Computational Social Science
Week 6 - A

Fariba Karimi

Prof. Dr.
Computer Science Department, TU Graz

e-mail: karimi@tugraz.at
web: www.networkinequality.com

TU Graz, 14.11.2023

Recap of the past weeks

- Brief history of quantitative (social) sciences
- Social impact theory
- Collective behavior through the lens of social psychology (Mirta Galesic)
- Social behavior and trends (Lisette Espin-Noboa)
- Social media sentiment analysis (May Pellert)

Overview of today‘s lecture

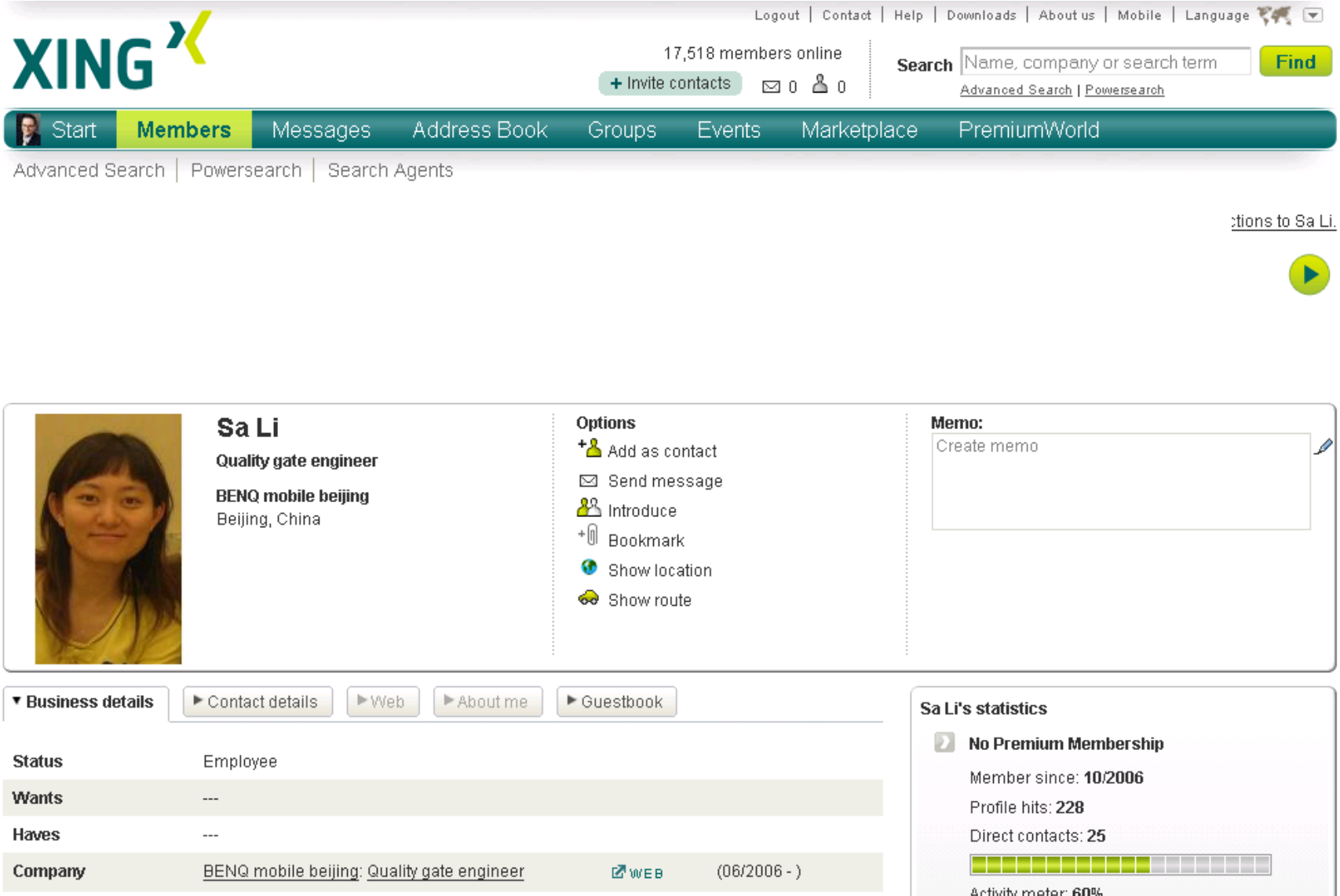
Theme: Properties of social networks and why they are important to study them.

