Web Structures and Algorithms, CMU

Social Search and the Small World Phenomenon: Experiment and Theory

Collective Dynamics Group, Columbia University

D. J. Watts, M. E. J. Newman, R. Muhamad, P. S. Dodds

➤ NSF
➤ Legg Mason
➤ McDonnell Foundation
➤ Office of Naval Research

ISERP
Dept. of Sociology
Columbia Earth Institute

Outline:

Social search: The Small World Phenomenon

1. History of the small-world problem.

2. Previous work on network search.

3. Current model.

4. Online Experiment.

The problem:

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?

Social Search:

A small slice of the pie:

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

Social Search:

A small slice of the pie:

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

A. Yes (apparently):
The small world phenomenon
or
"Six Degrees of Separation."

The problem:

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:

Two significant features characterize a small-world network:

1. Short paths exist.

The problem:

Two significant features characterize a small-world network:

1. Short paths exist.

and

2. People are good at finding them.

Previous work—short paths:

Connected random networks
have short average path lengths:

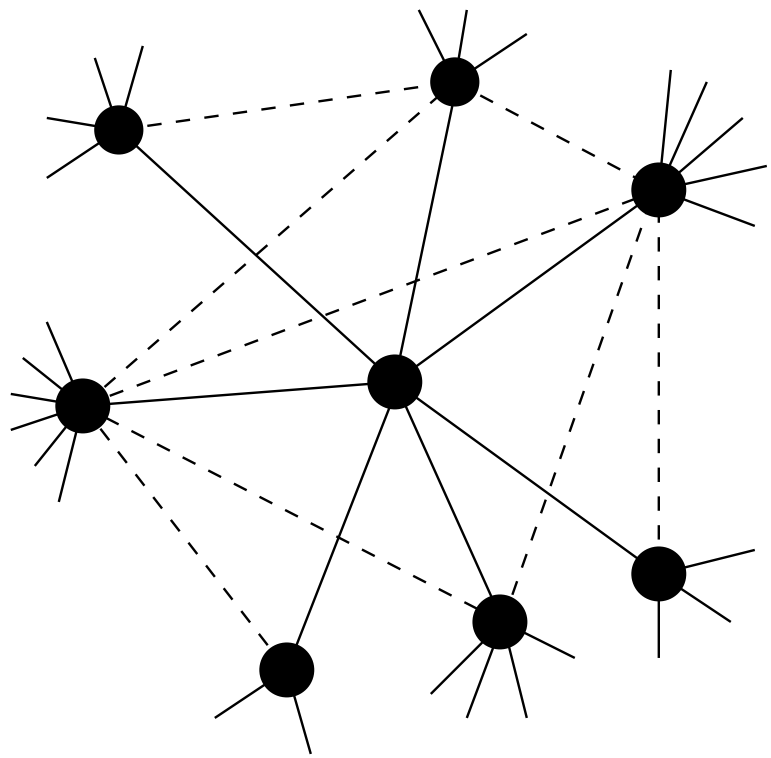
$\langle d_{AB} \rangle \sim \log(N)$

N = population size,
d_{AB} = distance between nodes A and B.

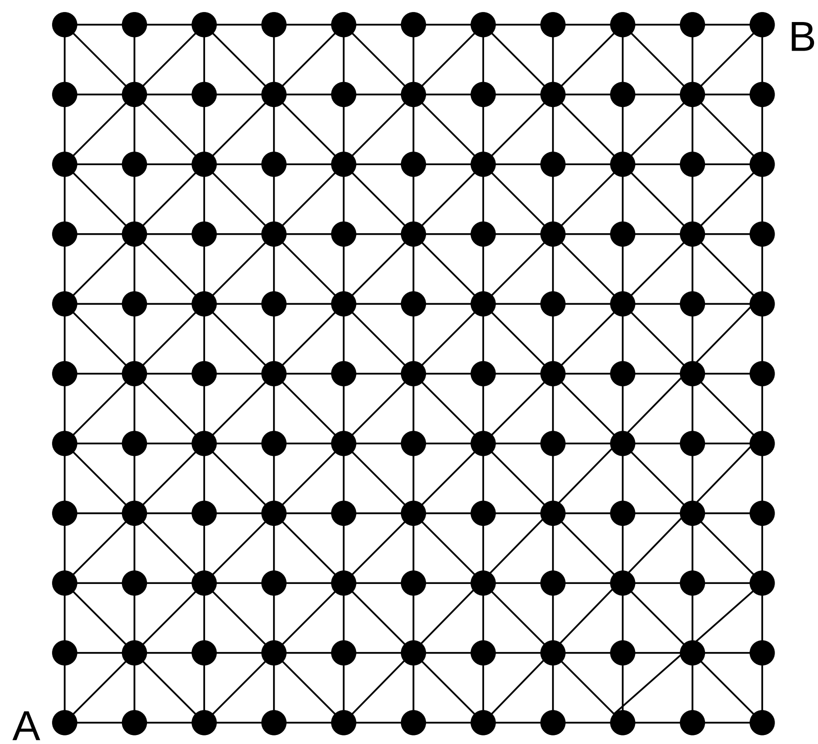
But: social networks aren't random.

