Dynamic and Social Network Analysis

Lecture 5

Miray Kas
Bilkent University
Computer Engineering Department

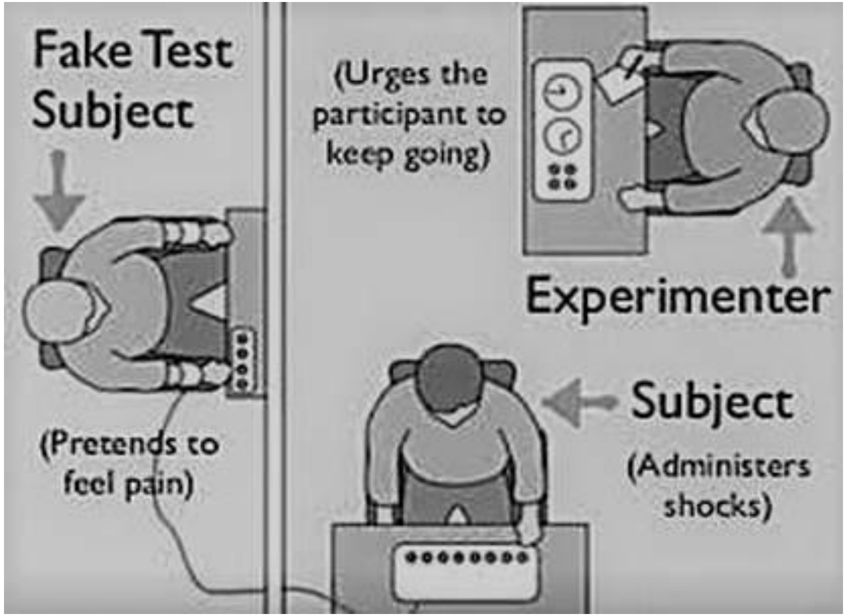
Small World Phenomenon

Stanley Milgram (1933-1984)

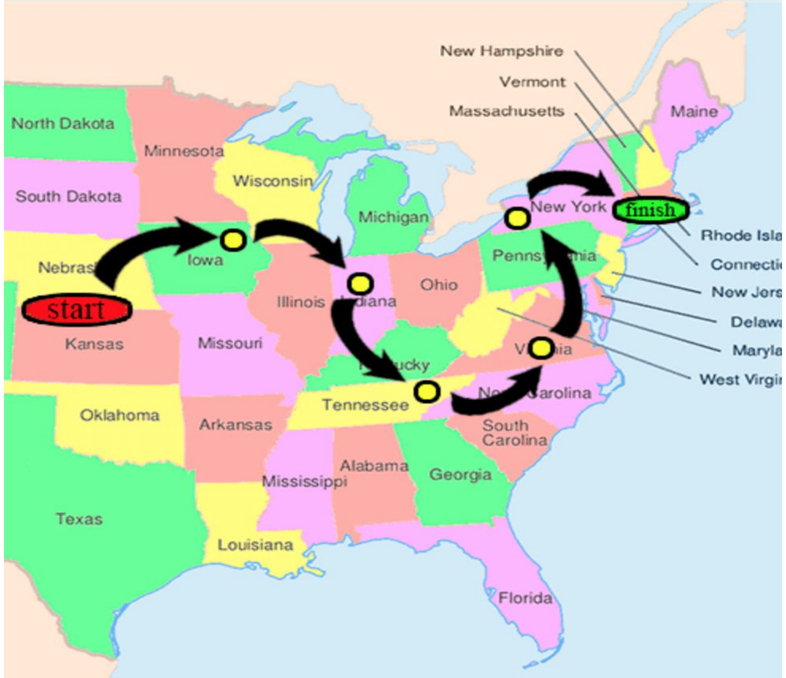
- American Social Psychologist
- Born to a Jewish family in New York
  - His parents moved to the US after World War I.
- His extended family affected by the holocaust
  - Relatives that survived Holocaust moved to their house
- Famous for his experiments
  - Likely denied tenure due to his Obedience to Authority experiment
- Very influential, one of the key psychologists of 20th century

Obedience to Authority (1963)

- Participants thought they were giving shocks to the learner
- Most of them went to the highest level of shock!



Milgram's Chain Letter Experiment (1967)



A target in Boston, MA

- Starters in Omaha, Nebraska
- Name, address & some personal info

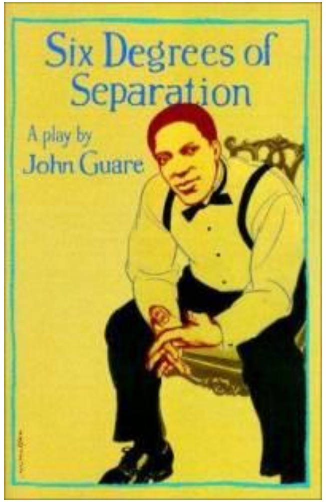
Pass down the letter to a personal acquaintance until the target is reached.

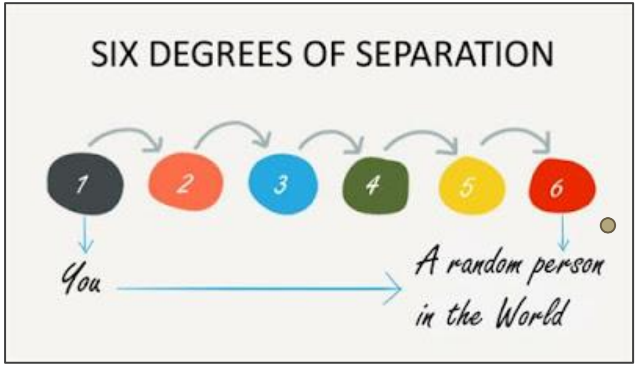
Conclusions from Milgram's Chain Letter Experiment

- Very short paths between arbitrary pairs of nodes exist
  - The average complete chain length around **6.5**
  - Later called ***Six-Degrees of Separation***
- Most correspondence chains were incomplete
  - **Case-1:** 232 of the 296 letters never reached the target.
  - **Case-2:** Only 24 letters out of 160 reached the target.
- Most of the required data was missing
- Individuals using purely local information are able to find the targets somehow.

