SOCIAL NETWORK ANALYTICS

Small World Networks/Small World Phenomenon

Prakash C O
Department of Computer Science and Engineering

SOCIAL NETWORK ANALYTICS

Small World Networks/Small World Phenomenon

Prakash C O

Department of Computer Science and Engineering

Small World Phenomenon

- Everyone talks about the small world phenomenon, but truly what is it?

- There are three landmark papers:

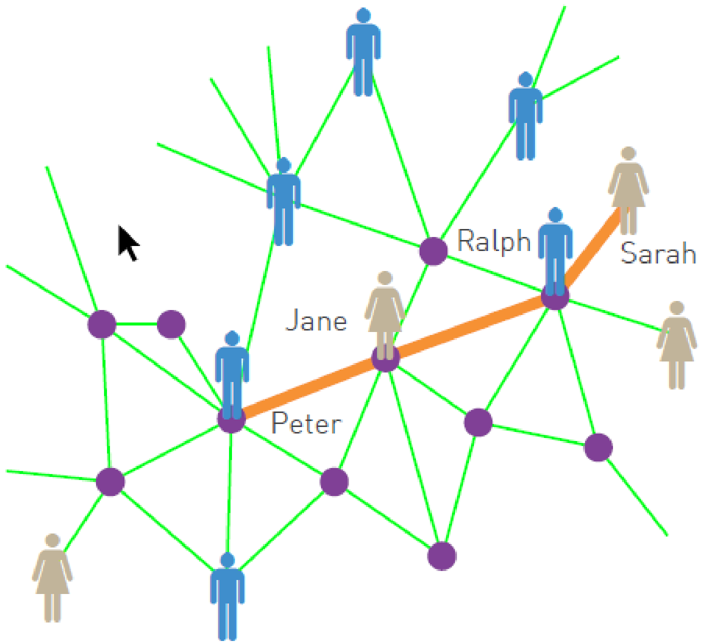
1. The Small World Problem - **Stanley Milgram(1967)**
2. Collective Dynamics Of Small-World Networks – **Duncan Watts and Strogatz(1998)**
3. The Small-World Phenomenon: An Algorithmic Perspective - **Jon Kleinberg(2000)**

Small World Networks

- Many research experiments measurements have shown that many real-world networks share two fundamental properties,
  - **Small-world phenomenon** roughly stating that distances in real-world networks are quite small, and
  - **Scale-free phenomenon** roughly stating that the degrees in real-world networks show an enormous amount of variability.

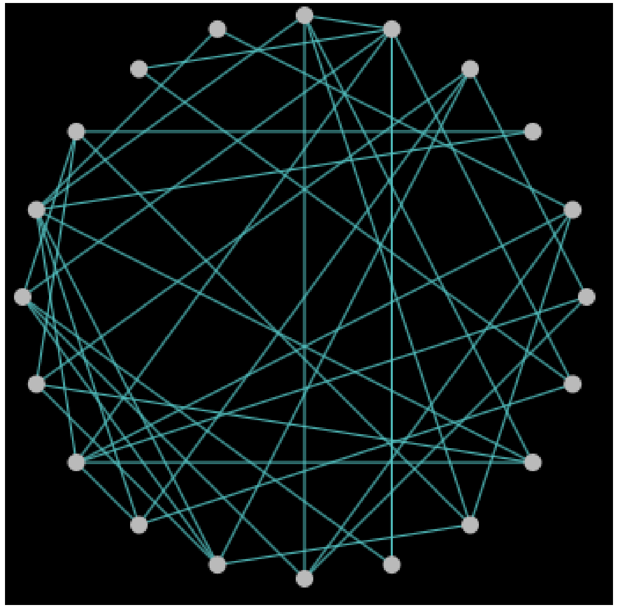
Small World Networks

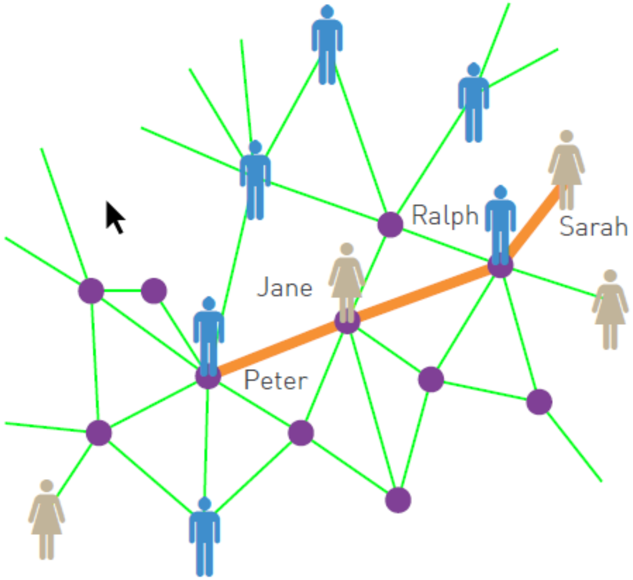
- **"Small world" phenomenon**
  - The idea that **a person is only a couple of connections away from any other person in the world**.
  - The principle that **we are all linked by short chains of acquaintances**



➢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 or steps.**

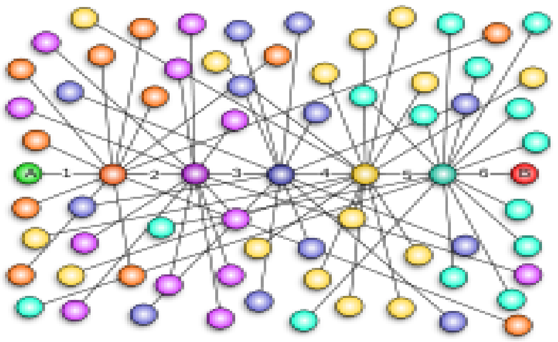




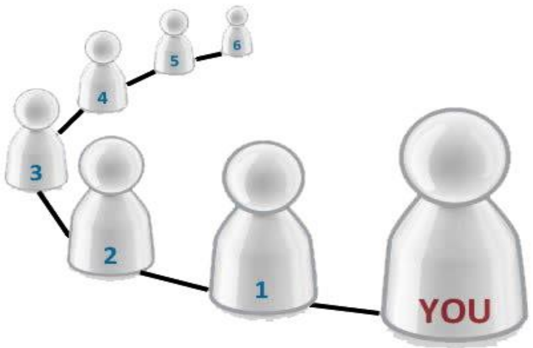
Small World Networks

- We call networks as ‘small-world’ networks, by analogy with the **small-world Phenomenon,** popularly known as **six degrees of separation**.
- **Six degrees of separation** was first proposed in 1929 by the Hungarian writer, Frigyes Karinthy, in a short story called “**Chains**.”

