The Abdus Salam International Centre for Theoretical Physics

United Nations Educational, Scientific and Cultural Organization

International Atomic Energy Agency

SMR.1656 - 12

School and Workshop on Structure and Function of Complex Networks

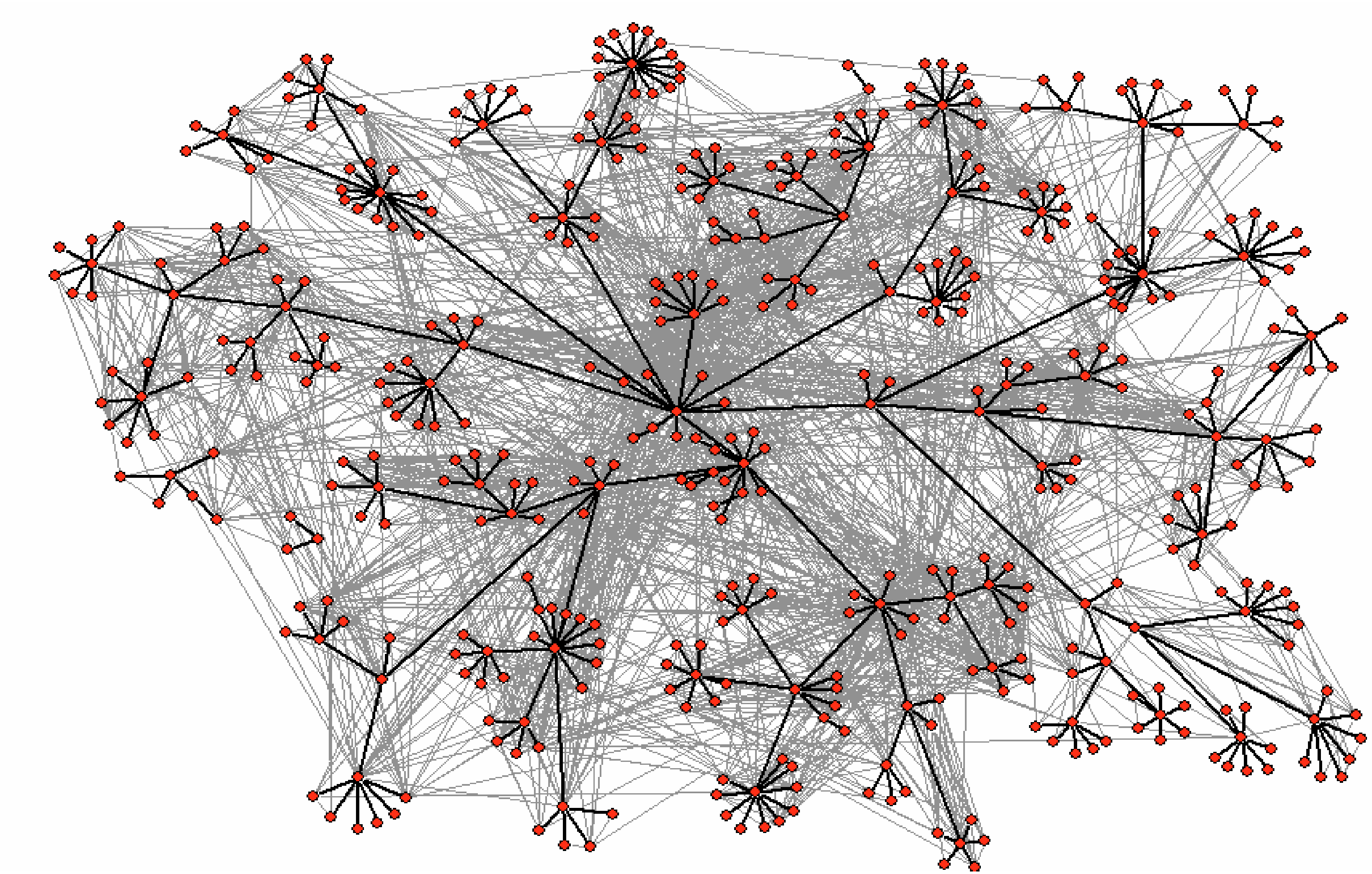
16 - 28 May 2005

Search in Structured Networks

Lada A. ADAMIC
HP Laboratories
1501 Page Mill Road
MS 1139
Palo Alto CA 94024
U.S.A.