- Small world
- Non-randomness in social networks
- Many clusters

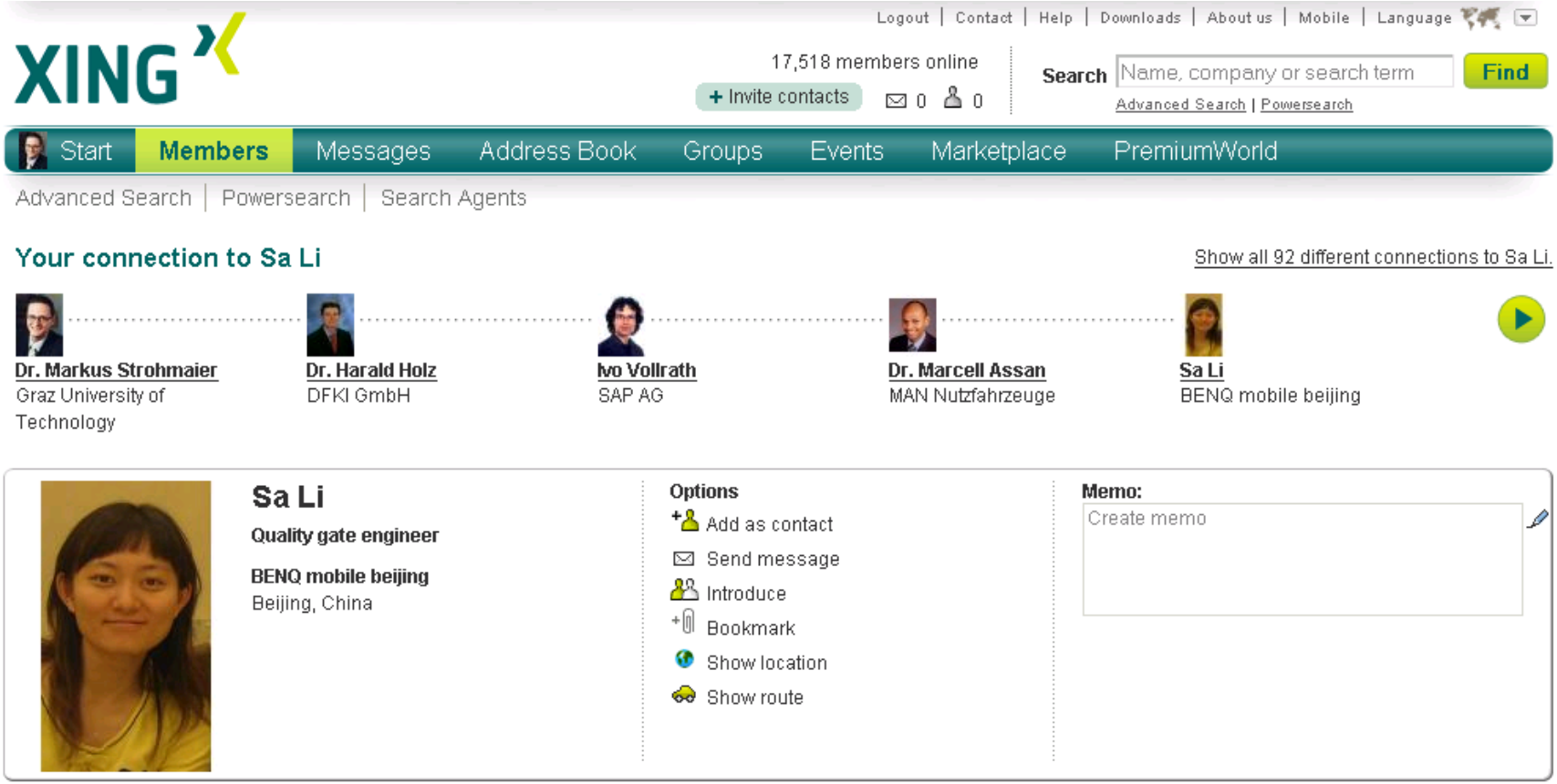
Part 1

SHAPE OF OUR SOCIAL NETWORKS: SMALL WORLD

Do I know somebody in …?

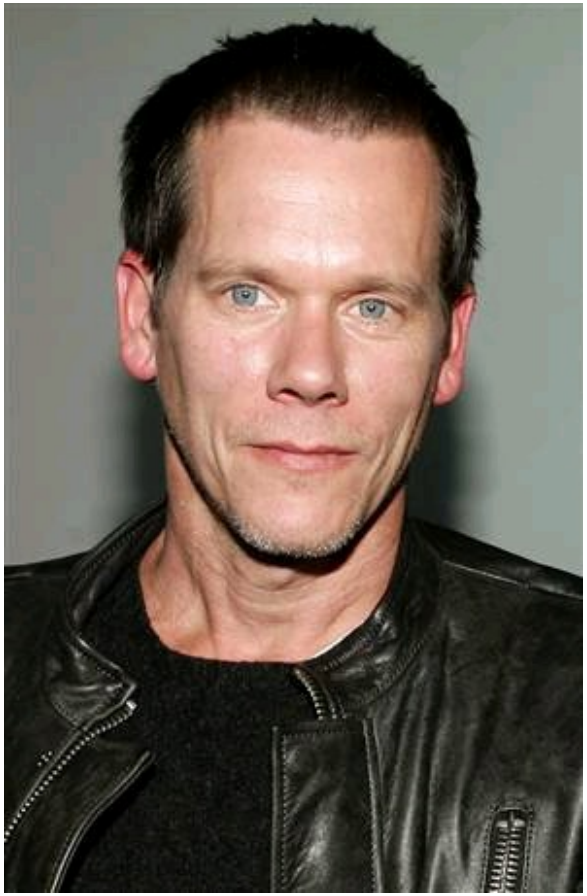


Do I know somebody in …?