Frigyes Karinthy

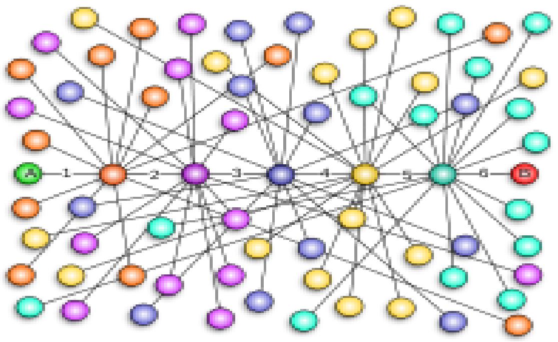


The "six degrees of separation" model

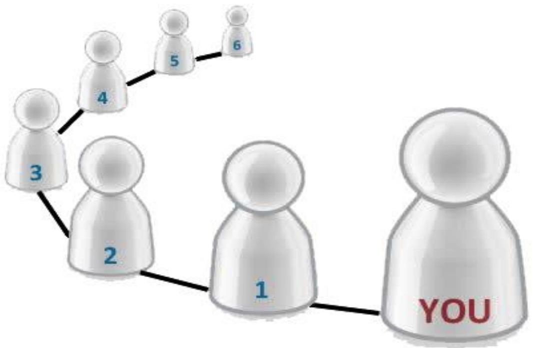


Small World Networks

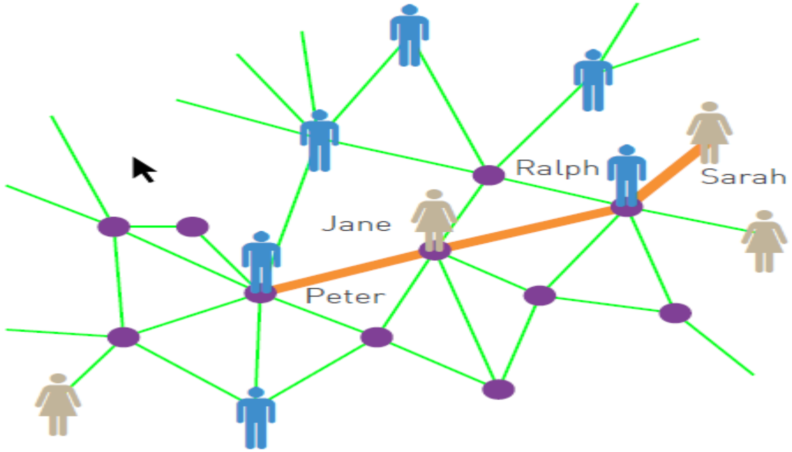
- **Six degrees of separation** is the idea that all people are six, or fewer, social connections away from each other.
- **Six degrees of separation** is also known as **the 6 Handshakes rule**. As a result, **a chain of "a friend of a friend" statements can be made to connect any two people in a maximum of six steps.**



The "six degrees of separation" model



Small World Networks



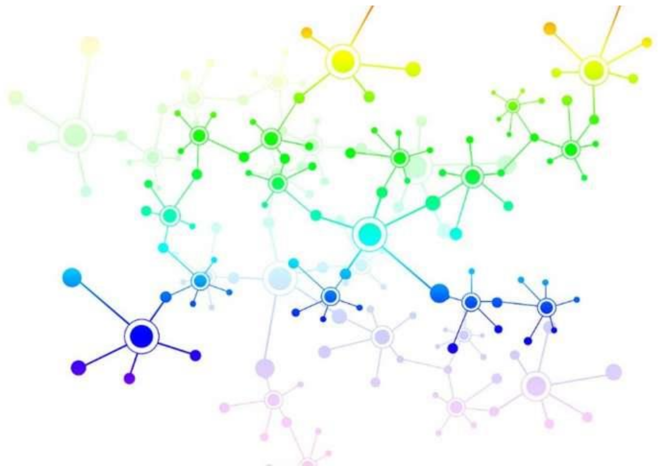
Six Degrees of Separation

- According to six degrees of separation **two individuals, anywhere in the world, can be connected through a chain of six or fewer acquaintances**.
- This means that while Sarah does not know Peter, she knows Ralph, who knows Jane and who in turn knows Peter. Hence Sarah is three handshakes, or three degrees from Peter.
- In the language of network science, six degrees of separation, also called the small world property, means that **the distance between any two nodes in a network is unexpectedly small.**

➢The Small-World Property

1. **The network has relatively few "long-distance" links** but **there are short paths between most pairs of nodes, usually created by "hubs".**

2. Most real-world complex networks seem to have the small-world property!



Small-world Networks examples

- ➢ The neural network of the worm C elegans,
- ➢ The power grid of the western United States, and
- ➢ The collaboration graph of film actors.

Small World Networks

- **The small world network model combines two important basic social network properties/ideas**:

  1. **Homophily** (the tendency to associate to those similar to ourselves) and
  2. **Weak ties** (the links to acquaintances that connect us to parts of the network that would otherwise be far away).

Small world network:

- Several other properties are often associated with small-world networks.
  - **Typically there is an over-abundance of *hubs* – nodes in the network with a high number of connections. These hubs serve as the common connections mediating the short path lengths between other edges**.
  - By analogy, the small-world network of airline flights has a small mean-path length (i.e. between any two cities you are likely to have to take three or fewer flights) because many flights are routed through hub cities.

Small World Networks

Small world network:

- **Homophily creates high clustering** while the **weak ties produce the branching structure that reaches many nodes in a few steps**.

- **Networks with short average path lengths** and **high clustering coefficients are considered small world networks.**

Small World Networks

- A small-world network is defined to be a network where **the typical distance *L* between two randomly chosen nodes grows proportionally to the logarithm of the number of nodes *N* in the network**, that is:

$$L \propto \log N$$

- In the context of a social network, this results in the small world phenomenon of strangers being linked by a short chain of acquaintances.
- Many empirical graphs show the small-world effect, including social networks, wikis such as Wikipedia, gene networks, and even the underlying architecture of the Internet.

Small World Networks

Small-world properties are found in many real-world phenomena:

- Transportation networks in ground, air or sea;
- Biology network such as food webs, gene network, protein network, neuron network, metabolism network, immune network;
- Technology network like the Internet, electric power grids, wireless network, cable network, telephone call graphs;
- Various social networks.

SOCIAL NETWORK ANALYTICS

Small-World Networks

Watts and Strogatz Small World Model

Prakash C O

Department of Computer Science and Engineering

Small World Networks – Watts and Strogatz Model

- A certain category of small-world networks were identified as a class of random graphs by Duncan Watts and Steven Strogatz in 1998.

- They noted that **graphs could be classified according to two independent structural features**, namely
  1. the clustering coefficient, and
  2. average node-to-node distance (also known as average shortest path length).

- Purely random graphs, built according to the **Erdős–Rényi (ER) model, exhibit**
  1. a small average shortest path length (varying typically as the logarithm of the number of nodes) and
  2. a small clustering coefficient.

Small World Networks – Watts and Strogatz Model

- Watts and Strogatz measured that in fact many **real-world networks** have
  1. a small average shortest path length, and
  2. a clustering coefficient significantly higher than expected by random chance.

- **Watts and Strogatz then proposed a novel graph model in 1998, currently named the Watts and Strogatz model, with**
  1. **a small average shortest path length, and**
  2. **a large clustering coefficient.**

Small World Networks – Watts and Strogatz Model

- The ER random graphs do not have two important properties observed in many real-world networks:

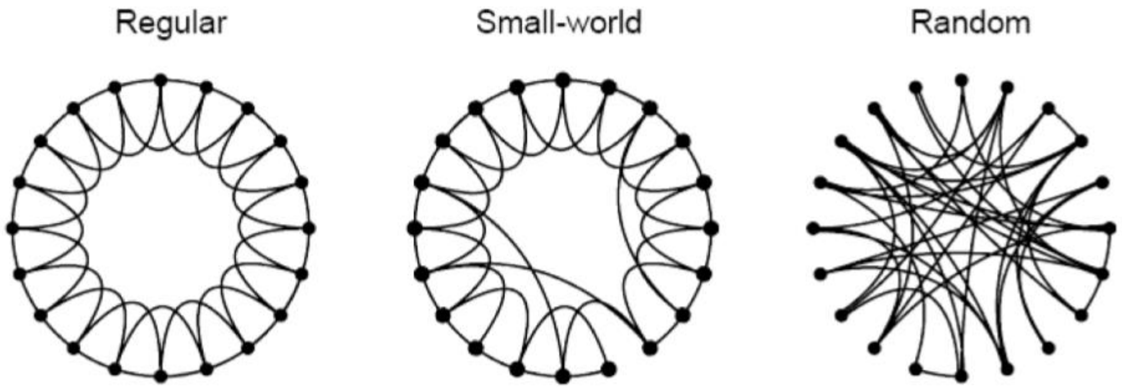
  1. **ER random graphs do not generate local clustering and triadic closures**.
     Instead, because they have a constant, random, and independent probability of two nodes being connected, ER graphs have a low clustering coefficient.

