Social Networks: Structure and Properties

Advanced Computer Networks

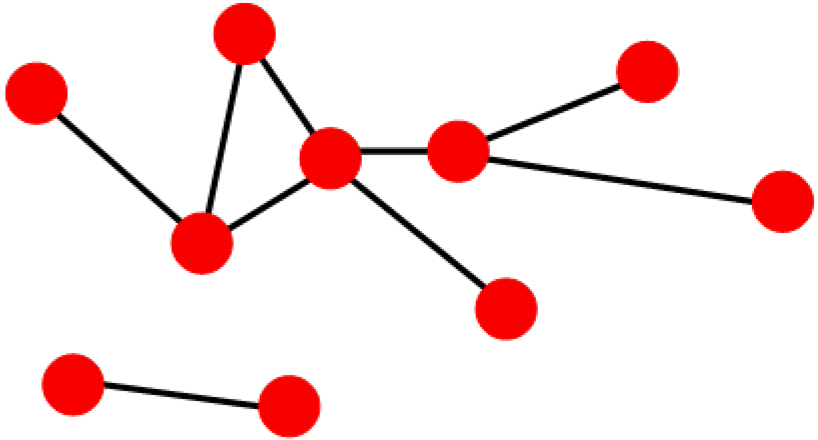
Summer Semester 2013

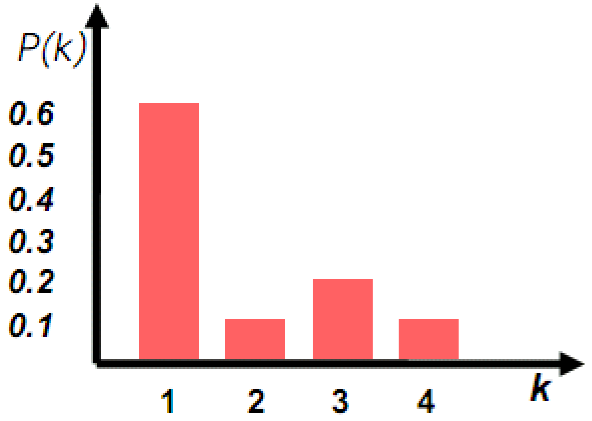
How to Characterize Networks?

- How many neighbors does a node have?
  - Degree distribution
- How far apart are nodes in the network?
  - Distance (the shortest path)
  - Network diameter
  - Average path length
- How close a set of nodes connect with each other?
  - Community
  - Clustering coefficient

Degree Distribution

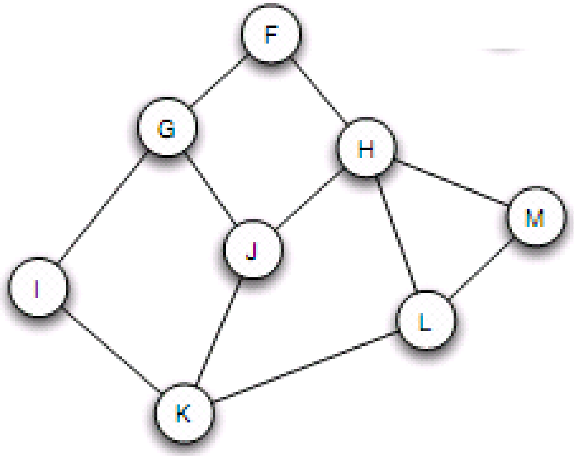
- Degree distribution P(k)
  - Probability that a randomly chosen nodes has degree k
  - N_k: number of nodes with degree k
  - $P(k)=N_k/N$





Path Length

- **Distance**: the number of edges along the shortest path connecting the nodes
  - If two nodes are disconnected, the distance is infinite
- **Diameter**: the maximum distance between any pair of nodes in the graph
- **Average path length**:



computer NETWORKS

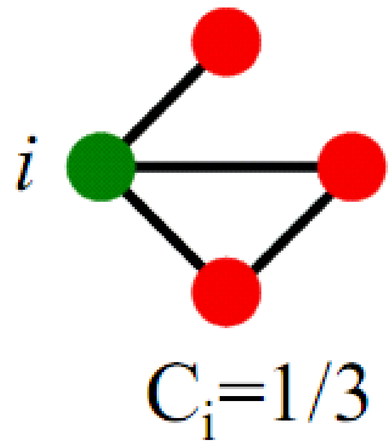
Clustering Coefficient

- Evaluate how the neighbors of a node are connected
- For node i with degree k, assume the number of edges between the neighbors of i is e, the clustering coefficient of i is

