

707.000
Web Science and Web Technology
„The Small World Problem“

Markus Strohmaier

Univ. Ass. / Assistant Professor
Knowledge Management Institute
Graz University of Technology, Austria

e-mail: markus.strohmaier@tugraz.at
web: <http://www.kmi.tugraz.at/staff/markus>

Overview

Topics

- Definition of the Small World Problem
- Results from a social experiment
- The importance of „weak ties“

Knowledge Management Institute TU Graz

Do I know somebody in ...?

Logout | Contact | Help | Downloads | About us | Mobile | Language


XING[®] 17,518 members online

+ Invite contacts | Search [Name, company or search term] Find


Start | **Members** | Messages | Address Book | Groups | Events | Marketplace | PremiumWorld

Advanced Search | Powersearch | Search Agents


Your connection to Sa Li Show all 92 different connections to Sa Li




Dr. Markus Strohmaier
Graz University of Technology




Dr. Harald Holz
DFKI GmbH




Ivo Vollrath
SAF AG



Dr. Marcell Assan
MAN Nutzfahrzeuge



Sa Li
BENQ mobile beijing



Sa Li
Quality gate engineer
BENQ mobile beijing
Beijing, China

Options

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

Memo:
Create memo

Business details | Contact details | Web | About me | Guestbook

Status	Employee
Wants	---
Has	---
Company	BENQ mobile beijing: Quality gate engineer (06/2006 -)

Sa Li's statistics

- No Premium Membership
- Member since: 10/2006
- Profile hits: 220
- Direct contacts: 25

Knowledge Management Institute TU Graz

The Bacon Number

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



Markus Strohmaier



The Kevin Bacon Game

The oracle of Bacon

www.oracleofbacon.org



The Bacon Number [Watts 2002]

TABLE 3.1 DISTRIBUTION 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 Erdős Number

Who was Erdős?

<http://www.oakland.edu/enp/>

A famous 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)
- ~ 1500 Publications

The Erdős Number

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 ->
Convenient: Amenable to computational analysis

What is my Erdős Number?

→ 5

me -> S. Easterbrook -> A. Finkelstein -> D. Gabbay ->
S. Shelah -> P. Erdős

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



1933-1984

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

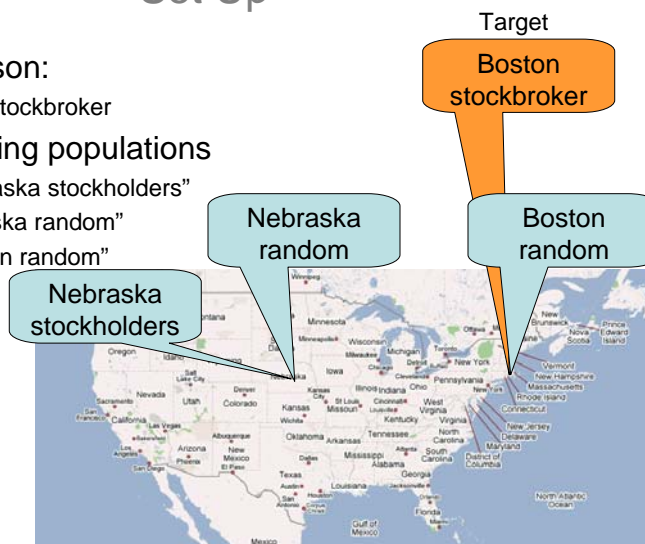
- starter group was only allowed to send the document to people they know and
- was urged to choose the next recipient in a way as to advance the progress of the document toward the target

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?