  2. **ER random graphs do not account for the formation of hubs**.
     Formally, the degree distribution of ER graphs converges to a Poisson distribution, rather than a power law observed in many real-world, scale-free networks.

- The Watts and Strogatz model was designed as the simplest possible model that addresses the **first** of the two limitations.

Small World Networks – Watts and Strogatz Model

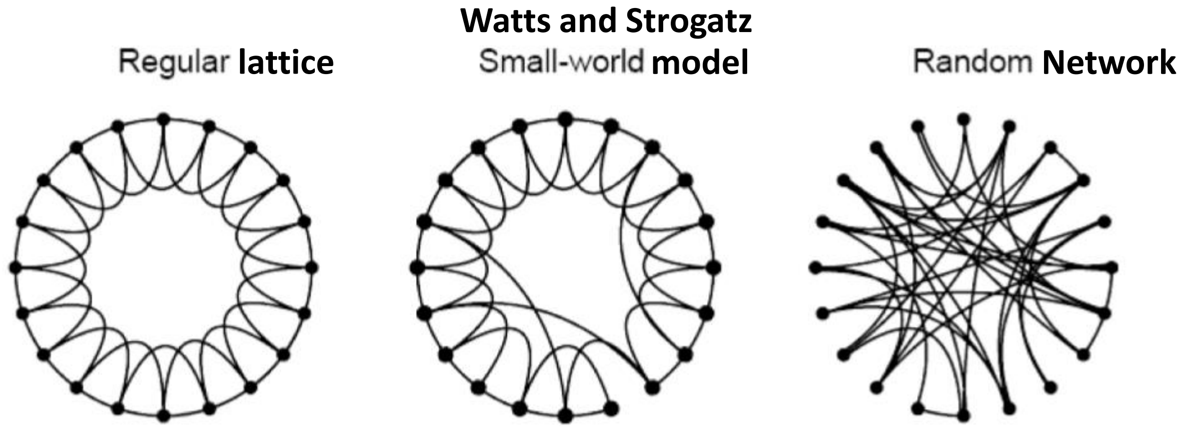
- The Watts and Strogatz model accounts for clustering while retaining the short average path lengths of the ER model.
  It does so by interpolating between a randomized structure close to ER graphs and a regular ring lattice.



- Consequently, the model is able to at least partially explain the "small-world" phenomena in a variety of networks, such as
  the power grid, neural network of C. elegans, networks of movie actors, or fat-metabolism communication in budding yeast.

Small World Networks – Watts and Strogatz Model

- **Watts and Strogatz model** exhibit
  - **a small average shortest path length, as random networks** and
  - **relevant clustering, as regular lattices**

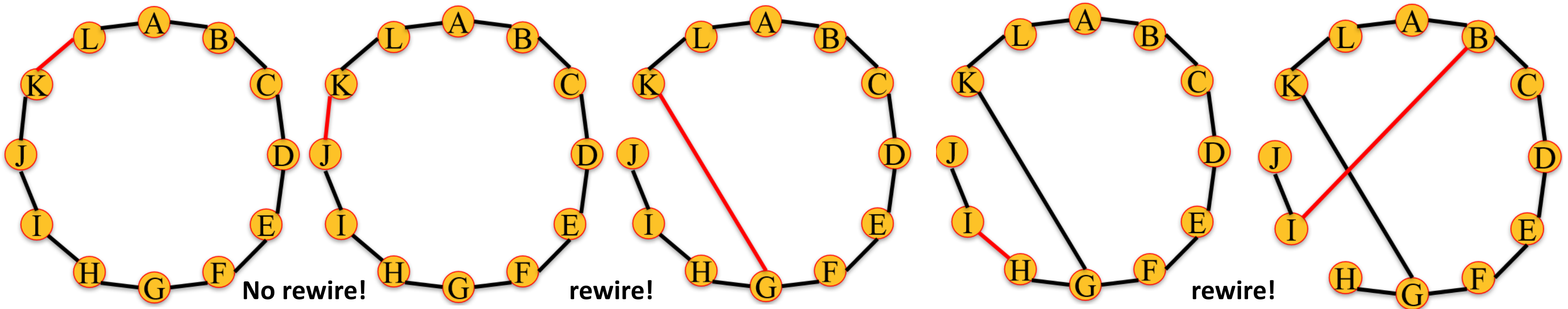


- **Watts and Strogatz model is built by** simply
  - **Re-wiring at random a small percentage of the regular edges**
  - **This is enough to dramatically shorten the average path length**, without destroying clustering

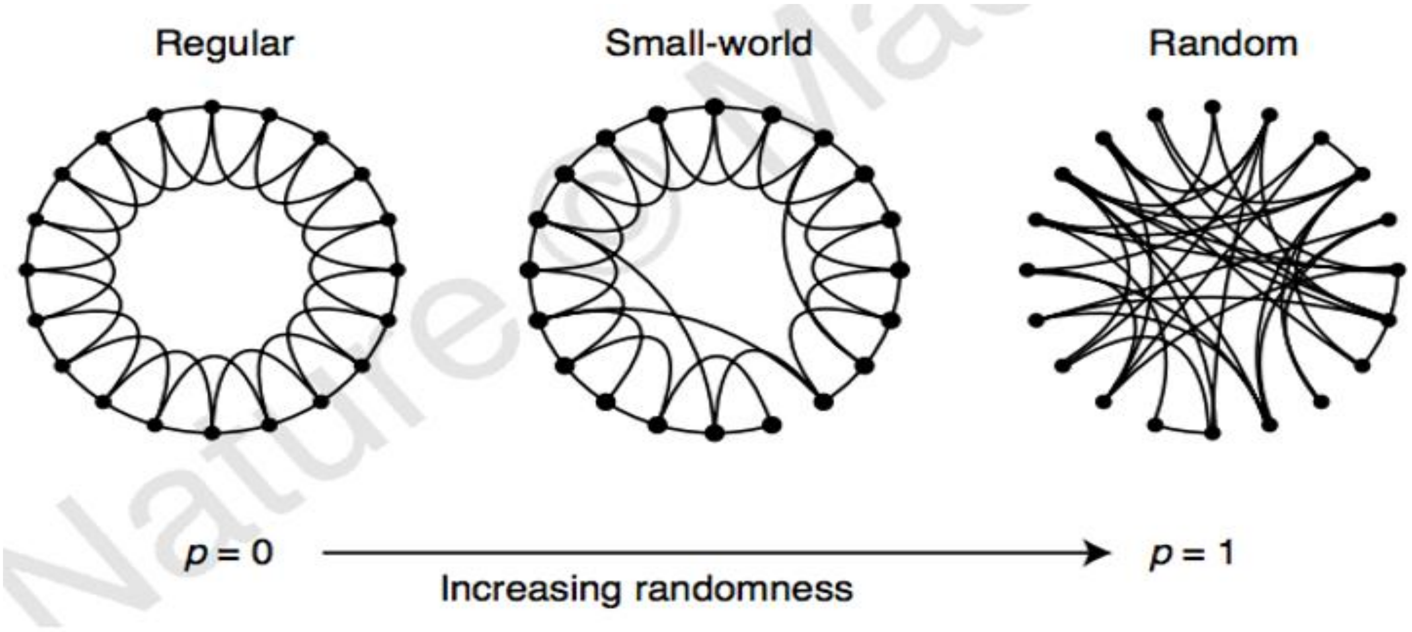
Algorithm - Watts and Strogatz Model

1. Start with a ring of n nodes, where each node is connected to its k(assumed to be an integer) nearest neighbors, $k/2$ on each side.
2. Fix a parameter $p \in [0,1]$
3. Consider each edge (u,v) and rewire with probability p. Rewiring is done by replacing (u,v) with (u,w) where w is chosen uniformly at random from all possible nodes while avoiding self loops (u≠w) and link duplication (there is no edge (u,w') with w=w' at this point in the algorithm)

