Small World Problem
Lecture: Social Web and Bibliometrics

University of Koblenz-Landau,
SS 2013

York Sure-Vetter

“every person on earth is connected to any other person through a chain of acquaintances not longer than 6”? 
Overview

Topics

• Definition of the Small World Problem
• Results from a social experiment
Do I know somebody in …?
The Bacon Number

http://www.imdb.com/name/nm0000102/

Kevin Bacon

Date of Birth: 8 July 1958, Philadelphia, Pennsylvania, USA

Mini Biography: Kevin Bacon's early training as an actor came from The Manning Street.

Trivia: His line, "I am a G-damn genius," is quoted in both "Hollow Man" (2000)

Awards: Nominated for Golden Globe. Another 8 wins & 7 nominations

Alternate Names: The Bacon Brothers / Kevin Bacon III / the bacon brothers

Overview

Filmography

3. Saving Angels (2007) (completed) ... Brent
4.Rails & Ties (2007) ...
5. Death Sentence (2007) ... Nick Hume
6. The Art of Breathing (2007) ... Live
7. Where the Truth Lies (2005) ... Larry
8. Beauty Shop (2005) ... Jorge

Bibliometrics - York Sure-Vetter and Andreas Strotmann
Six Degrees of Kevin Bacon

Six Degrees of Kevin Bacon is a trivia game based on the concept of the small world phenomenon and rests on the assumption that any individual can be linked through his or her film roles to actor Kevin Bacon within six steps. The name of the game is a play on the "six degrees of separation" concept. In 2007, Bacon started a charitable organization named SixDegrees.org.

The game requires a group of players to try to connect any individual to Kevin Bacon as quickly as possible and in as few links as possible. The fantasy author-editor Richard Gilliam devised his Movie Links online game in 1990, and it was played extensively on GENie for years before the quite similar Six Degrees of Kevin Bacon game was promoted in 1994. Gilliam's game is much more difficult in that a player is required to find the shortest number of movies linking actors as diverse as, say, John Candy and Chris Farley, rather than continual links to the same specific actor.
OracleOfBacon.org
The Bacon Number
[Watts 2002]

<table>
<thead>
<tr>
<th>BACON NUMBER</th>
<th>NUMBER OF ACTORS</th>
<th>CUMULATIVE TOTAL NUMBER OF ACTORS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>1,550</td>
<td>1,551</td>
</tr>
<tr>
<td>2</td>
<td>121,661</td>
<td>123,212</td>
</tr>
<tr>
<td>3</td>
<td>310,365</td>
<td>433,577</td>
</tr>
<tr>
<td>4</td>
<td>71,516</td>
<td>504,733</td>
</tr>
<tr>
<td>5</td>
<td>5,314</td>
<td>510,047</td>
</tr>
<tr>
<td>6</td>
<td>652</td>
<td>510,699</td>
</tr>
<tr>
<td>7</td>
<td>90</td>
<td>510,789</td>
</tr>
<tr>
<td>8</td>
<td>38</td>
<td>510,827</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>510,828</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>510,829</td>
</tr>
</tbody>
</table>
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)
- ~ 1500 Publications
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
Convenient: Amenable to computational analysis

What is my Erdös Number?
⇒ 4
me -> Pascal Hitzler -> Guo-Qiang Zhang -> E. Rodney Canfield -> Paul Erdös
The Erdös Number

PS: Don‘t forget to check this comic out …

http://www.xkcd.com/599/
Stanley Milgram

- A social psychologist
- Yale and Harvard University

- Study on the Small World Problem, beyond well defined communities and relations (such as actors, scientists, …)

- Controversial: The Obedience Study

- What we will discuss today: „An Experimental Study of the Small World Problem“
Introduction

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.

[ Milgram 1967, according to http://www.ils.unc.edu/dpr/port/socialnetworking/theory_paper.html#2 ]
An Experimental Study of the Small World Problem [Travers and Milgram 1969]

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
• Define a single target person and a group of starting persons
• Generate an acquaintance chain from each starter to the target

Experimental Set Up
• Each starter receives a document
• was asked to begin moving it by mail toward the target
• Information about the target: name, address, occupation, company, college, year of graduation, wife’s name and hometown
• Information about relationship (friend/acquaintance) [Granovetter 1973]

Constraints
• starter group was only allowed to send the document to people they know and
• 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:
  - A Boston stockbroker
- Three starting populations
  - 100 “Nebraska stockholders”
  - 96 “Nebraska random”
  - 100 “Boston random”
Results I

• How many of the starters would be able to establish contact with the target?
  – 64 out of 296 reached the target