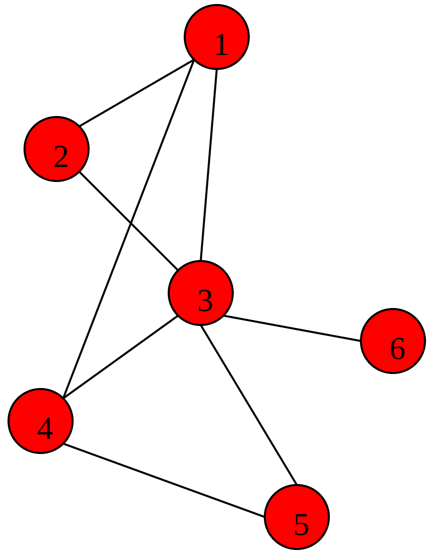
$$C_i = \frac{e}{k(k-1)/2}$$

- Average clustering coefficient

$$C = \frac{1}{N}\sum_{i}^{N} C_i$$



Key Network Properties



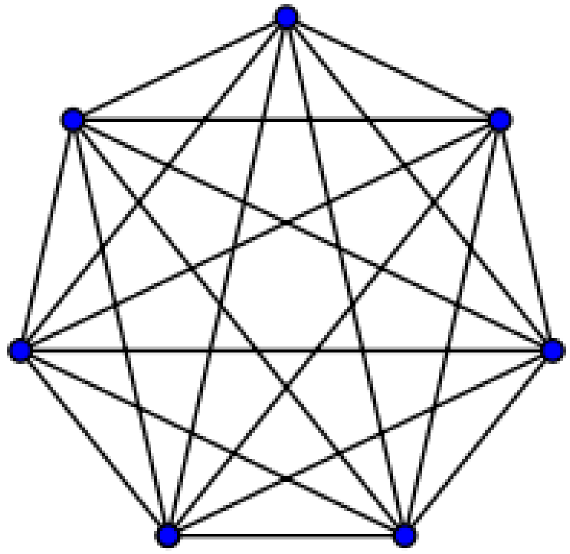
- Degree distribution: P(k)
- Path length: h
- Clustering coefficient: C

- Example:
- P(k):
  - P(1)= ; P(2)= ;P(3)= ; P(4)= ; P(5)= .
- h_max= ; h_avg= ;
- C(i): C(1)= ; C(3)= ; C_avg= .

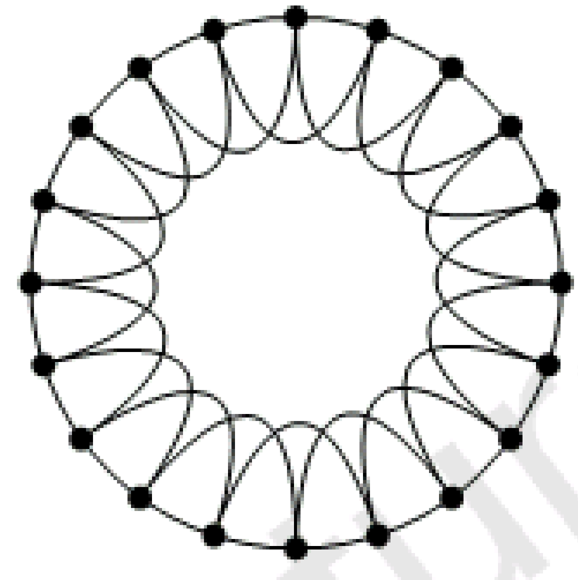
computer NETWORKS

Complete Graph

- Degree distribution: P(k)=N-1
- Path Length:
  - Diameter: 1
  - Average path length: 1
- Clustering coefficient
  - C=1
  - Average clustering coefficient: 1



Regular Lattice



- Degree distribution:

$$P(k) = \begin{cases} 1, & k = 4 \\ 0, & \text{otherwise} \end{cases}$$

- Path length:
  - Diameter: $h_{max} = \frac{N}{4}$
  - Average: for node, its distance to other nodes are: 1, 1, 2, 2, 3, 3, …, N/4, N/4.
  - So

$$h_{avg} = \frac{2 \times (1 + 2 + \cdots + N/4)}{N/2} = \frac{2\frac{(1+N/4)*N/4}{2}}{N/2} = 1/2 + \frac{N}{8}$$

- Clustering coefficient $C_i = \frac{e}{k(k-1)/2}$
  - C=2*3/(4*3)=1/2 for N>6
- Summary: constant degree, constant clustering coefficient, but average path is O(N)