Six Degrees of Separation

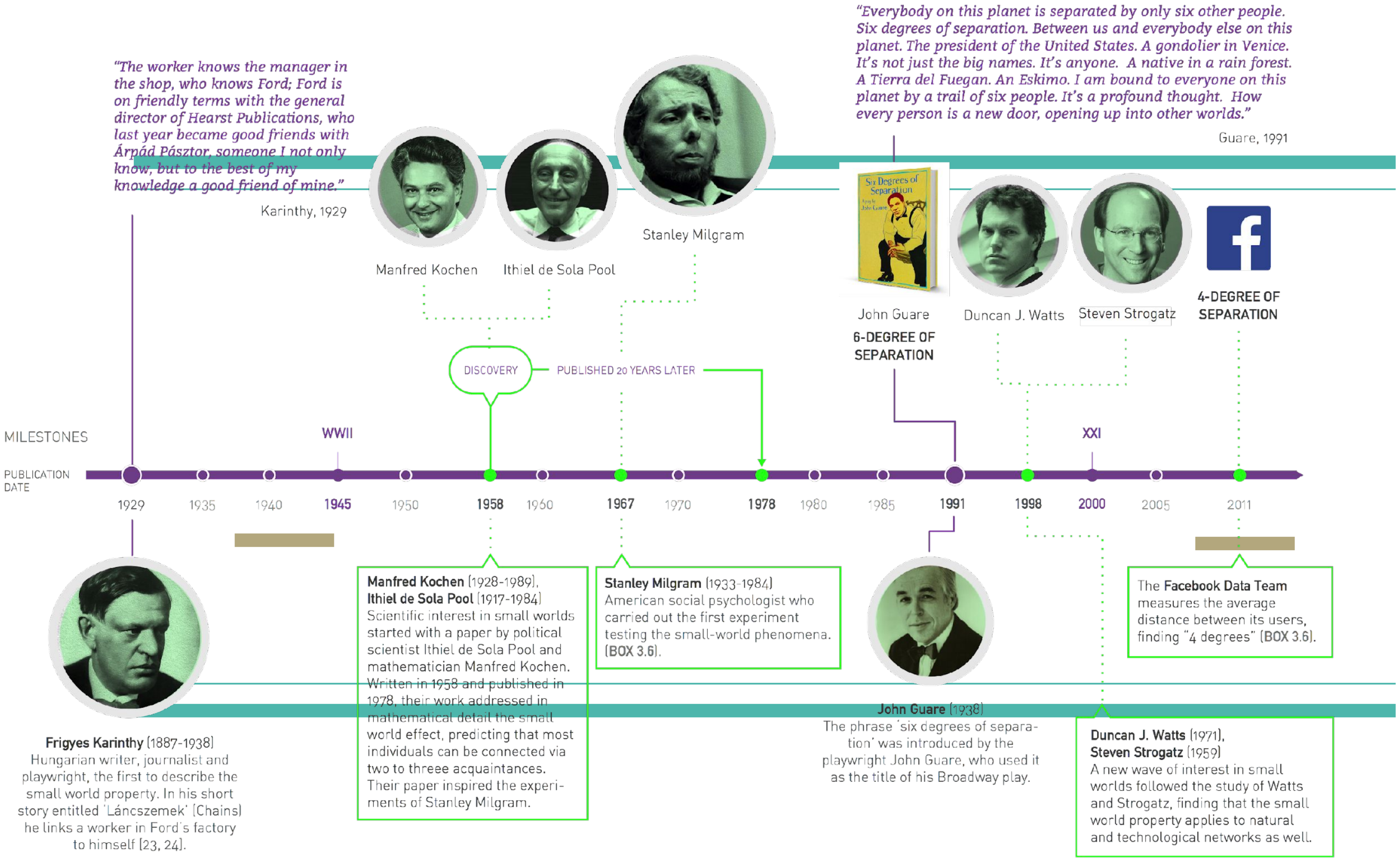
All people on average are 6 hops away from one another





How is it even possible?

The phrase was invented by the playwright John Guare, who used it as the title of his Broadway play in 1991. It was later turned into a movie with the same title.


"The worker knows the manager in the shop, who knows Ford; Ford is on friendly terms with the general director of Hearst Publications, who last year became good friends with Árpád Pásztor, someone I not only know, but to the best of my knowledge a good friend of mine."
Karinthy, 1929
Manfred Kochen
Ithiel de Sola Pool
Stanley Milgram
"Everybody on this planet is separated by only six other people. Six degrees of separation. Between us and everybody else on this planet. The president of the United States. A gondolier in Venice. It's not just the big names. It's anyone. A native in a rain forest. A Tierra del Fuegan. An Eskimo. I am bound to everyone on this planet by a trail of six people. It's a profound thought. How every person is a new door, opening up into other worlds."
Guare, 1991
Six Degrees of Separation
John Guare
6-DEGREE OF SEPARATION
Duncan J. Watts
Steven Strogatz
4-DEGREE OF SEPARATION
DISCOVERY
PUBLISHED 20 YEARS LATER
MILESTONES
WWII
XXI
PUBLICATION DATE
1929
1935
1940
1945
1950
1958
1960
1967
1970
1978
1980
1985
1991
1998
2000
2005
2011
Frigyes Karinthy (1887-1938)
Hungarian writer, journalist and playwright, the first to describe the small world property. In his short story entitled 'Láncszemek' (Chains) he links a worker in Ford's factory to himself [23, 24].
Manfred Kochen (1928-1989), **Ithiel de Sola Pool** (1917-1984)
Scientific interest in small worlds started with a paper by political scientist Ithiel de Sola Pool and mathematician Manfred Kochen. Written in 1958 and published in 1978, their work addressed in mathematical detail the small world effect, predicting that most individuals can be connected via two to threee acquaintances. Their paper inspired the experiments of Stanley Milgram.
Stanley Milgram (1933-1984)
American social psychologist who carried out the first experiment testing the small-world phenomena. (BOX 3.6).
John Guare (1938)
The phrase 'six degrees of separation' was introduced by the playwright John Guare, who used it as the title of his Broadway play.
The **Facebook Data Team** measures the average distance between its users, finding "4 degrees" (BOX 3.6).
Duncan J. Watts (1971), **Steven Strogatz** (1959)
A new wave of interest in small worlds followed the study of Watts and Strogatz, finding that the small world property applies to natural and technological networks as well.

Chain Attrition in “Small World” Networks

(Discusses “*Social Search in “Small-World” Experiments*” by By Sharad Goel, Roby Muhamad, and Duncan Watts)

Why are the chains incomplete?

Chain Attrition

- Remember in Milgram's experiment only 20% of the letters made it to the target.

- White,1970
  - Lack of motivation, unrelated to the process/topology

- Dodds, 2003
  - Removed last step attrition assumption

Why are the chains incomplete?

Chain Attrition

- White, 1
  - Lack /topology

Implicitly assumes **'homogeneity'** among individuals

- Dodds,
  - Removed last step attrition assumption

General Features of Chain Attrition

- Reflects variability among individuals.
  - Not merely a function of unrelated extrinsic factors.
- Should take into account heterogeneity of individuals
- Highly related with the 'Social Capital'.
  - Higher 'Social Capital', higher continuance properties.

Modern Imitation of Milgram's Experiment

Experiment Setup

- Email chains
- Mostly from the US and Western Europe
- Predominantly white and Christian
- Largely young, college educated, middle-class professionals

Results

- Typical small world results
- Short chain lengths
- Low completion rates

Understanding the Participants

- 66 parameters to model heterogeneity among individuals and their social capital
- Gender and Race
- Others in 9 different categories

Category
Target
Age
Relationship Origin
Income
Work Position
Work Field
Reason for Choosing Recipient
Relationship Strength
Education Level

Social Capital & Contributing to the Experiment

Attributes	Attributes
Age Under 17 18-29 30-39 40-49 50-59 Above 60	**Income** More than \$100,000 \$50,000 - \$100,000 \$25,000 - \$49,999 \$2,000 - \$24,999 Less than \$2000
Education Level Graduate school College/University High school Elementary school	**Relationship Strength** Extremely close Very close Fairly close Casually Not close
Work Field Media/Advertising/Arts Education/Science IT/Telecommunication Government Other	**Reason for Choosing Recipient** Profession Education Work brings contact Geography Other
Work Position Specialist/Technical Student Other Unemployed/Retired Executive/Manager	**Relationship Origin** Work School Internet Mutual friend Relative Other

- Higher social capital
  - Higher continuance
- Graduate education
  - 4% more likely to pass
- Earning over \$100K,
  - 2% more likely to pass
- Earning less than \$25K
  - 1% less likely to pass

Social Capital & Contributing to the Experiment

Attributes
Age
Under 17
18-29
30-39
40-49
50-59
Above 60
Education Level
Graduate school
College/University
High school
Elementary school
Work Field
Media/Advertising/Arts
Education/Science
IT/Telecommunication
Government
Other
Work Position
Specialist/Technical
Student
Other
Unemployed/Retired
Executive/Manager

Attributes	
Income	
More than $100,000	
$50,000 - $100,000	
$25,000 - $49,999	
$2,000 - $24,999	
Less	
Rel	
Extr	
Very	
Fair	
Casu	
Not	
Rea	
Rec	
Prof	
Education	
Work brings contact	
Geography	
Other	
Relationship Origin	
Work	Relative
School	Other
Internet	
Mutual friend	

Homogeneity assumption is invalid

- Higher social capital
  - Higher continuance
- e education
  - ore likely to pass
- over $100K,
  - 2% more likely to pass
- Earning less than $25K
  - 1% less likely to pass

Conclusions

- Chain attrition is a function of social capital.
  - Heterogeneous across individuals
- Chain attrition is a typical problem of small-world experiments.
- Chain attrition changes the inferences made about a certain network.
  - The average path length is larger if chain attrition is taken into account.
- Chain attrition introduces ***missing data*** which should be taken into account.
  - If Milgram's experiments considered chain attrition, the average path length would be 8, not 6.