The Bacon Number

http://www.imdb.com/



The Kevin Bacon Game

The oracle of Bacon

www.oracleofbacon.org

The Bacon Number [Watts 2002]

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

Paul Edrös was 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?

More generally, how many handshakes away are two researchers? There is a link if they write a paper together (co-authorship link)

What is Erdös Number of your favorite professor to Tim Berners-Lee? https://www.csauthors.net/distance/

me -> -> P. Erdös ?

Stanley Milgram coined the term Small World

- A social psychologist
- Yale and Harvard University

1933-1984

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

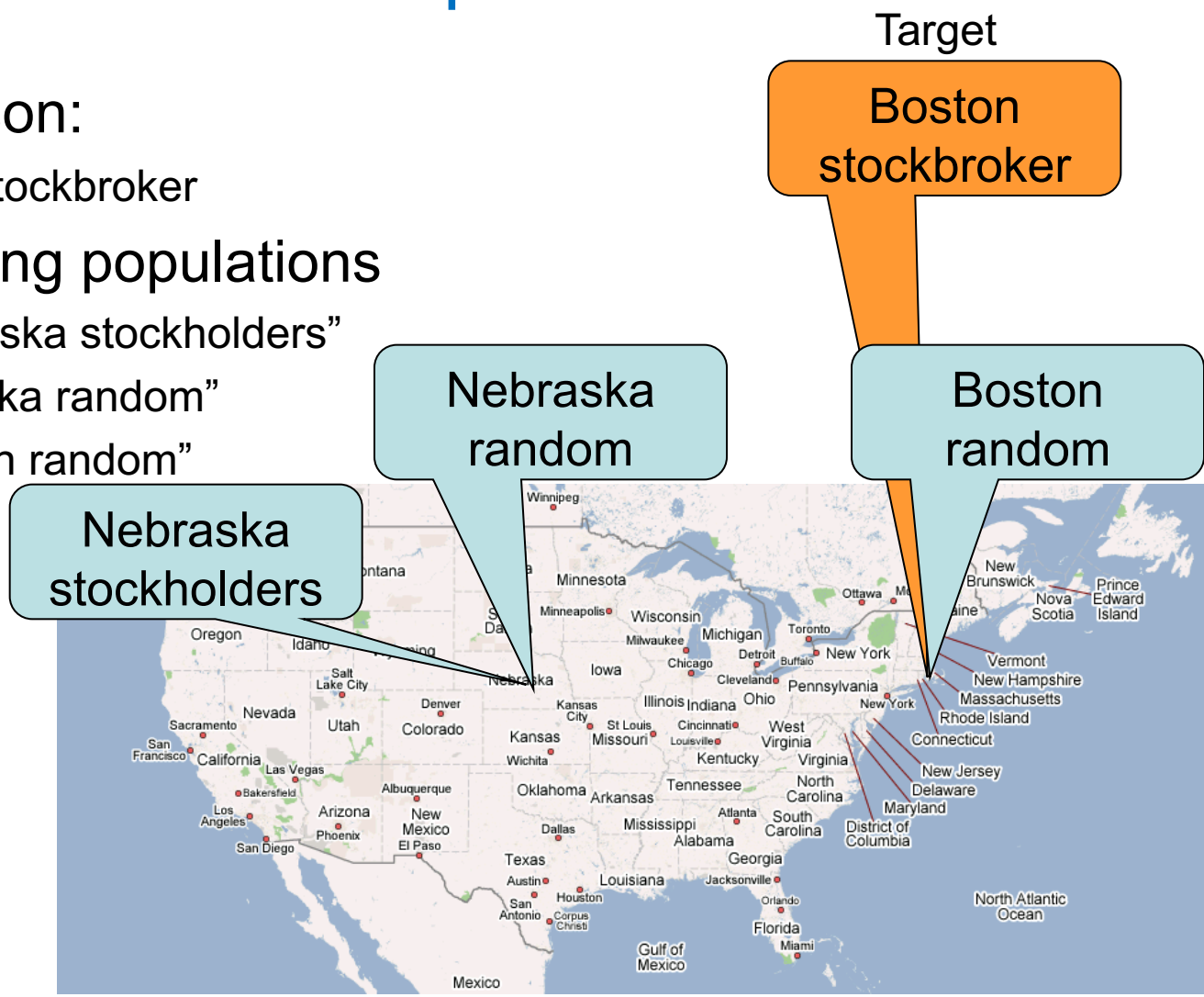
- starter group was only allowed to send the document to people they know and
- was urged to choose the next recipient in a way as to advance the progress of the document toward the target

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”