Example: $k = 2,\ p = 0.4$



Small World Networks – Watts and Strogatz Model



Watts and Strogatz, 1999

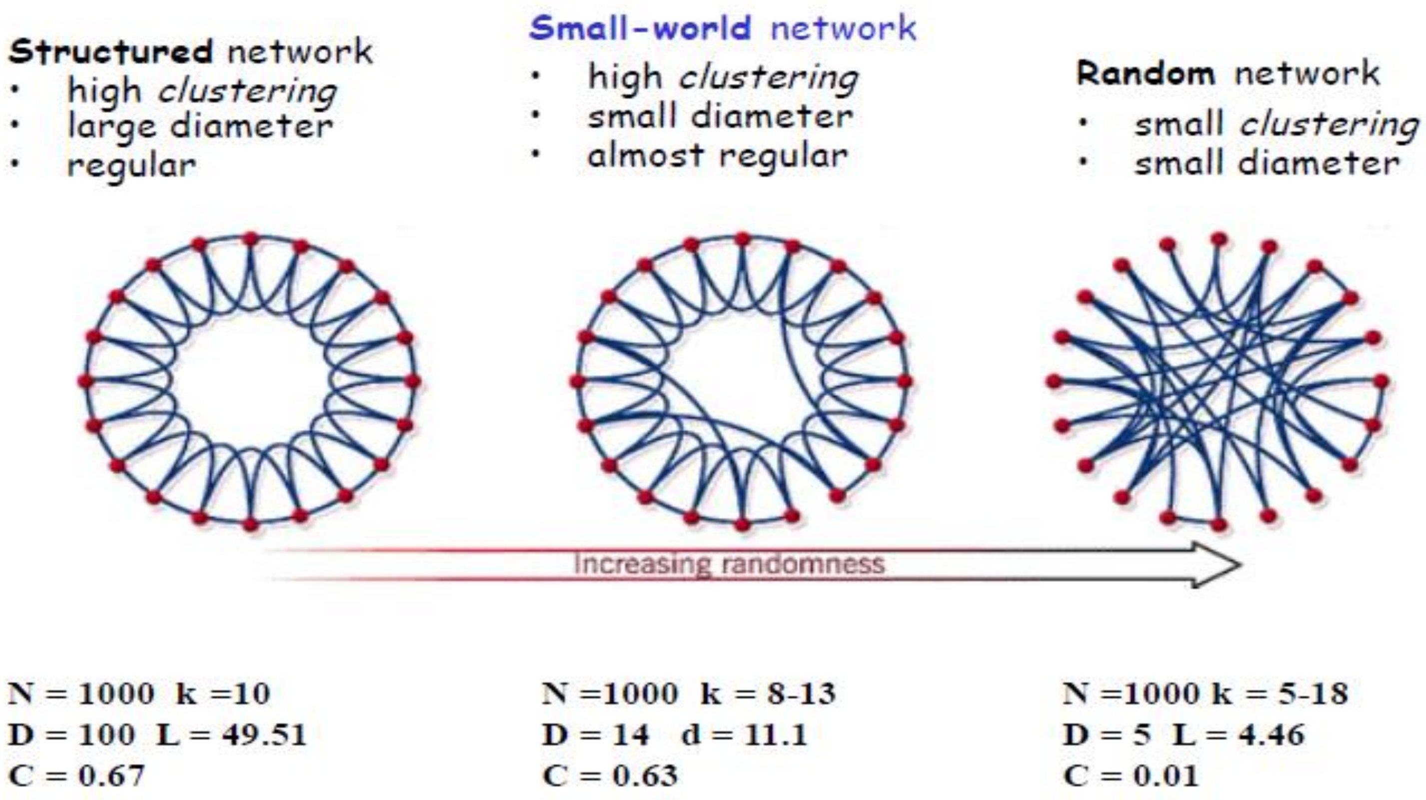
- **Regular Lattice (p=0):** no edge is rewired.
- **Small World Network (0<p<1):** Some edges are rewired. Network conserves some local structure but has some randomness.
- **Random Network (p=1):** all edges are rewired.

Watts-Strogatz model: Generating small world graphs

Select a fraction p of edges reposition one of their endpoints

Add a fraction p of additional edges leaving underlying lattice intact

Small World Networks – Watts and Strogatz Model



N = Number of nodes
k = Number of neighbors (k edges per vertex)
D = (Diameter) Maximum distance between any pair of nodes
L = Average shortest Path length
***C* = (Clustering Coefficient) Are neighbors of a node also neighbors among them?**

Small World Networks – Watts and Strogatz Model

What is the average clustering coefficient and shortest path of a small world network?

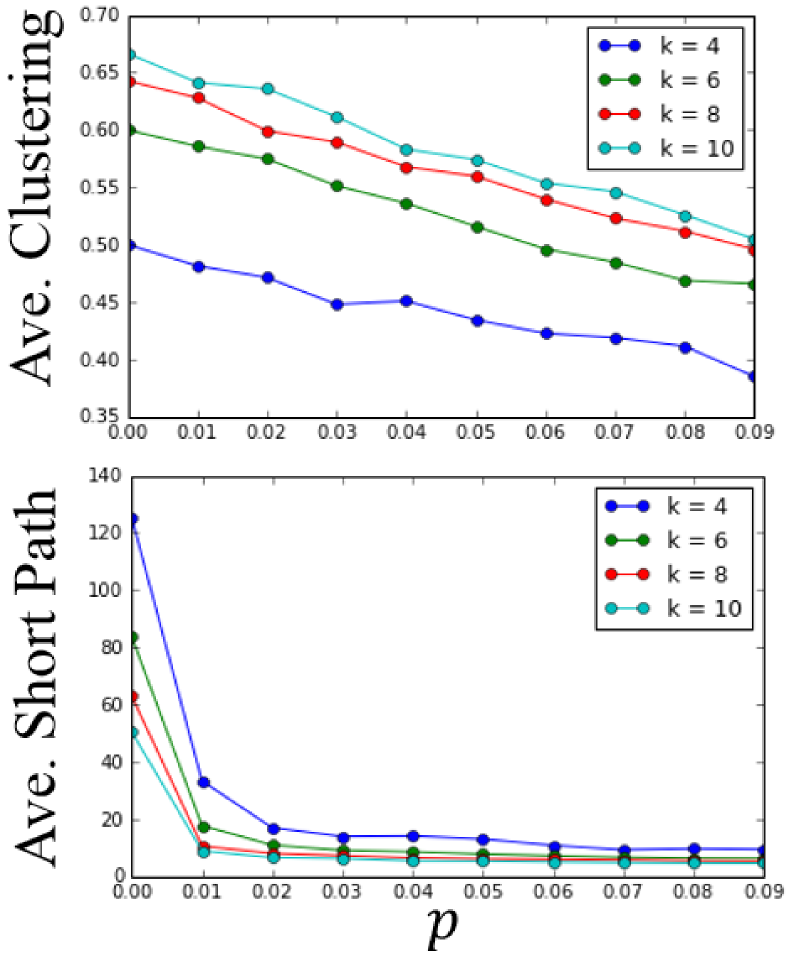
It depends on parameters k and p.

As p increases from 0 to 0.01:

- average shortest path decreases rapidly.
- average clustering coefficient decreases slowly.