Measuring Small World Phenomenon

- ***Diameter***
  - Have low diameters *
  - In real networks, the average distance between nodes rises very slowly.
- ***Transitivity***
  - Have high transitivity
  - A friend of a friend tends to be your friend
  - Have high clustering coefficient *
- ***Degree distribution***
  - Real networks’ degree distribution follow power law

Measuring Small World Phenomenon

Network	size	av. shortest path	Shortest path in fitted random graph	Clustering (averaged over vertices)	Clustering in random graph
Film actors	225,226	3.65	2.99	0.79	0.00027
MEDLINE co-authorship	1,520,251	4.6	4.91	0.56	1.8×10^{-4}
E.Coli substrate graph	282	2.9	3.04	0.32	0.026
C.Elegans	282	2.65	2.25	0.28	0.05

Comparison with random graph is used to determine whether a real network is a small world network

Can we mimic Small-World phenomenon in simulation?

Network Models

Random Networks & Mimicking Real Networks in Simulation

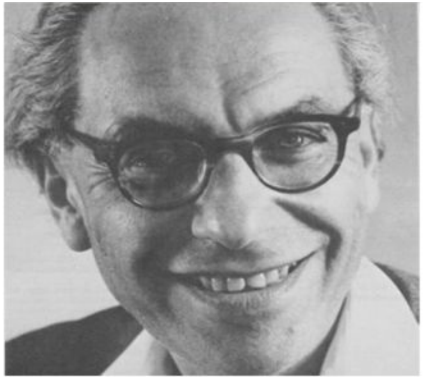
This section is partially adapted from Barabasi Network Science lecture slides.

Important Network Models

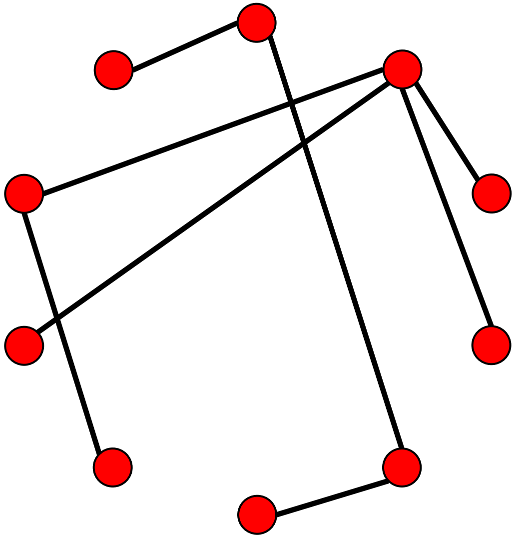
- Erdos - Renyi model (ER model)
- Watts - Strogatz model
- Barabasi - Albert model

Erdos - Renyi Random Network Model

Pál Erdös
(1913-1996)



Alfréd Rényi
(1921-1970)



Erdös-Rényi model (1960)

Connect with probability p

p=1/6 N=10

<k> ~ 1.5

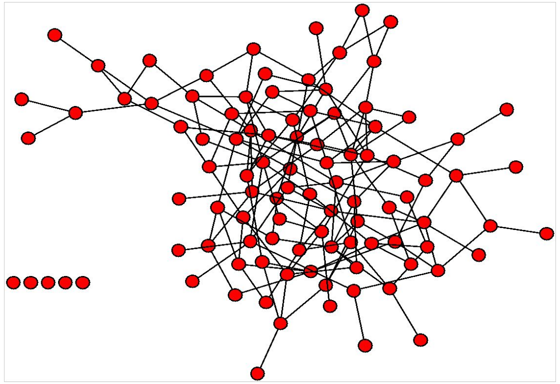
Erdos - Renyi Random Network Model

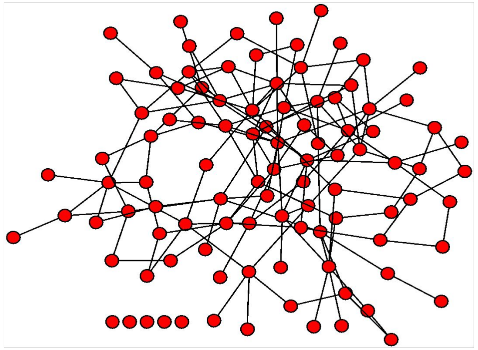
- The simplest, most well-studied and famous random graph model (Gilbert, 1959; Erdős & Rényi, 1961).
  - References:
    - Gilbert, E. N. (1959). Random graphs. *Ann. Math. Statist.*, *30*, 1141–1144.
    - Erdős, P., & Rényi, A. (1961). On the evolution of random graphs. *Bull. Inst. Internat. Statist.*, *38*, 343–347.

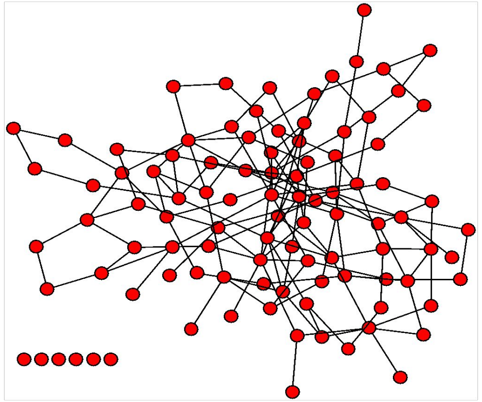
- Each possible edge appears independently and with identical probability *p*.

Erdos - Renyi Random Network Model

p=0.03   N=100





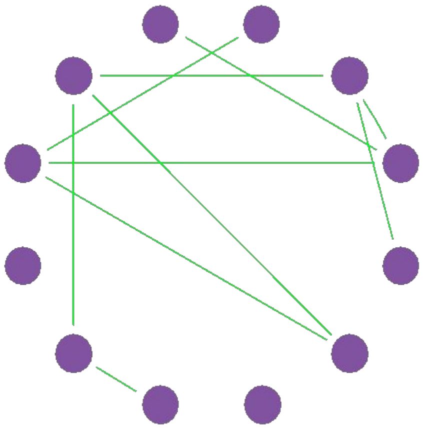


Erdos - Renyi Random Network Model

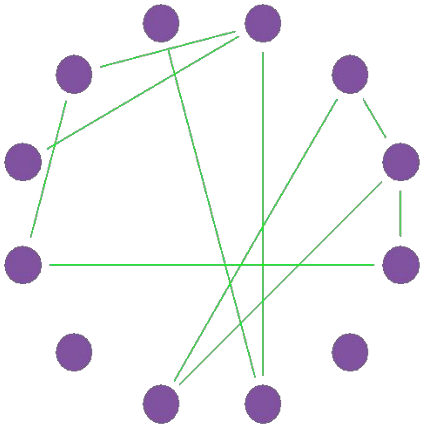
The number of links is variable!

