

# The Small-World Phenomenon

Complex Networks, CSYS/MATH 303, Spring, 2010

Prof. Peter Dodds

Department of Mathematics & Statistics  
Center for Complex Systems  
Vermont Advanced Computing Center  
University of Vermont



History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

# Outline

History

An online experiment

Previous theoretical work

An improved model

References

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

# Some problems for people thinking about people?:

## How are social networks structured?

- ▶ How do we define connections?
- ▶ How do we measure connections?
- ▶ (remote sensing, self-reporting)

## What about the dynamics of social networks?

- ▶ How do social networks evolve?
- ▶ How do social movements begin?
- ▶ How does collective problem solving work?
- ▶ How is information transmitted through social networks?

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

## A small slice of the pie:

- ▶ **Q.** Can people pass messages between distant individuals using only their existing social connections?
- ▶ **A.** Apparently yes...

## Handles:

- ▶ The Small World Phenomenon
- ▶ or “Six Degrees of Separation.”

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

## History

An online  
experiment

Previous  
theoretical work

An improved  
model

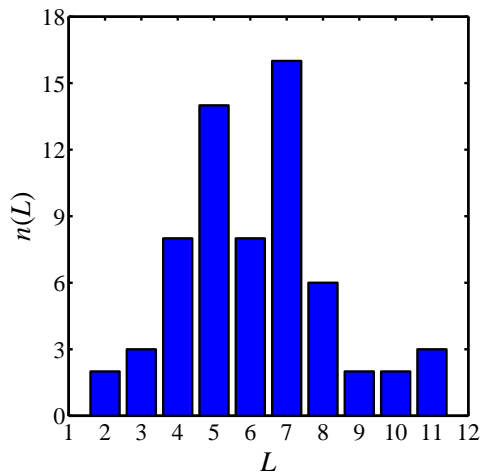
References

## Stanley Milgram et al., late 1960's:

- ▶ Target person worked in Boston as a stockbroker.
- ▶ 296 senders from Boston and Omaha.
- ▶ 20% of senders reached target.
- ▶ average chain length  $\simeq 6.5$ .

# The problem

Lengths of successful chains:



From Travers and  
Milgram (1969) in  
Sociometry:<sup>[4]</sup>  
“An Experimental  
Study of the Small  
World Problem.”

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

## History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

Two features characterize a social 'Small World':

1. Short paths exist  
and
2. People are good at finding them.

## Milgram's small world experiment with e-mail [2]

The screenshot displays the Small World project website. On the left, there are navigation menus for 'Events and News', 'Project Information', 'Research Team', and 'Web Development'. The main content area features a network diagram with circular nodes representing participants, connected by lines. The nodes include: Vijay (Delhi, India) worked at an engineering firm with Sameer (Kolkata, India) whose daughter is Prema (Berkeley, USA) who goes to school in California and plays soccer with Christie (Berkeley, USA) whose best friend from high school is William (New York, NY) who is studying medicine with Alice (New York, USA). The website also includes a 'home my small world chat FAQ related links' menu, 'login' and 'sign up' buttons, and introductory text about the project's goal to test the idea that any two people in the world can be connected via 'six degrees of separation'.

**Events and News**  
Duncan J. Watts's new book is out now!

**Project Information**  
In the Press  
Description  
Procedures  
Security and Privacy  
Articles/References  
Results

**Research Team**  
Duncan J. Watts  
Peter Dodds  
Roby Muhamad

**Web Development**  
Peter Housel

Vijay (Delhi, India) worked at an engineering firm with

home  
my small world  
chat  
FAQ  
related links

login

sign up

The **SMALL WORLD** project is an online experiment to test the idea that any two people in the world can be connected via 'six degrees of separation'.

Your objective is to get a message to a "target person", somewhere in the world, by forwarding the message to a friend of yours—someone who is "closer" to the target than you are. (If you happen know the target, you can of course send it to them)

If we have asked you to participate (you would have received a message from a friend of yours), you should [continue](#) the chain.

If you are just visiting us, sign up to start a new chain.