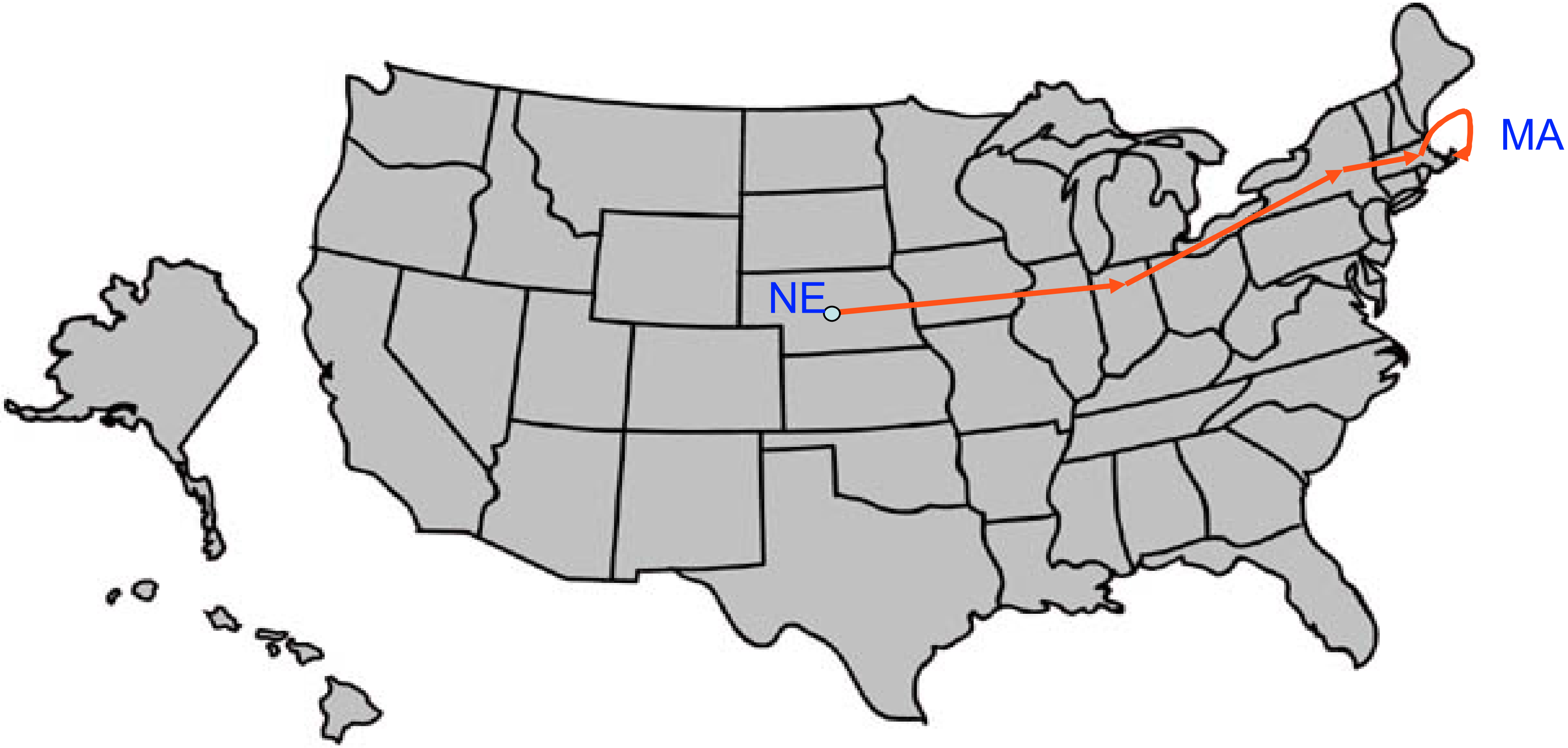
Search in structured networks

Lada Adamic



School on the Structure and Function of Complex Networks, Trieste, 2005

Small world experiments then



Milgram's experiment (1960's):

Given a target individual and a particular property, pass the message to a person you correspond with who is "closest" to the target.

