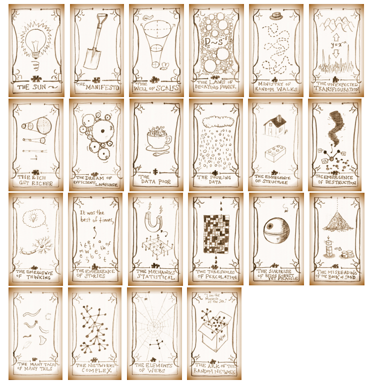
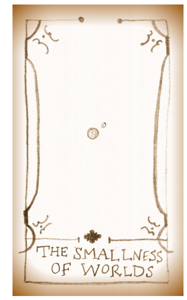
Small-world networks

Principles of Complex Systems | @pocsvox
CSYS/MATH 300, Fall, 2016 | #FallPoCS2016

Prof. Peter Dodds | @peterdodds

Dept. of Mathematics & Statistics | Vermont Complex Systems Center
Vermont Advanced Computing Core | University of Vermont





These slides are brought to you by:

People thinking about people:

How are social networks structured?

- How do we define and measure connections?
- Methods/issues of self-report and remote sensing.

What about the dynamics of social networks?

- How do social networks/movements begin & evolve?
- How does collective problem solving work?
- How does information move through social networks?
- Which rules give the best 'game of society?'

Sociotechnical phenomena and algorithms:

- What can people and computers do together? (google)
- Use Play + Crunch to solve problems. Which problems?

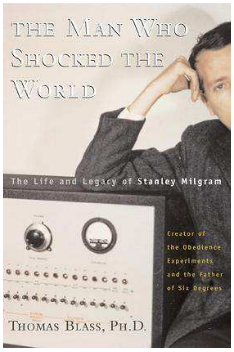
Outline

Social Search

A small slice of the pie:

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

Milgram's social search experiment (1960s)



http://www.stanleymilgram.com

- Target person = Boston stockbroker.
- 296 senders from Boston and Omaha.
- 20% of senders reached target.
- chain length $\simeq 6.5$.

Popular terms:

- The Small World Phenomenon;
- "Six Degrees of Separation."

You may already be a winner in NSA's "three-degrees" surveillance sweepstakes!

NSA's probes could cover hundreds of millions of Americans. Thanks, Kevin Bacon.

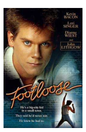
by Sean Gallagher - July 18 2013, 4:00pm EDT

BIG DATA 109

Aurich Lawson

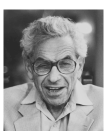
- Many people are within three degrees from a random person ...

Six Degrees of Kevin Bacon:



- It's a game:
  "Kevin Bacon is the Center of the Universe"
- The Oracle of Bacon

Six Degrees of Paul Erdös:

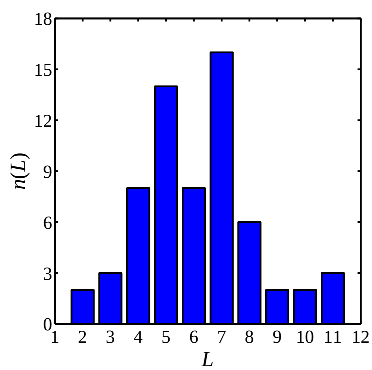


- Academic papers.
- Erdös Number
- Erdös Number Project

- So naturally we must have the Erdös-Bacon Number ...
- One computational Story Lab team member has EBN $< \infty$.
- Natalie Hershlag's (Portman's) EBN# = 5 + 2 = 7.

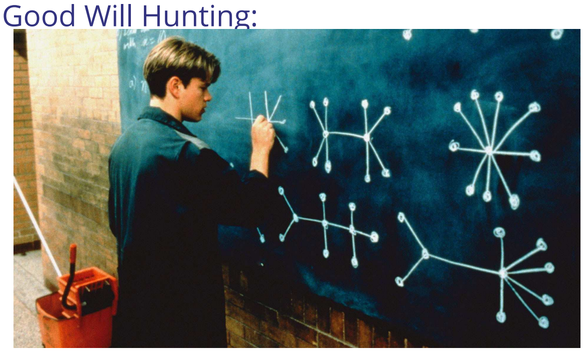
The problem

Lengths of successful chains:



From Travers and Milgram (1969) in Sociometry: [12]
"An Experimental Study of the Small World Problem."