Sameer (Kolkata, India) whose daughter

Prema (Berkeley, USA) goes to school in California and plays soccer with

Alice (New York, USA)

Christie (Berkeley, USA) whose best friend from high school

William (New York, NY) is studying medicine with

COLUMBIA UNIVERSITY

History

An online experiment

Previous theoretical work

An improved model

References



# Social search—the Columbia experiment

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

- ▶ 60,000+ participants in 166 countries
- ▶ 18 targets in 13 countries including
  - ▶ a professor at an Ivy League university,
  - ▶ an archival inspector in Estonia,
  - ▶ a technology consultant in India,
  - ▶ a policeman in Australia,  
and
  - ▶ a veterinarian in the Norwegian army.
- ▶ 24,000+ chains

# Social search—the Columbia experiment

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

- ▶ Milgram's participation rate was roughly 75%
- ▶ Email version: Approximately 37% participation rate.
- ▶ Probability of a chain of length 10 getting through:

$$.37^{10} \simeq 5 \times 10^{-5}$$

- ▶  $\Rightarrow$  384 completed chains (1.6% of all chains).

# Social search—the Columbia experiment

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

- ▶ Motivation/Incentives/Perception matter.
- ▶ If target *seems* reachable  
⇒ participation more likely.
- ▶ Small changes in attrition rates  
⇒ large changes in completion rates
- ▶ e.g., ↘ 15% in attrition rate  
⇒ ↗ 800% in completion rate

# Social search—the Columbia experiment

## Successful chains disproportionately used

- ▶ weak ties (Granovetter)
- ▶ professional ties (34% vs. 13%)
- ▶ ties originating at work/college
- ▶ target's work (65% vs. 40%)

## ...and disproportionately avoided

- ▶ hubs (8% vs. 1%) (+ no evidence of funnels)
- ▶ family/friendship ties (60% vs. 83%)

## Geography → Work

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

# Social search—the Columbia experiment

Senders of successful messages showed **little absolute dependency** on

- ▶ age, gender
- ▶ country of residence
- ▶ income
- ▶ religion
- ▶ relationship to recipient

Range of completion rates for subpopulations:

30% to 40%

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

# Social search—the Columbia experiment

Nevertheless, some weak discrepancies do exist...

## An above average connector:

Norwegian, secular male, aged 30-39, earning over \$100K, with graduate level education working in mass media or science, who uses relatively weak ties to people they met in college or at work.

## A below average connector:

Italian, Islamic or Christian female earning less than \$2K, with elementary school education and retired, who uses strong ties to family members.

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

# Social search—the Columbia experiment

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

Mildly bad for continuing chain:

choosing recipients because “they have lots of friends” or  
because they will “likely continue the chain.”

Why:

- ▶ Specificity important
- ▶ Successful links used relevant information.  
(e.g. connecting to someone who shares same  
profession as target.)

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

## Basic results:

- ▶  $\langle L \rangle = 4.05$  for all completed chains
- ▶  $L_*$  = Estimated 'true' median chain length (zero attrition)
- ▶ Intra-country chains:  $L_* = 5$
- ▶ Inter-country chains:  $L_* = 7$
- ▶ All chains:  $L_* = 7$
- ▶ Milgram:  $L_* \simeq 9$



# Previous work—short paths

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

- ▶ Connected **random networks** have short average path lengths:

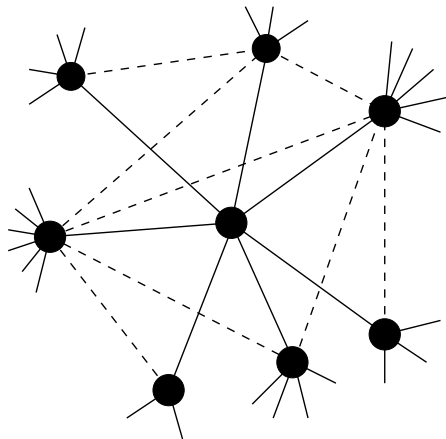
$$\langle d_{AB} \rangle \sim \log(N)$$

$N$  = population size,

$d_{AB}$  = distance between nodes  $A$  and  $B$ .

- ▶ **But: social networks aren't random...**

# Previous work—short paths



Need “**clustering**” (your friends are likely to know each other):

History

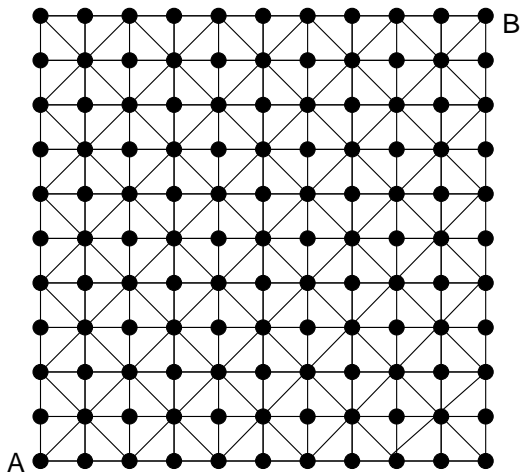
An online  
experiment

Previous  
theoretical work

An improved  
model

References

# Non-randomness gives clustering



$d_{AB} = 10 \rightarrow$  too many long paths.