• How many intermediaries would be required to link starters with the target?
  – Well, that depends: the overall mean 5.2 links
  – Through hometown: 6.1 links
  – Through business: 4.6 links
  – Boston group faster than Nebraska groups
  – Nebraska stockholders not faster than Nebraska random
Results II

- What form would the distribution of chain lengths take?
Results III

- Incomplete chains

What reasons can you think of for incomplete chains?
Results IV

• Common paths
• Also see: Gladwell’s “Law of the few”

"The Law of the Few", or, as Gladwell states, "The success of any kind of social epidemic is heavily dependent on the involvement of people with a particular and rare set of social gifts." According to Gladwell, economists call this the "80/20 Principle, which is the idea that in any situation roughly 80 percent of the 'work' will be done by 20 percent of the participants."
6 degrees of separation

- So is there an upper bound of six degrees of separation in social networks?

  - Extremely hard to test
  - In Milgram’s study, ~2/3 of the chains didn’t reach the target
  - 1/3 random, 1/3 blue chip owners, 1/3 from Boston
  - Danger of loops (mitigated in Milgram’s study through chain records)
  - Target had a “high social status” [Kleinfeld 2000]
Question: How can we distinguish small world networked from other networks?

Answer: Duncan J. Watts and Steven Strogatz introduced the clustering coefficient in 1998 to determine whether a graph is a small world network.
(Local) Clustering Coefficient

The **local clustering coefficient** of a *vertex* in a *graph* quantifies how close its *neighbors* are to being a *clique* (complete graph).

Examples:

- $c = 1$
- $c = \frac{1}{3}$
- $c = 0$

Network Average Clustering Coefficient

The clustering coefficient for the whole network is given by Watts and Strogatz as the average of the local clustering coefficients of all the vertices \( n \):

\[
\bar{C} = \frac{1}{n} \sum_{i=1}^{n} C_i.
\]

Follow up work (2008)

- 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
- Largest social network constructed and analyzed to date (2008)
Follow up work (2008)

- the **clustering coefficient** decays very slowly with exponent $-0.37$ with the degree of a node and the average clustering coefficient is 0.137.

- This result suggests that clustering in the Messenger network is much higher than expected—that **people with common friends also tend to be connected**.

![Graphs](image-url)

Figure 19: (a) Clustering coefficient; (b) distribution of connected components. 99.9% of the nodes belong to the largest connected component.
Follow up work (2008)

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.
- 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.
- Result: there are about “7 degrees of separation” among people.
- long paths exist in the network; in the experiment paths up to a length of 29 have been found.
Wikipedia on „Small World Networks“

In mathematics, physics and sociology 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. A small world network, where nodes represent people and edges connect people that know each other, captures the **small world phenomenon of strangers being linked by a mutual acquaintance**.

Many empirical graphs are well modeled by small-world networks. Social networks, the connectivity of the Internet, and gene networks all exhibit small-world network characteristics.

Network Average Clustering Coefficient and Small World Networks

The clustering coefficient for the whole network is given by Watts and Strogatz as the average of the local clustering coefficients of all the vertices $n$:

$$\bar{C} = \frac{1}{n} \sum_{i=1}^{n} C_i.$$

A graph is considered small-world, if its average clustering coefficient $\bar{C}$ is significantly higher than a random graph constructed on the same vertex set, and if the graph has approximately the same mean-shortest path length as its corresponding random graph.

Small Worlds
http://www.infosci.cornell.edu/courses/info204/2007sp/

• Every pair of nodes in a graph is connected by a path with an extremely small number of steps (low diameter)
• Two principle ways of encountering small worlds
  – Dense networks
  – Sparse networks with well-placed connectors
Small Worlds
[Newman 2003]

• The small-world effect exists, if
  – “The number of vertices within a distance $r$ of a typical central vertex grows exponentially with $r$ (the larger it get, the faster it grows) $x(t) = x_0 e^{kt}$

In other words:
  – Networks are said to show the small-world effect if the value of $l$ (avg. shortest distance) scales logarithmically or slower with network size for fixed mean degree $e^{\ln(x)} = x$ if $x > 0$

![Graph showing the small-world effect](image)

Example for base e
Formalizing the Small World Problem
[Watts and Strogatz 1998]

The small-world phenomenon is assumed to be present when

\[ L > L_{\text{random}} \text{ but } C >> C_{\text{random}} \]

Or in other words: We are looking for networks where local clustering is high and global path lengths are small

What’s the rationale for the above formalism?

One potential answer:
Cavemen and Solaris Worlds
The Solaris World
Random Social Connections

How do random social graphs differ from „real“ social networks?

http://vimeo.com/9669721

Solaris is a Random Graph

In mathematics, a random graph is a graph that is generated by some random process.