Previous work—short paths:

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

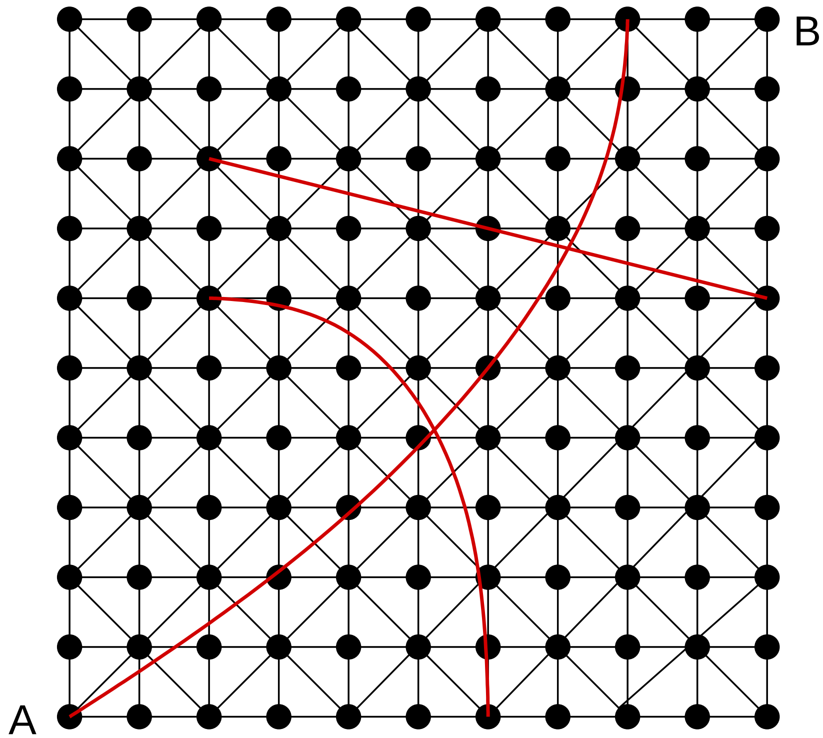


Non-randomness gives clustering:



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

Randomness + regularity:



Now have $d_{AB} = 3$

$\langle d \rangle$ decreases overall

Previous work—short paths:

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

Small-world networks found everywhere:

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

Previous work—short paths:

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

Small-world networks found everywhere:

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

Very weak requirements:
local regularity
+
random short cuts.

Previous work—finding short paths:

But are these short cuts findable?

Previous work—finding short paths:

But are these short cuts findable?

No.

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

Previous work—finding short paths:

What can a local search method 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),
"Navigation in a small world."

Allowed to vary:

1. local search algorithm,
and
2. network structure.

Previous work—finding short paths:

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

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

Previous work—finding short paths:

Theoretical optimal search:

1. "Greedy" algorithm.

2. $\alpha = d$.

Search time grows like $\log^2(N)$.

For $\alpha \neq d$, polynomial factor N^β appears.

But: social networks aren't lattices plus links.

Previous work—finding short paths:

If networks have hubs can also search well (Adamic et al.)

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

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:

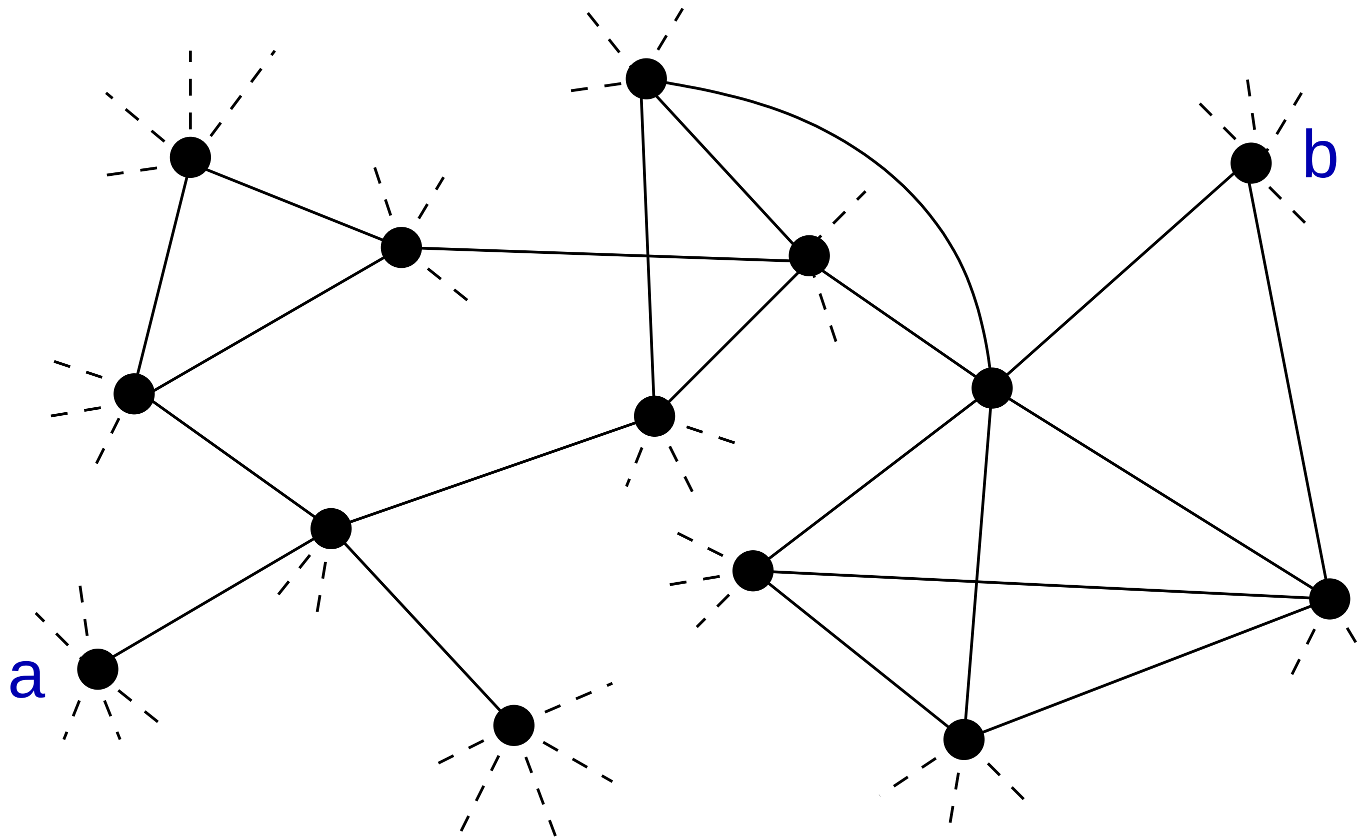
If there are no hubs and no underlying lattice, how can search be efficient?

Which friend is closest to the target?

What does ‘closest’ mean?

