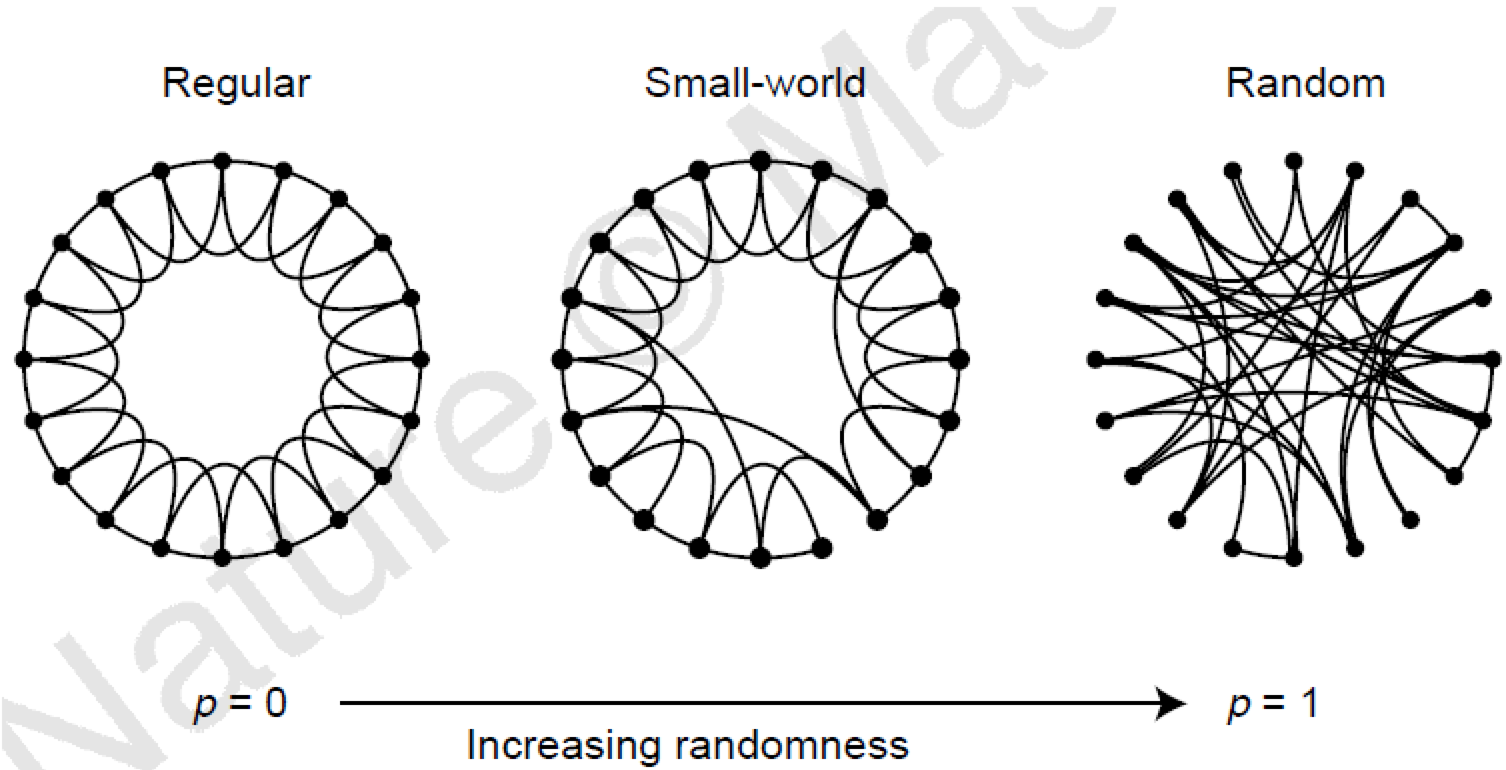
- *n* nodes are arranged on a ring
- Each node connects to k of its closest neighbors
- Each link is independently "rewired" to be random with probability *p*



- *p* controls randomness

n = 9, *k* = 2, *p* = 1/3

The effect of *p*



The effect of p

Measures of randomness and structure:

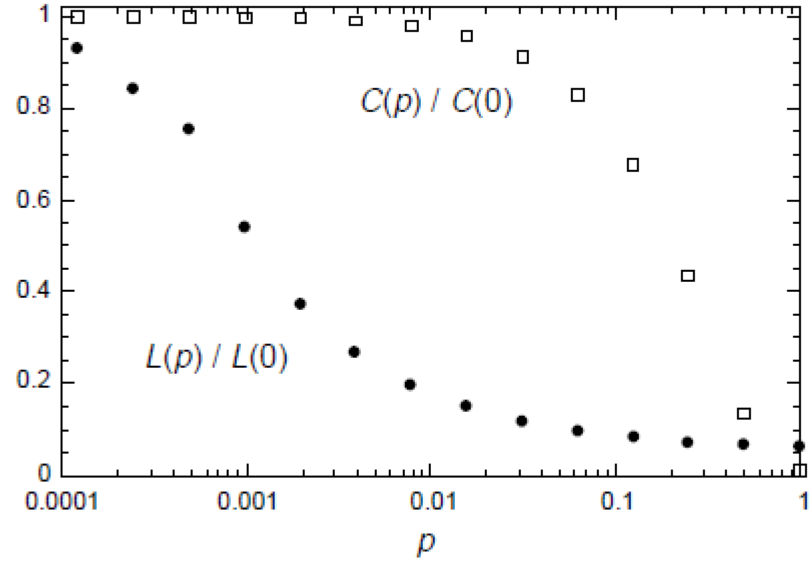
- $L(p)$ = average distance between 2 nodes
- $C(p)$ = clustering coefficient

As $p \to 0$

- $L(p) \to n/k$, $C(p) \to 3/4$

As $p \to 1$

- $L(p) \to \ln(n)/\ln(k)$, $C(p) \to k/n$



There is a large range of p where L(p) is almost like random but C(p) is much better than random

Interpretation of Links

- Structured links are local
- Random links are long-range

- Local links reflect **strong ties**
- Long-range links reflect **weak ties**

- Social networks consist of a mixture of structured and local links

Preferential Attachment

- A model for the creation of social networks
- We start with a network
- In every step a new node arrives with 2 slots
- Each slot connects to a node with prob. proportional to **degree**