Small World Problem

Computational Social Systems 1 (VU) (706.616)

Denis Helic

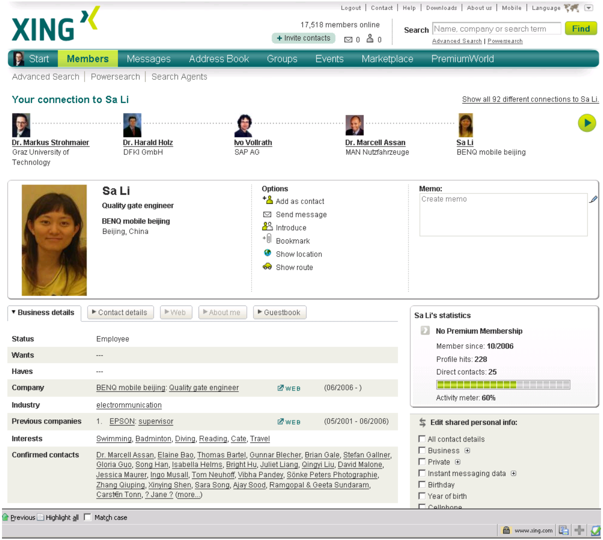
ISDS, TU Graz

March 28, 2020

Outline

1. Introduction
2. Small World Experiment
3. Small world networks
4. Small World Networks: Intuition
5. Small World Networks: Formalization
6. Distance Contraction in Sparse Networks
7. Small World Phenomenon in Empirical Networks
8. Alternative Small World Models
9. Emergence of Small World Networks
10. Applications

Do I know somebody in...?



The Bacon Number

The Kevin Bacon Game

The Bacon Number

TABLE 3.1 DISTRUBUTION OF ACTORS ACCORDING TO BACON NUMBER

BACON NUMBER	NUMBER OF ACTORS	CUMULATIVE TOTAL NUMBER OF ACTORS
0	1	1
1	1,550	1,551
2	121,661	123,212
3	310,365	433,577
4	71,516	504,733
5	5,314	510,047
6	652	510,699
7	90	510,789
8	38	510,827
9	1	510,828
10	1	510,829

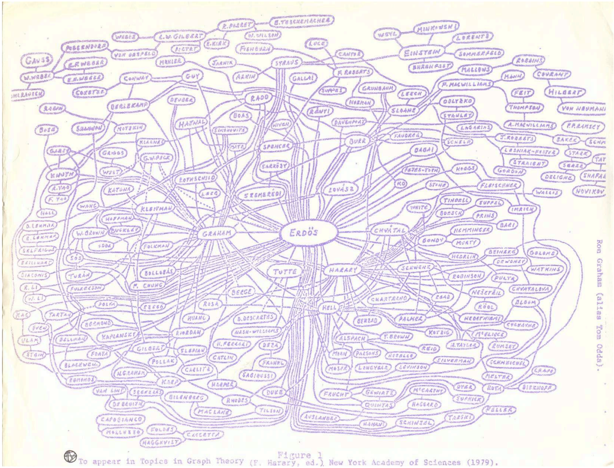
The Erdös Number

- Who was Erdös?
- http://www.oakland.edu/enp/
- A famous Hungarian Mathematician, 1913-1996
- Erdös posed and solved problems in number theory and other areas and founded the field of discrete mathematics
- 511 co-authors (Erdös number 1)

The Erdös Number

- The Erdös Number
- Through how many research collaboration links is an arbitrary scientist connected to Paul Erdös?
- What is a research collaboration link?
- Per definition: Co-authorship on a scientific paper
- What is my Erdös Number? 4
- me $\rightarrow$ H. Maurer $\rightarrow$ W. Kuich $\rightarrow$ N. Sauer $\rightarrow$ P. Erdös
- http://www.ams.org/mathscinet/collaborationDistance.html

Erdös network



Stanley Milgram

- A social psychologist
- Yale and Harvard University
- Study on the Small World Problem
- Controversial: The Obedience Study
- What we will discuss today: “An Experimental Study of the Small World Problem”

Small world problem

Small world

The simplest way of formulating the small-world problem is: Starting with any two people in the world, what is the likelihood that they will know each other?

A somewhat more sophisticated formulation, however, takes account of the fact that while person X and Z may not know each other directly, they may share a mutual acquaintance - that is, a person who knows both of them. One can then think of an acquaintance chain with X knowing Y and Y knowing Z. Moreover, one can imagine circumstances in which X is linked to Z not by a single link, but by a series of links, X-A-B-C-D...Y-Z. That is to say, person X knows person A who in turn knows person B, who knows C... who knows Y, who knows Z.

Small world experiment

- A Social Network Experiment tailored towards demonstrating, defining, and measuring inter-connectedness in a large society (USA)
- A test of the modern idea of "six degrees of separation"
- Which states that: every person on earth is connected to any other person through a chain of acquaintances not longer than 6

Experiment

- Goal
  1. Define a single target person and a group of starting persons
  2. Generate an acquaintance chain from each starter to the target
- Experimental Set Up
  1. Each starter receives a document
  2. Each starter was asked to begin moving it by mail toward the target
  3. Information about the target: name, address, occupation, company, college, year of graduation, wife's name and hometown
  4. Information about relationship (friend/acquaintance)

Experiment

- Constraints
  1. Starter group was only allowed to send the document to people they know and
  2. Starter group was urged to choose the next recipient in a way as to advance the progress of the document toward the target

Questions

- How many of the starters would be able to establish contact with the target?
- How many intermediaries would be required to link starters with the target?
- What form would the distribution of chain lengths take?