Erdos - Renyi Random Network Model

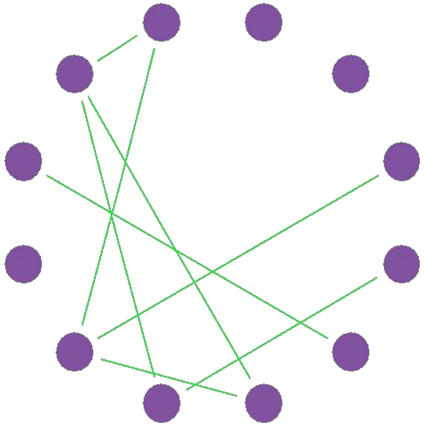
p=1/6   N=12



L = 8



L = 10



L = 7

Number of Links in a Network

P(L): the probability to have exactly *L* links in a network of *N* nodes & probability *p*

The maximum number of links in a network of N nodes.

$$P(L) = \begin{pmatrix} \begin{pmatrix} N \\ 2 \end{pmatrix} \\ L \end{pmatrix} p^{L} (1-p)^{\frac{N(N-1)}{2} - L}$$

Number of different ways we can choose L links among all potential links.

Binomial distribution...

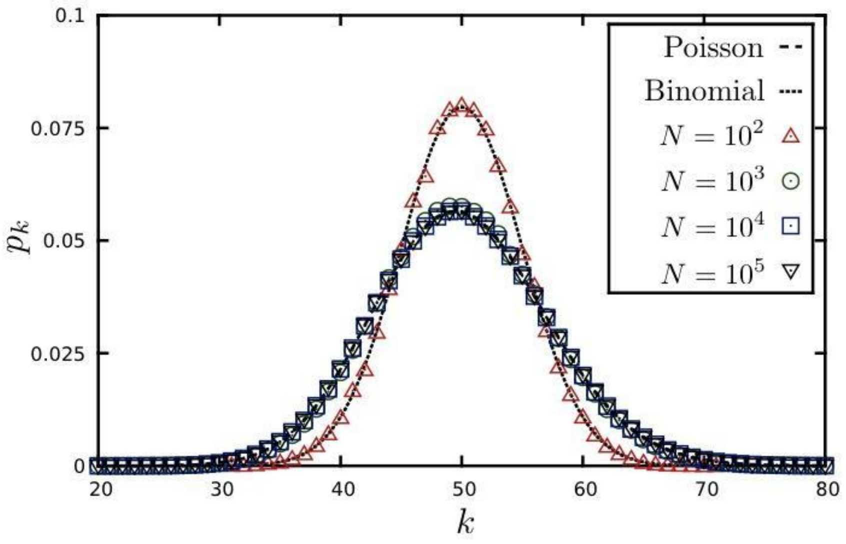
Degree Distribution of a Random Network

$$P(k) = e^{-<k>} \frac{<k>^k}{k!}$$

- ***N*** = number of nodes
- ***k*** = degree
- ***P^k*** = probability of having degree *k*

- As the network size increases, the distribution becomes increasingly **narrow** - we are increasingly confident that the degree of a node is in the vicinity of <*k*>.

- For large *N*, small *k*, we arrive at **Poisson Distribution**

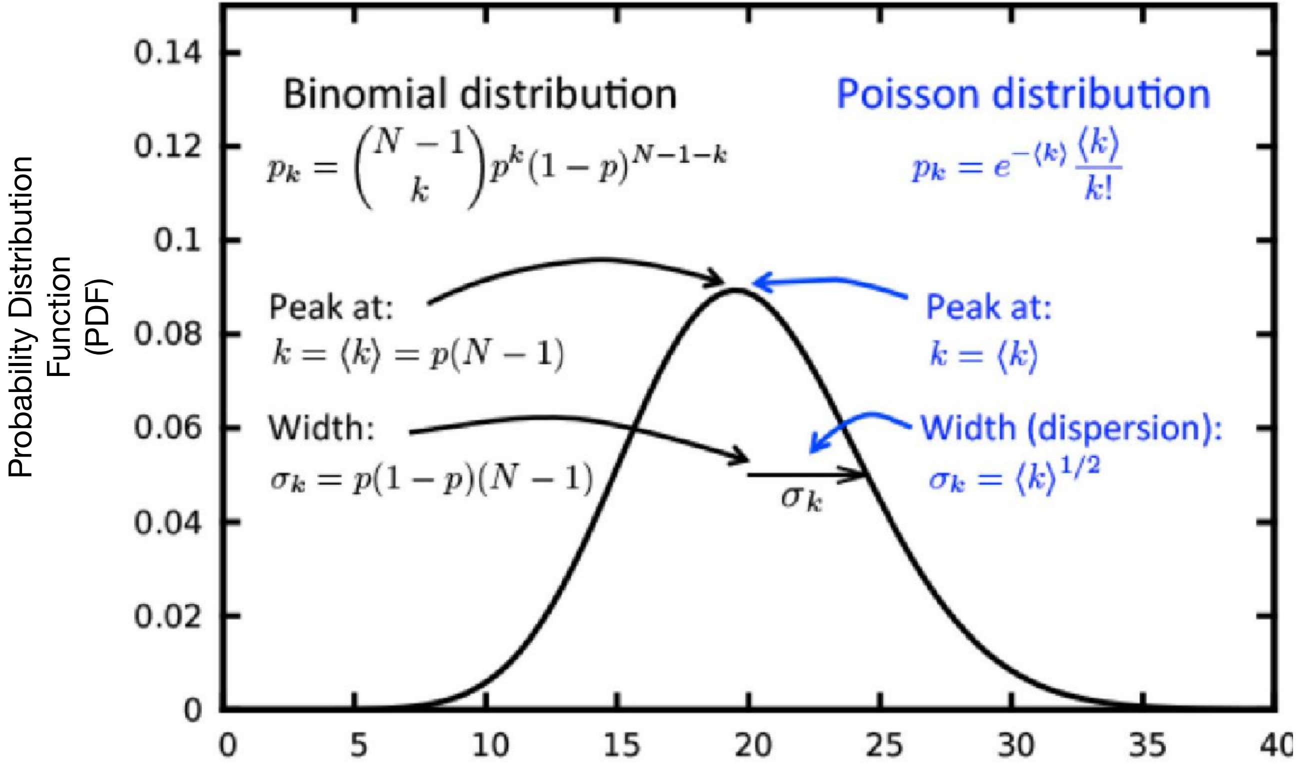


<*k*> = 50

Degree Distribution of a Random Network

Exact Result
-binomial

Large N limit
-Poisson



Let's Face the Reality!

Real Networks are not random!

Real Networks are not Poisson!

In real networks the degrees vary far more widely than predicted by random network theory

Real Networks do not resemble Erdos-Renyi!