Good Will Hunting:



- Boardwork by Dan Kleitman, EBN# = 1 + 2 = 3.
- See Kleitman's sidebar in Mark Saul's Movie Review (Notices of the AMS, Vol. 45, 1998.)

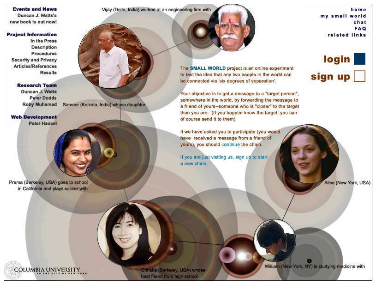
The problem

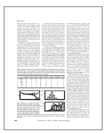
Two features characterize a social 'Small World':

1. Short paths exist, (= Geometric piece)
   and
2. People are good at finding them. (= Algorithmic piece)

Social Search

Milgram's small world experiment with email:





"An Experimental study of Search in Global Social Networks"
Dodds, Muhamad, and Watts,
Science, **301**, 827–829, 2003. [6]

Social search—the Columbia experiment

- 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

We were lucky and contagious (more later):

"Using E-Mail to Count Connections", Sarah Milstein, New York Times, Circuits Section (December, 2001)

All targets:

Table S1

Target	*City*	*Country*	*Occupation*	*Gender*	*N*	N_c *(%)*	r (r_0)	*<L>*
1	Novosibirsk	Russia	PhD student	F	8234	20(0.24)	64 (76)	4.05
2	New York	USA	Writer	F	6044	31 (0.51)	65 (73)	3.61
3	Bandung	Indonesia	Unemployed	M	8151	0	66 (76)	n/a
4	New York	USA	Journalist	F	5690	44 (0.77)	60 (72)	3.9
5	Ithaca	USA	Professor	M	5855	168 (2.87)	54 (71)	3.84
6	Melbourne	Australia	Travel Consultant	F	5597	20 (0.36)	60 (71)	5.2
7	Bardufoss	Norway	Army veterinarian	M	4343	16 (0.37)	63 (76)	4.25
8	Perth	Australia	Police Officer	M	4485	4 (0.09)	64 (75)	4.5
9	Omaha	USA	Life Insurance Agent	F	4562	2 (0.04)	66 (79)	4.5
10	Welwyn Garden City	UK	Retired	M	6593	1 (0.02)	68 (74)	4
11	Paris	France	Librarian	F	4198	3 (0.07)	65 (75)	5
12	Tallinn	Estonia	Archival Inspector	M	4530	8 (0.18)	63(79)	4
13	Munich	Germany	Journalist	M	4350	32 (0.74)	62 (74)	4.66
14	Split	Croatia	Student	M	6629	0	63 (77)	n/a
15	Gurgaon	India	Technology Consultant	M	4510	12 (0.27)	67 (78)	3.67
16	Managua	Nicaragua	Computer analyst	M	6547	2 (0.03)	68 (78)	5.5
17	Katikati	New Zealand	Potter	M	4091	12 (0.3)	62 (74)	4.33
18	Elderton	USA	Lutheran Pastor	M	4438	9 (0.21)	68 (76)	4.33
Totals					98,847	384 (0.4)	63 (75)	4.05

Social search—the Columbia experiment

- 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).

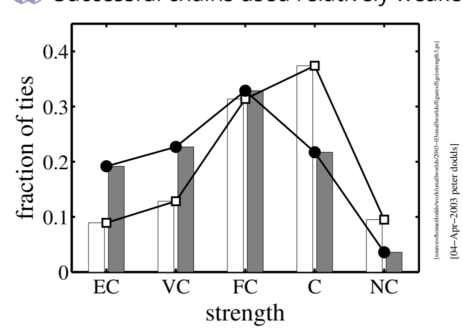
Social search—the Columbia experiment

- Motivation/Incentives/Perception matter.
- If target *seems* reachable
  $\Rightarrow$ participation more likely.
- Small changes in attrition rates
  $\Rightarrow$ large changes in completion rates
- e.g., $\searrow$ 15% in attrition rate
  $\Rightarrow$ $\nearrow$ 800% in completion rate

Social search—the Columbia experiment

Comparing successful to unsuccessful chains:

- Successful chains used relatively weaker ties:



Social search—the Columbia experiment

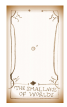
Successful chains disproportionately used:

- Weak ties, Granovetter [7]
- 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 $\rightarrow$ Work

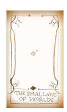


Social search—the Columbia experiment

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%



Social search—the Columbia experiment

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.)