Set Up

- Target person
  1. A Boston stockbroker
- Three starting populations
  1. 100 “Nebraska stockholders”
  2. 96 “Nebraska random”
  3. 100 “Boston random”

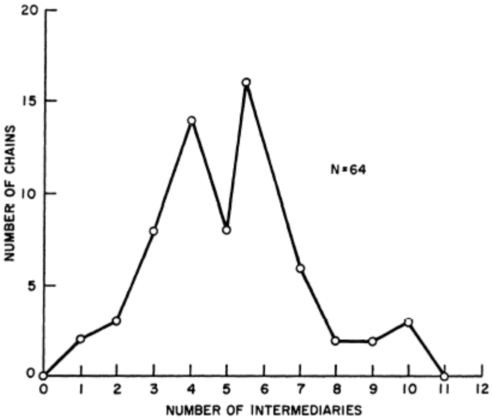
Set Up


Nebraska random
Nebraska stockholders
Target
Boston stockbroker
Boston random

Results

- How many of the starters would be able to establish contact with the target?
  1. 64 out of 296 reached the target
- How many intermediaries would be required to link starters with the target?
  1. Well, that depends: the overall mean 5.2 links
  2. Through hometown: 6.1 links
  3. Through business: 4.6 links
  4. Boston group faster than Nebraska groups
  5. Nebraska stockholders not faster than Nebraska random
- What form would the distribution of chain lengths take?

Chain length distribution



Results

- What have been common strategies?
  1. Geography
  2. Profession
- What are the common paths?
  1. See e.g. Gladwell's "Law of the few"

Common paths and gatekeepers

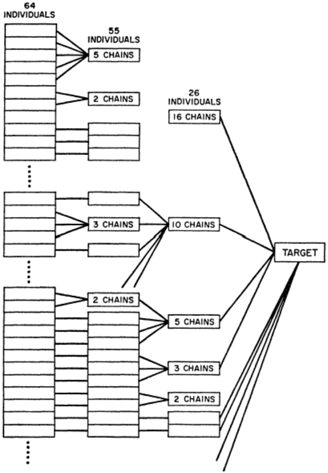


FIGURE 3

Common Paths Appear as Chains Converge on the Target

Conclusions and 6 degrees of separation

- So is there an upper bound of six degrees of separation in social networks?
  1. Extremely hard to test
  2. In Milgram's study, 2/3 of the chains did not reach the target
  3. 1/3 random, 1/3 blue chip owners, 1/3 from Boston
  4. Danger of loops (mitigated in Milgram's study through chain records)
  5. Target had a "high social status"

Follow up work

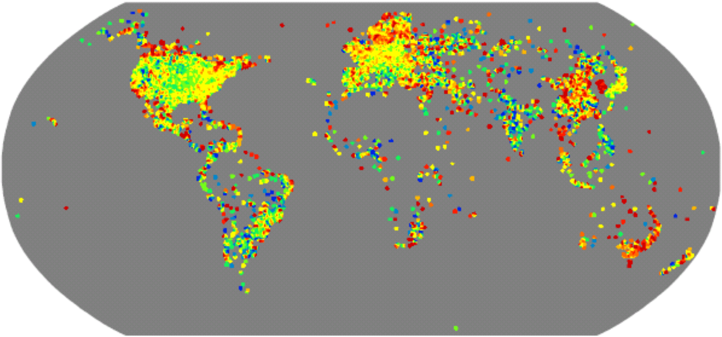


Figure 10: Number of users at a particular geographic location. Color represents the number of users. Notice the map of the world appears.

Follow up work

- Horvitz and Leskovec study 2008
- 30 billion conversations among 240 million people of Microsoft Messenger
- Communication graph with 180 million nodes and 1.3 billion undirected edges

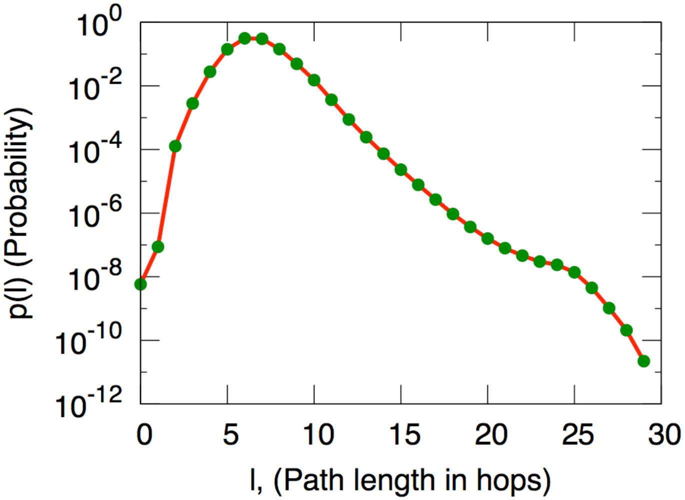
Follow up work

- Approximation of "Degrees of separation"
- Random sample of 1000 nodes
- For each node the shortest paths to all other nodes was calculated. The average path length is 6.6, median at 7
- Result: a random pair of nodes is 6.6 hops apart on the average, which is half a link longer than the length reported by Travers/Milgram

Follow up work

- 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.
- Finding that there are about "7 degrees of separation" among people
- Long paths exist in the network; paths up to a length of 29

Chain length distribution



Follow up work at Facebook

- Facebook study: February 2016
- https://research.facebook.com/blog/three-and-a-half-degrees-of-separation/
- 1.5 billion users
- $\sim$ 3.5 degrees of separation
- Approximative algorithms for calculating average distance

Small world networks

- Every pair of nodes is connected by a path with an extremely small number of steps
- Low diameter ℓ_{max}
- Low average distance $\overline{\ell}$

Small world networks

Small world networks

The small-world effect exists, if the number of nodes within a distance s of a typical central vertex grows exponentially with s. In other words, networks are said to show the small-world effect if the value of $\overline{\ell}$ scales logarithmically or slower with network size for a fixed average degree $\overline{k}$.