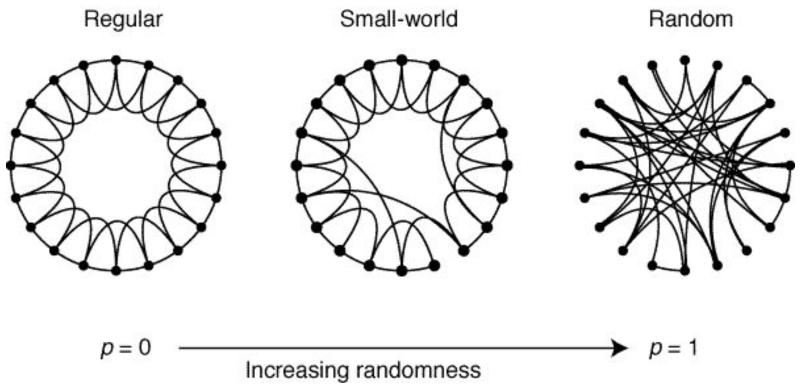
Watts and Strogatz Small World Graph Model

- Published in Nature in 1998
    - Watts, D., Strogatz, S. Collective dynamics of ‘small-world’ networks. *Nature* 393, 440–442 (1998). https://doi.org/10.1038/30918

- Reconciles two ideas:
    - Small Average Shortest Paths
    - High Clustering (friends of friends tend to be friends)

Watts and Strogatz Proposed Approach

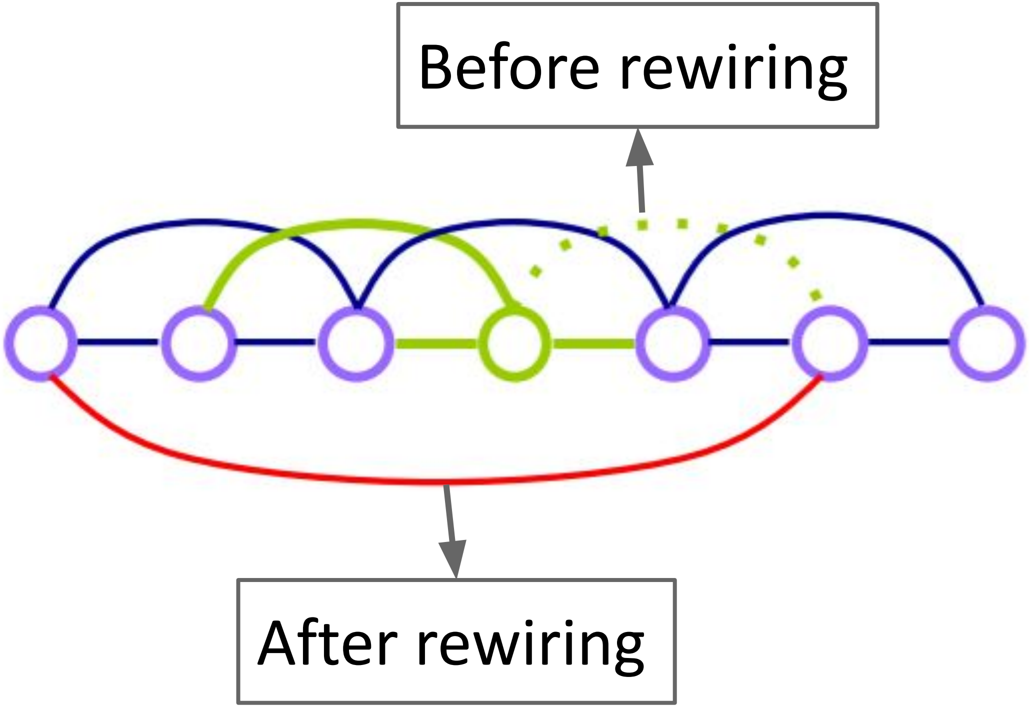
- Cross over from regular lattices to random



- Parametrized and Tunable (Probability *p* of **rewiring**)

- Achieves a small world graph with
  - Small average shortest paths
  - High Clustering (which is not obeyed by random graphs)

A Closer Look at Rewiring Idea



This process Introduces short cuts in the lattice

Image from http://snap.stanford.edu/class/cs224w-2015/slides/05-smallworlds.pdf

Generating Watts & Strogatz Small World Graph

Two ways of constructing

a)

Select a fraction ***p*** of links and reposition their end points

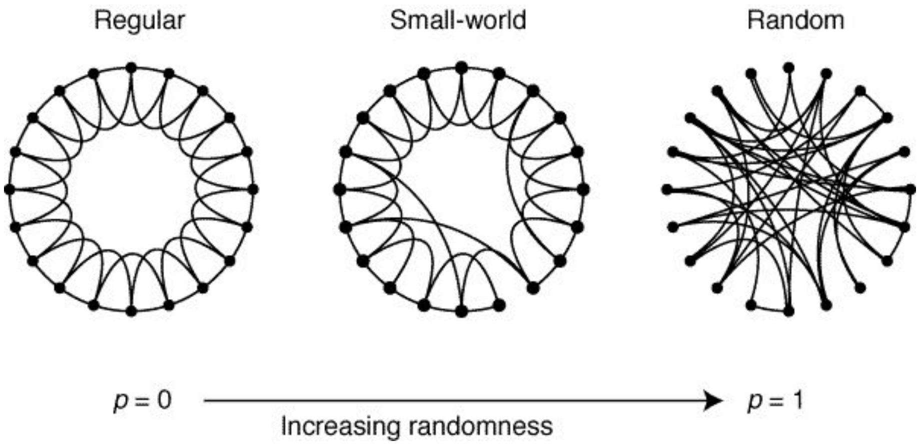
b)

Add a fraction ***p*** of links on top of existing lattice

Common: Disallow self-links. Disallow multiple links

Watts and Strogatz Original Model

- *p* is between 0 and 1
  - *p = 0 (Regular Lattice)*
  - *p = 1 (Random Graph)*



- Parameter p (probability of rewiring a link) is tunable
- Each node has *K >= 4* neighbors

- **Somewhere between 0 and 1 for *p*, you get a small-world.**

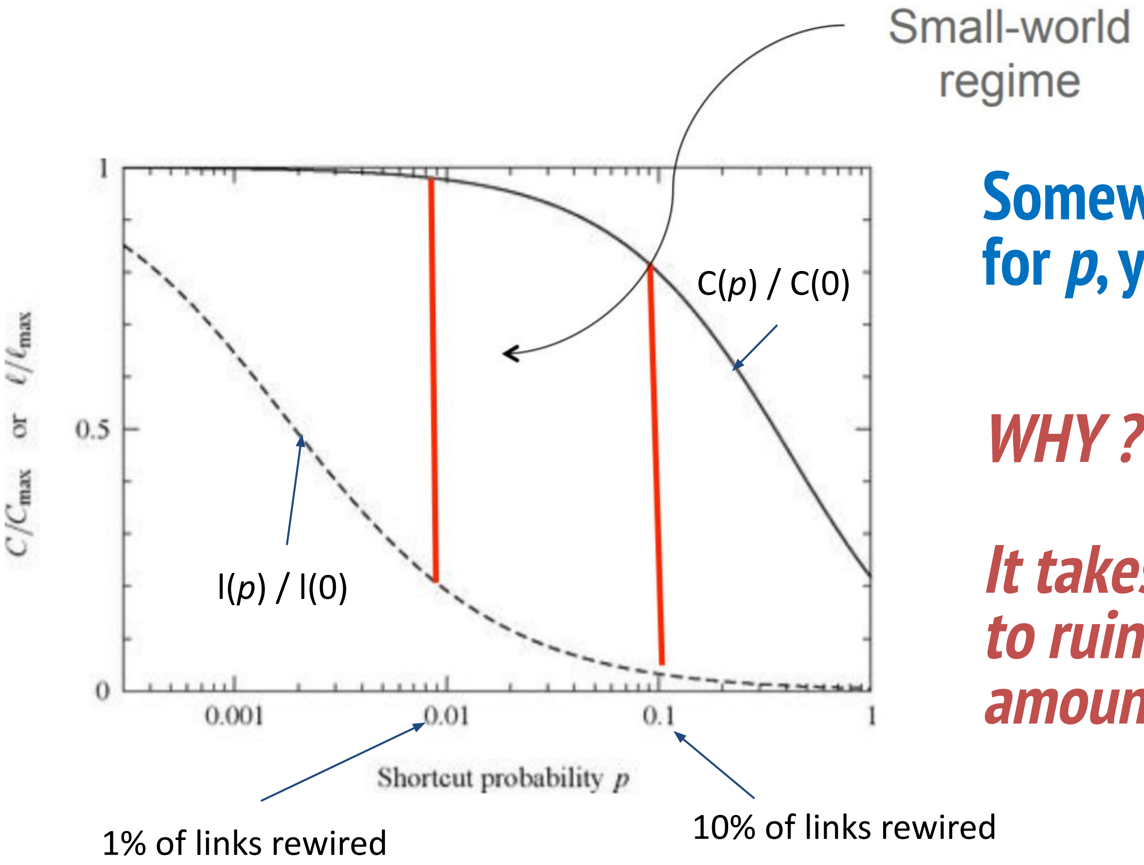
Somewhere between 0 and 1 for *p*, you get a Small-World