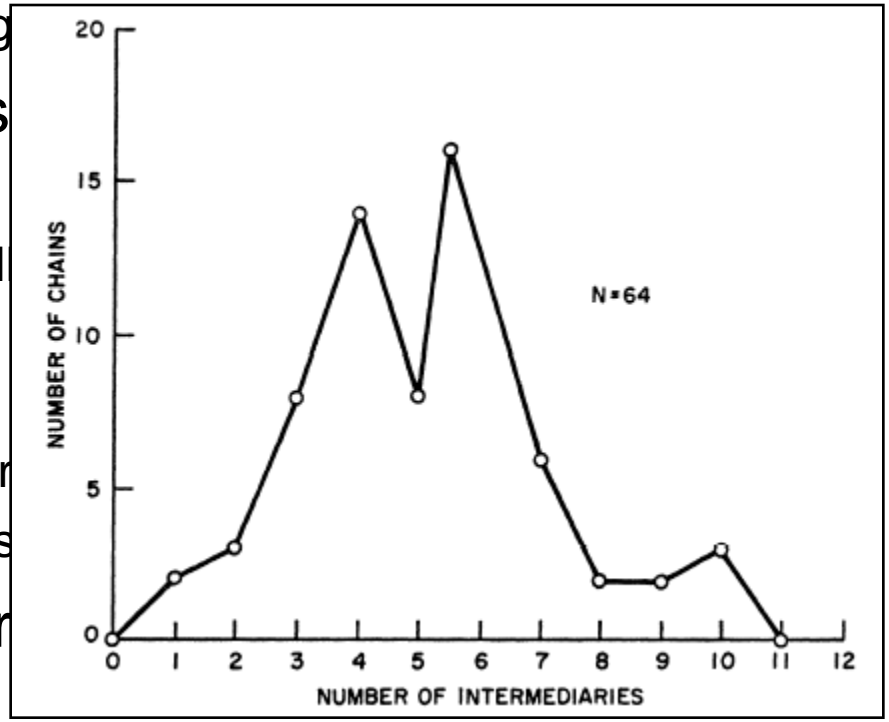


Results I

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

Results I

- How many of the starters would be able to establish contact with the target?
  - 64 out of 296 reached the targ
- How many intermediaries starters with the target?
  - Well, that depends: the overall
  - Through hometown: 6.1 links
  - Through business: 4.6 links
  - Boston group faster than Nebr
  - Nebraska stockholders not fas
- What form would the distr take?



Results II

- Incomplete chains

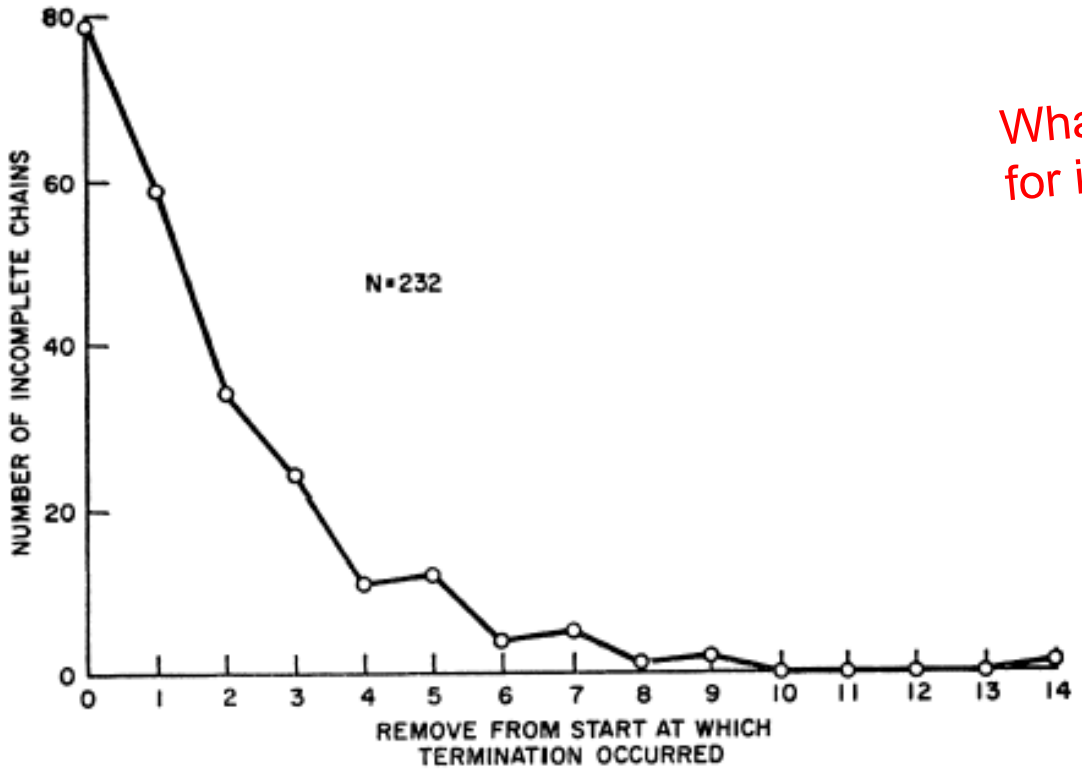


FIGURE 2

Lengths of Incomplete Chains

What reasons can you think of for incomplete chains?