- According to this definition a random graph is a small-world network since $\overline{\ell}_{random} \approx \frac{ln(n)}{ln(\overline{k})}$
- However, there are other properties that a (realistic) small-world network must possess

Small world networks

- When we would perceive a network as small:
  1. $\overline{\ell}$ scales maximally logarithmically with n for a fixed $\overline{k}$
  2. The network itself is large in the sense that it contains $n \gg 1$ nodes
  3. The network is sparse, i.e. $n \gg \overline{k}$
  4. The network is decentralized, i.e. there is no central node that connects to almost every other node
  5. The network is highly clustered, i.e. many of our friends are also friends of each other

Small world networks

- Explanations to the criteria:
  1. If $\bar{\ell}$ scales e.g. linearly then for e.g. $n = 10^6$ the network is clearly not small
  2. If n is small as in e.g. a social network of a small town then there is a high chance that everyone knows each other and therefore $\bar{\ell}$ is small, i.e. (1) is trivially satisfied
  3. A person has on average a couple hundreds of friends among e.g. $8 \cdot 10^9$ people in the world, i.e. it is a sparsely connected network

Small world networks

- Explanations to the criteria:
  4. Even if some people are better connected than the others, there are physical constraints on the number of (mutual) connections (this criteria can be expressed as $n \gg k_{max}$
  5. Fraction of friends who are also friends of each other is significantly higher than in a random network. Otherwise our friends will be equally likely to come from a different country, occupation, etc. (this eliminates a random network from being a small-world network)
- We are looking for networks where local clustering is high and global path lengths are small

The Cavemen World: highly clustered social connections

The Solaris World: random social connections

Small world networks

- Two seemingly contradictory requirements for the Small World Phenomenon:
  1. Network should display a large clustering coefficient, so that a node's friends will know each other (as in Caveman world)
  2. It should be possible to connect two people chosen at random via chain of only a few intermediaries (as in Solaria world)

A random graph model

- Now we will analyze an ensemble of networks, i.e. a special random graph model
- Recollect that an ensemble defines a probability distribution over all possible graphs
- We will characterize networks in terms of $\bar{\ell}$ and C
- In order to decide if a network is "‘small"’ or "‘large"’ we need to determine the ranges over which $\bar{\ell}$ and C vary

Duncan Watts

This part of the slides is based on the paper "‘Networks, Dynamics, and the Small-World Phenomenon"’ by D. Watts

Constraints for the model

- All networks need to satisfy the following constraints:
  1. We fix n
  2. Graph is sparse but sufficiently dense, i.e. $1 \ll \bar{k} \ll n$. Wide range of structures are possible
  3. The graphs are connected

Extremal properties

- What is the largest value for C and in what kind of a graph we observe such C?

Extremal properties

- What is the largest value for C and in what kind of a graph we observe such C?
- $C = 1$ in a complete graph ($\bar{k} = n - 1$)
- What is the minimal value for C and which graph has that?

Extremal properties

- What is the largest value for C and in what kind of a graph we observe such C?
- $C = 1$ in a complete graph ($\bar{k} = n - 1$)
- What is the minimal value for C and which graph has that?
- $C = 0$ in an empty graph ($\bar{k} = 0$)
- What is $\bar{\ell}$ in those two graphs?

Extremal properties

- What is the largest value for C and in what kind of a graph we observe such C?
- $C = 1$ in a complete graph ($\bar{k} = n - 1$)
- What is the minimal value for C and which graph has that?
- $C = 0$ in an empty graph ($\bar{k} = 0$)
- What is $\bar{\ell}$ in those two graphs?
- Complete graph: $\bar{\ell} = 1$, empty graph $\bar{\ell} = \infty$
- These are theoretical extremal points

Extremal properties

- However, it is obvious how $\bar{\ell}$ and C will change when we start with an empty graph and add more links

Extremal properties

- However, it is obvious how $\bar{\ell}$ and C will change when we start with an empty graph and add more links
- $\bar{\ell}$ will go down and C will increase
- A more interesting question is how these statistics change when we rearrange a fixed number of links m
- We arrive at the last constraint for our model:
    4. m is fixed

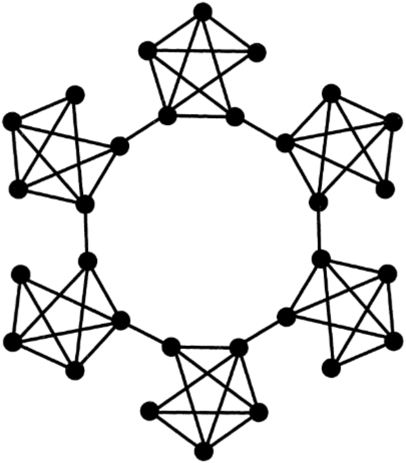
Small world networks

- For such graphs, we will now try to answer these questions:
  1. What is the most clustered graph and what are its C and $\bar{\ell}$?
  2. What graph has the smallest $\bar{\ell}$ and what are its C and $\bar{\ell}$?
  3. What is the relation between C and $\bar{\ell}$ in sparse graphs?