- Metrics for measuring small world graphs
  - Average shortest path length (mean distance)
  - Clustering coefficient

- When simulating p between 0 and 1, they discovered a region
  - Fast decrease of mean distance
  - Constant clustering

Rapid drop of mean distance due to appearance of shortcuts

Change in Clustering Coefficient and Average Path Length



Somewhere between 0 and 1 for *p*, you get a Small-World

WHY ???

It takes a lot of randomness to ruin clustering, but a small amount to create shortcuts!

Hubs, Scale Free Networks
&
Preferential Attachment

MSN Messenger: Real Network Example

- MSN Messenger network investigated in 2008 (pdf)
  - Planetary-scale views on a large instant-messaging network | Proceedings of the 17th international conference on World Wide Web

- 1 month activity
- 245 million users logged in
- 180 million users engaged in conversations
- More than 30 billion conversations
- More than 255 billion exchanged messages

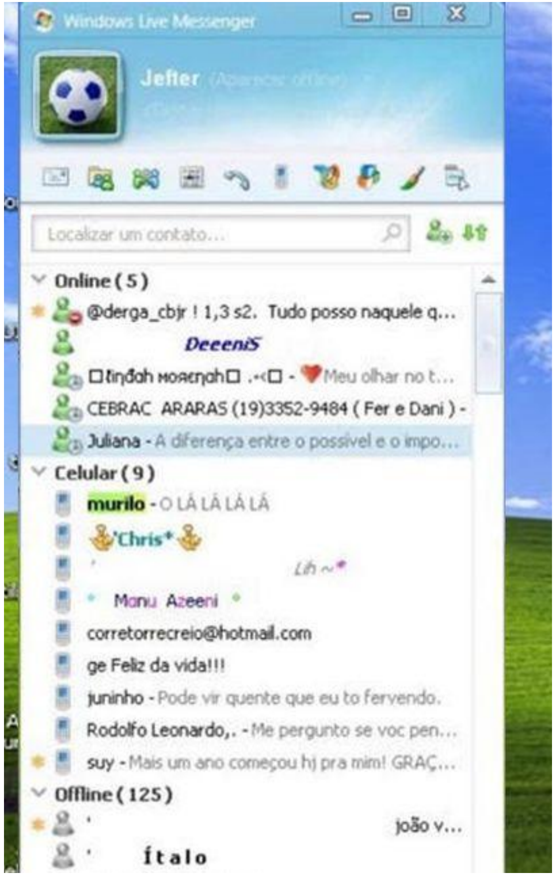


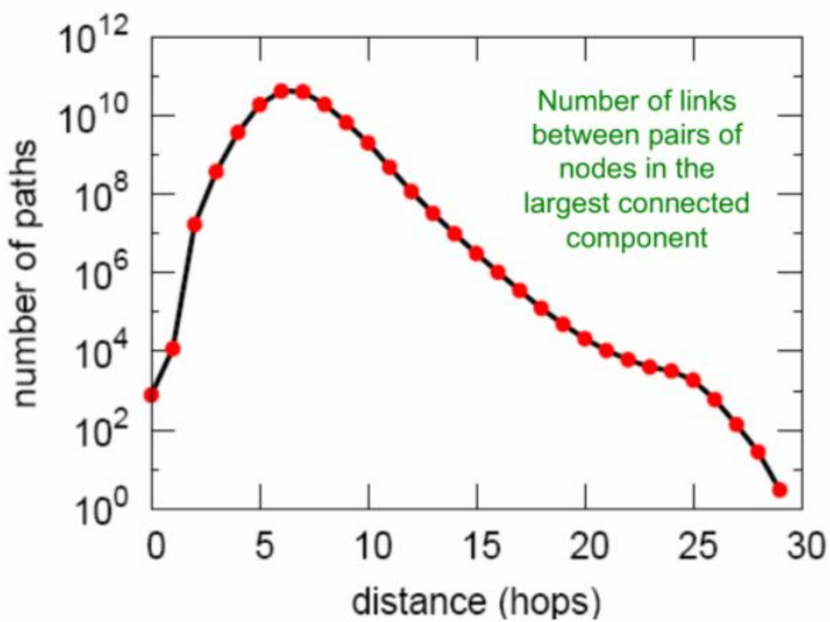
Image from:
https://www.world-today-news.com/msn-messenger-turned-22-what-could-you-do-in-the-legendary-messenger-program/

MSN Messenger: Communication vs Buddy Graph

- **Communication graph**
  - 180M nodes 1.3B edges
  - An undirected edge between two users that communicated
  - 99% of comms btw 2 people
- **Buddy graph**
  - 240M nodes, 9.1B edges
  - Two people are connected with an undirected edge if they appear on each other's contact list
  - 50 buddies on average per account

Image from:
https://www.world-today-news.com/msn-messenger-turned-22-what-could-you-do-in-the-legendary-messenger-program/

MSN Messenger: Do we really have 6 degrees of separation?