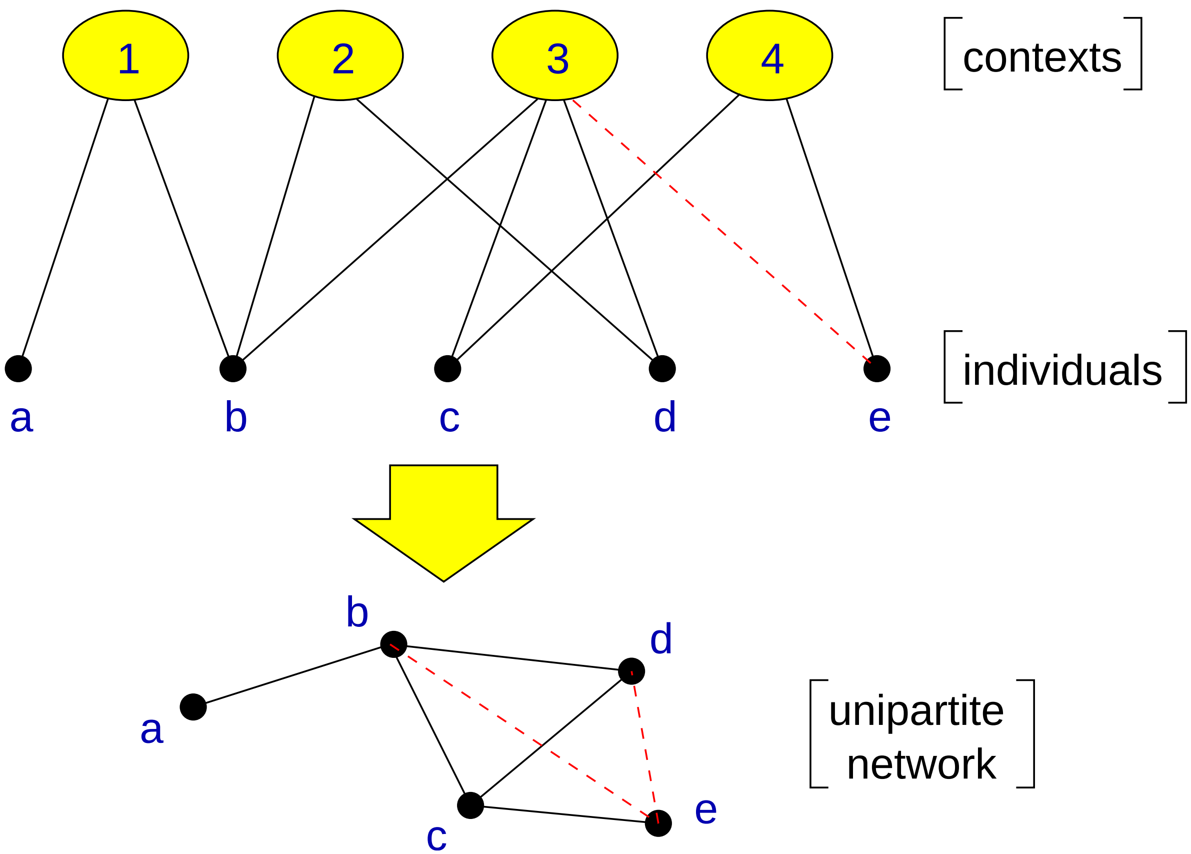
How to measure ‘social distance’?

The problem—Bare networks:



How can a reach b using local information?

Social distance—Bipartite networks:



The model:

One approach: incorporate identity.

The model:

One approach: incorporate identity.

Identity is formed from attributes such as:

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

The model:

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.

The model:

One approach: incorporate identity.

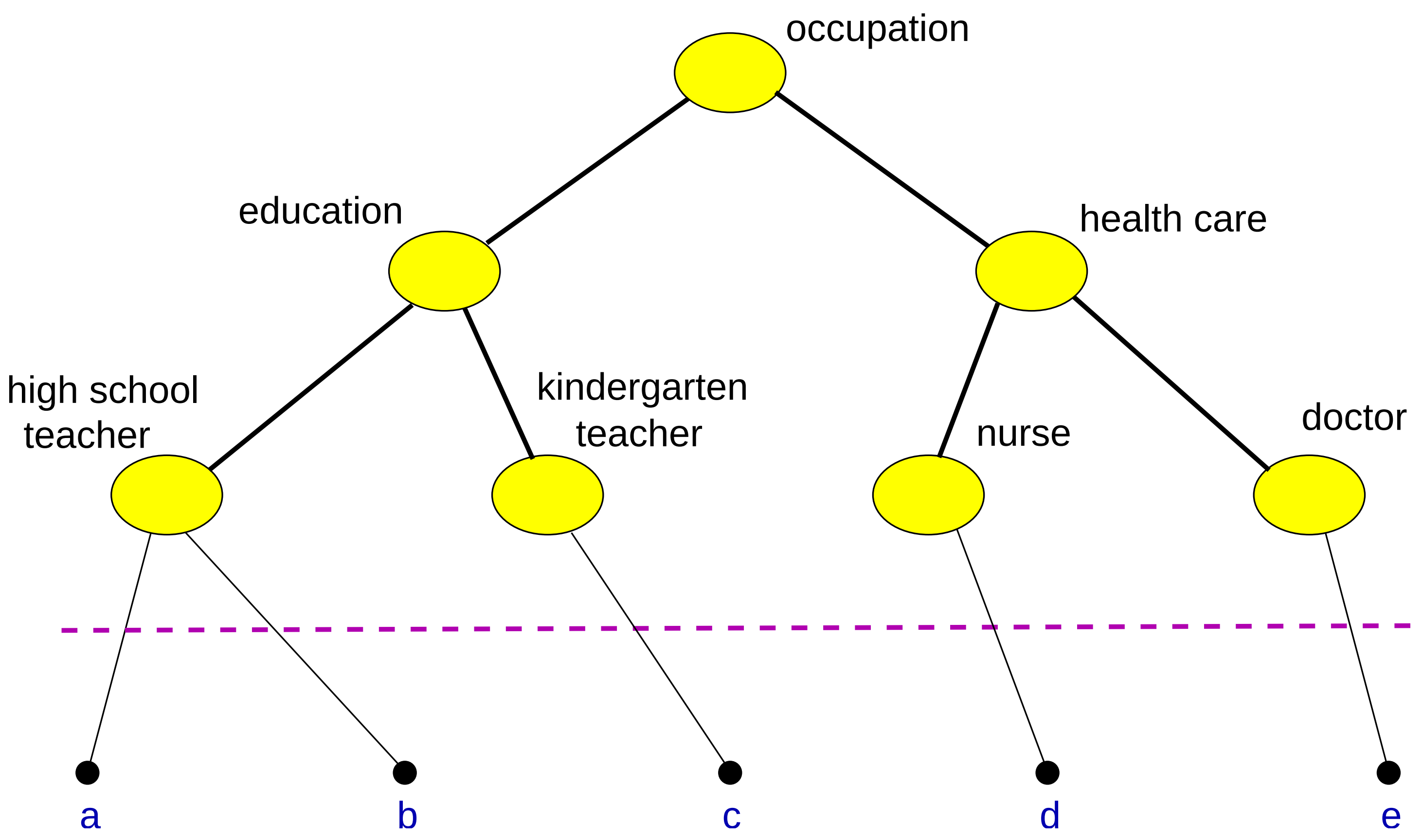
Identity is formed from attributes such as:

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

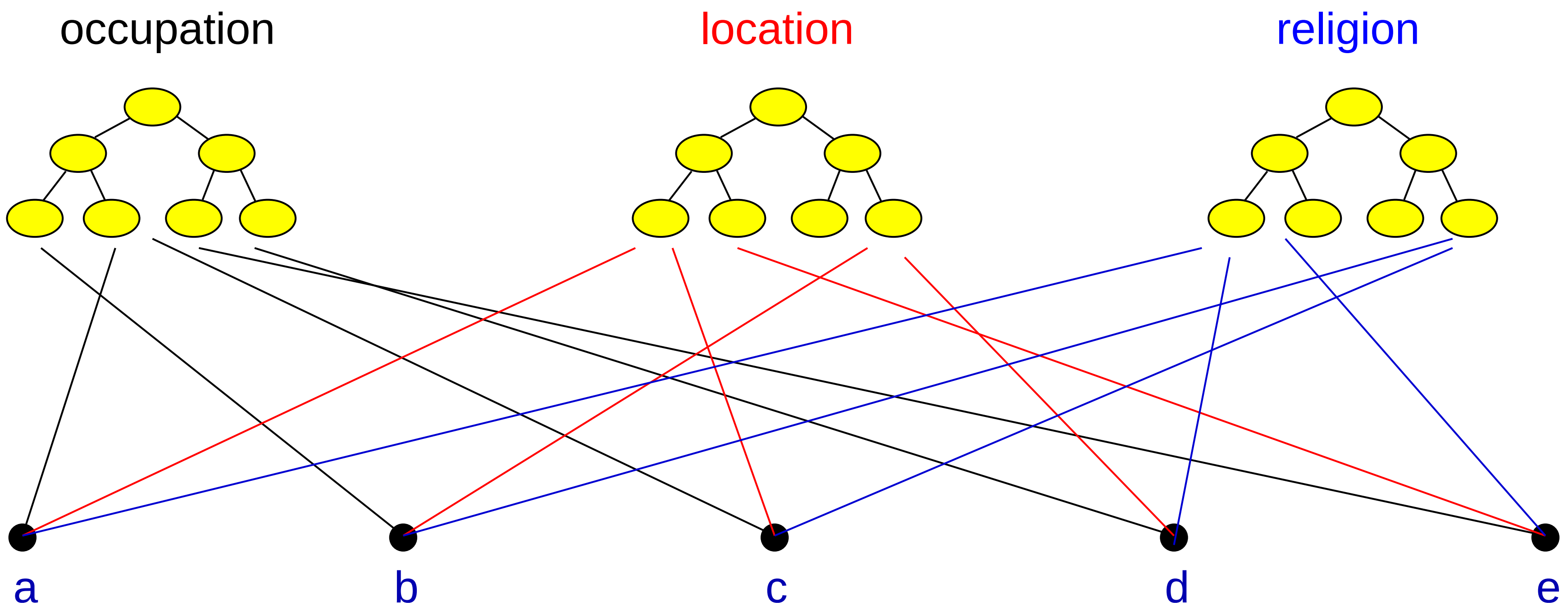
Groups are formed by people with at least one similar attribute.

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

Social distance—Context distance:



Social distance—Generalized context space:



The model:

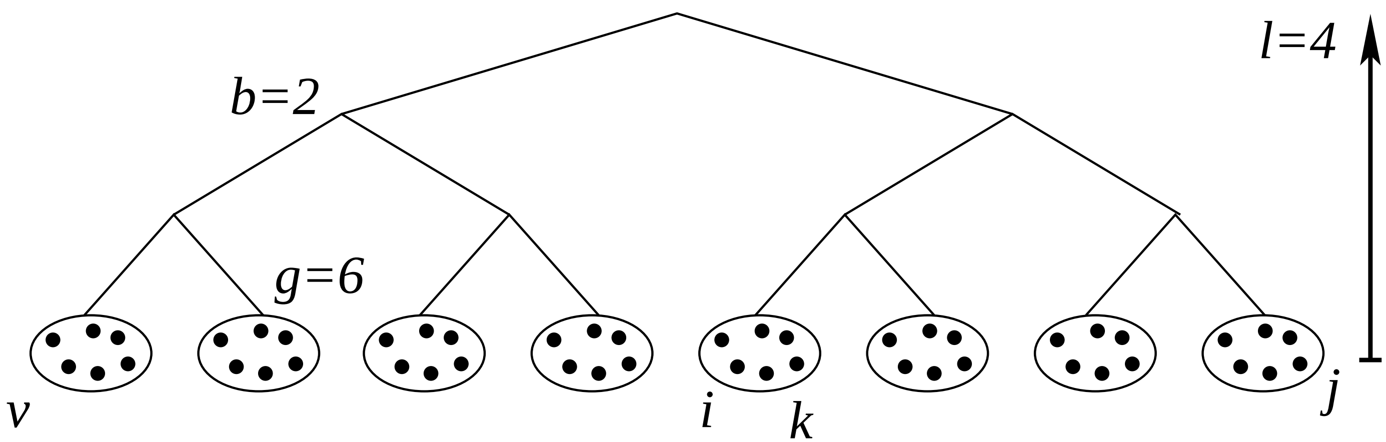
Six propositions about social networks:

P1: Individuals have identities and belong to various groups that reflect these identities.