Random Graph

- Degree distribution: Binomial distribution

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

- Average path length: $O(\log n)$

- Clustering coefficient: $C = p = \bar{k}/n$

Other Properties

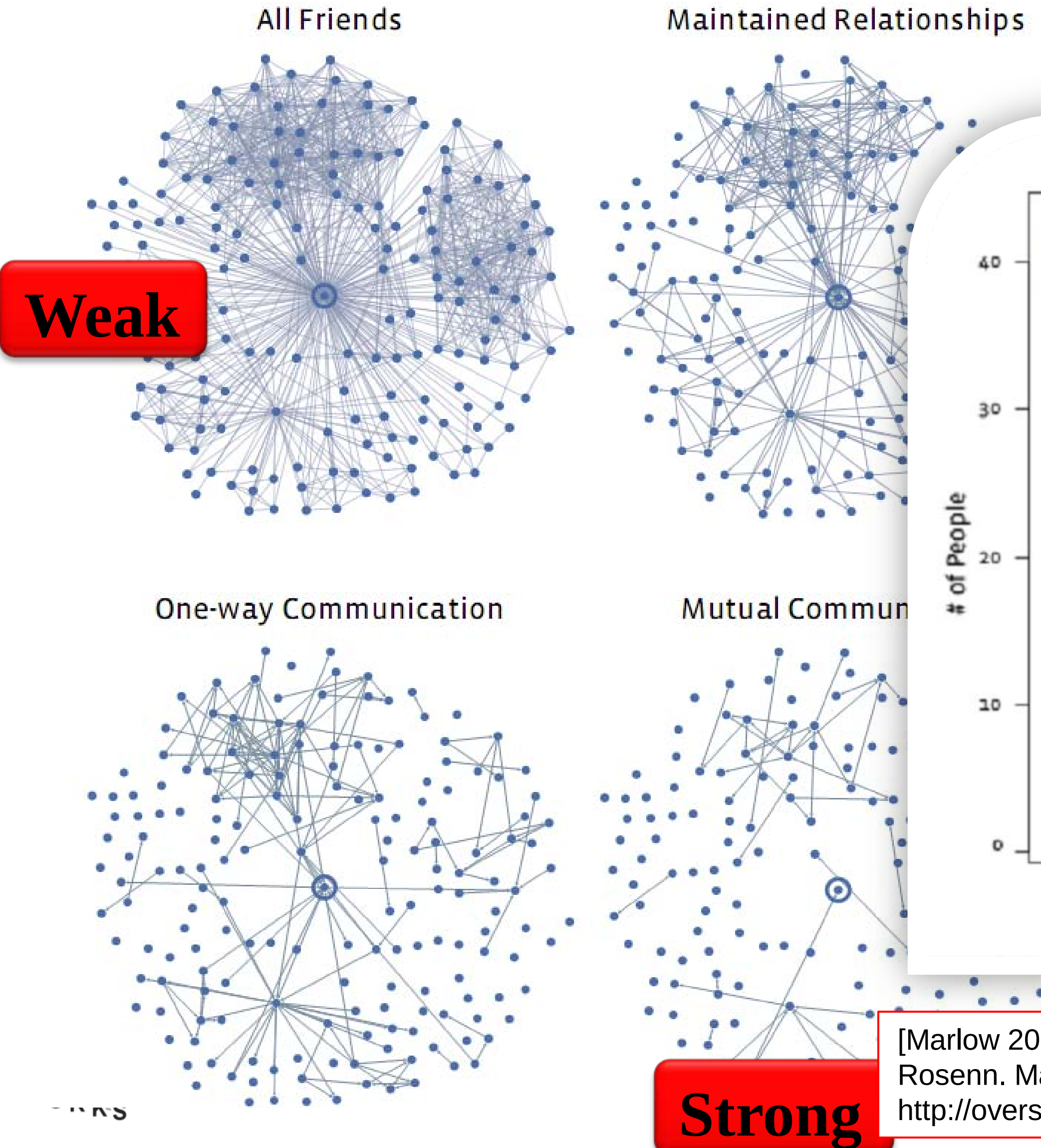
- Strength of social ties
- Centrality

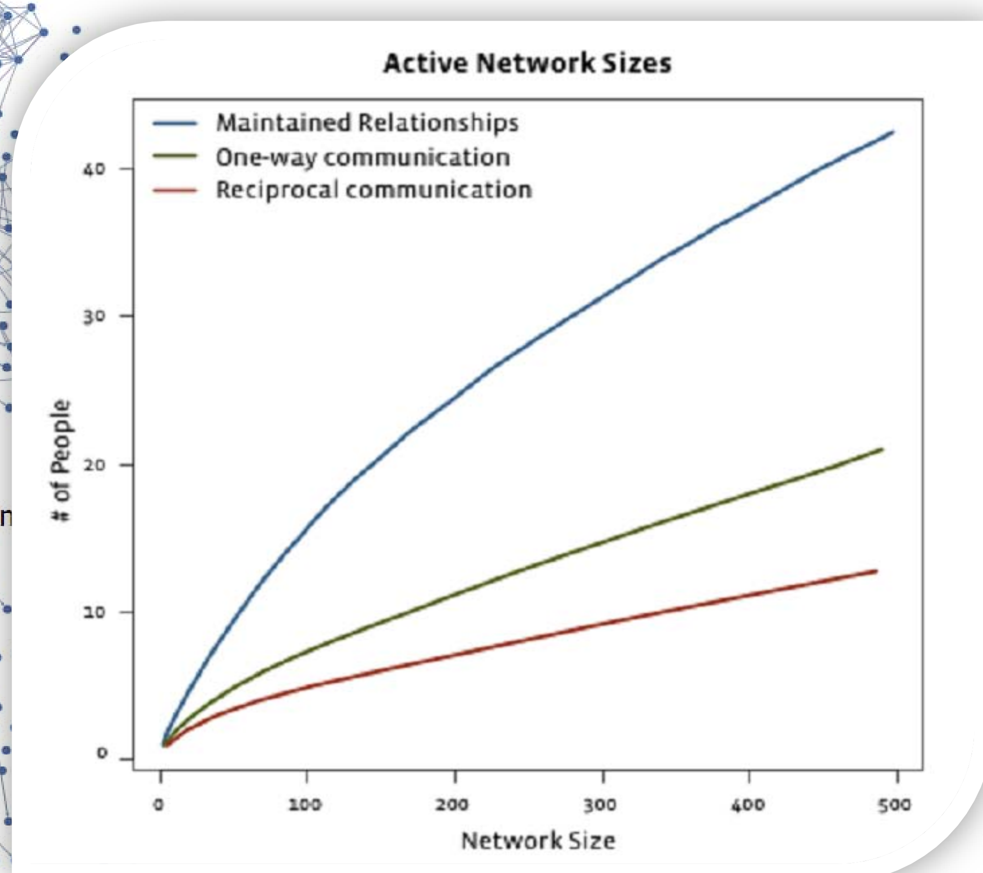
Strength of Social Ties

- Homophily: people love those who are like themselves—“similarity begets friendship”
- Strong ties: the stronger social connections-“friends”
- Weak ties: the weaker social connections-“acquaintances”
- The strength of weak ties [Granovetter 1973]
  - The strength of a tie was measured by the number of times that individuals had interacted in a past year.
    - strong = at least twice a week,
    - medium= less than twice a week but more than once a year
    - weak = once a year or less
  - For whom had found their most recent job through a social contact,
    - 16.7% through a strong tie
    - 55.7 percent through a medium tie
    - 27.6 percent through a weak tie

[Granovetter 1973] Granovetter, M. (1973) The Strength of Weak Ties, American Journal of Sociology, 78, 1360-1380.