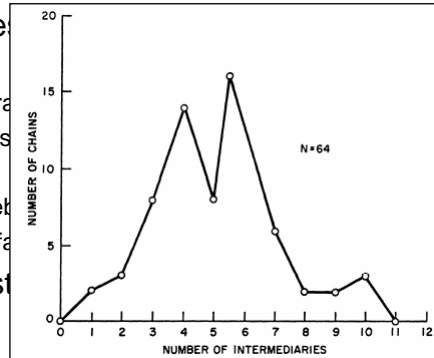
Set Up

- Target person:
 - A Boston stockbroker
- Three starting populations
 - 100 “Nebraska stockholders”
 - 96 “Nebraska random”
 - 100 “Boston random”



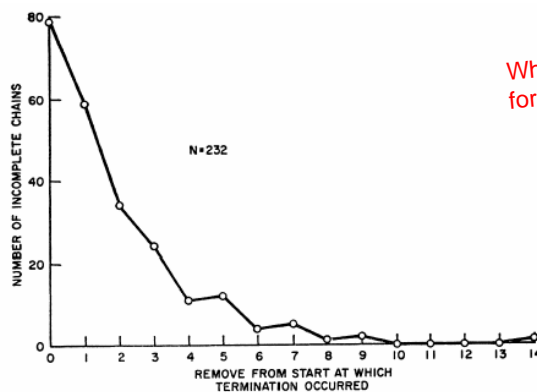
Results I

- How many of the starters would be able to establish contact with the target?
 - 64 (or 29%) out of 296 reached the target
- How many intermediaries starters with the target?
 - Well, that depends: the overall average is 5.1
 - Through hometown: 6.1 links
 - Through business: 4.6 links
 - Boston group faster than Net
 - Nebraska stakeholders not fast
- What form would the distribution take?



Results II

- Incomplete chains



What reasons can you think of for incomplete chains?

FIGURE 2
Lengths of Incomplete Chains

Results III

- Common paths
- Also see:
Gladwell's "Law of the few"

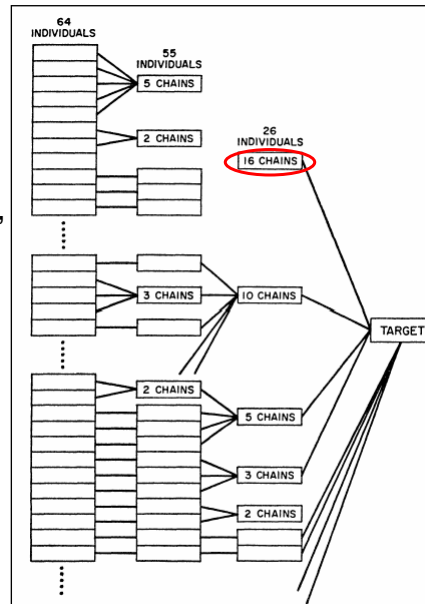


FIGURE 3

Common Paths Appear as Chains Converge on the Target

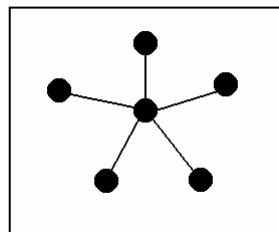
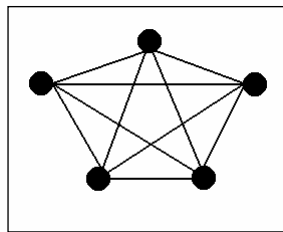
6 degrees of separation

- 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]

Small Worlds

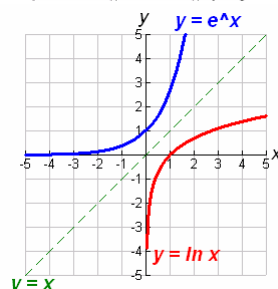
<http://www.infosci.cornell.edu/courses/info204/2007sp/>

- 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 $x > 0$



Example for base e

Contemporary Software

- Where does the small-world phenomenon come into play in contemporary software, in organizations, ..?
- Xing, LinkedIn, Myspace, Facebook, FOAF, ...
- Business Processes, Information and Knowledge Flow

How do Small World Networks form?