An instance of a network of 1000 nodes, k=6, and p=0.04 has:

- 8.99 average shortest path.
- 0.53 average clustering coefficient.



Small World Networks – Watts and Strogatz Model

- **Watts and strogatz model Illustration in NetLogo**

1. http://ccl.northwestern.edu/netlogo/

   Go to File/Model Library/Networks/Small Worlds

2. **SmallWorldWS** (NetLogo 4.0 or 4.1.3)

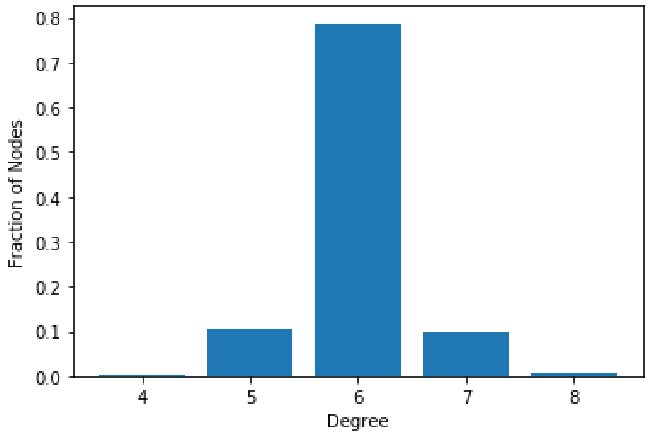
   http://www.ladamic.com/netlearn/NetLogo4/SmallWorldWS.html

Small World Networks – Watts and Strogatz Model

- watts_strogatz_graph(n, k, p) returns a small world network with n nodes, starting with a ring lattice with each node connected to its k nearest neighbors, and rewiring probability p.

Program 1: Small world network degree distribution:

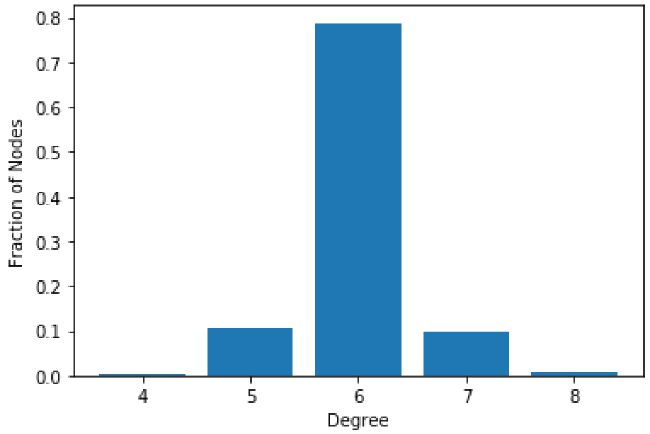
```
G = nx.watts_strogatz_graph(1000,6,0.04)
degrees = G.degree()
degree_values = sorted(set(dict(degrees).values()))
histogram = [list(dict(degrees).values()).count(i)/float(nx.number_of_nodes(G)) for i in degree_values]
plt.bar(degree_values,histogram)
plt.xlabel('Degree')
plt.ylabel('Fraction of Nodes')
plt.show()
```



Program 2: WattsStrogatz.py - Small world network demo

Small World Model in NetworkX

- Small world network: 1000 nodes, k = 6, and p = 0.04
- No power law degree distribution.
- Since most edges are not rewired, most nodes have degree of 6.
- Since edges are rewired uniformly at random, no node accumulated very high degree, like in the preferential attachment model



Small World Networks

Duncan J. Watts & Steven H. Strogatz,
Nature 393, 440-442 (1998)

Real life networks are clustered, large C, but have small average distance L.

	L	L_{rand}	C	C_{rand}	N
WWW	3.1	3.35	0.11	0.00023	153127
Actors	3.65	2.99	0.79	0.00027	225226
Power Grid	18.7	12.4	0.080	0.005	4914
C. Elegans	2.65	2.25	0.28	0.05	282

Table 1 Empirical examples of small-world networks

	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

Small World Networks

- Facebook's data team released two papers in Nov. 2011
  - 721 million users with 69 billion friendship links
  - Average distance of 4.74
- The study of **5.2 billion** Twitter friendships
  - Sysomos reports the average distance is 4.67 (2010)
  - people are 4 steps apart, nearly everyone is 5 steps or less
- Bakhshandeh et al. (2011) report an average distance of 3.435 among 1,500 random Twitter users

Assignment 1:

Find Small world phenomenon: Business applications in the following social networks?

- Facebook, Google+, Orkut, Friendster
- LinkedIn:
- Spoke, VisiblePath

Assignment 2: Paper reading

- **The Anatomy of the Facebook Social Graph** - Johan Ugander, Brian Karrer, Lars Backstrom, Cameron Marlow
- **Four Degrees of Separation** - Lars Backstrom, Paolo Boldi, Marco Rosa, Johan Ugander, Sebastiano Vigna
- **Three and a half degrees of separation** - **Research at Facebook**

Assignment 3: Activity

1. **Upload a social network (e.g. your Facebook social network into Gephi and visualize it).**
   - Use Selenium as a web scraping tool to obtain a list of all your friends and store it as a JSON file. Then visit each of their profiles programmatically and scrape the list of your mutual friends you had with them.
   - With so many hours of scraping you will be managed to obtain a list of all your friends with a list of all the mutual friends you shared with them.
   - Now you had N JSON files, one file for each friend which contained a list of friends you had in common with them.
   - Gephi requires two files to plot the network — 'nodes' and 'edges' — the former listing all the nodes with a unique ID and some additional data about each node, the latter listing all the connections between the nodes with additional information about the connection such as the directionality and weightage.
   - Use the number of messages you had exchanged with a person to be included in the visualization to see where your close friends lie in the network. Use your Facebook archive data to compute this.
   - All of the data, thus was distilled into 2 neat files — **nodes.csv** and **edges.csv**

https://towardsdatascience.com/visualising-my-facebook-network-clusters-346bac842a63

SOCIAL NETWORK ANALYTICS

Small-World Networks

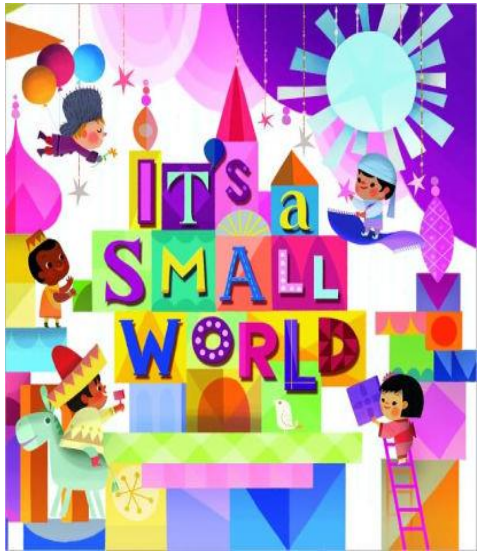
Stanley Milgram's Experiment and Six degrees of Kevin Bacon game