History

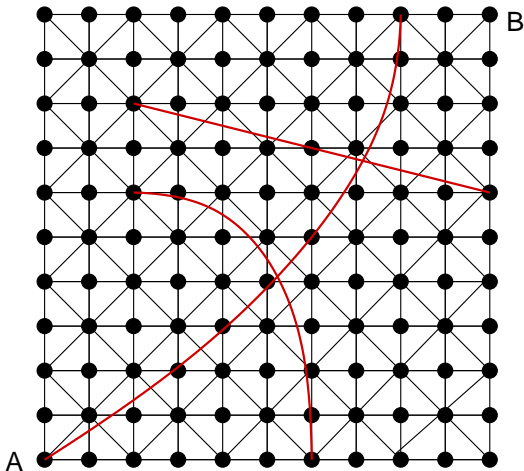
An online experiment

Previous theoretical work

An improved model

References

# Randomness + regularity



Now have  $d_{AB} = 3$

$\langle d \rangle$  decreases overall

History

An online experiment

Previous theoretical work

An improved model

References

# Small-world networks

Introduced by

Watts and Strogatz (Nature, 1998) [6]

“Collective dynamics of ‘small-world’ networks.”

Small-world networks were found everywhere:

- ▶ neural network of *C. elegans*,
- ▶ semantic networks of languages,
- ▶ actor collaboration graph,
- ▶ food webs,
- ▶ social networks of comic book characters,...

Very weak requirements:

- ▶ **local regularity** + random **short cuts**

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

# Toy model

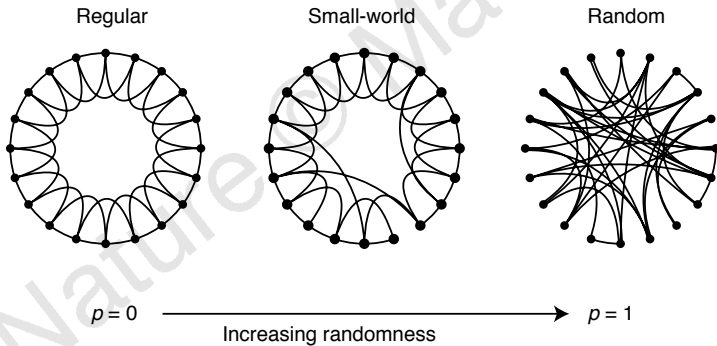
History

An online experiment

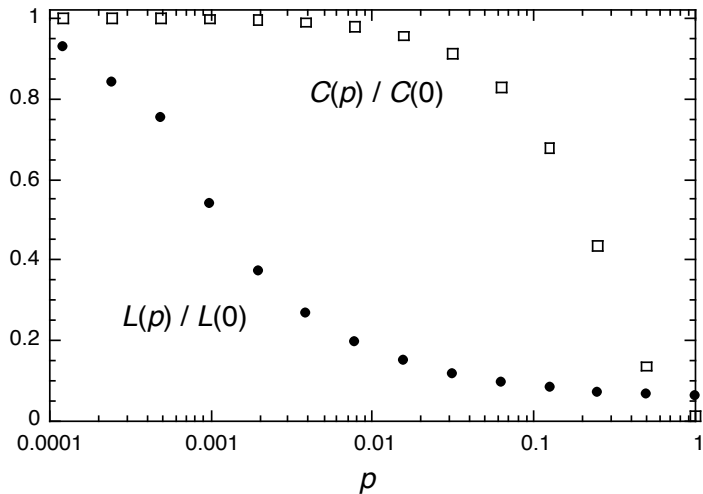
Previous theoretical work

An improved model

References



# The structural small-world property



History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

# Previous work—finding short paths

But are these short cuts findable?

No.

Nodes **cannot** find each other quickly  
with **any local search method**.

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References



# Previous work—finding short paths

- ▶ What can a local search method reasonably use?
- ▶ How to find things without a map?
- ▶ Need some measure of distance between friends and the target.