[...] A random graph is obtained by starting with a set of $n$ vertices and adding edges between them at random.

Solaris World

The Solaris World can be described as:

\[\text{The influence of existing friends on new friendships is indistinguishable from random chance.}\]

See e.g.: [http://www.rule110.org/amhso/results/small-worlds-intro.pdf](http://www.rule110.org/amhso/results/small-worlds-intro.pdf)
The Cave Men World
Highly Clustered Social Connections

See e.g.: http://mathworld.wolfram.com/CavemanGraph.html
Cave Men World

The Cave Men World can be described as:

\textit{Everybody you know knows everybody else you know, but no one else.}

However, there are some „weaker“ forms possible, e.g. the „connected Cave Men World“:

\textit{Everybody you know knows everybody else you know and some know also other people.}
Solaris World vs. Cave Men World

Actually in most cases the truth lies somewhere in between.

Both worlds are specific models of the „alpha-Modell“ for social network graphs: The alpha-Modell is designed to construct a network similar to real social networks: new edges are formed based on a function of the currently existing network.

Your current friends determine to a certain extent your new friends.
Informal definition: SWP exists when every pair of nodes in a graph is connected by a path with an extremely small number of steps.

Problem: Does not take searchability into account. Random networks are hard to search with local knowledge.

Two seemingly contradictory requirements for the Small World Phenomenon:

- It should be possible to connect two people chosen at random via chain of only a few intermediaries (as in Solaria world)
- Network should display a large clustering coefficient, so that a node's friends will know each other (as in Caveman world)
Formalizing the Small World Problem

[Watts 2003]

Two seemingly contradictory requirements for the Small World Phenomenon:

- It should be possible to connect two people chosen at random via chain of only a few intermediaries (as in Solaria world)
- Network should display a large clustering coefficient, so that a node’s friends will know each other (as in Caveman world)

Under which conditions can these two requirements be reconciled?
Formalizing the Small World Problem

[Watts 2003]

- Page 76 -82
- The alpha parameter
- Path length: computed only over nodes in the same connected component

![Diagram showing path length as a function of alpha (\(\alpha\)). At the critical alpha value, many small clusters join to connect the entire network, whose length then shrinks rapidly.](image)
Formalizing the Small World Problem

[Watts 2003]

• Page 76 -82

• Comparison between path length and clustering coefficient

Small World Phenomenon exists when

$L > L_{\text{random}}$ but $C >> C_{\text{random}}$

Q: Why does this area not qualify to represent a small world network?

A: Not all components are connected yet (unconnected caves)
Demo – Small Worlds the Alpha Model

http://kmi.tugraz.at/staff/markus/demos/sw-alpha.htm

Small World Simulation - The Alpha Model
Examples for Small World Networks

Table 1 Empirical examples of small-world networks

<table>
<thead>
<tr>
<th></th>
<th>$L_{\text{actual}}$</th>
<th>$L_{\text{random}}$</th>
<th>$C_{\text{actual}}$</th>
<th>$C_{\text{random}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Film actors</td>
<td>3.65</td>
<td>2.99</td>
<td>0.79</td>
<td>0.00027</td>
</tr>
<tr>
<td>Power grid</td>
<td>18.7</td>
<td>12.4</td>
<td>0.080</td>
<td>0.005</td>
</tr>
<tr>
<td>C. elegans</td>
<td>2.65</td>
<td>2.25</td>
<td>0.28</td>
<td>0.05</td>
</tr>
</tbody>
</table>

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 of this graph, which includes $\sim$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: $L > L_{\text{random}}$ but $C > C_{\text{random}}$. 
Contemporary Software

• Where does the small-world phenomenon come into play in contemporary software, in organizations, ..?

• Wer-kennt-Wen, Xing, LinkedIn, Myspace, Facebook, FOAF, …

• Business Processes, Information and Knowledge Flow
Question: How do Small World Networks form?

Answer: especially by …

  Preferential Attachment,
  Assorciative Mixing,
  Disassortativity, and
  Weak Ties
Preferential Attachment  
[Barabasi 1999]

„The rich getting richer“

Preferential Attachment refers to the high probability of a new vertex to connect to a vertex that already has a large number of connections

Example:
1. a new website linking to more established ones
2. a new individual linking to well-known individuals in a social network
Preferential Attachment Example

Which node has the highest probability of being linked by a new node in a network that exhibits traits of preferential attachment?

A small example network with eight vertices and ten edges.