Prakash C O

Department of Computer Science and Engineering

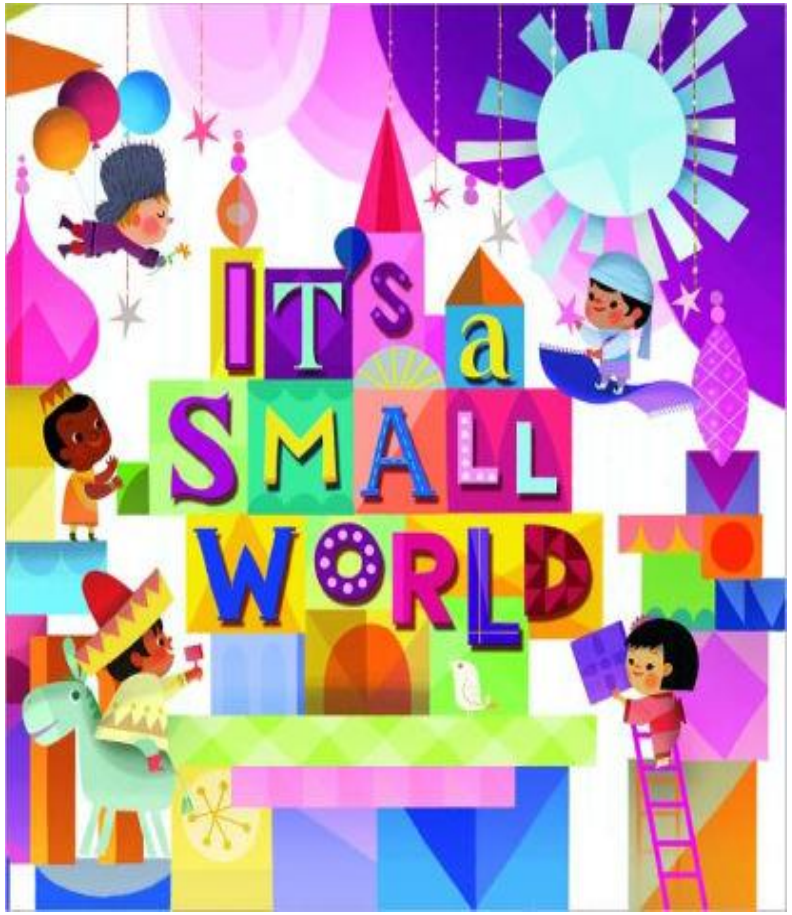
Small World Networks

- The **small-world experiment** comprised several experiments conducted by Stanley Milgram and other researchers **examining the average path length for social networks of people in the United States**.
- The research was groundbreaking in that it suggested that **human society is a small-world-type network** characterized by short path-lengths.



The Small-World Phenomenon

- The world is small in the sense that “short” paths exists between almost any two people.
- How short are these paths?
- How can we measure their length?



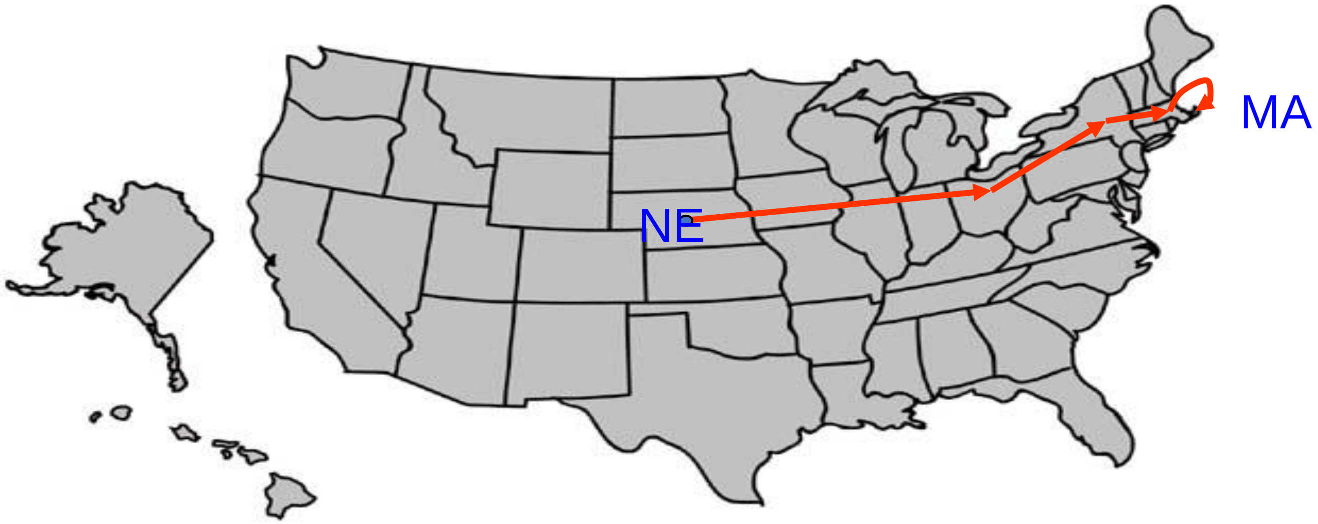
Small World Networks

Stanley Milgram, "The Small World Problem", Psychology Today, 1967, Vol. 2, 60-67

Milgram Small World Experiment

1. Milgram picked randomly 296 individuals(source persons) in Nebraska

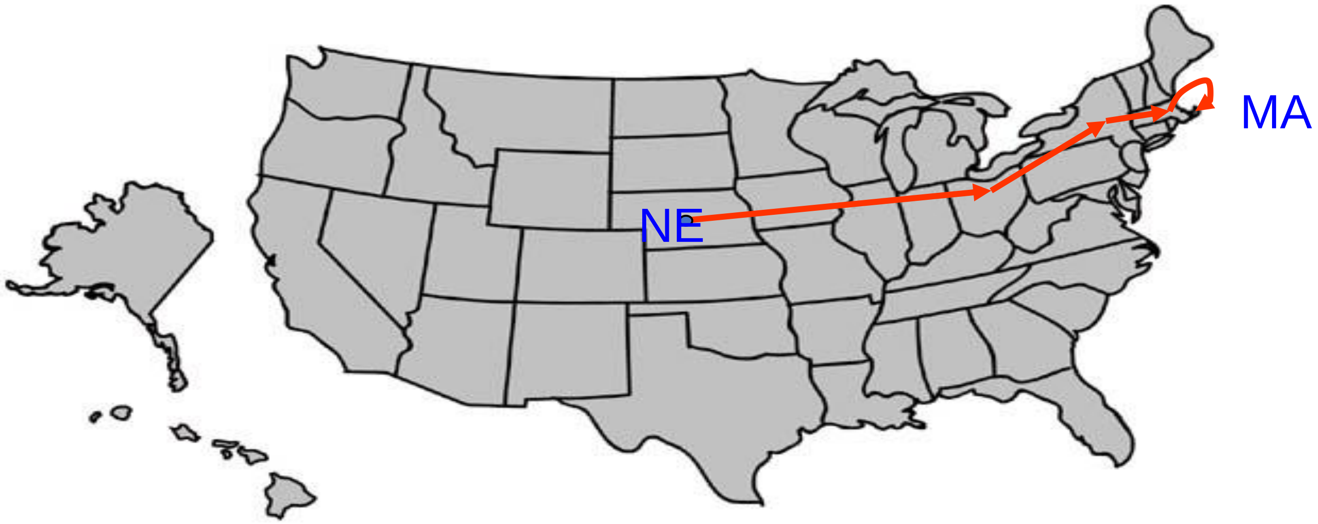
Stanley Milgram
(1933-1984)



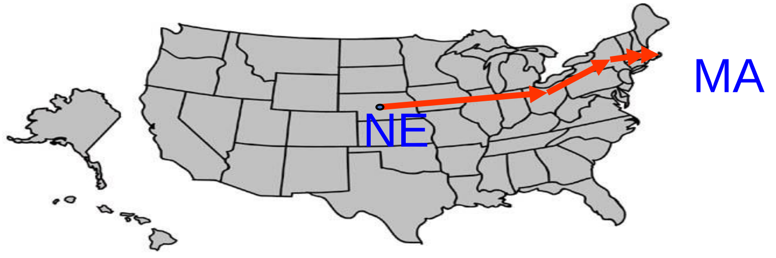
Stanley Milgram, "The Small World Problem", Psychology Today, 1967, Vol. 2, 60-67

Milgram Small World Experiment

2. Everyone was given a letter to be delivered to a target person (stockbroker) in Boston (Massachusetts).



Stanley Milgram, "The Small World Problem", Psychology Today, 1967, Vol. 2, 60-67



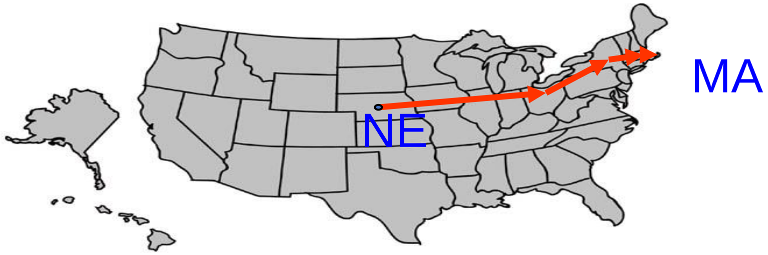
Milgram Small World Experiment

3. Instructions:

- “If you know the target person, give the letter to him, otherwise give it to someone who you think is closer to the target.”
- Some information about the target, such as address, and occupation, was provided to the source.