Sparse graph with the largest C

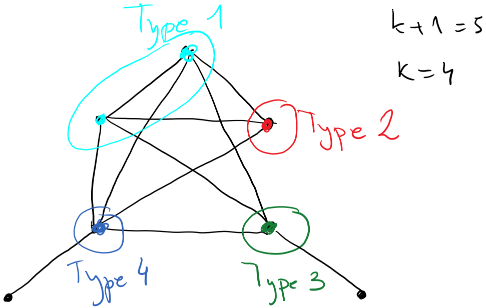
- We construct a *caveman world*
- It consists of $\frac{n}{\bar{k}+1}$ isolated caves (cliques)
- All $\bar{k}+1$ nodes in each cave are connected to all other nodes in that cave
- This graph has $C = 1$
- But it is not connected!
- We rewire a single link from each clique and connect it to a neighboring clique
- We form a loop

Sparse graph with the largest C



C in caveman world

- We first calculate the clustering coefficient of a single cave
- Later, we will use that result to obtain the total clustering coefficient by taking into account the number of caves



C in caveman world

- Each cave has $(k+1)$ nodes, there are $\frac{n}{k+1}$ caves altogether
- We have 4 types of nodes in each cave
- We have $(k-2)$ of nodes of type 1, and one node each of types 2, 3, and 4
- We recollect that the number of possible links between neighbors of a node with degree k is $\frac{k(k-1)}{2}$

$$
\begin{aligned}
C_1 &= \frac{\frac{k(k-1)}{2} - 1}{\frac{k(k-1)}{2}} = \frac{\frac{k(k-1)-2}{\cancel{2}}}{\frac{k(k-1)}{\cancel{2}}} = 1 - \frac{2}{k(k-1)} \\
C_2 &= 1 \\
C_3 &= \frac{\frac{(k-1)(k-2)}{2}}{\frac{k(k-1)}{2}} = \frac{\frac{\cancel{(k-1)}(k-2)}{\cancel{2}}}{\frac{k\cancel{(k-1)}}{\cancel{2}}} = 1 - \frac{2}{k}
\end{aligned}
$$

C in caveman world

$$
\begin{aligned}
C_4 &= \frac{\frac{k(k-1)-2}{\cancel{2}}}{\frac{(k+1)k}{\cancel{2}}} = \frac{k^2-k-2}{(k+1)k} = \frac{k^2+k-2k-2}{(k+1)k} \\
&= \frac{k(k+1)-2(k+1)}{(k+1)k} = \frac{\cancel{(k+1)}(k-2)}{\cancel{(k+1)}k} = 1-\frac{2}{k}
\end{aligned}
$$

C in caveman world

- Now we sum clustering coefficients of all nodes in a single cave

$$
\begin{aligned}
\sum_{cave} C_i &= \left(1 - \frac{2}{k(k-1)}\right)(k-2) + 2\left(1 - \frac{2}{k}\right) + 1 \\
&= k - 2 - (k-2)\frac{2}{k(k-1)} + 3 - \frac{4}{k} \\
&= (k+1) - \frac{2}{k-1} + \frac{4}{k(k-1)} - \frac{4}{k} \\
&= (k+1) + \frac{-2k + 4 - 4(k-1)}{k(k-1)} = (k+1) + \frac{-6k+8}{k(k-1)} \\
&= (k+1) - \frac{6\cancel{k}}{\cancel{k}(k-1)} + \frac{8}{k(k-1)} \\
&= (k+1) - \frac{6}{k-1} + \frac{8}{k(k-1)}
\end{aligned}
$$

C in caveman world

- The third term is $O(k^{-2})$ and under the assumption that $k \gg 1$ can be ignored
- Thus, the approximated sum of clustering coefficients over a cave:

$$\sum_{cave} C_i \approx (k+1) - \frac{6}{k-1}$$

C in caveman world

- The total clustering coefficient is then sum over all caves divided by the number of nodes

$$
\begin{aligned}
C_{caveman} &\approx \frac{\cancel{n}}{k+1}((k+1) - \frac{6}{k-1})\frac{1}{\cancel{n}} \\
&= 1 - \frac{6}{k^2-1}
\end{aligned}
$$

- Again, assuming $k \gg 1$ the clustering coefficient is close to 1 as we expected

$\overline{\ell}$ in caveman world

- The average distance in a caveman world is composed of a local distance within the caves and a global distance between the caves
- We first calculate $\overline{\ell}_{local}$
- One link is missing and therefore we have 1 pair of nodes at distance 2 and the remaining $(\frac{k(k+1)}{2} - 1)$ pairs at distance 1:

$$
\begin{aligned}
\overline{\ell}_{local} &= \frac{2}{k(k+1)}\left[(\frac{k(k+1)}{2} - 1) \cdot 1 + 1 \cdot 2\right] \\
&= \frac{2}{k(k+1)}\left[(\frac{k(k+1)}{2} + 1\right] \\
&= 1 + \frac{2}{k(k+1)}
\end{aligned}
$$

- Assuming $k \gg 1$ we have $\overline{\ell}_{local} \approx 1$

$\overline{\ell}$ in caveman world

- To calculate $\overline{\ell}_{global}$ we first abstract caves as simple nodes
- In this way $n' = \frac{n}{k+1}$ caves are ordered into a topological ring
- For simplicity (without loss of generality) we assume that n' is even
- Then $\overline{\ell}_{global}$ determines average distance between caves
- We start by calculating sum of distances of a single node i from the ring

$\bar{\ell}$ in caveman world

$$
\begin{aligned}
\sum_j \ell_{ij} &= 2 \cdot 1 + 2 \cdot 2 + 2 \cdot 3 + \cdots + 2(\frac{n'}{2} - 1) + 1 \cdot \frac{n'}{2} \\
&= 2(1 + 2 + 3 + \cdots + \frac{n'}{2}) - \frac{n'}{2} \\
&= \cancel{2}\frac{(\frac{n'}{2} + 1)\frac{n'}{2}}{\cancel{2}} - \frac{n'}{2} \\
&= \frac{n'}{2}\left(\frac{n'}{2} + \cancel{1} - \cancel{1}\right) \\
&= \frac{n'^2}{4}
\end{aligned}
$$

$\overline{\ell}$ in caveman world

- Then, we sum over all nodes and divide by the total number of pairs (in both directions):

$$\begin{aligned}\overline{\ell}_{global} &= \cancel{n'}\frac{n'^2}{4}\frac{1}{\cancel{n'}(n'-1)} = \frac{n'^2}{4(n'-1)} \\ &= \frac{\left(\frac{n}{k+1}\right)^2}{4\left(\frac{n}{k+1}-1\right)}\end{aligned}$$