[Newman 2003]
Assortative Mixing (or Homophily) [Newman 2003]

Assortative Mixing refers to selective linking of nodes to other nodes who share some common property

- E.g. degree correlation
  high degree nodes in a network associate preferentially with other high-degree nodes
- E.g. social networks
  nodes of a certain type tend to associate with the same type of nodes (e.g. by race)
Assortative Mixing (or Homophily) [Newman 2003]

FIG. 8 Friendship network of children in a US school. Friendships are determined by asking the participants, and hence are directed, since A may say that B is their friend but not vice versa. Vertices are color coded according to race, as marked, and the split from left to right in the figure is clearly primarily along lines of race. The split from top to bottom is between middle school and high school, i.e., between younger and older children. Picture courtesy of James Moody.
Disassortativity
[Newman 2003]

Disassortativity refers to selective linking of nodes to other nodes who are different in some property

• E.g. the Web
  low degree nodes tend to associate with high degree nodes

• … do you have other examples?
But …

Isn’t all of this an oversimplification of the world of social systems?

- Ties/relationships vary in intensity
- People who have strong ties tend to share a similar set of acquaintances
- Ties change over time
- Nodes (people) have different characteristics, and they are *actors*
- …
The Strength of Weak Ties
[Granovetter 1973]

The strength of an interpersonal tie is a (probably linear) combination of
– the amount of time,
– the emotional intensity,
– the intimacy, and
– the reciprocal services which characterize the tie

Can you give examples of strong / weak ties?

Mark Granovetter,
Stanford University
The Strength of Weak Ties and Mutual Acquaintances [Granovetter 1973]

Consider:
Two arbitrarily selected individuals A and B and
The set $S = C,D,E$ of all persons with ties to either or both of them.

Hypothesis:
The stronger the tie between A and B, the larger the proportion of individuals in S to whom they will both be tied.

Theoretical corroboration:
Stronger ties involve larger time commitments – probability of B meeting with some friend of A (who B does not know yet) is increased.
The stronger a tie connecting two individuals, the more similar they are.
The Strength of Weak Ties
[Granovetter 1973]

The forbidden triad

![Forbidden Triad Diagram](image)

Why is it called the forbidden triad?
Bridges
[Granovetter 1973]

A bridge is a line in a network which provides the only path between two points.

In social networks, a bridge between A and B provides the only route along which information or influence can flow from any contact of A to any contact of B.

Which edge represents a bridge? Why?
Bridges and Strong Ties
[Granovetter 1973]

Example:
1. Imagine the strong tie between A and B
2. Imagine the strong tie between A and C
3. Then, the forbidden triad implies that a tie exists between C and B
   (it forbids that a tie between C and B does not exist)
4. From that follows, that A-B is not a bridge (because there is another path
   A-B that goes through C)

Why is this interesting?
⇒ Strong ties can be a bridge ONLY IF neither
   party to it has any other strong ties
⇒ Highly unlikely in a social network of any size
⇒ Weak ties suffer no such restriction, though
   they are not automatically bridges
⇒ But, all bridges are weak ties
In Reality ....
[Granovetetter 1973]

it probably happens only rarely, that a specific tie provides *the only path* between two points

**Local bridges**: the shortest path between its two points (other than itself)

- **Bridges are efficient paths**
- **Alternatives are more costly**
- **Local bridges of degree n**
- **A local bridge is more significant as its degree increases**

![Diagram](image-url)

**Fig. 2.**—Local bridges. \(a, \text{Degree 3}; \ b, \text{Degree 13.}\) —— = strong tie; ——— = weak tie.
In Reality …

Strong ties can represent *local* bridges BUT they are weak (i.e. they have a low degree)

Homework: Why?

What’s the degree of the local bridge A-B?
Implications of Weak Ties
[Granovetter 1973]

- Those weak ties, that are local bridges, create more, and shorter paths.
- The removal of the average weak tie would do more damage to transmission probabilities than would that of the average strong one.
- **Paradox:** While *weak ties* have been denounced as generative of alienation, *strong ties*, breeding local cohesion, lead to overall fragmentation.

Completion rates in Milgram’s experiment were reported higher for acquaintance than friend relationships [Granovetter 1973]
Implications of Weak Ties
[Granovetter 1973]

- Example: Spread of information/rumors in social networks
  - Studies have shown that people rarely act on mass-media information unless it is also transmitted through personal ties [Granovetter 2003, p 1274]
  - Information/rumors moving through strong ties is much more likely to be limited to a few cliques than that going via weak ones, bridges will not be crossed

How does information spread through weak ties?
Summary

What are Small Word Networks?
   Examples: Bacon, Erdös, Milgram …
Properties
Solaris vs. Cave Men

How do SWNs form?
   Preferential Attachment
   Assorciative Mixing
   Disassortativity
   Weak Ties