Preferential Attachment [Barabasi 1999]

„The rich getting richer“

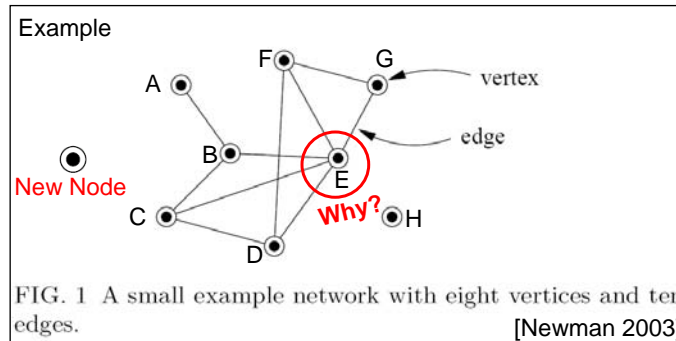
Preferential Attachment refers to the high probability of a new vertex to connect to a vertex that already has a large number of connections

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?

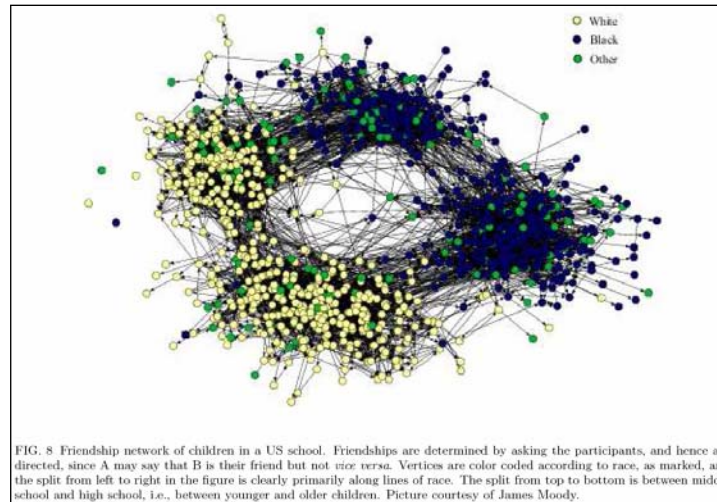


Assortative Mixing (or Homophily) [Newman 2003]

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]



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

Network Resilience [Newman 2003]

The resilience of networks with respect to vertex removal and network connectivity.

If vertices are removed from a network, the typical length of paths between pairs of vertices will increase – vertex pairs will be disconnected.

Examples:

1. Deletion of a hub
2. Deletion of a leaf node element

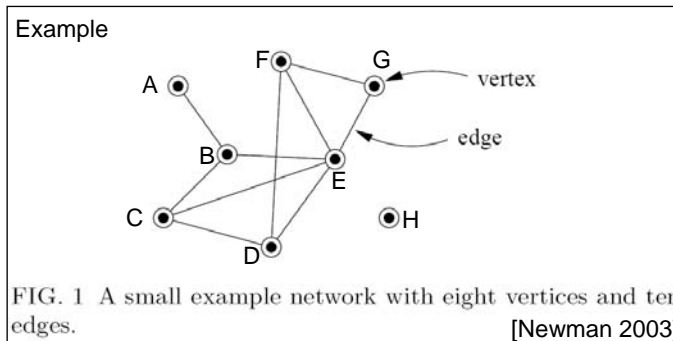
The web is highly resilient against random failure of vertices, but highly vulnerable to deliberate attack on its highest-degree vertices

Network Resilience [Newman 2003]

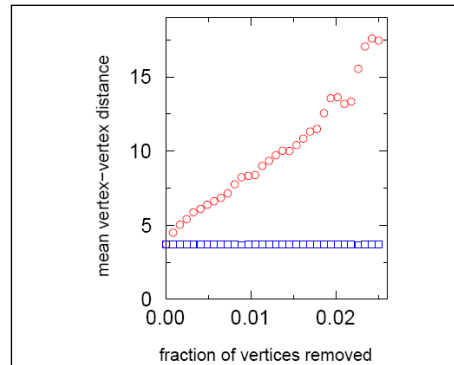
Delete a random node, what happens to the network?