  Anyone subsequently receiving the letter would be given the same instructions, and the chain of communication would continue until the target was reached.

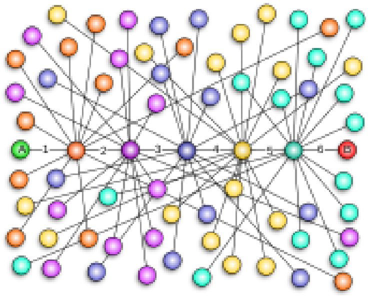
Stanley Milgram, "The Small World Problem", Psychology Today, 1967, Vol. 2, 60-67



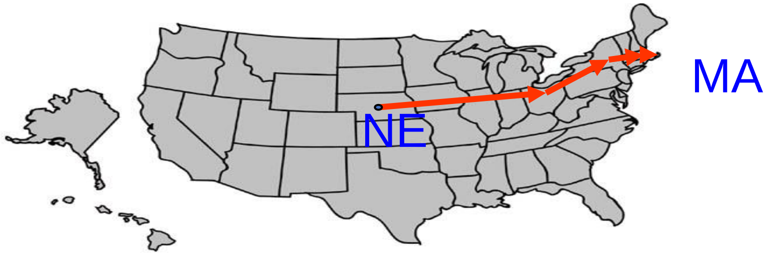
Milgram Small World Experiment

4. Results:

- 64 out of the 296 letters reached the target.
- **The average number of intermediate steps in a successful chain was found to lie between five and six** (consistent with the phrase “six degrees of separation”)



Small World Networks

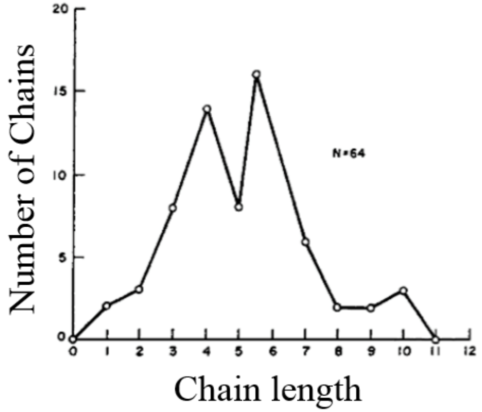


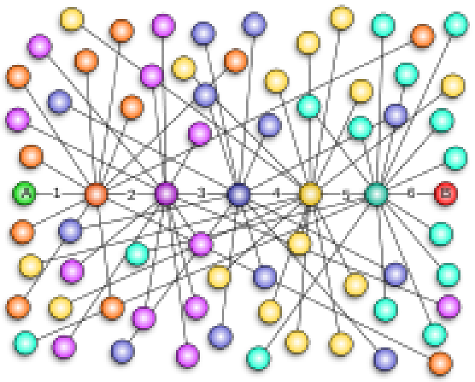
Stanley Milgram, "The Small World Problem", Psychology Today, 1967, Vol. 2, 60-67

Milgram Small World Experiment

Key points:

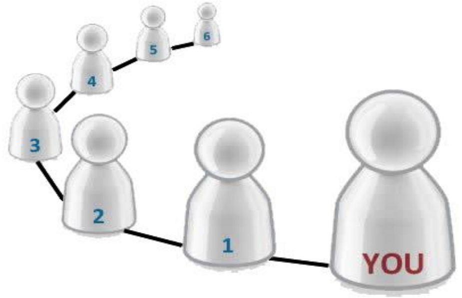
- **A relatively large percentage (>20%) of letters reached target.**
- **Paths were relatively short.**
- **People were able to find these short paths.**

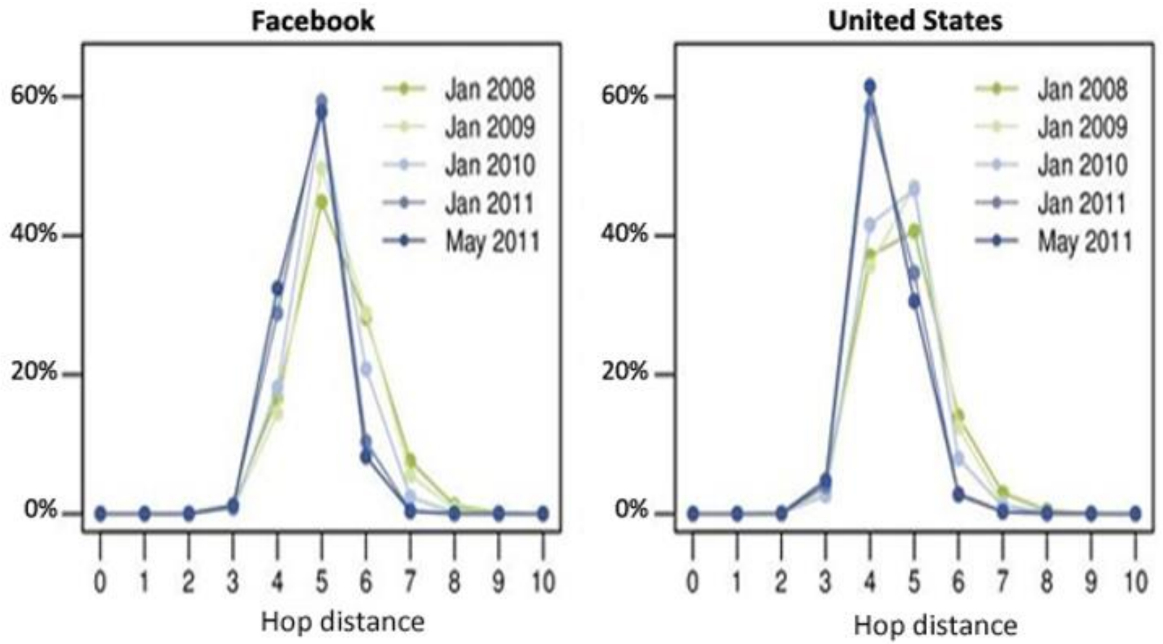




Small World of Facebook

- **Global network:** average path length in 2008 was 5.28 and in 2011 it was 4.74.
- Path are even shorter if network is restricted to US only.
- Three and a half degrees of separation **- Research at Facebook**