Social search—the Columbia experiment

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: $L_* \simeq 9$

Usefulness:

Harnessing social search:

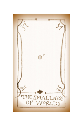
- Can distributed social search be used for something big/good?
- What about something evil? (Good idea to check.)
- What about socio-inspired algorithms for information search? (More later.)
- For real social search, we have an incentives problem.
- Which kind of influence mechanisms/algorithms would help propagate search?
- Fun, money, prestige, ... ?
- Must be 'non-gameable.'

Red balloons:

A Grand Challenge:

- 1969: The Internet is born (the ARPANET—four nodes!).
- Originally funded by DARPA who created a grand Network Challenge for the 40th anniversary.
- Saturday December 5, 2009: DARPA puts 10 red weather balloons up during the day.
- Each 8 foot diameter balloon is anchored to the ground somewhere in the United States.
- Challenge: Find the latitude and longitude of each balloon.
- Prize: $40,000.

*DARPA = Defense Advanced Research Projects Agency.



Where the balloons were:

Finding red balloons:

The winning team and strategy:

- MIT's Media Lab won in less than 9 hours.[9]
- Pickard et al. "Time-Critical Social Mobilization,"[9] Science Magazine, 2011.
- People were virally recruited online to help out.
- Idea: Want people to both (1) find the balloons, and (2) involve more people.
- Recursive incentive structure with exponentially decaying payout:
  - $2000 for correctly reporting the coordinates of a balloon.
  - $1000 for recruiting a person who finds a balloon.
  - $500 for recruiting a person who recruits the balloon finder, ...
  - (Not a Ponzi scheme.)
- True victory: Colbert interviews Riley Crane

Finding balloons:

Clever scheme:

- Max payout = $4000 per balloon.
- Individuals have clear incentives to both
  1. involve/source more people (spread), and
  2. find balloons (goal action).
- Gameable?
- Limit to how much money a set of bad actors can extract.

Extra notes:

- MIT's brand helped greatly.
- MIT group first heard about the competition a few days before. Ouch.
- A number of other teams did well.
- Worthwhile looking at these competing strategies.[9]

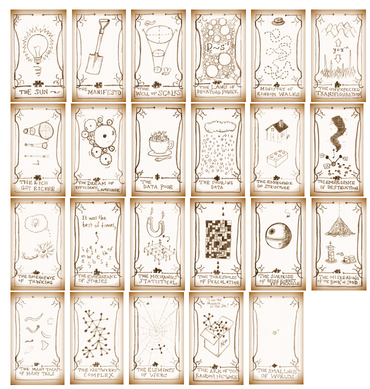
Collective Detective:

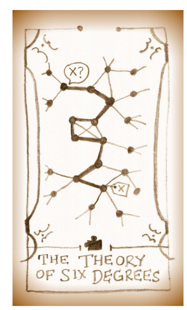
- Finding an errant panda


Once again, social media proved to be a powerful dragnet. Around 1:15 p.m., a Washingtonian posted a picture on Twitter of Rusty in a patch of weeds in the Adams Morgan district, not far from the 163-acre zoo, which was created in 1889 by an act of Congress. "Red panda in our neighborhood," wrote Ashley Foughty, who identified herself as a singer, actress and traveler. "Please come save him!"

Another neighbor posted a photograph of two zoo workers, one in safari shorts standing on a rooftop, one holding a giant butterfly net. Soon the zoo announced: "Rusty the red panda has been recovered, crated & is headed safely back to the National Zoo!"

- Nature News: "Crowdsourcing in manhunts can work: Despite mistakes over the Boston bombers, social media can help to find people quickly" by Philip Ball (April 26, 2013)
- Motherboard, Vice: One Degree of Separation in the Forever War by Brian Castner (November 11, 2015)





The social world appears to be small... why?