P2: Individuals break down the world into a hierarchy of categories.

The model:

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



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

The model:

P3: Individuals are more likely to know each other the closer they are within a hierarchy.

Construct z connections for each node using

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

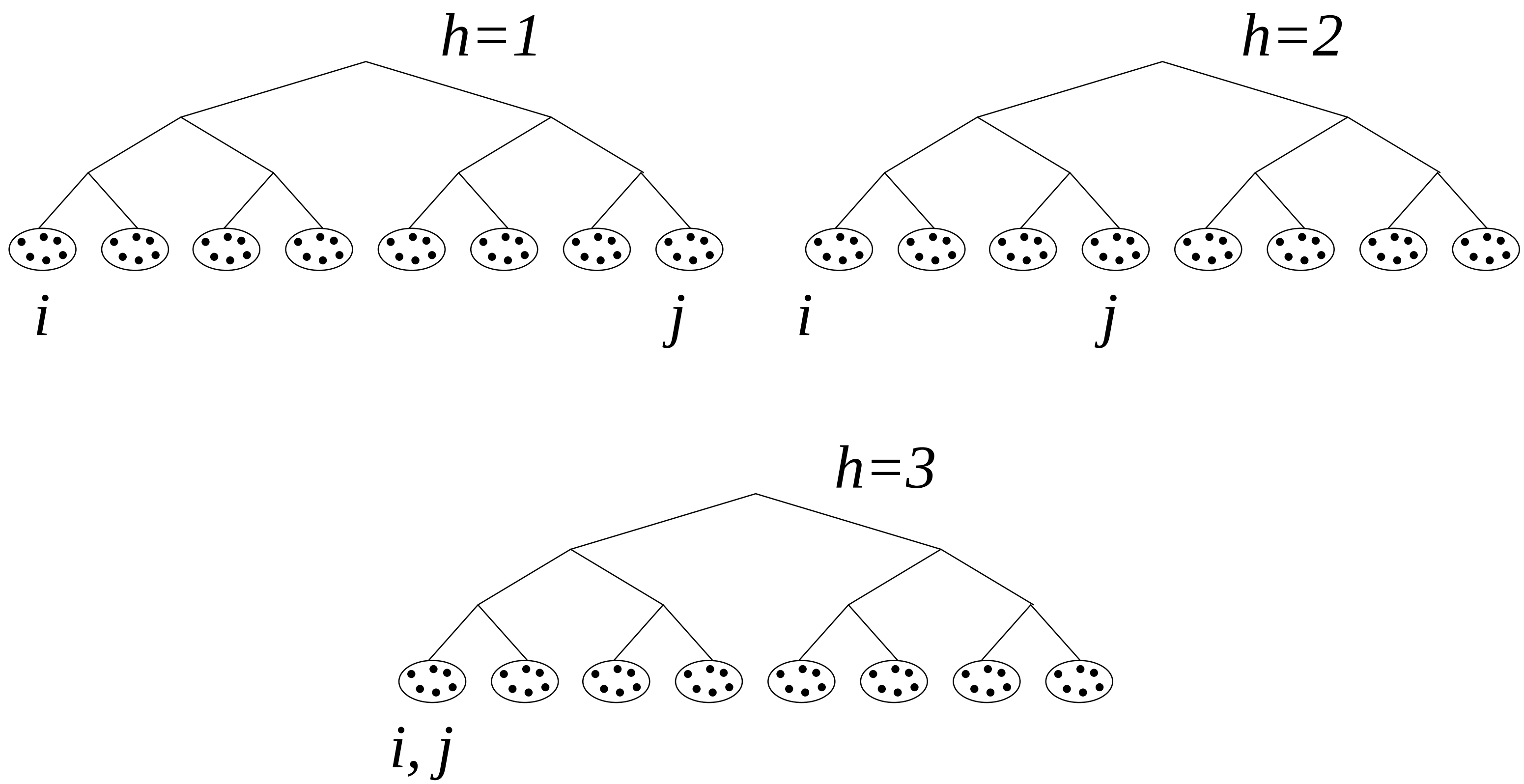
$\alpha = 0$: random connections.

α large: local connections.

The model:

P4: Each attribute of identity $\equiv$ hierarchy.

The model:



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

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

The model:

P5: "Social distance" is the minimum distance between two nodes in all hierarchies.

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

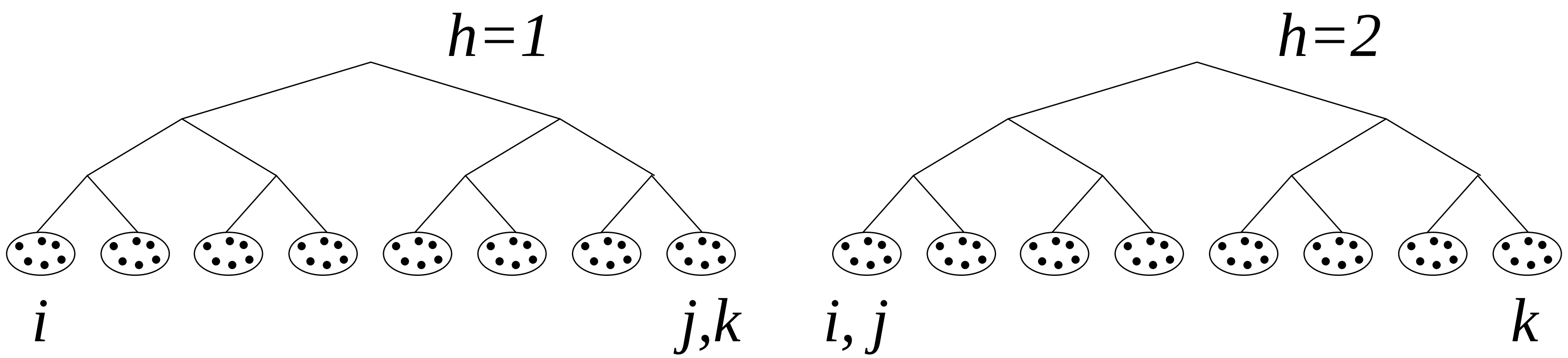
Previous slide:

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

$$\Rightarrow y_{ij} = 1.$$

The model:

Triangle inequality doesn't hold:



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

The model:

P6: 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.