- Assuming $1 \ll k \ll n$ we have $\left(\frac{n}{k+1}-1\right) \approx \frac{n}{k+1}$
- This gives $\overline{\ell}_{global} \approx \frac{n}{4(k+1)}$

$\overline{\ell}$ in caveman world

- Now we want to calculate average distance for nodes i and j from two different caves
- Starting at i we first need to go outside its cave: for this we need $\overline{\ell}_{local}$ steps on average
- To reach the cave in which j is situated we need to make $2\overline{\ell}_{global} - 1$ steps (one step to reach a cave and another to go through and -1 because we do not go through the last cave)
- To reach j in the last cave we need another $\overline{\ell}_{local}$ steps on average

$\bar{\ell}$ in caveman world

- Altogether we have:

$$
\begin{aligned}
\ell_{ij} &= 2\bar{\ell}_{local} + 2\bar{\ell}_{global} - 1 \\
&\approx 2 + 2\frac{n}{4(k+1)} - 1 = \frac{n}{2(k+1)} + 1
\end{aligned}
$$

- Assuming $1 \ll k \ll n$ we have $\ell_{ij} \approx \frac{n}{2(k+1)}$

$\bar{\ell}$ in caveman world

- Now we need to count how many pairs of nodes are from the same cave and how many from different caves

$$\#(local - pairs) = \frac{k\cancel{(k+1)}}{2}\frac{n}{\cancel{k+1}} = \frac{nk}{2}$$

- Then $\#(global - pairs)$ is the total number of pairs minus $\#(local - pairs)$:

$$\#(global - pairs) = \frac{n(n-1)}{2} - \frac{nk}{2} = \frac{n(n-k-1)}{2}$$

$\bar{\ell}$ in caveman world

- Finally, we have all terms to calculate $\bar{\ell}_{caveman}$:

$$\begin{aligned}\bar{\ell}_{caveman} &\approx \frac{\cancel{2}}{\cancel{n}(n-1)}\left[\frac{\cancel{n}k}{\cancel{2}}\cdot 1+\frac{\cancel{n}(n-k-1)}{\cancel{2}}\frac{n}{2(k+1)}\right]\\ &= \frac{k}{n-1}+\frac{n(n-k-1)}{2(k+1)(n-1)}\end{aligned}$$

- Assuming $k \ll n$ we have $(n-1)\approx n$, $(n-k-1)\approx n$ and $\frac{k}{n-1}\approx 0$
- After canceling: $\bar{\ell}_{caveman} \approx \frac{n}{2(k+1)}$

Sparse graph with the largest C

- Thus, for the connected caveman graph we have:

$$
\begin{aligned}
C_{caveman} &\approx 1 - \frac{6}{(\bar{k})^2 - 1} \\
\bar{\ell}_{cavemen} &\approx \frac{n}{2(\bar{k}+1)}
\end{aligned}
$$

- $C_{caveman}$ tends to 1 for sufficiently large $\bar{k} \ll n$
- $\bar{\ell}_{cavemen}$ scales linearly with n

Sparse graph with the smallest $\bar{\ell}$

- Studies have shown that no general structure possesses the smallest $\bar{\ell}$
- Cerf et al. "'A lower bound on the average shortest path length in regular graphs"'
- A good approximation can be achieved by a random graph $G(n, p)$
- Bolobas, Bela. "'Random Graphs"'

Sparse graph with the smallest $\overline{\ell}$

- $\overline{\ell}_{random} \approx \frac{ln(n)}{ln(\overline{k})}$
- $C_{random} \approx \frac{\overline{k}}{n-1}$
- For large n
- Scaling of $\overline{\ell}_{random}$ is logarithmic in n
- Sparsity: $\overline{k} \ll n \implies C_{random}$ is very small

Conclusions from extreme cases

- C is a simple measure of *local order* in a graph
- Large C as in caveman graphs indicates a strong local order
- On the other hand, a random graph is locally disordered
- Intuition 1: Highly clustered (locally ordered) graphs will have long average distances (linear scaling)
- Intuition 2: Graphs with small average distances will have also a small clustering coefficient (no clustering)
- No small-worlds after all?
- But many studies observed them!

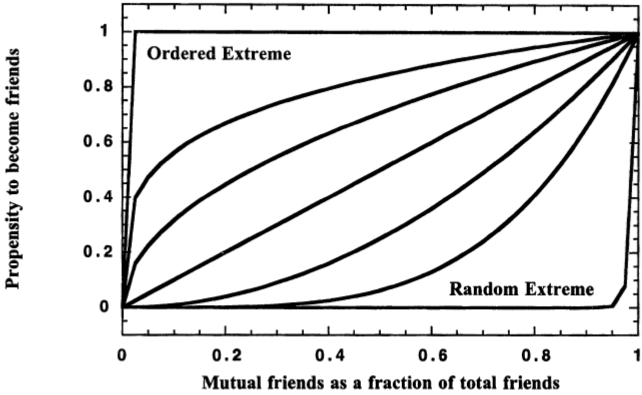
Modeling the real world

- Both extreme cases are not very realistic
- One extreme case: total order, in which two people become friends only if they share a common friend
- Another extreme case: randomness, in which two people become friends regardless of connections that they already have
- The real world is somewhere between those two extreme cases
- But we do not know exactly where is the reality: it has both of these mechanisms but we do not know to what extents
- Let us model **all** the situations in between
- Keep the model simple: we will introduce a single parameter

Probability of becoming friends

- The question is: how already existing links influence creation of new links
- We want to model all intermediate stages between order and randomness with a single parameter
- Total order: new friends only if mutual friends (Caveman world)
- Randomness: new friends completely autonomously (Solaria World)