Tie Strength on Facebook [Marlow 2009]



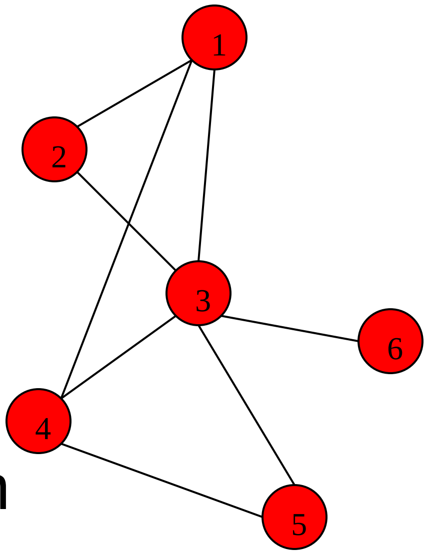


[Marlow 2009] Cameron Marlow, Lee Byron, Tom Lento, and Itamar Rosenn. Maintained relationships on Facebook, 2009. On-line at http://overstated.net/2009/03/09/maintained-relationships-on-facebook.

Centrality

- Measure how important a given node related to the overall network.
- Many different measures of centrality have been developed
  - Degree - how connected a node is
  - Closeness - how easily a node can reach other nodes
  - Betweenness - how important a node is in terms of connecting other nodes
  - Neighbors' characteristics - how important, central, or influential a node's neighbors are

- Degree Centrality: $d_i(g)/(n-1)$



- Closeness Centrality: $(n-1)/\sum_{j \neq i} \ell(i,j)$
  - Where $\ell(i,j)$ is the length of shortest path between i and j

- Betweenness Centrality:

$$Ce_i^B(g) = \sum_{k \neq j : i \notin \{k,j\}} \frac{P_i(kj)/P(kj)}{(n-1)(n-2)/2}$$

# of shortest paths between k and j that i lies on

# of shortest paths between k and j

# of node pairs except i

Attention: Guideline for registration with ITIS

page | discussion | view source | history

navigation
- Main page
- Teaching
- Theses and Projects
- Feedback
- NET homepage
- Recent changes

search

Go Search

toolbox
- What links here
- Related changes
- Special pages
- Printable version
- Permanent link

Advanced Computer Networks (Summer 2013)

Contents [hide]

Details

For ITIS students, please follow these guidelines for exam registration. Put in the module number as exam number and send an extra registration mail to David, indicating that you wish to take the exam until July 4th. If there are any difficulties with the ITIS examination office, please let us know.

Workload/ECTS Credits:	120h, 6 ECTS (old PO), 6 ECTS (new PO), 5 (ITIS)
Module:	M.Inf.1222.Mp, **M.Inf.1223.Mp** (see Networking Study Plan), M.Inf.221.3C1 OR 3.17: Selected Topics in Advanced Networking (ITIS)
Lecturer:	Prof. Xiaoming Fu, Dr. Wenzhong Li
Teaching assistant:	David Koll
Time:	start: 11.04.2013, Thursdays 10-12
Place:	3.101
UniVZ	[1]

Power-law Distribution

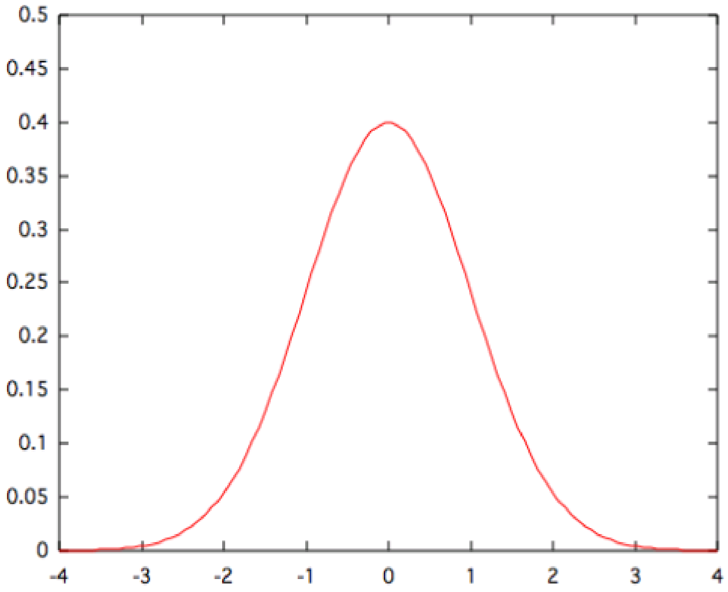
Questions

- How many neighbors does a node have? (degree)
- How far apart are nodes in the network? (distance)
- How close a set of nodes connect with each other? (clustering coefficient)

- Popularity of nodes in social networks
  - Imbalance
  - 20% of web pages receives 80% visits
  - Celebrities in Twitter have millions of fans
  - A few rich people own a large amount of wealth
- Node degree distribution
  - What fraction of all nodes have degree k? P(k)=?
  - Normal distribution? – for random graph
    P(k)=exponential function of -k