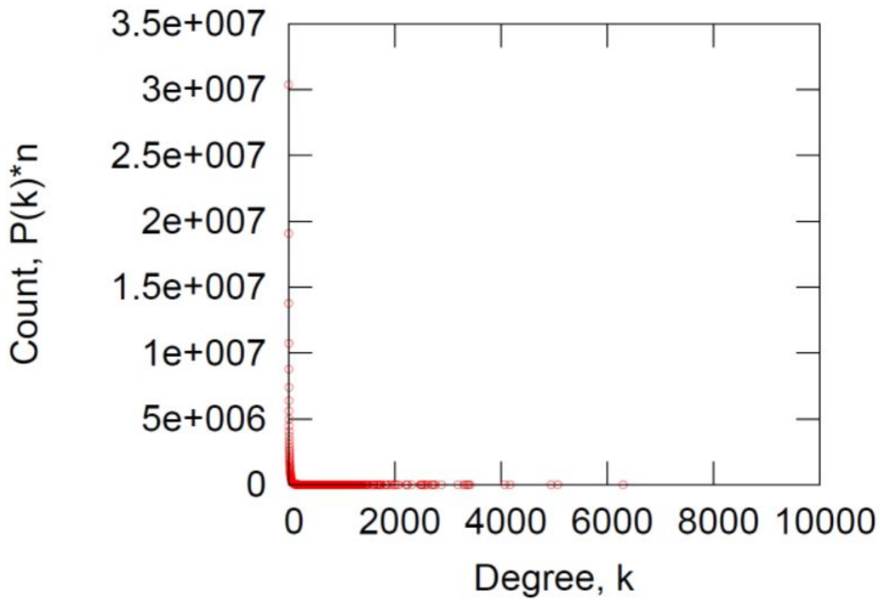


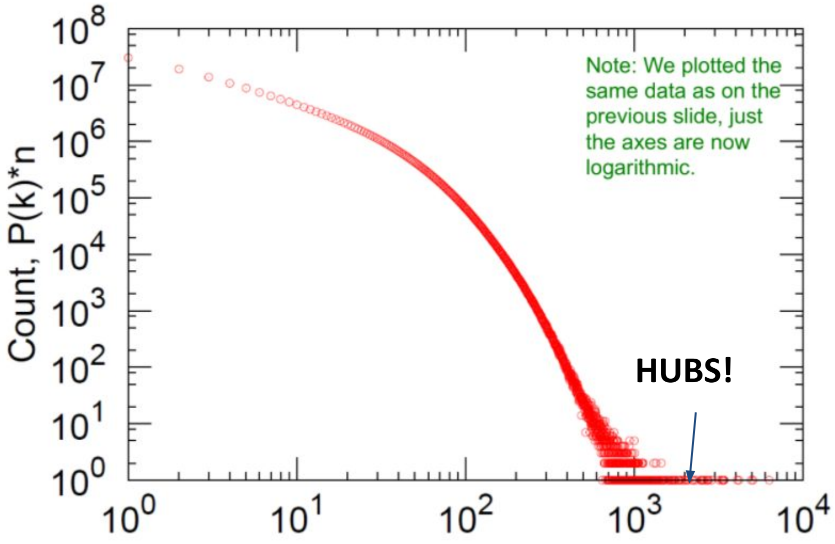
Avg. path length **6.6**
90% of the nodes can be reached in < 8 hops

"…. The average path length is 6.6. This result means that a random pair of nodes in the Messenger network is 6.6 hops apart on the average, which is half a link longer than the length measured by Travers and Milgram. The 90th percentile (effective diameter [16]) of the distribution is 7.8. 48% of nodes can be reached within 6 hops and 78% within 7 hops. So, we might say that, via the lens provided on the world by Messenger, we find that there are about "7 degrees of separation" among people. We note that long paths, i.e., nodes that are far apart, exist in the network; we found paths up to a length of 29. …."

Images from CS224W: Analysis of Networks Jure Leskovec, Stanford University
Quote from [Planetary-scale views on a large instant-messaging network | Proceedings of the 17th international conference on World Wide Web]

MSN Messenger: Communication Degree Distribution

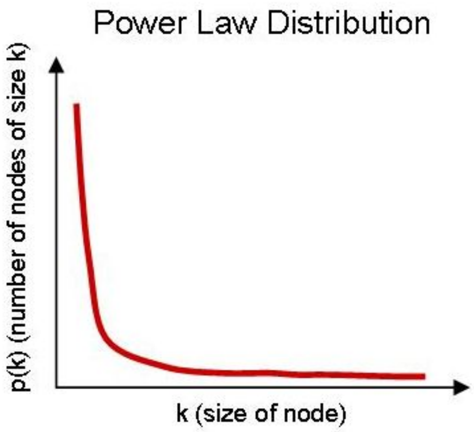




Average degree:14.4
Heavily skewed distribution. Fat-tailed.
Most nodes have very low degrees, very few nodes are HUBs

Images from CS224W: Analysis of Networks Jure Leskovec, Stanford University

Scale Free Networks



- Their degree distributions follow power laws

- This is due to the existence of hubs
  - Random networks – hubs are forbidden (most nodes have comparable degrees)
  - Scale-free network, hubs occur naturally (and are expected to occur) and they are large (e.g. k_{max} is very high).
- Many real life networks are fat-tailed (or heavy tailed), although they may not be precisely following power laws statistically

Power Laws in Economy

Pareto principle

- In Italy, a 19th century economist (Pareto) noticed that 80% of land is owned by 20% of the population.

Global Income

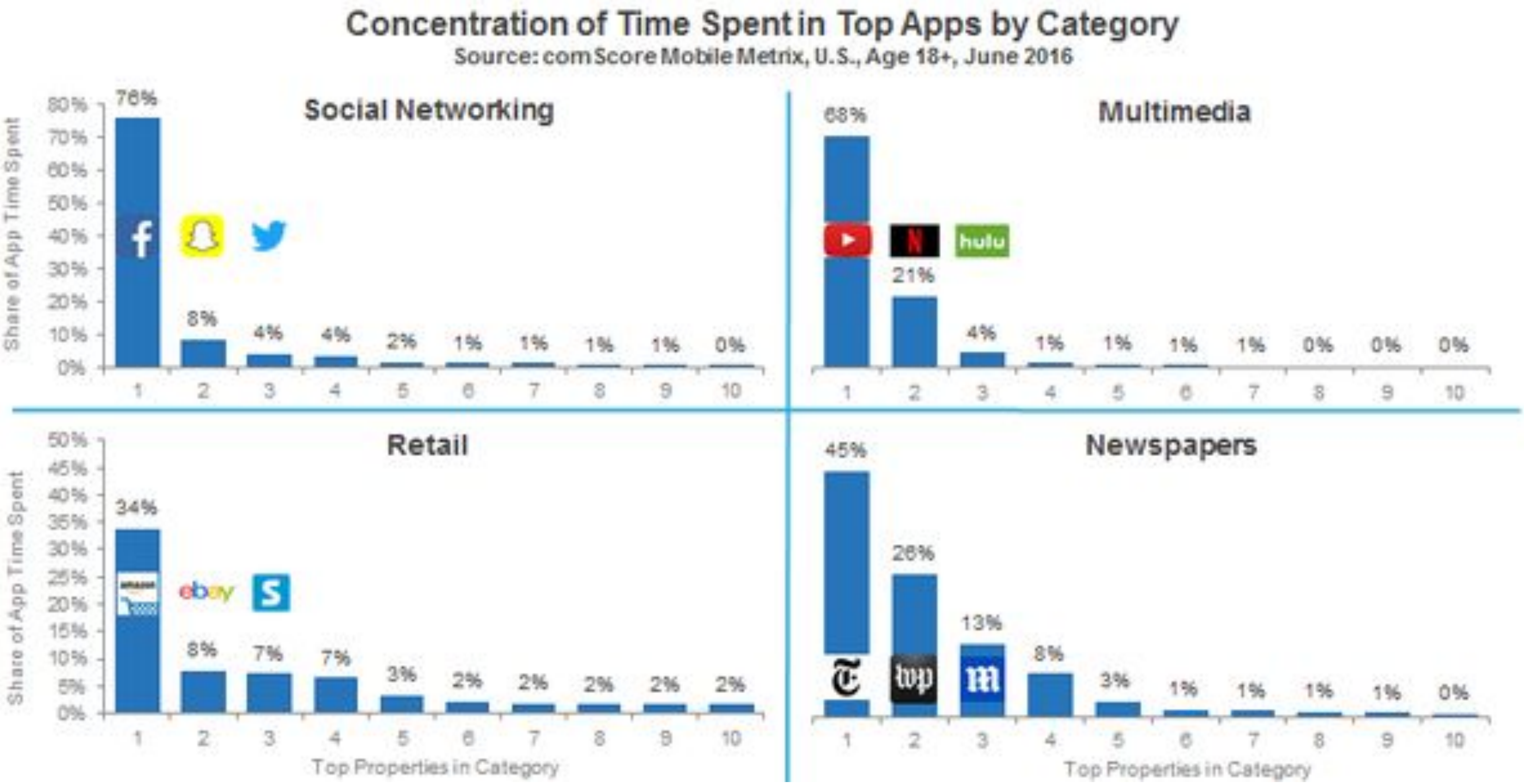
- GDP distribution in 1989 (Based 1992 UN report)