Theory: how do we understand the small world property?

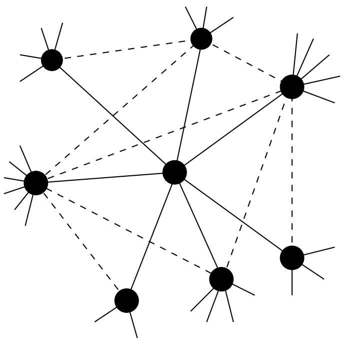
- Connected random networks have short average path lengths:

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

  N = population size,

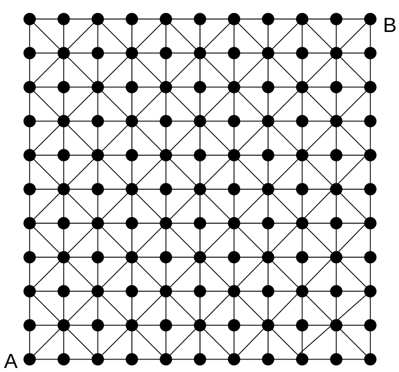
  d_{AB} = distance between nodes A and B.
- But: social networks aren't random...

Simple socialness in a network:



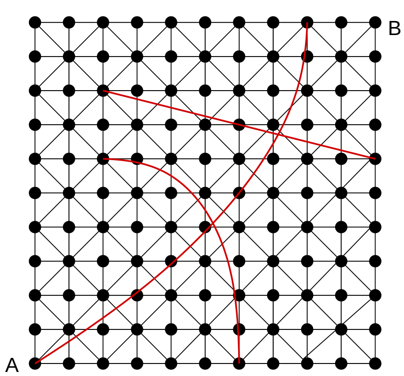
Need "clustering" (your friends are likely to know each other):

Non-randomness gives clustering:



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

Randomness + regularity



Now have $d_{AB} = 3$

$\langle d \rangle$ decreases overall

Small-world networks

Introduced by Watts and Strogatz (Nature, 1998) [14]
"Collective dynamics of 'small-world' networks."

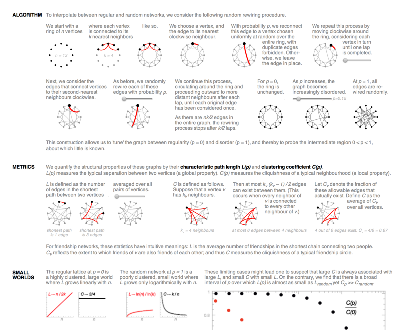
Small-world networks were found everywhere:

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

Very weak requirements:

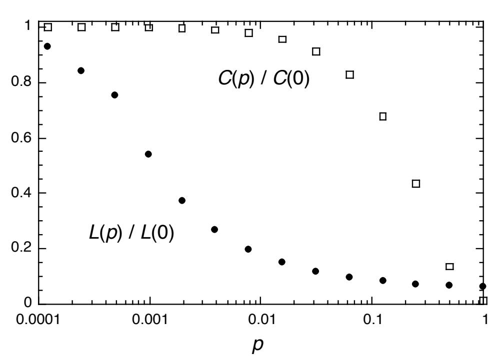
- local regularity + random short cuts

Papers should be apps:



- Bret Victor's Scientific Communication As Sequential Art
- Interactive figures and tables = windows into large data sets (empirical or simulated).

The structural small-world property:



- $L(p)$ = average shortest path length as a function of p
- $C(p)$ = average clustring as a function of p

Previous work—finding short paths

But are these short cuts findable?

Nope. [8]

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

Need a more sophisticated model...

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

Previous work—finding short paths

Jon Kleinberg (Nature, 2000) [8]
"Navigation in a small world."

Allowed to vary:

1. local search algorithm
   and
2. network structure.

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

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

- $\alpha = 0$: random connections.
- α large: reinforce local connections.
- $\alpha = d$: connections grow logarithmically in space.

Previous work—finding short paths

Theoretical optimal search:

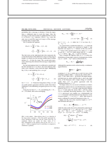
- "Greedy" algorithm.
- Number of connections grow logarithmically (slowly) in space: $\alpha = d$.
- Social golf.