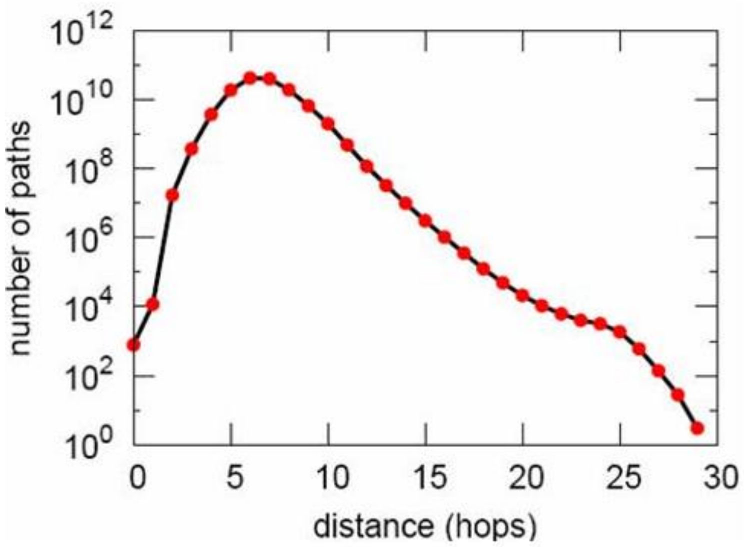




[Backstrom et al. 2012]

Small World of Instant Message

- Nodes: 240 million active users on **Microsoft Instant Messenger.**
- Edges: Users engaged in two-way communication over a one-month period.
- Estimated median path length of 7.



[Leskovec and Horvitz, 2008]

Six degrees of Kevin Bacon – game

About the Oracle of Bacon

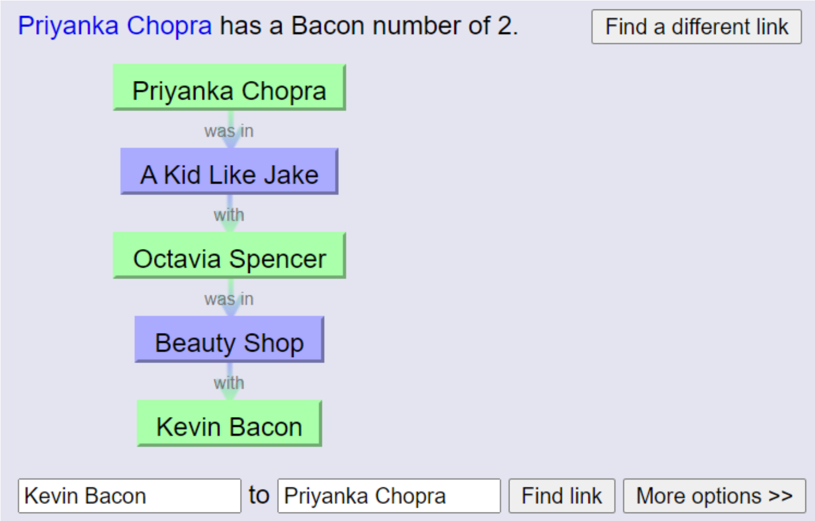
- This is the most comprehensive version of the Kevin Bacon game on the web.
- **The object of the game** is to start with any actor or actress who has been in a movie and connect them to Kevin Bacon in the smallest number of links possible.
- Two people are linked if they've been in a movie together.
- The Oracle of Bacon not considered links through television shows, made-for-tv movies, writers, producers, directors, etc.

https://oracleofbacon.org/

Six degrees of Kevin Bacon – game

- For example, you might wonder how Priyanka Chopra can be connected to Kevin Bacon. One answer is that:

Six degrees of Kevin Bacon – game



- Then we can count how many links were necessary and assign the actor or actress a **Bacon number**.
- Bacon numbers higher than 4 are very rare. In the example above, **Priyanka Chopra has a Bacon number of 2**.
- **The Oracle uses the data from Wikipedia and can give you the shortest path from every actor and actress that can be connected to Kevin Bacon**.

Six degrees of Kevin Bacon – game

How the Oracle of Bacon Works

- Every couple of weeks the Oracle downloads every English-language article from Wikipedia. Using an open-source script, They produce a JSON file with films and actors and actresses.
- There is a database service running at all times that stores the database file in memory. The service handles three different types of requests:
  - Find the link from Actor A to Actor B.
  - How good a "center" is a given actor?
  - Who are all the people with an Actor A number of N?

There are several PHP pages -- one for each of the above types of queries -- that run on the Oracle of Bacon web server, which all connect to the database service using TCP.

The database service uses a breadth-first search (BFS) to find the shortest path between pairs of actors.

Quiz Q:

1) Ignore for the time being the fact that many of your friends' friends are your friends as well. If everyone has 500 friends, the average person would have how many friends of friends?

Small World Networks

Quiz Q:

2) If the network were completely cliquish, that is all of your friends of friends were also directly your friends, what would be true?

PES UNIVERSITY ONLINE

Quiz Q:

3) What does it mean to be 1, 2, 3 hops apart on Facebook, Twitter, LinkedIn, Google Plus?

Small World Networks

Videos:

- The power of the small world phenomenon - Richard Olsen – TEDxZurich
- The Science of Six Degrees of Separation

Assignment: Activity

1. **Upload a social network (e.g. your Facebook social network into Gephi and visualize it).**
   - Use Selenium as a web scraping tool to obtain a list of all your friends and store it as a JSON file. Then visit each of their profiles programmatically and scrape the list of your mutual friends you had with them.
   - With so many hours of scraping you will be managed to obtain a list of all your friends with a list of all the mutual friends you shared with them.
   - Now you had N JSON files, one file for each friend which contained a list of friends you had in common with them.
   - Gephi requires two files to plot the network — ‘nodes’ and ‘edges’ — the former listing all the nodes with a unique ID and some additional data about each node, the latter listing all the connections between the nodes with additional information about the connection such as the directionality and weightage.
   - Use the number of messages you had exchanged with a person to be included in the visualization to see where your close friends lie in the network. Use your Facebook archive data to compute this.
   - All of the data, thus was distilled into 2 neat files — **nodes.csv** and **edges.csv**

https://towardsdatascience.com/visualising-my-facebook-network-clusters-346bac842a63

References

1. Social Network Analysis: **Lada Adamic,** University of Michigan.
2. The Small World Problem - Stanley Milgram (1967)
3. Collective Dynamics Of Small-World Networks – Duncan Watts and Strogatz (1998)
4. The Small-World Phenomenon: An Algorithmic Perspective - Jon Kleinberg (2000)
5. http://snap.stanford.edu/class/cs224w-2015/slides/05-smallworlds.pdf
6. Applied Social Network Analysis in Python | Michigan Online
7. https://www.cs.rpi.edu/~slotag/classes/FA16/slides/lec06-social.pdf
8. Wikipedia – Current Literature

THANK YOU

Prakash C O
Department of Computer Science and Engineering
coprakasha@pes.edu
+91 98 8059 1946

Algorithm - Watts and Strogatz Model

Given the desired number of nodes N, the mean degree K (assumed to be an even integer), and a special parameter β, satisfying $0 \leq \beta \leq 1$ and $N \gg K \gg \ln N \gg 1$, the model constructs an undirected graph with N nodes and $\frac{NK}{2}$ edges in the following way:

1. Construct a regular ring lattice, a graph with N nodes each connected to K neighbors, $K/2$ on each side. That is, if the nodes are labeled $0 \ldots N-1$, there is an edge (i, j) if and only if
$$0 < |i - j| \bmod \left(N - 1 - \frac{K}{2}\right) \leq \frac{K}{2}.$$
2. For every node $i = 0, \ldots, N-1$ take every edge connecting i to its $K/2$ rightmost neighbors, that is every edge $(i, j \bmod N)$ with $i < j \leq i + K/2$, and rewire it with probability β. Rewiring is done by replacing $(i, j \bmod N)$ with (i, k) where k is chosen uniformly at random from all possible nodes while avoiding self-loops $(k \neq i)$ and link duplication (there is no edge (i, k') with $k' = k$ at this point in the algorithm).