Delete the node with the highest degree, what happens to the network?

What is the node on which the network's connectivity depends on most?



Network Resilience [Newman 2003]



Removal of high degree nodes first

Removal of random nodes

FIG. 7 Mean vertex-vertex distance on a graph representation of the Internet at the autonomous system level, as vertices are removed one by one. If vertices are removed in random order (squares), distance increases only very slightly, but if they are removed in order of their degrees, starting with the highest degree vertices (circles), then distance increases sharply. After Albert *et al.* [15].

Connectivity of the Web [Newman 2003, Broder et al 2000]

What does it need to destroy the connectivity of the web?

According to Broder et al 2000, you need to remove all vertices with a degree greater than five.

Because of the highly skewed degree distribution of the web, the fraction of vertices with degree greater than five is only a small fraction of all vertices.

But ...

Isn't all of this an over simplification 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
- The reciprocal services which characterize the tie



Mark Granovetter,
Stanford University

Can you give examples of strong / weak ties?

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.

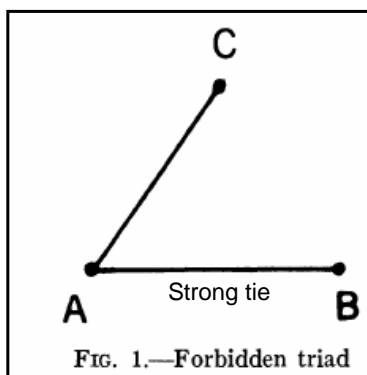
Theoretical corroboration:

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

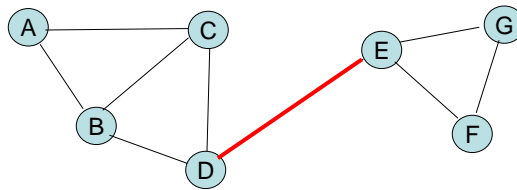


Why is it called the
forbidden triad?

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

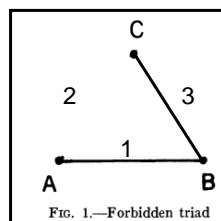


Which edge represents a bridge?
Why?

Bridges and Strong Ties [Granovetter 1973]

Example:

1. Imagine the strong tie between A and B
2. Imagine the strong tie between B and C
3. Then, the forbidden triad **implies** that a tie **exists** between C and A
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 [Granovetter 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

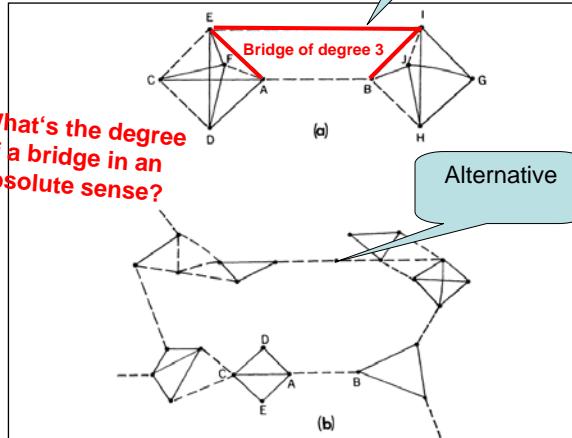
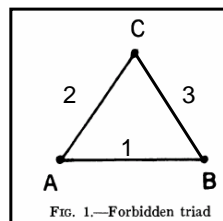


FIG. 2.—Local bridges. a, Degree 3; b, Degree 13. — = strong tie; - - - = weak tie.

In Reality ...

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

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

What are sources
of weak
ties/bridges?

Can you identify some
implications for social
networks on the web / for
search in these networks?

How does this relate to
Milgram's experiment?

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?

Next Week

We will have a look at

Network theory and terminology including (excerpt)

- Degree
- Degree distributions
- Clustering Co-efficients
- Random networks
- Scale Free networks
- And others

Any questions?

See you next week!