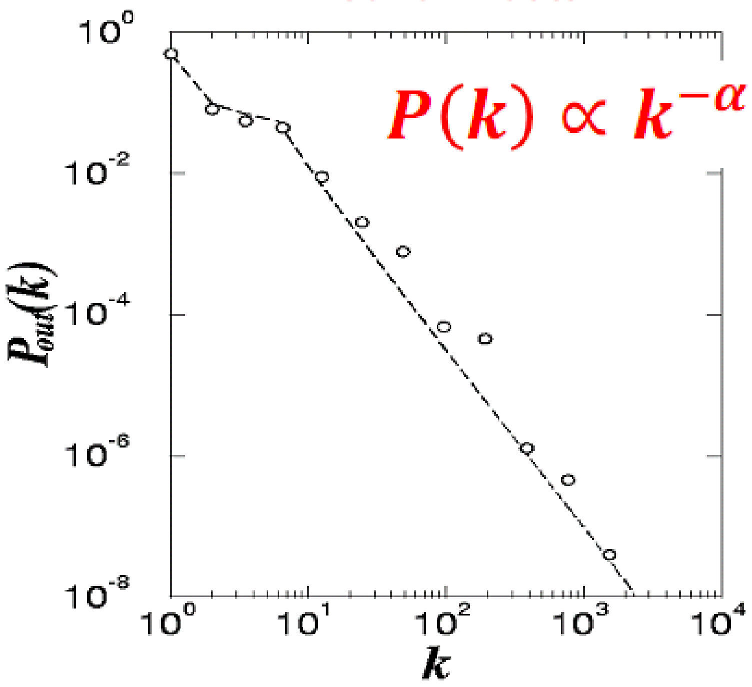
Guess: Normal distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Real network: Power-law

Found in data



- Phenomenon: popularity seems to exhibit extreme imbalances
- Observations:
  - The fraction of telephone numbers that receive k calls per day is roughly proportional to 1/k^2;
  - The fraction of books that are bought by k people is roughly proportional to 1/k^3;
  - The fraction of scientific papers that receive k citations in total is roughly proportional to 1/k^3
  - => The number of monthly downloads for each song at a large on-line music site is proportional to 1/k^c for some constant c

- Power-law distribution
  - Let f(k) be the fraction of items have value k

$$f(k) = zk^{-\alpha}$$

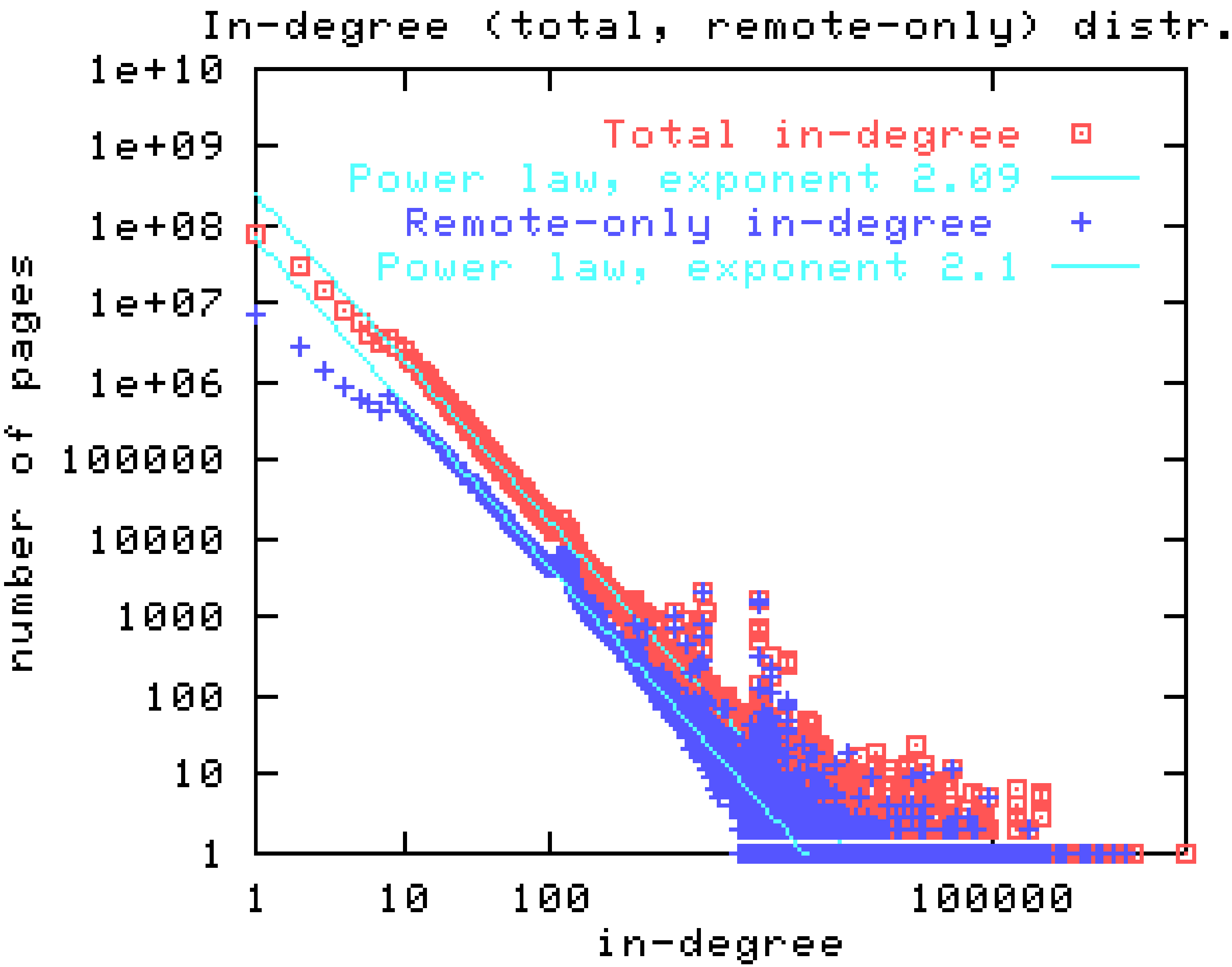
where α and z are constants

- Taking logarithms of both sides

$$log f(k) = log z - \alpha log k$$

- Testing for power-law distribution
  - If we draw k and f(k) in “log-log” scale, it shows a straight line

