Class 8: Small World Networks
(Chapter 3 in Textbook)

Boleslaw Szymanski
I: Subcritical \(<k\) < 1

II: Critical \(<k\) = 1

III: Supercritical \(<k\) > 1

IV: Connected \(<k\) > \(\ln N\)

\(N=100\)
Unique giant component: $N_G \sim (p-p_c)N$

$\rightarrow$ GC has loops.

Cluster size distribution: exponential

$$p(s) \sim s^{-3/2}e^{-\langle k \rangle s+(s-1)\ln \langle k \rangle}$$
(b) Subcritical Regime
- No giant component
- Cluster size distribution: $p_s \sim s^{-\gamma}$
- Size of the largest cluster: $N_{0} \sim \ln N$
- The clusters are trees

(c) Critical Point
- No giant component
- Cluster size distribution: $p_s \sim s^{-\gamma}$
- Size of the largest cluster: $N_{0} \sim N^{1/2}$
- The clusters may contain loops

(d) Supercritical Regime
- Single giant component
- Cluster size distribution: $p_s \sim s^{-\gamma}$
- Size of the giant component: $N_{0} \sim (p - p_c)N$
- The small clusters are trees
- Giant component has loops

(e) Connected Regime
- Single giant component
- No isolated nodes or clusters
- Size of the giant component: $N_{0} \sim N$
- Giant component has loops
Real networks are supercritical
The measurements indicate that real networks extravagantly exceed the $\langle k \rangle = 1$ threshold. Sociologists estimate that an average person has around 1,000 acquaintances; a typical neuron is connected to dozens of other neurons, some to thousands; in our cells, each molecule takes part in several chemical reactions, some, like water, in hundreds.

The average degree of real networks is well beyond the $\langle k \rangle = 1$ threshold, implying they all have a giant component.

Do we have single component (if $\langle k \rangle > \ln N$), or multiple components (if $\langle k \rangle < \ln N$)? For social networks this requires $\langle k \rangle \geq \ln(7 \times 10^9) \approx 22.7$; so nearly two dozens acquaintances per person; with $\langle k \rangle \approx 1,000$ this is clearly satisfied. Most real networks do not satisfy this criteria, e.g., the Internet implying some routers are disconnected, so of little utility!

Most real networks are in the supercritical regime. This means that these networks have a giant component, but it coexists with many disconnected components and nodes, but only if real networks are accurately described by the Erdős-Rényi model, i.e. are random.

Today, we will further discuss the structure of real networks, we will understand why real networks can stay connected despite failing the $k > \ln N$ criteria.
Small worlds
SIX DEGREES

small worlds

Frigyes Karinthy, 1929
Stanley Milgram, 1967

Peter

Jane

Ralph

Sarah

Frigyes Karinthy, 1929
Stanley Milgram, 1967
1929: "Look, Selma Lagerlöf just won the Nobel Prize for Literature, thus she is bound to know King Gustav of Sweden, after all he is the one who handed her the Prize, as required by tradition. King Gustav, to be sure, is a passionate tennis player, who always participates in international tournaments. He is known to have played Mr. Kehrling, whom he must therefore know for sure, and as it happens I myself know Mr. Kehrling quite well.”

"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 Arpad Pasztor, someone I not only know, but to the best of my knowledge a good friend of mine. So I could easily ask him to send a telegram via the general director telling Ford that he should talk to the manager and have the worker in the shop quickly hammer together a car for me, as I happen to need one."
HOW TO TAKE PART IN THIS STUDY

1. ADD YOUR NAME TO THE ROSTER AT THE BOTTOM OF THIS SHEET, so that the next person who receives this letter will know who it came from.

2. DETACH ONE POSTCARD. FILL IT AND RETURN IT TO HARVARD UNIVERSITY. No stamp is needed. The postcard is very important. It allows us to keep track of the progress of the folder as it moves toward the target person.

3. 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.

4. IF YOU DO NOT KNOW THE TARGET PERSON ON A PERSONAL BASIS, DO NOT TRY TO CONTACT HIM DIRECTLY. INSTEAD, MAIL THIS FOLDER (POSTCARDS AND ALL) TO A PERSONAL ACQUAINTANCE WHO IS MORE LIKELY THAN YOU TO KNOW THE TARGET PERSON. You may send the folder to a friend, relative or acquaintance, but it must be someone you know on a first name basis.
"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."
WWW: 19 DEGREES OF SEPARATION

Image by Matthew Hurst
Blogosphere
DISTANCES IN Small Word Nets

Small Word Nets tend to have a tree-like topology with almost constant node degrees.

\[ N = 1 + \langle k \rangle + \langle k \rangle^2 + \ldots + \langle k \rangle^{d_{\text{max}}} = \frac{\langle k \rangle^{d_{\text{max}}+1}}{\langle k \rangle - 1} - 1 \approx \langle k \rangle^{d_{\text{max}}} \]

\[ d_{\text{max}} = \frac{\log N}{\log \langle k \rangle} \]

\( \langle k \rangle \) nodes at distance one \((d=1)\).

