Small World Problem
Web Science (VU) (707.000)

Denis Helic
KTI, TU Graz
Mar 16, 2015
Do I know somebody in...?

Sa Li
Quality Gate Engineer
BENO Mobile Beijing
Beijing, China

Options
- Add as contact
- Send message
- Introduce
- Bookmark
- Show location
- Show route

Status: Employee
Wants: --
Haves: --
Company: BENO Mobile Beijing, Quality Gate Engineer
Industry: Electronics Industry
Previous companies: 1. EPSON, supervisor
Interests: Swimming, Badminton, Cooking, Reading, Call, Travel

Sa Li's statistics:
- No Premium Membership
- Profile hits: 220
- Direct contacts: 25
- Activity meter: 60%

Edit shared personal info:
- All contact details
- Business
- Private
- Instant messaging data
- Birthday
- Year of birth
- Maintenance

Denis Helic (KTI, TU Graz)
Small-World
Mar 16, 2015
Introduction

The Bacon Number

Kevin Bacon

Overview

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

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

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

Awards: Nominated for Golden Globe, Another 8 wins & 7 nominations more>

Alternate Names: The Bacon Brothers / Kevin Bacon II / the Bacon Brothers

Filmography

Jump to filmography as: Actor, Director, Producer, Soundtrack, Thanks, Self, Archive Footage

Actor:
2. Frost/Nixon (2008) [Running] ... Jack Brennan
3. Saving Angels (2007) [completed] ... Brant
5. Death Sentence (2007) ... Nick Humce
6. The Air I Breathe (2007) ... Love
7. Where the Truth Lies (2006) ... Lanny
8. Beauty Shop (2005) ... Jorge

Denis Helic (KTI, TU Graz)
The Kevin Bacon Game
The Bacon Number

<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 Erdős Number

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

- Through how many research collaboration links is an arbitrary scientist connected to Paul Erdős?
- What is a research collaboration link?
- Per definition: Co-authorship on a scientific paper
- What is my Erdős Number? 4
- me → H. Maurer → W. Kuich → N. Sauer → P. Erdős

http://www.ams.org/mathscinet/collaborationDistance.html
Erdös network

One of the largest such computational studies was performed by Jure Leskovec and Eric Horvitz [273]. They analyzed the 240 million active user accounts on Microsoft Instant Messenger, building a graph in which each node corresponds to a user, and there is an edge between two users if they engaged in a two-way conversation at any point during a month-long observation period. As employees of Microsoft at the time, they had access to a complete snapshot of the system for the month under study, so there were no concerns about missing data. This graph turned out to have a giant component containing almost all of the nodes, and the distances within this giant component were very small. Indeed, the distances in the Instant Messenger network closely corresponded to the numbers from Milgram's experiment, with an estimated average distance of 6.6, and an estimated median
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 stockholders
- Nebraska random
- Boston random
- Boston stockbroker
- Target
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

![Diagram showing the distribution of chain lengths]

- Number of chains on the y-axis
- Number of intermediaries on the x-axis
- N = 64
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 didn’t 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
- Largest social network constructed and analyzed to date
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
Figure 2.11: The distribution of distances in the graph of all active Microsoft Messenger user accounts, with an edge joining two users if they communicated at least once during a month-long observation period [273].

Step connections to CEOs and political leaders don’t yield immediate payoffs on an everyday basis, but the existence of all these short paths has substantial consequences for the potential speed with which information, diseases, and other kinds of contagion can spread through society, as well as for the potential access that the social network provides to opportunities and to people with very different characteristics from one’s own. All these issues—and their implications for the processes that take place in social networks—are rich enough that we will devote Chapter 20 to a more detailed study of the small-world phenomenon and its consequences.

One reason for the current empirical consensus that social networks generally are “small worlds” is that this has been increasingly confirmed in settings where we do have full data on the network structure. Milgram was forced to resort to an experiment in which letters served as “tracers” through a global friendship network that he had no hope of fully mapping on his own; but for other kinds of social network data where the full graph structure is known, one can just load it into a computer and perform the breadth-first search procedure to determine what typical...
Small world networks

- Every pair of nodes is connected by a path with an extremely small number of steps (low diameter)
- Nodes are clustered, e.g. people share friends
- Two principle ways of encountering small worlds
  1. Dense networks
  2. Sparse networks with well-placed connectors
The small-world effect exists, if the number of nodes within a distance \( r \) of a typical central vertex grows exponentially with \( r \). In other words, networks are said to show the small-world effect if the value of \( \ell \) (avg. shortest distance) scales logarithmically or slower with network size for fixed average degree.
The small-world phenomenon is assumed to be present when:

\[ \ell \approx \ell_{\text{random}} \]

\[ C \gg C_{\text{random}} \]
Small world networks

- 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
The Cavemen World: highly clustered social connections
Two seemingly contradictory requirements for the Small World Phenomenon:

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

Figure 3.1. Two extreme kinds of interaction rules. In the top curve (caveman world), even a single mutual friend implies that A and B are highly likely to meet. In the bottom curve (Solaria world), all interactions are equally unlikely, regardless of how many friends A and B share.
Watts’ $\alpha$-model

Figure 3.2. Between the two extremes, a whole family of interaction rules exists, each one specified by a particular value of the tuneable parameter $\alpha$. When $\alpha = 0$, we have a caveman world; when $\alpha$ becomes infinite, we have Solaria.
Watts’ $\alpha$-model

Figure 3.3. 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.
Watts’ $\alpha$-model

- NetLogo Example
- [http://kti.tugraz.at/staff/socialcomputing/courses/webscience/SmallWorld.nlogo](http://kti.tugraz.at/staff/socialcomputing/courses/webscience/SmallWorld.nlogo)
Figure 3.4. Comparison between path length ($L$) and clustering coefficient ($C$). The region between the curves, where $L$ is small and $C$ is large (shaded), represents the presence of 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:** $L \gg L_{\text{random}}$ but $C \gg C_{\text{random}}$. 
Watts’ $\beta$-model

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{watts_beta_model}
\caption{Construction of the beta model. The links in a one-dimensional, periodic lattice are randomly rewired with probability beta ($\beta$). 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 $B_{\text{new}}$).}
\end{figure}
Watts’ $\beta$-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’ $\beta$-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](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
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?