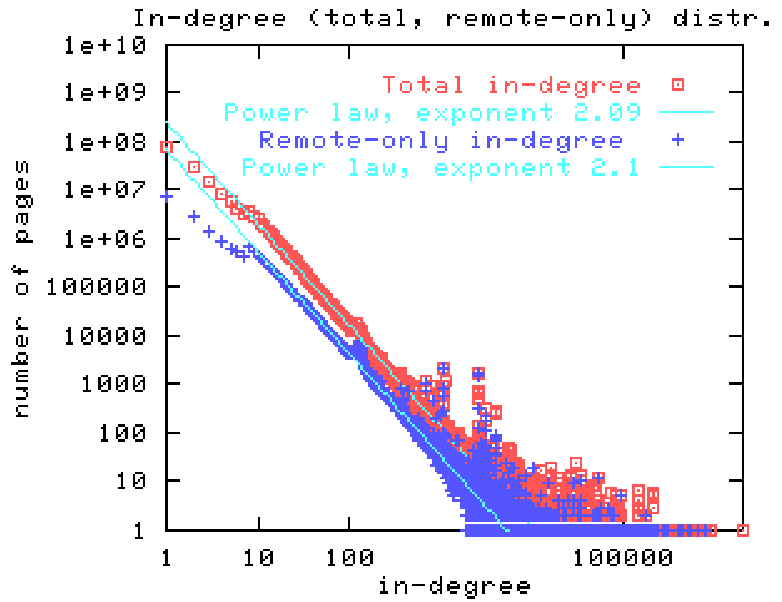
Node Degree of Websites



Estimating power-Law Exponent

- Estimating α ? $$log f(k) = log z - \alpha log k$$
- Simple method: fit a line on log-log axis

$$\min_{\alpha}(\log(y) - \alpha \log(x))^2$$



- For power-law distribution $f(k) = zk^{-\alpha}$
- Estimating the normalizing constant

$$P(x) = z\, x^{-\alpha} \qquad z=?$$

- According to definition

$P(x)$ is a distribution: $\int P(x)dx = 1$

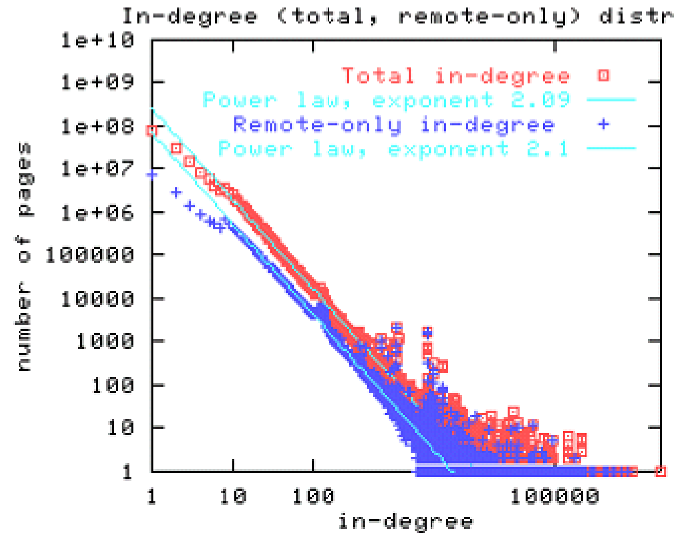
- Thus, let

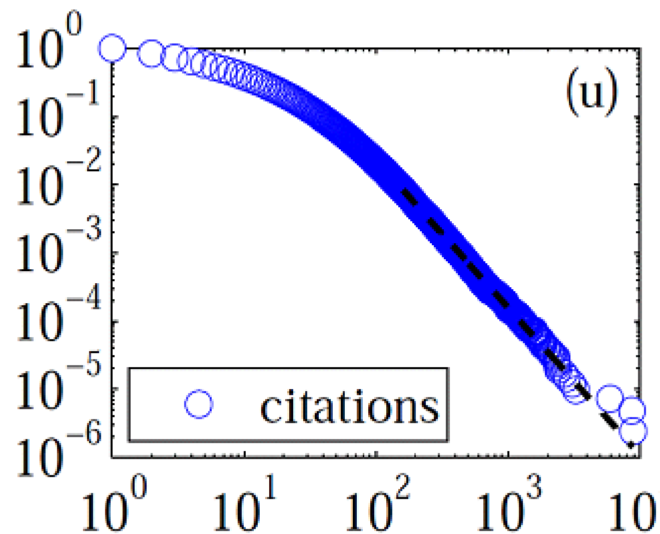
$$1 = \int_{x_{min}}^{\infty} P(x)dx = z \int_{x_m}^{\infty} x^{-\alpha} dx$$

- z can be obtained by solving the above equation

$$f(k) = zk^{-\alpha}$$

- **Power-law degree exponent is typically $2 < \alpha < 3$**
  - Web graph:
    - $\alpha_{in} = 2.1$, $\alpha_{out} = 2.4$ [Broder et al. 00]
  - Autonomous systems:
    - $\alpha = 2.4$ [Faloutsos³, 99]
  - Actor-collaborations:
    - $\alpha = 2.3$ [Barabasi-Albert 00]
  - Citations to papers:
    - $\alpha \approx 3$ [Redner 98]
  - Online social networks:
    - $\alpha \approx 2$ [Leskovec et al. 07]





computer NETWORKS

Power-laws Are Everywhere

- Pareto, 1897 – Wealth distribution
- Lotka, Alfred J. 1926 - The frequency of publications by authors in a given field
- Zipf 1940s –Word frequency
- Simon 1950s – City populations
- Holger Ebel 2002- Email Network
- Who-talks-to-whom network
- Number of friends in Facebook
- …