## Some possible knowledge:

- ▶ Target's identity
- ▶ Friends' popularity
- ▶ Friends' identities
- ▶ Where message has been

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

# Previous work—finding short paths

Jon Kleinberg (Nature, 2000) <sup>[3]</sup>  
“Navigation in a small world.”

Allowed to vary:

1. local search algorithm  
and
2. network structure.

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

# Previous work—finding short paths

## Kleinberg's Network:

1. Start with regular  $d$ -dimensional cubic lattice.
2. Add local links so nodes know all nodes within a distance  $q$ .
3. Add  $m$  short cuts per node.
4. Connect  $i$  to  $j$  with probability

$$p_{ij} \propto d_{ij}^{-\alpha}.$$

- ▶  $\alpha = 0$ : random connections.
- ▶  $\alpha$  large: reinforce local connections.
- ▶  $\alpha = d$ : same number of connections at all scales.

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

# Previous work—finding short paths

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

## Theoretical optimal search:

- ▶ “Greedy” algorithm.
- ▶ Same number of connections at all scales:  $\alpha = d$ .

Search time grows slowly with system size (like  $\log^2 N$ ).

**But: social networks aren't lattices plus links.**

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

- ▶ If networks have hubs can also search well: Adamic et al. (2001) <sup>[1]</sup>

$$P(k_i) \propto k_i^{-\gamma}$$

where  $k$  = degree of node  $i$  (number of friends).

- ▶ Basic idea: get to hubs first (airline networks).
- ▶ But: hubs in social networks are limited.

# The problem

History

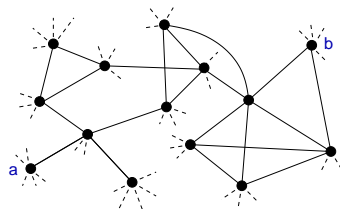
An online  
experiment

Previous  
theoretical work

An improved  
model

References

If there are no hubs and no underlying lattice, how can search be efficient?



Which friend of **a** is closest to the target **b**?

What does 'closest' mean?

What is 'social distance'?

# The model

One approach: incorporate **identity**.

(See “Identity and Search in Social Networks.” Science, 2002, Watts, Dodds, and Newman<sup>[5]</sup>)

Identity is formed from attributes such as:

- ▶ Geographic location
- ▶ Type of employment
- ▶ Religious beliefs
- ▶ Recreational activities.

**Groups** are formed by people with at least one similar attribute.

Attributes  $\Leftrightarrow$  Contexts  $\Leftrightarrow$  Interactions  $\Leftrightarrow$  Networks.

History

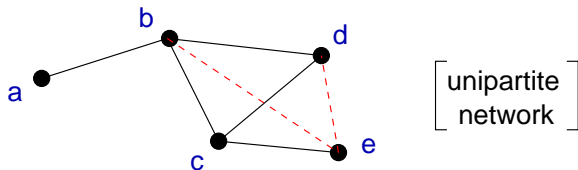
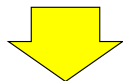
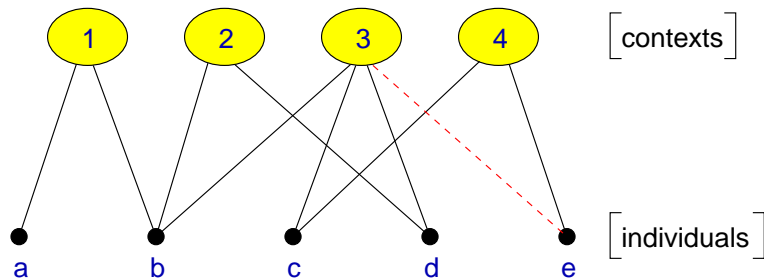
An online  
experiment