The model:

Define q as probability of an arbitrary message chain reaching a target.

Definition of a searchable network:

Any network for which

$$\boxed{q \geq r}$$

for a desired r.

The model:

If message chains fail at each node with probability p, require

$$q = \langle (1-p)^L \rangle_L \geq r.$$

where $L =$ length of message chain.

The model:

If message chains fail at each node with probability p, require

$$q = \langle (1-p)^L \rangle_L \geq r.$$

where L = length of message chain.

Approximation:

$$\langle L \rangle \leq \ln r / \ln (1-p).$$

The model:

If message chains fail at each node with probability p, require

$$q = \langle (1-p)^L \rangle_L \geq r.$$

where L = length of message chain.

Approximation:

$$\langle L \rangle \leq \ln r / \ln (1-p).$$

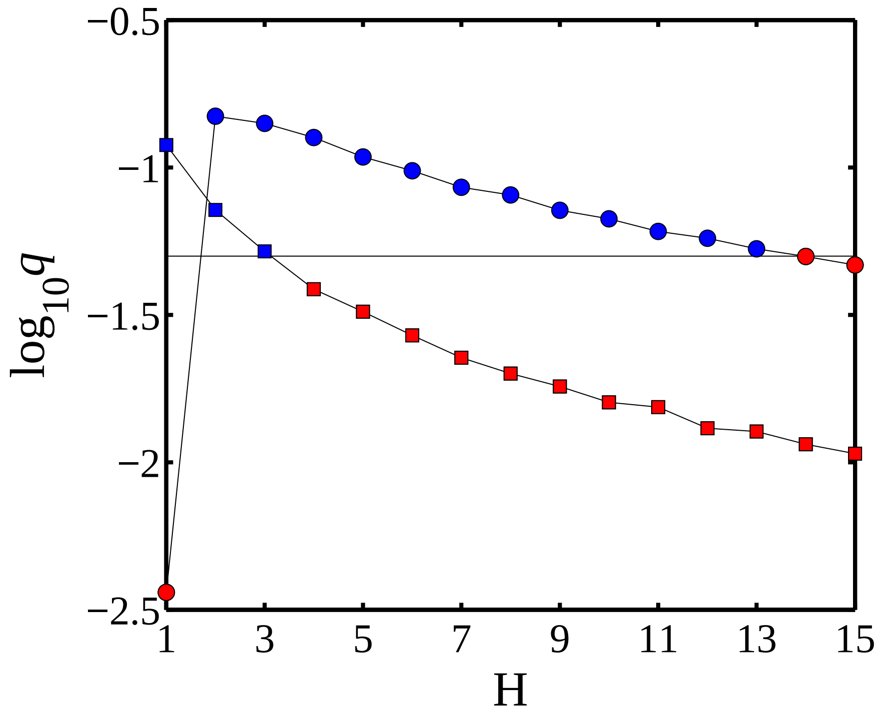
For $r = 0.05$ and $p = 0.25$,

$$\langle L \rangle \lesssim 10$$

independent of N.

The model-results:

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

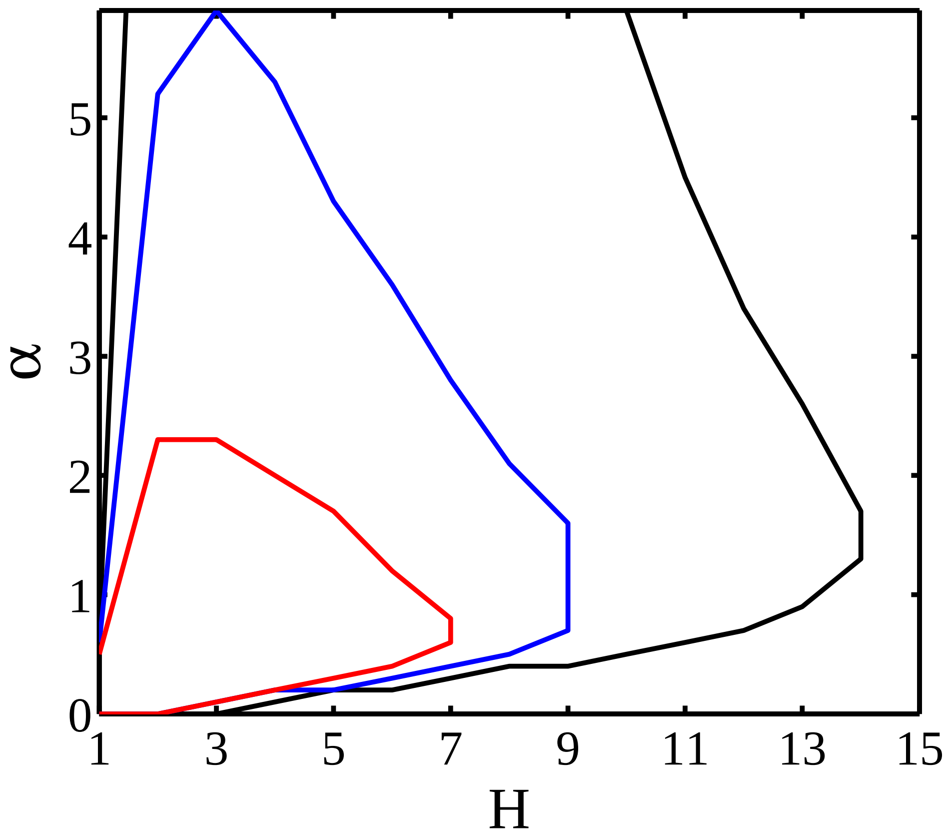


$q \geq r$

$q < r$

$r = 0.05$

The model-results:



$N = 102400$

$N = 204800$

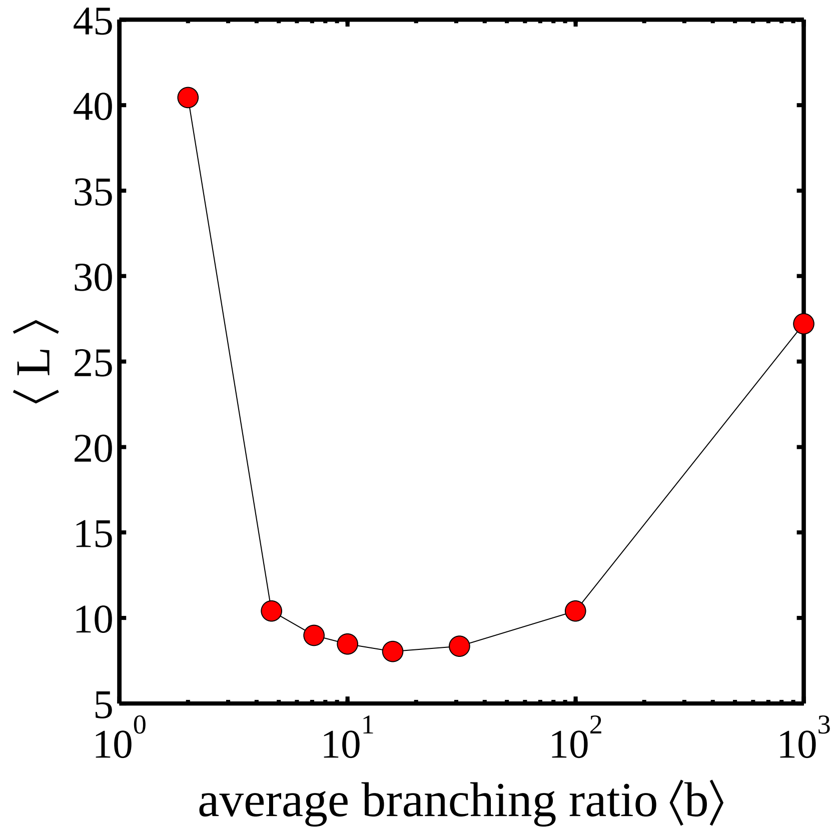
$N = 409600$

$p = 0.25$, $r = 0.05$

$p = 0.25$, $r = 0.05$

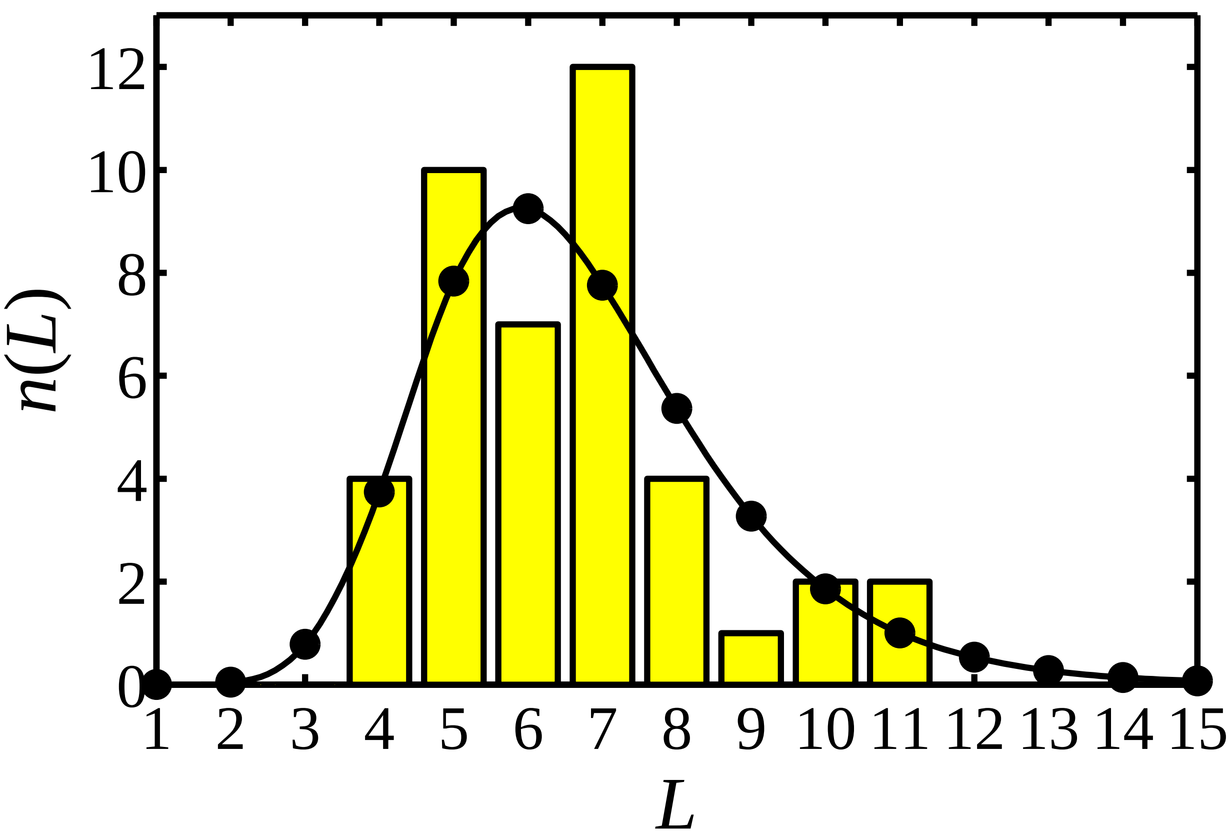
The model-results:

$N \simeq 10^8$:



The model-results:

Milgram's Nebraska-Boston data:



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

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

Conclusions:

- Bare networks are typically unsearchable.
- Paths are findable if nodes understand how network is formed.
- Importance of identity (interaction contexts).