The Small-World Phenomenon

Questions

- How many neighbors does a node have? (degree)
- How far apart are nodes in the network? (distance)
- How close a set of nodes connect with each other? (clustering coefficient)

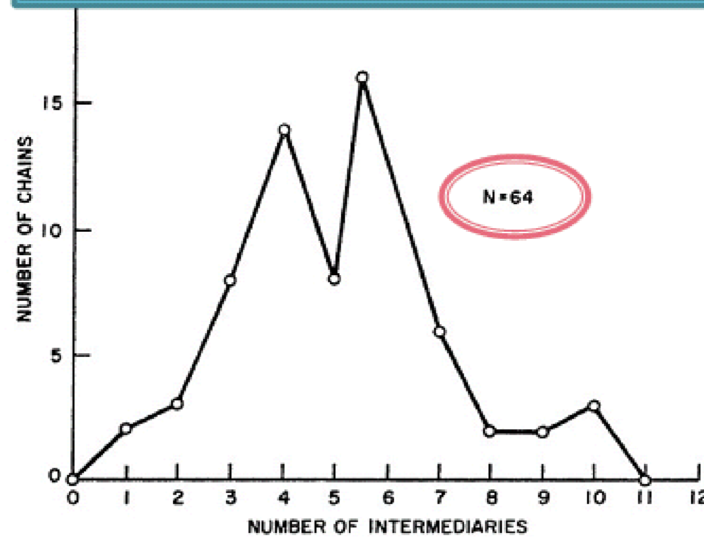
Six Degrees of Separation

- What is the typical shortest length between any two people in human society?
  - Global measurement is impossible
  - Sampling
- Experiment
  - Milgram 1967
  - Idea: ask randomly chosen “starter” individuals to try to forward a letter to a designated “target” person

Milgram's Experiment [1967]

- Procedure
  - The target person
    - A stockbroker who worked in Boston and lived in Sharon, Massachusetts
  - The starting person
    - Randomly picked 300 people in Omaha, Nebraska and Wichita, Kansas
  - The target's name, address, occupation, and some personal information are provided
  - Rules: "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 … it must be someone you know on a first-name basis".
  - The names of the person who forward the letter are attached

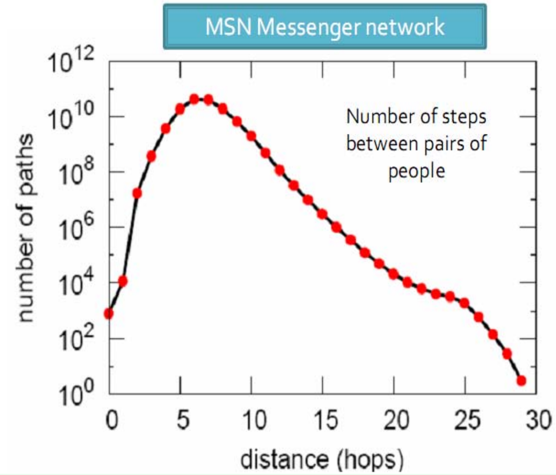
Milgram's small world experiment



- How many steps did it take?
  - 64 letters reached the target
  - It took 6.2 steps on average
- Short paths exist! -- Six Degrees of Separation
- Similar results are verified in other social networks like actor network, email network, who-talks-to-whom network (MSN), Facebook …