What does it mean?

- Imagine you know on average 100 people very well with their first name. everyone on average know 100 persons very well. How many people are you connected to with 3 handshakes?

Algorithmic vs. topological definition

- Milgram explored the algorithmic definition of the small-world problem
  - Social search: shortest path that ordinary people can find given their ***local topological information*** about the underlying social graph

- Large-scale web social da
  exploration of the topologi
  problem
  - Shortest possible path between two individuals, with a global knowledge of the social network

Are algorithmic and topological definitions the same?

See also *"How small is the world, really?"* by D. J. Watts
bit.ly/1oRZvIX

J. Kleinberg. *The small-world phenomenon: An algorithmic perspective*, 2000

Algorithmic vs. topological applications

- Topological distance: in which conditions does it matter?
- Human search (algorithmic distance): in which condition does it matter?

See also *"How small is the world, really?"* by D. J. Watts
bit.ly/1oRZvIX

Follow up work (2008)

http://arxiv.org/PS_cache/arxiv/pdf/0803/0803.0939v1.pdf

- 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
- One of the largest social network constructed and analyzed to date (2008)

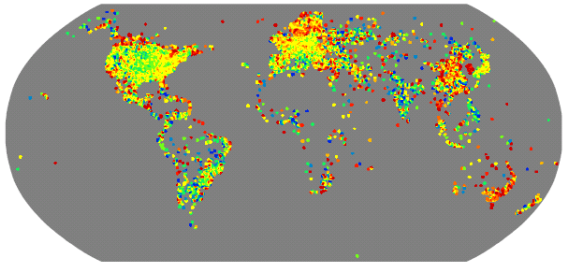


Figure 10: Number of users at a particular geographic location. Color represents the number of users. Notice the map of the world appears.

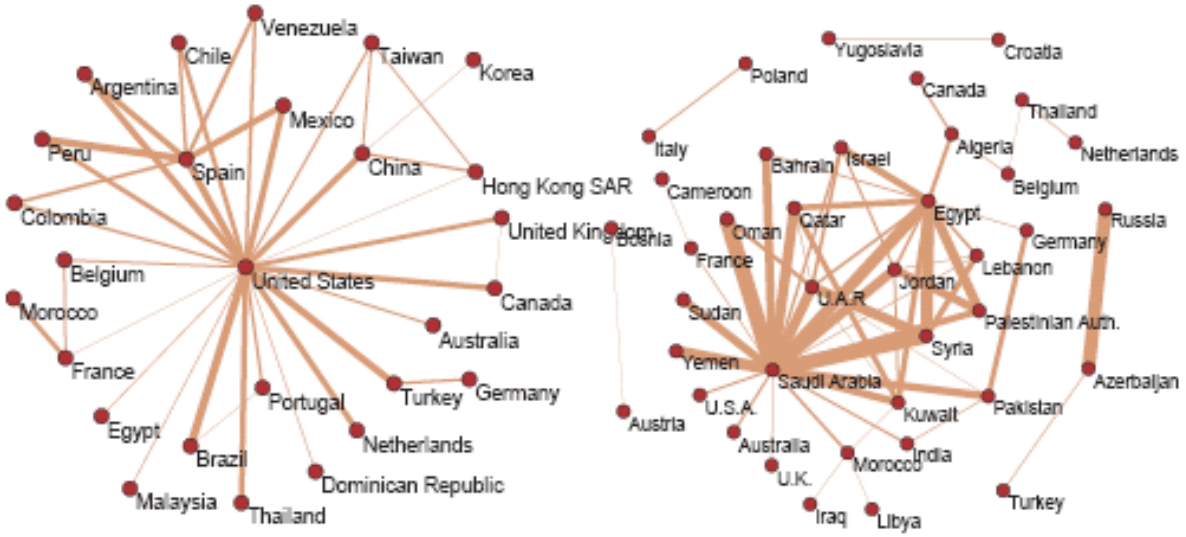


Figure 14: (a) Communication among countries with at least 10 million conversations in June 2006. (b) Countries by average length of the conversation. Edge widths correspond to logarithms of intensity of links.

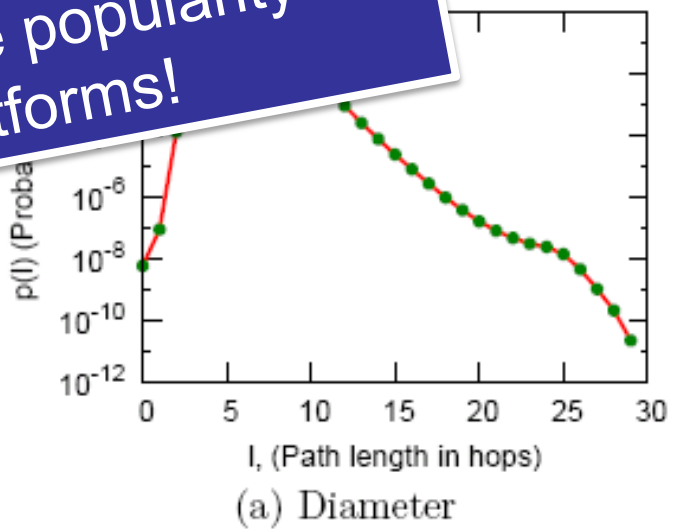
Follow up work (2008)

http://arxiv.org/PS_cache/arxiv/pdf/0803/0803.0939v1.pdf

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 leng is 6.6. median at 7.
- Result: a random pair of apart on the average, wh than the length reported b ilgram.
- 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.
- we find that there are about “7 degrees of separation” among people.
- long paths exist in the network; we found paths up to a length of 29.