Applications:

- Improved social network models.
- Construction of peer-to-peer networks.
- Construction of searchable information databases.

Social search:

The Small World Phenomenon

Online Experiment

Social search—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

Social search—Experiment:

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—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—Experiment:

Successful chains disproportionately used

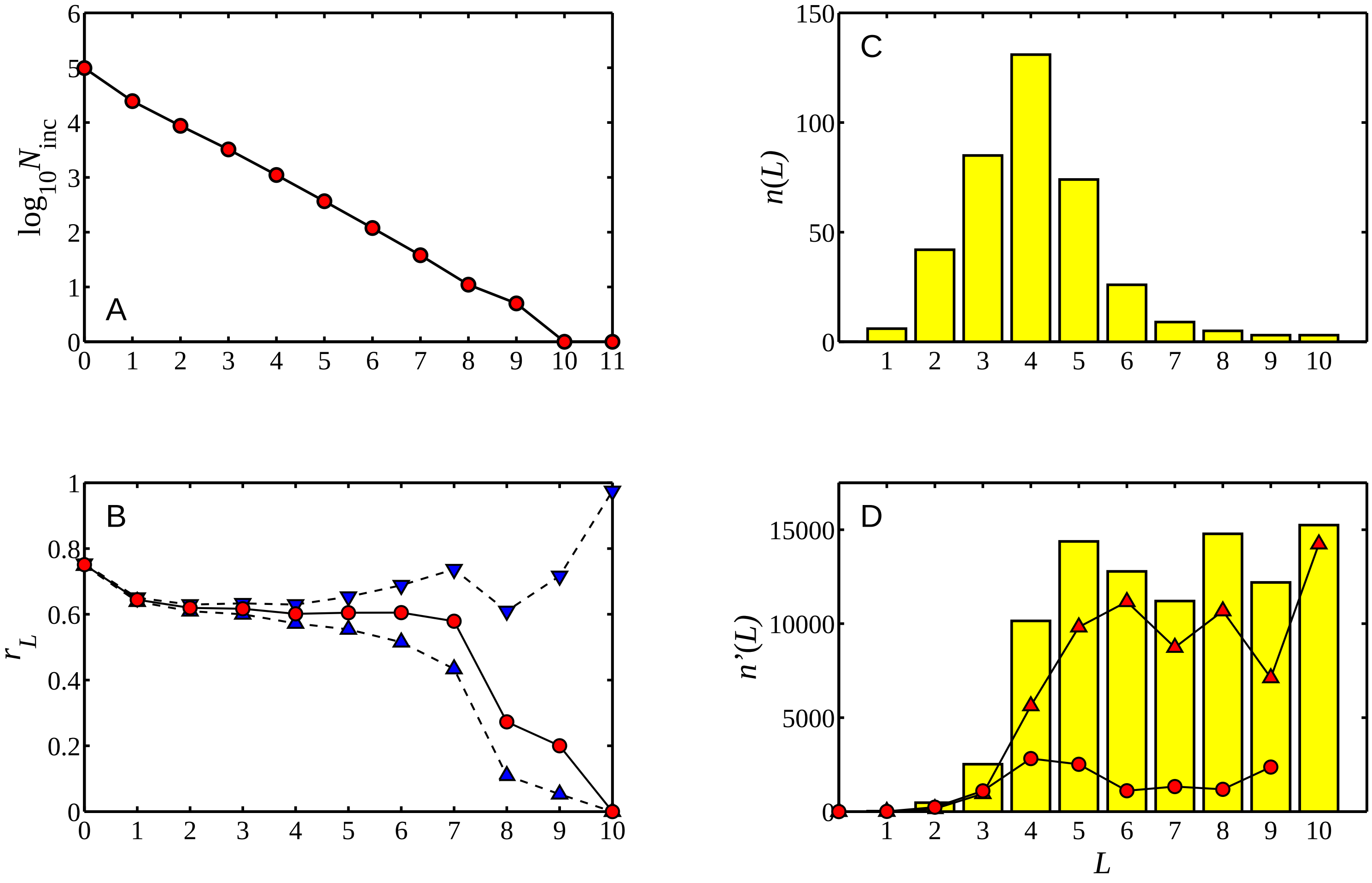
- weak ties (Granovetter)
- professional ties (34% vs. 13%)
- ties originating at work/college
- target's work (65% vs. 40%)

. . . and disproportionately avoided

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

Geography $\rightarrow$ Work

Social search—Experiment:



Social search—Experiment:

$\langle L \rangle = 4.05$ for all completed chains

Social search—Experiment:

$\langle L \rangle = 4.05$ for all completed chains

$L_* =$ Estimated 'true' median chain length

Intra-country chains: $L_* = 5$

Inter-country chains: $L_* = 7$

All chains: $L_* = 7$

Social search—Experiment:

$\langle L \rangle = 4.05$ for all completed chains

$L_* =$ Estimated ‘true’ median chain length

Intra-country chains: $L_* = 5$

Inter-country chains: $L_* = 7$

All chains: $L_* = 7$

Milgram: $L_* \simeq 8$–9

Social search—Experiment:

Other experiments:

1. Small World Experiment II (now running)

2. The People Finder project

3. Expert search

Social search—Experiment:

Other experiments:

1. Small World Experiment II (now running)

2. The People Finder project

3. Expert search

4. The universal badness of Spam...

<u>*References:*</u>

D. J. Watts, P. S. Dodds, & M. E. J. Newman.
"Identity and Search in Social Networks"
Science, **296**, 1302–1305, 2002.

P. S. Dodds, R. Muhamad, & D. J. Watts.
"An Experimental study of Search in Global Social Networks"
Science, **301**, 827–829, 2003.