Previous  
theoretical work

An improved  
model

References

# Social distance—Bipartite affiliation networks



History

An online  
experiment

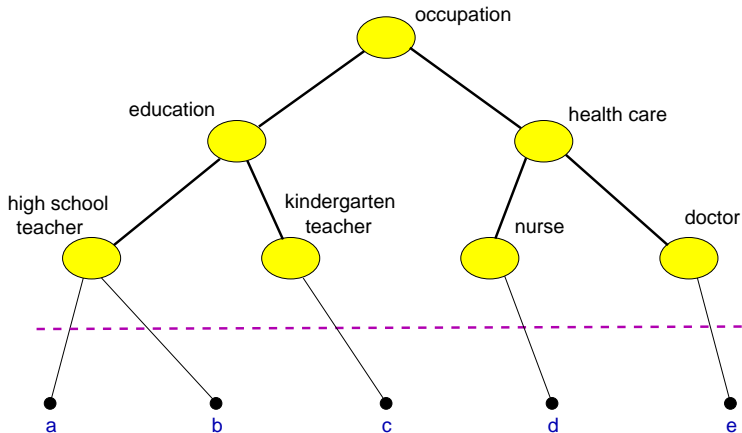
Previous  
theoretical work

An improved  
model

References



# Social distance—Context distance



History

An online  
experiment

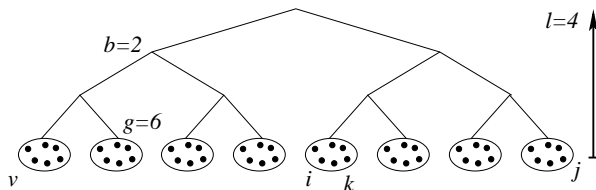
Previous  
theoretical work

An improved  
model

References

# The model

Distance between two individuals  $x_{ij}$  is the height of lowest common ancestor.



$$x_{ij} = 3, x_{ik} = 1, x_{iv} = 4.$$

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

- ▶ Individuals are more likely to know each other the closer they are within a hierarchy.
- ▶ Construct  $z$  connections for each node using

$$p_{ij} = c \exp\{-\alpha x_{ij}\}.$$

- ▶  $\alpha = 0$ : random connections.
- ▶  $\alpha$  large: local connections.

# Social distance—Generalized context space

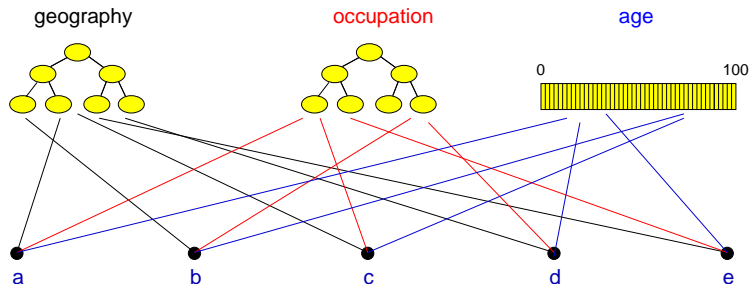
History

An online  
experiment

Previous  
theoretical work

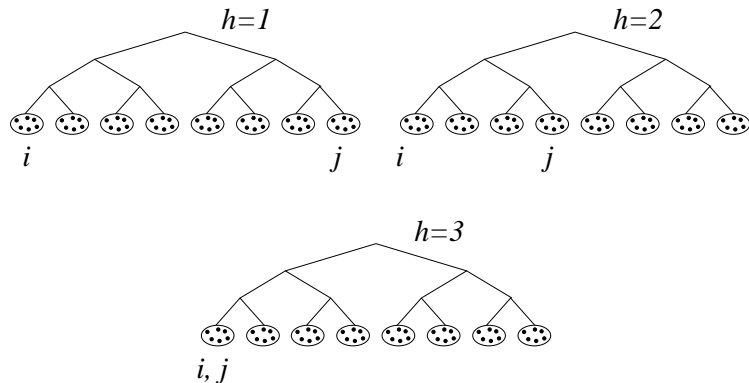
An improved  
model

References



(Blau & Schwartz, Simmel, Breiger)

# The model



$$\vec{v}_i = [1 \ 1 \ 1]^T, \vec{v}_j = [8 \ 4 \ 1]^T$$

$$x_{ij}^1 = 4, x_{ij}^2 = 3, x_{ij}^3 = 1.$$

Social distance:

$$y_{ij} = \min_h x_{ij}^h.$$

History

An online experiment

Previous theoretical work

An improved model

References

History

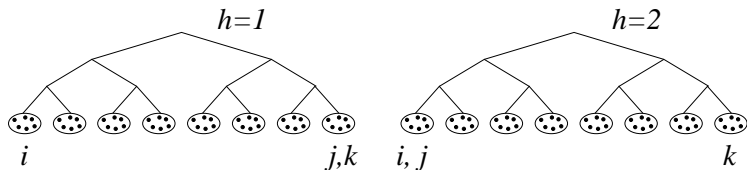
An online  
experiment

Previous  
theoretical work

An improved  
model

References

Triangle inequality doesn't hold:



$$y_{ik} = 4 > y_{ij} + y_{jk} = 1 + 1 = 2.$$

History

An online  
experiment

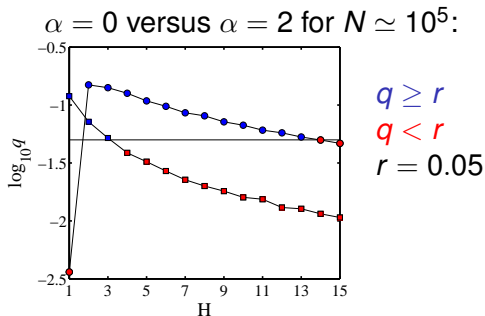
Previous  
theoretical work

An improved  
model

References

- ▶ Individuals know the identity vectors of
  1. themselves,
  2. their friends,  
and
  3. the target.
- ▶ Individuals can estimate the social distance between their friends and the target.
- ▶ Use a greedy algorithm + allow searches to fail randomly.

# The model-results—searchable networks



$q$  = probability an arbitrary message chain reaches a target.

- ▶ A few dimensions help.
- ▶ Searchability decreases as population increases.
- ▶ Precise form of hierarchy largely doesn't matter.

History

An online  
experiment

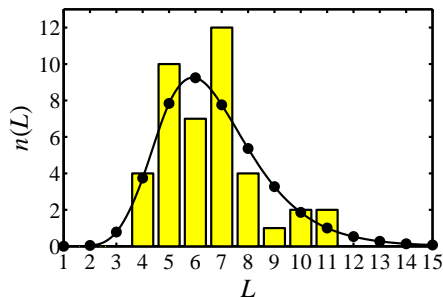
Previous  
theoretical work

An improved  
model

References



Milgram's Nebraska-Boston data:



Model parameters:

- ▶  $N = 10^8$ ,
- ▶  $z = 300, g = 100$ ,
- ▶  $b = 10$ ,
- ▶  $\alpha = 1, H = 2$ ;
  
- ▶  $\langle L_{\text{model}} \rangle \simeq 6.7$
- ▶  $L_{\text{data}} \simeq 6.5$

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

## Adamic and Adar (2003)

- ▶ For HP Labs, found probability of connection as function of organization distance well fit by exponential distribution.
- ▶ Probability of connection as function of real distance  $\propto 1/r$ .

# Social Search—Real world uses

- ▶ Tags create identities for objects
- ▶ Website tagging: `http://www.del.icio.us`
- ▶ (e.g., Wikipedia)
- ▶ Photo tagging: `http://www.flickr.com`
- ▶ Dynamic creation of metadata plus links between information objects.
- ▶ Folksonomy: collaborative creation of metadata

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

History

An online  
experiment

Previous  
theoretical work

An improved  
model

References

## Recommender systems:

- ▶ Amazon uses people's actions to build effective connections between books.
- ▶ Conflict between 'expert judgments' and tagging of the hoi polloi.

- ▶ Bare networks are typically unsearchable.
- ▶ Paths are findable if nodes understand how network is formed.
- ▶ Importance of identity (interaction contexts).
- ▶ Improved social network models.
- ▶ Construction of peer-to-peer networks.
- ▶ Construction of searchable information databases.

History





An online  
experiment

Previous  
theoretical work

An improved  
model

References

# References I

-  [1] L. Adamic, R. Lukose, A. Puniyani, and B. Huberman.  
Search in power-law networks.  
*Phys. Rev. E*, 64:046135, 2001. [pdf](#) (⊞)
-  [2] P. S. Dodds, R. Muhamad, and D. J. Watts.  
An experimental study of search in global social networks.  
*Science*, 301:827–829, 2003. [pdf](#) (⊞)
-  [3] J. Kleinberg.  
Navigation in a small world.  
*Nature*, 406:845, 2000. [pdf](#) (⊞)
-  [4] J. Travers and S. Milgram.  
An experimental study of the small world problem.  
*Sociometry*, 32:425–443, 1969. [pdf](#) (⊞)

[History](#)[An online  
experiment](#)[Previous  
theoretical work](#)[An improved  
model](#)[References](#)


History


An online  
experiment

Previous  
theoretical work

An improved  
model

References

 [5] D. J. Watts, P. S. Dodds, and M. E. J. Newman.  
Identity and search in social networks.  
*Science*, 296:1302–1305, 2002. [pdf](#) (田)

 [6] D. J. Watts and S. J. Strogatz.  
Collective dynamics of ‘small-world’ networks.  
*Nature*, 393:440–442, 1998. [pdf](#) (田)