This is before the popularity of social media platforms!



(a) Diameter

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
  Social networks have a low diameter.

- Question 1: Why social networks are small worlds? What **social mechanisms** create such small-world phenomena?

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)

- Question 1: Why social networks are small worlds? What **social mechanisms** create such small world phenomena?
- Question 2: **Why is it important** to understand the Small World properties of our social networks? What does it mean?

Formalizing the Small World Problem

[Watts and Strogatz 1998]

The small-world phenomenon is assumed to be present when local clustering is high and average path lengths are relatively small

average path length: *L*

clustering coefficient: *C*

Examples for Small World Networks

[Watts and Strogatz 1998]

Table 1 Empirical examples of small-world networks

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

	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: $L \geqslant L_{random}$ but $C \gg C_{random}$.

Formalizing the Small World Problem

[Watts and Strogatz 1998]

The small-world phenomenon is assumed to be present when

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

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

Formalizing the Small World Problem

[Watts 2003]

- Page 76 -82
- The alpha parameter

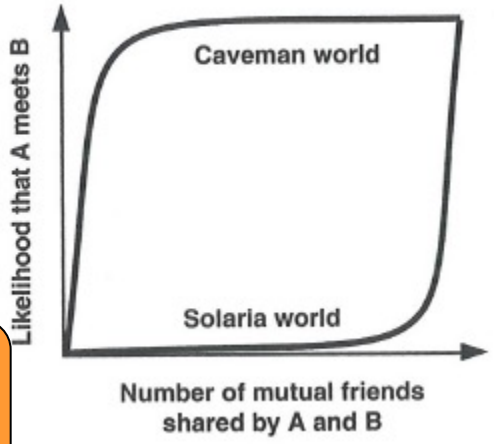


Figure 3.1. Two extreme kinds of interaction rules. In the top curve (caveman world), even a single mutual friend implies that A and B are highly likely to meet. In the bottom curve (Solaria world), all interactions are equally unlikely, regardless of how many friends A and B share.

Under which conditions can these two requirements be reconciled?

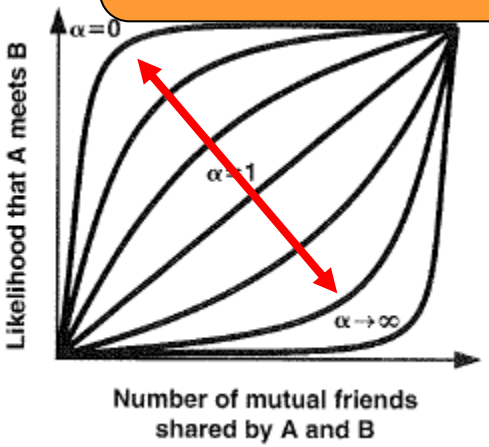


Figure 3.2. Between the two extremes, a whole family of interaction rules exists, each one specified by a particular value of the tuneable parameter alpha (α). When α = 0, we have a caveman world; when α becomes [illegible] we have Solaria.