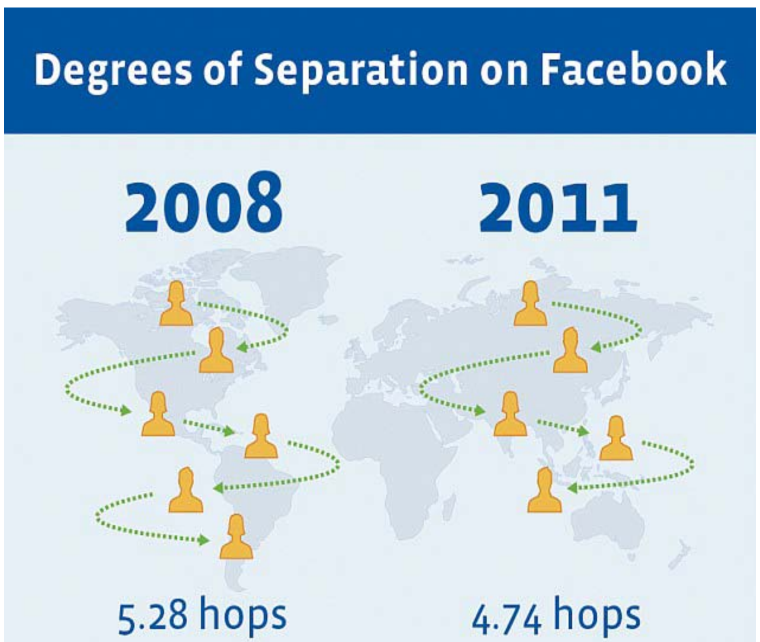
More Results

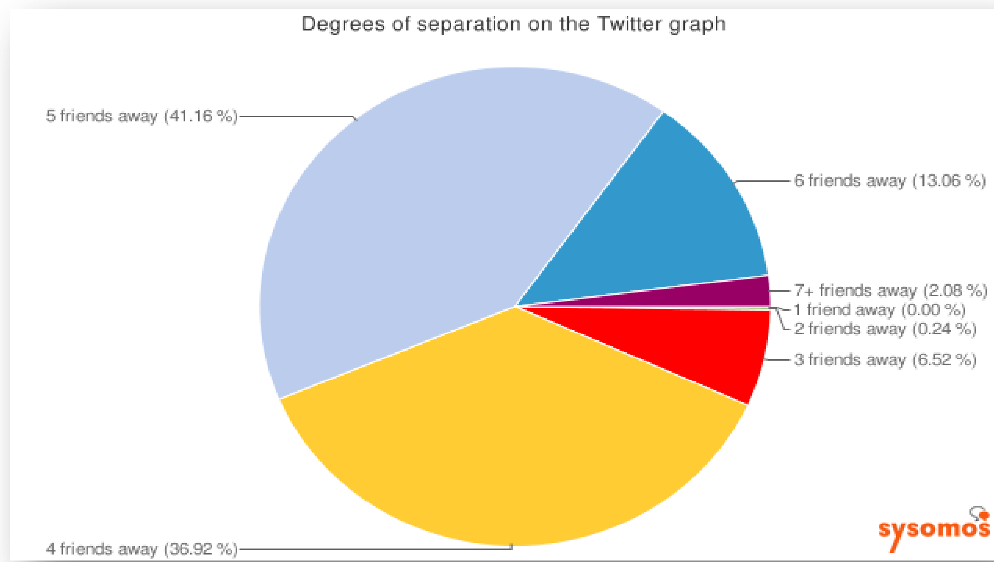
- Microsoft MSN (2006): 6.6 hops
- Facebook(2011): 4.74 hops
- Twitter (2010): 4.67 hops



Avg. path length **6.6**
90% of the people can be reached in < 8 hops

[Leskovec 2008] Jure Leskovec and Eric Horvitz. Worldwide buzz: Planetary-scale views on an instant-messaging network. In Proc. 17th International World Wide Web Conference, 2008.



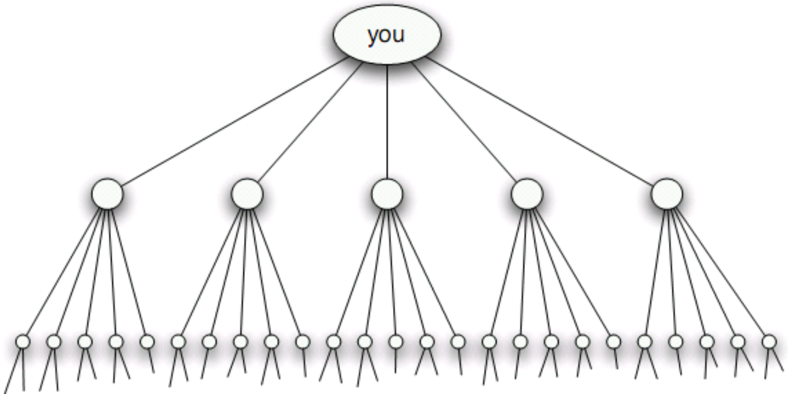


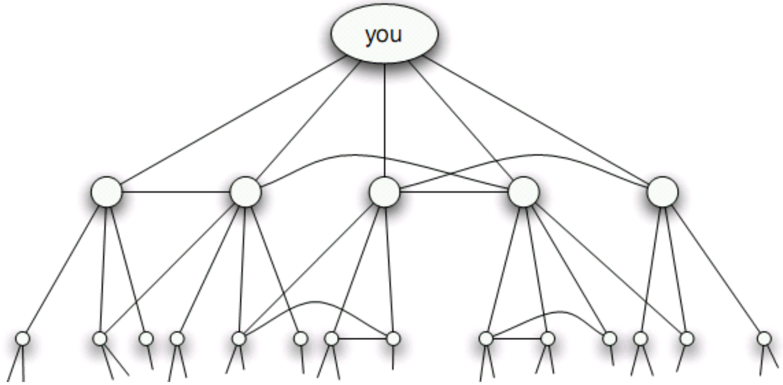
[Barnett 2012] Emma Barnett. “Facebook cuts six degree of seperation to four”. Telegraph. Retrieved 7 May 2012.

[Cheng 2010] Alex Cheng. ” Six Degrees of Separation, Twitter Style”. http://www.sysomos.com/insidetwitter/sixdegrees/, 2010

A Simple Explanation

- Suppose each person knows 100 other people on a first-name basis
  - Step 1: reach 100 people
  - Step 2: reach 100*100 people
  - …
  - Step 5: reach 100^5= 10 billion people
  - Ref: the world population is 7.019 billion (Wiki, 2012)
- The numbers are growing by powers of 100
- But it is not true for real network!!!
  - Triadic relationships are common
  - Social network is highly clustered, not the kind of massively branching structure.





Regular Network vs Small-World Network vs Random Network

- Regular network: high clustering, high diameter
- Random network: low clustering, low diameter
- Question
  - Is there a network inbetween the regular network and random network, with high clustering coefficient and low average path length?
- Small-World network: high clustering, low diameter

Small-world Network

- A small-world network is a type of mathematical graph in which most nodes are not neighbors of one another, but most nodes can be reached from every other by a small number of hops. [wiki]
- Properties: high clustering, low diameter
  - Clustering efficient: much larger than random network
  - Diameter: almost equal to random network
- Most of social networks are found to be small-world network.

Summary

<table>
<tr><th>Properties</th><th>Regular Network</th><th>Random Network</th><th colspan="2">Social Network</th></tr>
<tr><td>Degree Distribution</td><td>Constant</td><td>Normal distribution</td><td colspan="2">Power-law distribution</td></tr>
<tr><td>Path Length (Diameter)</td><td>High</td><td>Low</td><td>Low</td><td rowspan="2">Small-world</td></tr>
<tr><td>Clustering Coefficient</td><td>High</td><td>Low</td><td>High</td></tr>
</table>