Probability of becoming friends



Probability of becoming friends

- Order:
  1. Probability of becoming friends if no mutual friends is almost zero
  2. With one friend in common, probability of becoming friend jumps to almost one and stays there
- Randomness:
  1. No preference to become friends to anybody in particular
  2. If all friends are mutual, friends probability climbs to one

Probability of becoming friends

- In between the curve can take any of the infinite numbers of possible forms
- It needs to remain smooth and monotonically increasing
- Single tunable parameter $\alpha \in [0, \infty]$

$$R_{ij} = \begin{cases} 1 & m_{ij} \geq \overline{k} \\ \left(\frac{m_{ij}}{\overline{k}}\right)^{\alpha} (1-p) + p & \overline{k} > m_{ij} > 0 \\ p & m_{ij} = 0 \end{cases} \tag{1}$$

- R_{ij} probability of node i connecting to node j, m_{ij} the number of mutual friends, $\overline{k}$ is the average degree, and p is a baseline probability for a link (i, j)

Probability of becoming friends

- What do we have for a small α or $\alpha = 0$

Probability of becoming friends

- What do we have for a small α or $\alpha = 0$
- Total order
- What do we have for $\alpha \to \infty$

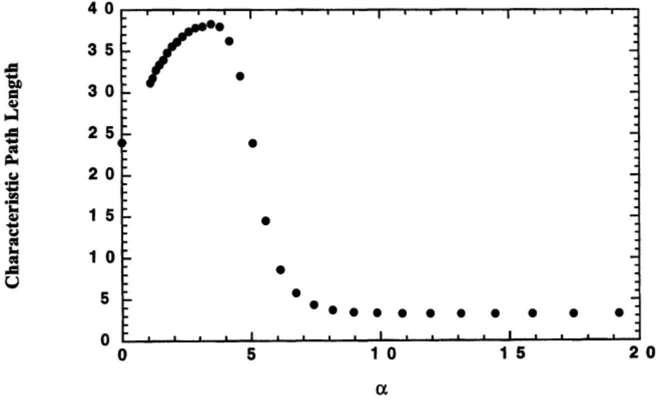
Probability of becoming friends

- What do we have for a small α or $\alpha = 0$
- Total order
- What do we have for $\alpha \to \infty$
- Random graph

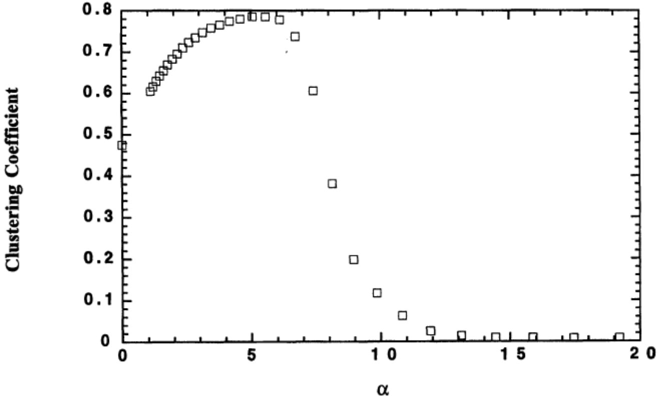
Numerical simulation

- We can not derive much analytically
- Thus, we construct a large number of graphs by fixing n and $\bar{k}$ and varying α
- For each α we create a number of graphs and calculate $\bar{\ell}$ and C
- We then plot averages of these experiments against α
- Another problem: if we start from an empty graph we will end up with an unconnected graph for small α
- We start with a ring

Numerical simulation



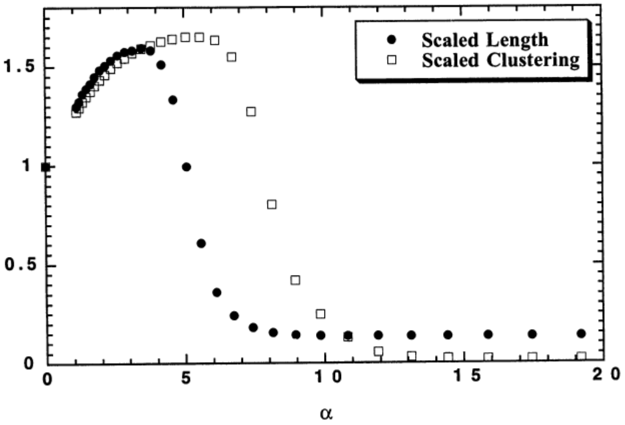
Numerical simulation



Numerical simulation

- For large α both statistics approach their expected random graph limit
- At $\alpha = 0$ both statistics are large and increase quickly to their maximum at small α
- Both statistics exhibit a sharp transition (phase transition) from their maximum values to their limits for large α

Numerical simulation



Numerical simulation

- Phase transitions of C and $\bar{\ell}$ are shifted with α
- The transition of C occurs with larger values of α
- Thus, there exists a class of graphs in a specific region of α for which $\bar{\ell}$ is small and C is large
- The region is limited by a smaller α for which phase transition in $\bar{\ell}$ occurs and a larger value of α for which phase transition for C occurs

Small world networks

Small world networks

The small-world phenomenon is present when:

$$\begin{aligned}\bar{\ell} &\approx \bar{\ell}_{random} \\ C &\gg C_{random}\end{aligned}$$

Short Recap

- Now we know that small networks exist
- We also understand something about ratio of order and randomness
- But we still do not know why there is a distance contraction
- While simultaneously clustering coefficient remains large
- Any ideas?