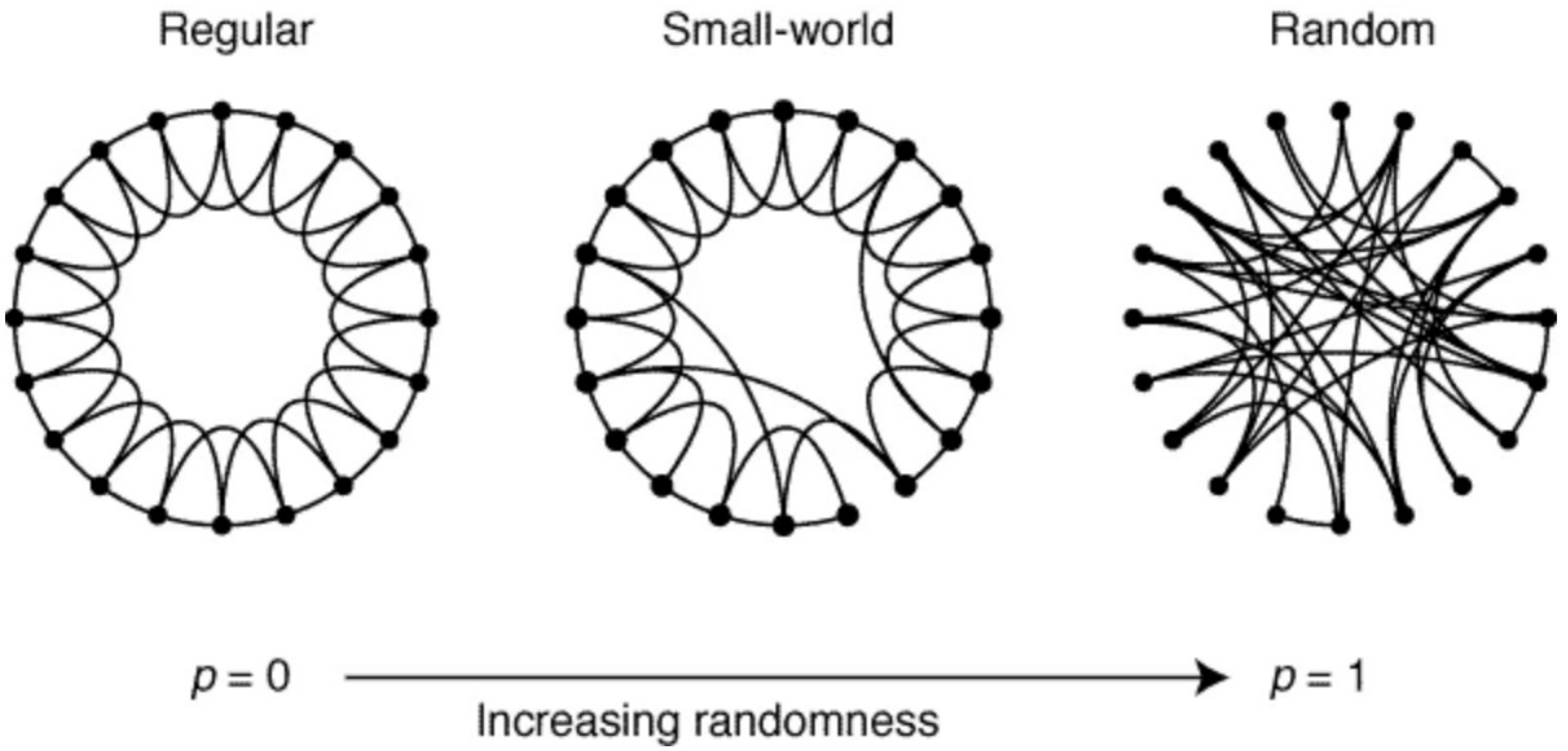
Two seemingly contradictory requirements for the Small World Phenomenon:

- 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)

Search-ability

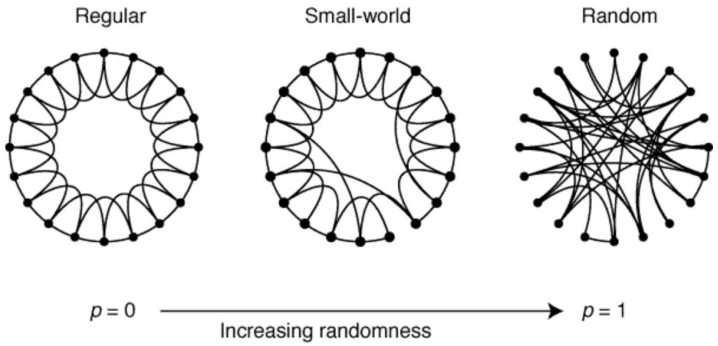
Formalizing the Small World Problem

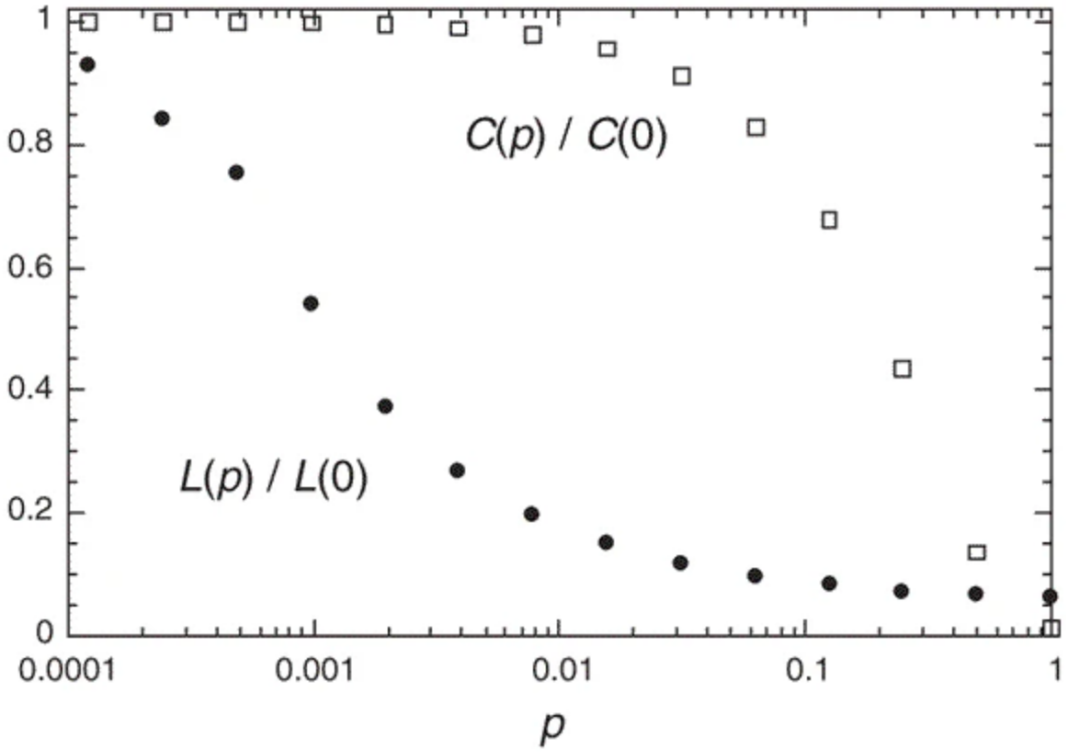
[Watts and Strogatz, Nature 1998]