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

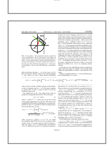
But: social networks aren't lattices plus links.

Advances for understanding Kleinberg's model:

"Kleinberg Navigation in Fractal Small World Networks"
Roberson and ben-Avrahma,
Phys. Rev. E, **74**, 017101, 2006. [10]



"Asymptotic behavior of the Kleinberg model"
Carmi et al.,
Phys. Rev. Lett., **102**, 238702, 2009. [4]



"Extended navigability of small world networks: Exact results and new insights"
Cartoza and De Los Rios,
Phys. Rev. Lett., **2009**, 238703, 2009. [5]

Previous work—finding short paths

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

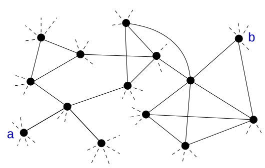
$$P(k_i) \propto k_i^{-\gamma}$$

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

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'?

Models

One approach: incorporate identity.

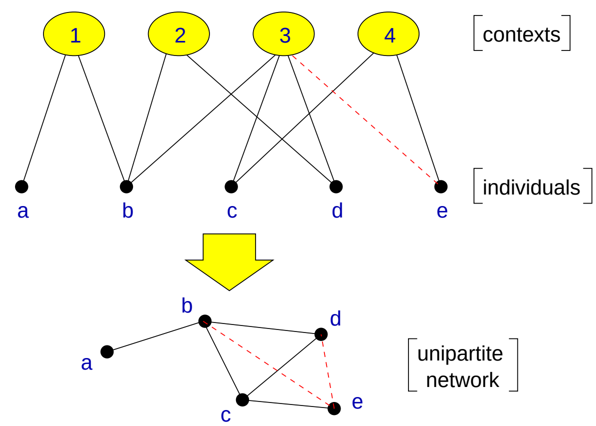
Identity is formed from attributes such as:

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

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

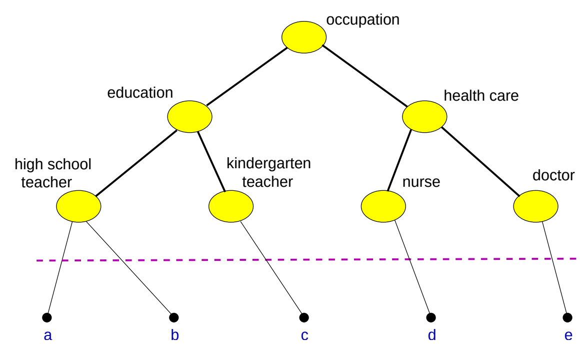
Attributes ⇔ Contexts ⇔ Interactions ⇔ Networks.

Social distance—Bipartite affiliation networks



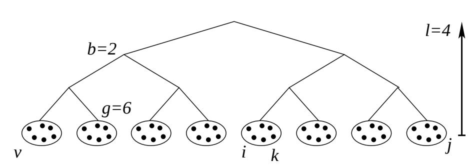
- Bipartite affiliation networks: boards and directors, movies and actors.

Social distance—Context distance



Models

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



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

Models

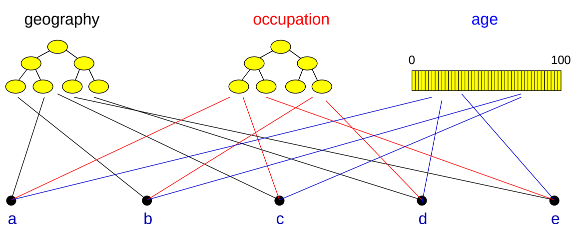
- Individuals are more likely to know each other the closer they are within a hierarchy.
- Construct z connections for each node using

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

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

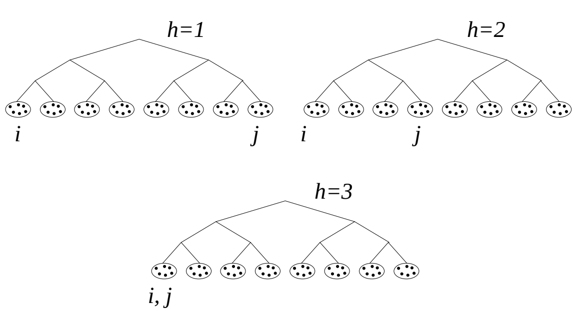
Models

Generalized affiliation networks



- Blau & Schwartz [2], Simmel [11], Breiger [3], Watts *et al.* [13]; see also Google+ Circles.