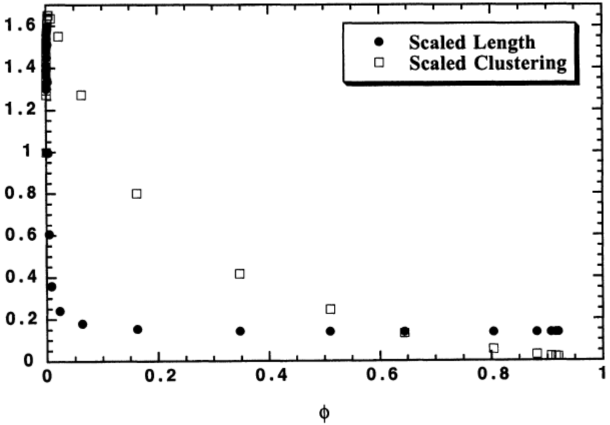
Theory of distance contraction

- Let us investigate how a newly created link contributes to the contraction of the average distance
- We start with $\alpha = 0$
- $R_{ij} = 1$ if $m_{ij} > 0$, i.e. if i and j share mutual friends they will be connected (triadic closure)
- Before the $\ell_{ij} = 2$, after $\ell_{ij} = 1$
- Little to no distance contraction globally
- In a random graph i and j that are close have the same chance to be connected by a new link as k and l that are far apart

Theory of distance contraction

- Now we define a range r of a link (i, j) as the distance ℓ_{ij} between i and j when the link has been deleted
- We define a link as a shortcut if its range $r > 2$
- Finally, we define Φ as the fraction of links that are shortcuts, i.e. $\Phi = \frac{\#shortcuts}{m}$
- Now we plot the evolution of C and $\bar{\ell}$ as the function of Φ

Theory of distance contraction



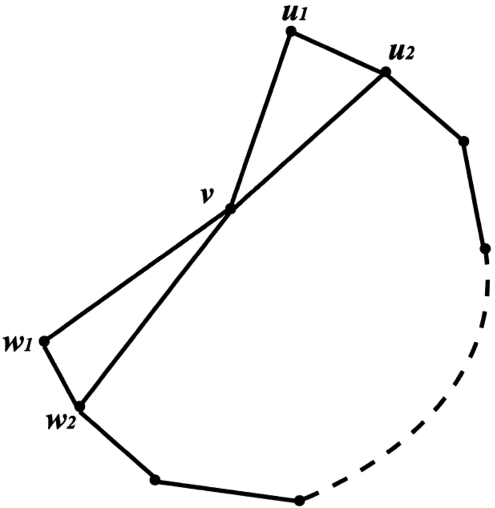
Theory of distance contraction

- We see that the distance contraction occurs already for a very small fraction of shortcuts
- Intuition is that for small Φ the average distance is large
- Introduction of even a single shortcut brings nodes close that were previously widely separated
- This shortcut reduces the distance not only between two nodes that are connected
- But also between their friends, friends of friends, etc.
- On the other hand, clustering coefficient does not drop dramatically since only a couple of triads are not closed

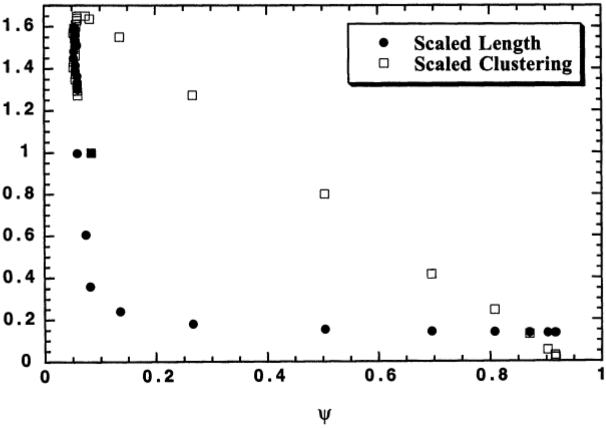
Theory of distance contraction

- Shortcuts are sufficient but not necessary
- Any link that brings two nodes closer together will do the job
- A contraction occurs when the second shortest path length between two nodes (sharing a common neighbor) is greater than two
- In other words, a contraction is a pair of nodes that share one and only one common friend
- We define Ψ as the fraction of contractions

Theory of distance contraction



Theory of distance contraction



Theory of distance contraction

- Common friend in a contraction $\implies$ common friend is pivotal
- Common friend in a contraction $\implies$ common friend is a local gatekeeper
- The other direction does not hold
- E.g. a pivotal node that is not a common friend
- Pair of nodes having more than one local gatekeepers

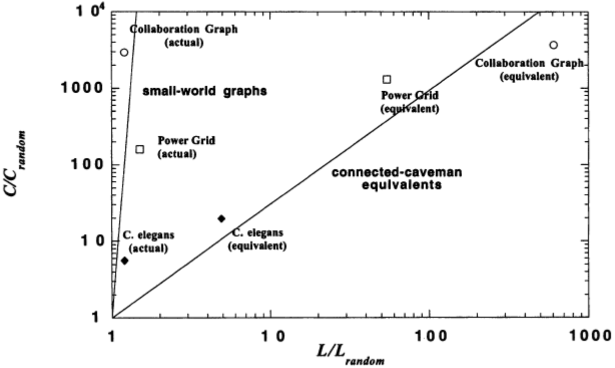
Small world in empirical networks

Table 1 Empirical examples of small-world networks

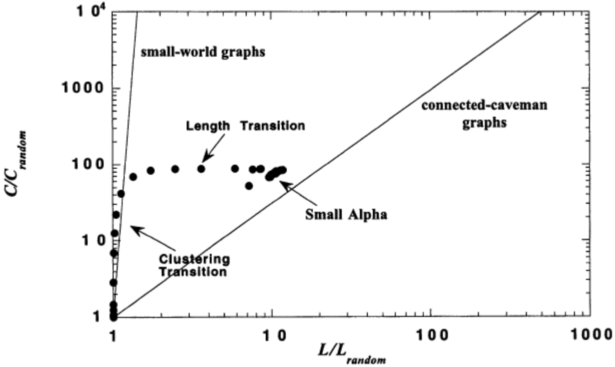
	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