Quantile of population	Income
Richest 20%	82.70%
Second 20%	11.75%
Third 20%	2.30%
Fourth 20%	1.85%
Poorest 20%	1.40%

Taxation

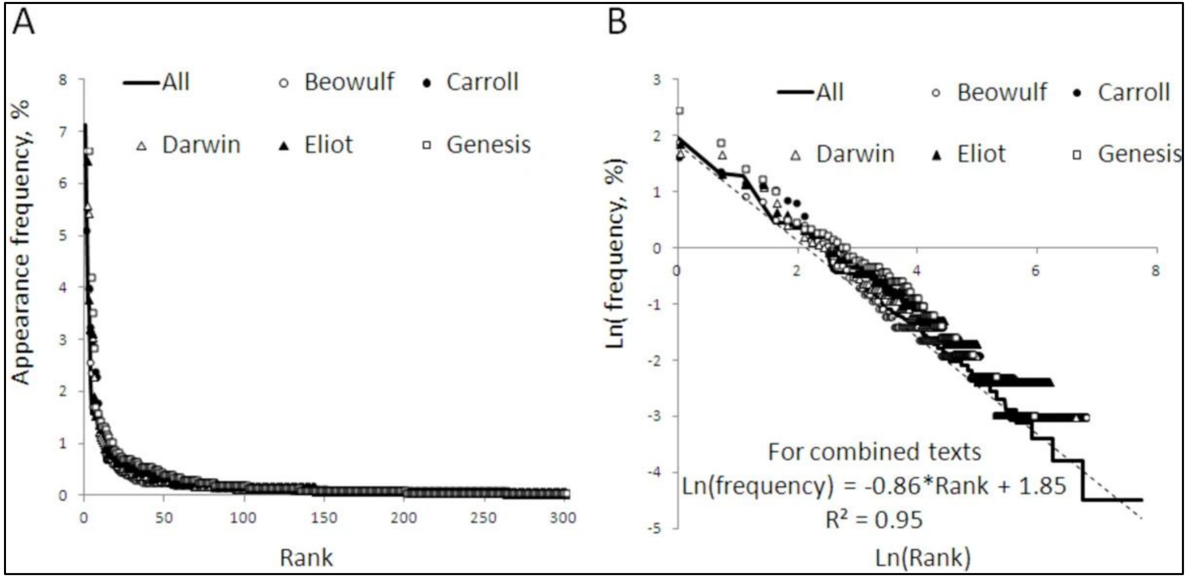
- Top 20% of Americans pay 87% of Income Tax (WSJ)

Power Laws (Power of Habit) in Mobile App Usage



Further reading:
https://www.comscore.com/fre/Perspectives/Blog/Part-2-Why-the-Power-of-Habit-Drives-Power-Law-Distributions-in-Mobile-App-Usage

Power Laws in Natural Languages



Written languages display power law behavior in word frequency.

Panel A: word appearance frequency as a function of word use rank in Beowulf, Through the Looking Glass (Lewis Carroll), The Origin of the Species (Charles Darwin), The Love Song of J Alfred Prufrock (TS Eliot), Genesis (King James Version), and the combined lexicon.

Panel B: natural logarithm (appearance frequency) versus natural logarithm (rank) for words in each text and the combined lexicon.

Source: Paladino, J.D., Crooke, P.S., Brackney, C.R. *et al.* Medical practices display power law behaviors similar to spoken languages. *BMC Med Inform Decis Mak* 13, 102 (2013). https://doi.org/10.1186/1472-6947-13-102

How do Scale Free Networks Emerge?

- The nodes need to have the capacity to link to an arbitrary number of other nodes.

- There should not be a limitation in the number of links a node can have, e.g. limitations on the size of the hubs.

Hubs present a striking difference between Erdos-Renyi random graphs and Scale free networks

Barabasi-Albert Preferential Attachment Model

- Real networks grow by addition of nodes
- Erdos-Renyi or Watts-Strogatz models keep *N* fixed.
- **Growth Model: Preferential Attachment**
  - Proposed by Barabasi & Albert in 1999
    - Barabási, A.-L.; R. Albert (1999). "Emergence of scaling in random networks". *Science*. **286** (5439): 509–512. arXiv:cond-mat/9910332
  - Model for generating networks with power-law degree distribution

New nodes prefer to connect to the more connected nodes

Most Real-Life Network are Dynamic

- Real networks grow! They are not static most of the time

Example:
Citation Network in Computational Linguistics area

2006			
	Paper citation network	Author citation network	Author collaboration network
n	8898	7849	7849
m	8765	137,007	41,362
2007			
	Paper citation network	Author citation network	Author collaboration network
n	9767	9421	9421
m	44,142	158,479	45,878

Table 1: Growth of citation volume

New nodes prefer to connect to the more connected nodes

Reference: https://1library.net/document/q5e6xdwq-the-acl-anthology-network.html?utm_source=related_list

Intuition Behind the Preferential Attachment Model

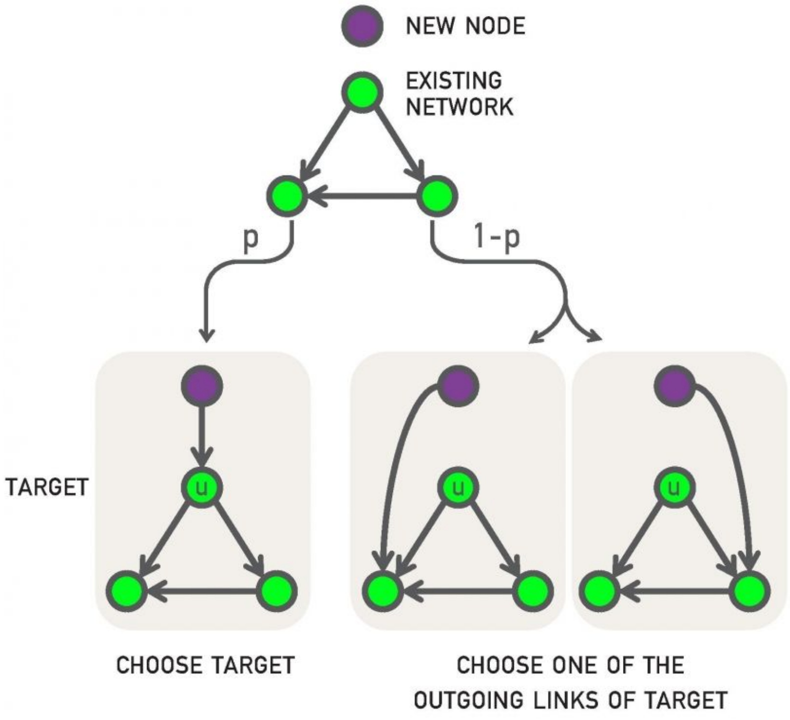
- While nodes in random networks randomly choose their interaction partner, in real networks new nodes prefer to link to the more connected nodes.

- The probability that a node connects to a node with *k* links is proportional to *k*.
  - *Something already big gets even bigger*
  - *Rich gets richer*

Barabási & Albert, Science 286, 509 (1999)

Copying Model

Copying Model



- ***Random Connection:*** With probability *p* the new node links to u.

- ***Copying:*** With probability 1-*p* we randomly choose an outgoing link of node u and connect the new node to the selected link's target. The new node "copies" one of the links of an earlier node

Social networks: *Copy your friend's friends.*
Citation Networks*: Copy references from papers we read.*

Next Lecture:

Structure: Groups & Community