\( \langle k \rangle^2 \) nodes at distance two \((d=2)\).

\( \langle k \rangle^3 \) nodes at distance three \((d=3)\).

\ldots

\( \langle k \rangle^d \) nodes at distance \(d\).
We will call the small world phenomena the property that the average path length or the diameter depends logarithmically on the system size. Hence, ”small” means that $\langle d \rangle$ is proportional to $\log N$, rather than $N$.

The $1/\log\langle k \rangle$ term implies that denser the network, the smaller will be the distance between the nodes.

$$d_{\text{max}} = \frac{\log N}{\log\langle k \rangle}$$

In most networks this offers a better approximation to the average distance between two randomly chosen nodes, $\langle d \rangle$, than to $d_{\text{max}}$.

$$< d > = \frac{\log N}{\log\langle k \rangle}$$
Given the huge differences in scope, size, and average degree, the agreement is excellent.
Why are small worlds surprising?  Surprising compared to what?

Network Science: Small Word Nets
Three, Four or Six Degrees?

For the globe’s social networks:

\[ \langle k \rangle \approx 10^3 \]

\[ N \approx 7 \times 10^9 \] for the world’s population.

\[ <d> = \frac{\ln(N)}{\ln\langle k \rangle} = 3.28 \]
“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

“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 Fuego. 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

Frigyes Karinthy [1887-1938]
Hungarian writer, journalist and playwright, the first to describe the small world property. In his short story entitled 'Láncszemeik' [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 three acquaintances. Their paper inspired the experiments of Stanley Milgram.

Stanley Milgram [1933-1964]
American social psychologist who carried out the first experiment testing the small-world phenomenon (Box 3.8).

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 3.8].

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.
Clustering coefficient
Since edges are independent and have the same probability $p$, we have:

\[
\langle L_i \rangle \approx p \frac{k_i(k_i-1)}{2}
\]

- The clustering coefficient of Small Word Nets is small.
- For fixed degree $C$ decreases with the system size $N$.
- $C$ is independent of a node’s degree $k$. 
C decreases with the system size $N$.

C is independent of a node’s degree $k$. 
Watts-Strogatz Model

(a) REGULAR
(b) SMALL-WORLD
(c) RANDOM

Increasing randomness

(d)

\( \langle C(p) \rangle \)

\( \langle C(0) \rangle \)

\( d(p) / d(0) \)

\( p \)
Real networks are not random
As quantitative data about real networks became available, we can compare their topology with the predictions of random graph theory.

Note that once we have $N$ and $<k>$ for a random network, from it we can derive every measurable property. Indeed, we have:

**Average path length:**

$$<l_{\text{rand}}> \approx \frac{\log N}{\log\langle k \rangle}$$

**Clustering Coefficient:**

**Degree Distribution:**

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

\[ <d> = \frac{\log N}{\log \langle k \rangle} \]

Real networks have short distances like Small Word Nets.
Prediction:

\( C_{\text{rand}} \) underestimates with orders of magnitudes the clustering coefficient of real networks.
THE DEGREE DISTRIBUTION

Prediction:

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

Data:

\[ P(k) \approx k^{-\gamma} \]
As quantitative data about real networks became available, we can compare their topology with the predictions of random graph theory. Note that once we have $N$ and $<k>$ for a random network, from it we can derive every measurable property. Indeed, we have:

**Average path length:**

$$< l_{\text{rand}} > \approx \frac{\log N}{\log <k>}$$

**Clustering Coefficient:**

**Degree Distribution:**

$$P(k) = e^{-<k>} \frac{<k>^k}{k!}$$
(B) Most important: we need to ask ourselves, are real networks random?

The answer is simply: NO

There is no network in nature that we know of that would be described by the random network model.
It is the reference model for the rest of the class.

It will help us calculate many quantities, that can then be compared to the real data, understanding to what degree is a particular property the result of some random process.

Patterns in real networks that are shared by a large number of real networks, yet which deviate from the predictions of the random network model.

In order to identify these, we need to understand how would a particular property look like if it is driven entirely by random processes.

While WRONG and IRRELEVANT, it will turn out to be extremly USEFUL!
Summary
1951, Rapoport and Solomonoff:

→ first systematic study of a random graph.
→ demonstrates the phase transition.

→ natural systems: neural networks; the social networks of physical contacts (epidemics); genetics.

Why do we call it the Erdos-Renyi random model?

Anatol Rapoport
1911- 2007

Edgar N. Gilbert
(b. 1923)
Erdos: 1,400 papers 507 coauthors

Einstein: EN=2
Paul Samuelson EN=5

ALB: EN: 3
Collaboration Network:

**Nodes:** Scientists  
**Links:** Joint publications

Physical Review:  

N=449,673  
L=4,707,958

See also Stanford Large Network database  
Network Science: Small Word Nets

Scale-free

Hierarchical
THE END