Characteristic path length L and clustering coefficient C for three real networks, compared to random graphs with the same number of vertices (n) and average number of edges per vertex (k). (Actors: $n = 225{,}226$, $k = 61$. Power grid: $n = 4{,}941$, $k = 2.67$. *C. elegans*: $n = 282$, $k = 14$.) The graphs are defined as follows. Two actors are joined by an edge if they have acted in a film together. We restrict attention to the giant connected component[16] of this graph, which includes ~90% of all actors listed in the Internet Movie Database (available at http://us.imdb.com), as of April 1997. For the power grid, vertices represent generators, transformers and substations, and edges represent high-voltage transmission lines between them. For *C. elegans*, an edge joins two neurons if they are connected by either a synapse or a gap junction. We treat all edges as undirected and unweighted, and all vertices as identical, recognizing that these are crude approximations. All three networks show the small-world phenomenon: $L \gtrsim L_{\text{random}}$ but $C \gg C_{\text{random}}$.

Small world in empirical networks



Small world in empirical networks



Watts' β-model

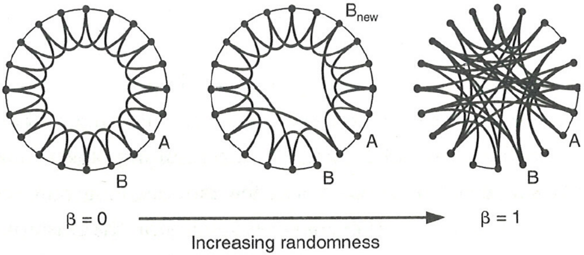


Figure 3.6. Construction of the beta model. The links in a one-dimensional, periodic lattice are randomly rewired with probability beta (β). When beta is zero (left), the lattice remains unchanged, and when beta is one (right), all links are rewired, generating a random network. In the middle, networks are partly ordered and partly random (for example, the original link from A to B has been rewired to B_{new}).

Watts' β-model

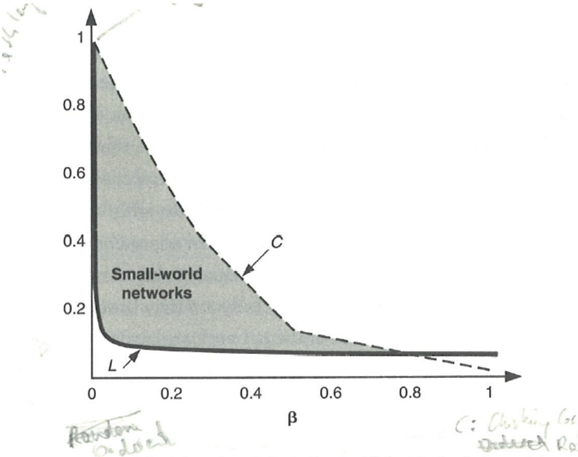


Figure 3.7. Path length and clustering coefficient in the beta model. As with the alpha model (see Figure 3.4), small-world networks exist when path length is small and the clustering coefficient is large (shaded region).

Watts' β-model

- NetLogo Example
- http://ccl.northwestern.edu/netlogo/models/SmallWorlds

Affiliation networks and small world

- Separate social and network structure
- Social structure implies two types of nodes, e.g. actors and movies
- Actors are connected to movies they acted in, and vice versa
- Co-Acting network is then constructed in the following way
  1. A given actor is connected to all actors from a given movie
  2. We repeat this procedure for all movies

Affiliation networks and small world

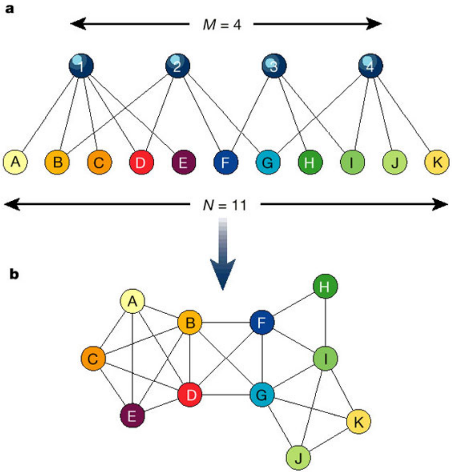


Figure: $\ell = 1.62$, $C = 0.7879$

Affiliation networks and small world

- Small-world networks arise naturally
- Divide nodes in two groups that reflect the social context, e.g. actors and movies, people and professions (hobbies), etc.
- But also information networks, e.g. tags and photos, hashtags and tweets in twitter, etc.
- Projection on one type of nodes, e.g. actors, people, tags, hashtags is always a small-world
- Why is this the case?

Affiliation networks and small world

- By definition every actor in a movie is connected to every other actor in that movie
- This is a fully connected clique of actors – local clustering is high
- Networks are then networks of overlapping cliques – locked together by actors acting in multiple movies
- By "randomly" connecting actors to movies we obtain a network with low diameter
- High local clustering + low diameter = small-world networks

Affiliation networks and small world

- NetLogo Example
- http://kti.tugraz.at/staff/socialcomputing/courses/webscience/SWAffiliation.nlogo

Applications and engineering

- We have learned what are small world networks and how they emerge?
- In what kind of applications can these new insights be applied?
- Many different possibilities, e.g. information retrieval on the Web – navigation
- Information diffusion in online social networks, e.g. viral marketing

Recommender systems

- In many recommender systems networks look very much like Cavemen World
- Isolated caves of similar and related items
- But almost no connections to other caves
- We have learned that a few (random) long-range links can turn such a world into a small world
- Serendipity in recommender networks
- How to have a surprise effect and connect various caves?