Formalizing the Small World Problem

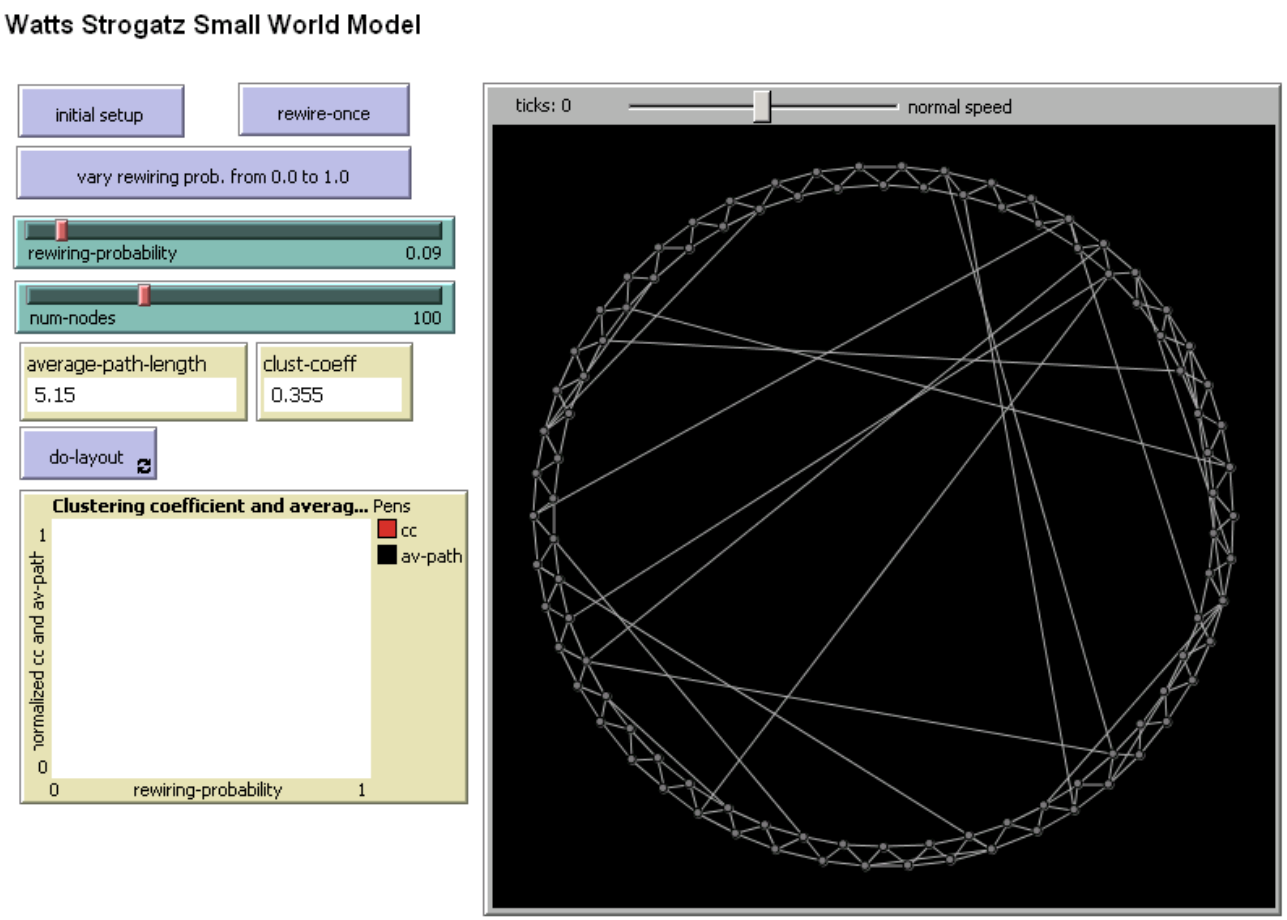
[Watts and Strogatz, Nature 1998]





Demo – Small Worlds

http://www.netlogoweb.org/launch



Contemporary Software

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

- Xing, LinkedIn, Facebook, …
- Business Processes, Information and Knowledge Flow

Any questions?

See you next week!