The model



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

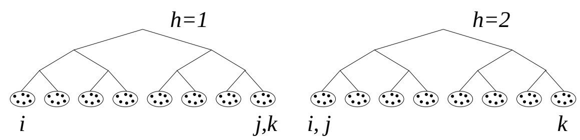
$x^1_{ij} = 4,\ x^2_{ij} = 3,\ x^3_{ij} = 1.$

Social distance:

$$y_{ij} = \min_h x^h_{ij}.$$

The model

Triangle inequality doesn't hold:



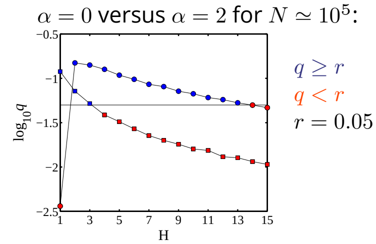
$$y_{ik} = 4 > y_{ij} + y_{jk} = 1 + 1 = 2.$$

The model

- 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

$\alpha = 0$ versus $\alpha = 2$ for $N \simeq 10^5$:

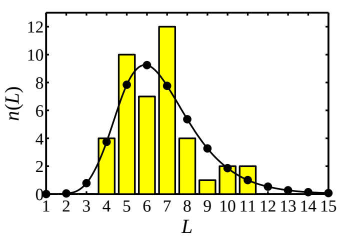


q = probability an arbitrary message chain reaches a target.

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

The model-results

Milgram's Nebraska-Boston data:



Model parameters:

- $N = 10^8$,
- $z = 300$, $g = 100$,
- $b = 10$,
- $\alpha = 1$, $H = 2$;

- $\langle L_{\text{model}} \rangle \simeq 6.7$
- $L_{\text{data}} \simeq 6.5$

Social search—Data

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: bitly.com
- (e.g., Wikipedia)
- Photo tagging: flickr.com
- Dynamic creation of metadata plus links between information objects.
- Folksonomy: collaborative creation of metadata

Social Search—Real world uses

Recommender systems:

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

Nutshell for Small-World Networks:

- 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.

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. M. Blau and J. E. Schwartz.
Crosscutting Social Circles.
Academic Press, Orlando, FL, 1984.

[3] R. L. Breiger.
The duality of persons and groups.
Social Forces, 53(2):181–190, 1974. pdf

[4] S. Carmi, S. Carter, J. Sun, and D. ben Avrahim.
Asymptotic behavior of the Kleinberg model.
Phys. Rev. Lett., 102:238702, 2009. pdf

References II

[5] C. C. Cartoza and P. De Los Rios.
Extended navigability of small world networks: Exact results and new insights.
Phys. Rev. Lett., 2009:238703, 2009. pdf

[6] P. S. Dodds, R. Muhamad, and D. J. Watts.
An experimental study of search in global social networks.
Science, 301:827–829, 2003. pdf

[7] M. Granovetter.
The strength of weak ties.
Am. J. Sociol., 78(6):1360–1380, 1973. pdf

[8] J. Kleinberg.
Navigation in a small world.
Nature, 406:845, 2000. pdf

References III

[9] G. Pickard, W. Pan, I. Rahwan, M. Cebrian, R. Crane, A. Madan, and A. Pentland.
Time-critical social mobilization.
Science, 334:509–512, 2011. pdf

[10] M. R. Roberson and D. ben Avrahma.
Kleinberg navigation in fractal small world networks.
Phys. Rev. E, 74:017101, 2006. pdf

[11] G. Simmel.
The number of members as determining the sociological form of the group. I.
American Journal of Sociology, 8:1–46, 1902.

References IV

[12] J. Travers and S. Milgram.
An experimental study of the small world problem.
Sociometry, 32:425–443, 1969. pdf

[13] D. J. Watts, P. S. Dodds, and M. E. J. Newman.
Identity and search in social networks.
Science, 296:1302–1305, 2002. pdf

[14] D. J. Watts and S. J. Strogatz.
Collective dynamics of 'small-world' networks.
Nature, 393:440–442, 1998. pdf