Milgram’s small world experiment

Target person worked in Boston as a stockbroker.

296 senders from Boston and Omaha.

20% of senders reached target.

average chain length = 6.5.

“Six degrees of separation”

Small world experiments now

email experiment
Dodds, Muhamad, Watts,
Science 301, (2003)

18 targets
13 different countries

24,163 message chains
384 reached their targets
average path length 4.0

image by Stephen G. Eick
http://www.bell-labs.com/user/eick/index.html
(unrelated to small world experiment…)

Small world experiment at Columbia

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%)

Strategy: Geography -> Work

Why study small world phenomena?

Curiosity:

- Why is the world small?
- How are people able to route messages?

Social Networking as a Business:

- Friendster, Orkut, MySpace
- LinkedIn, Spoke, VisiblePath

Six degrees of separation - to be expected

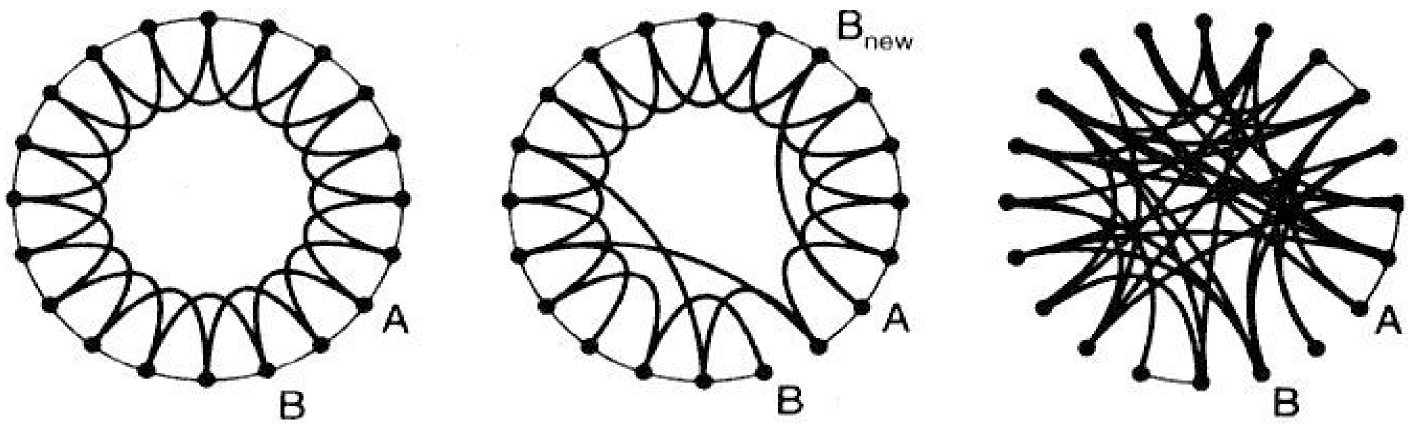
Pool and Kochen (1978) - average person has 500-1500 acquaintances

Ignoring clustering, other redundancy …

~ 10^3 first neighbors, 10^6 second neighbors, 10^9 third neighbors

But networks are clustered: my friends' friends tend to be my friends

Watts & Strogatz (1998) - a few random links in an otherwise clustered graph give an average shortest path close to that of a random graph



But how are people are able to find short paths?

How to choose among hundreds of acquaintances?

Strategy:
Simple greedy algorithm - each participant chooses correspondent who is closest to target with respect to the given property

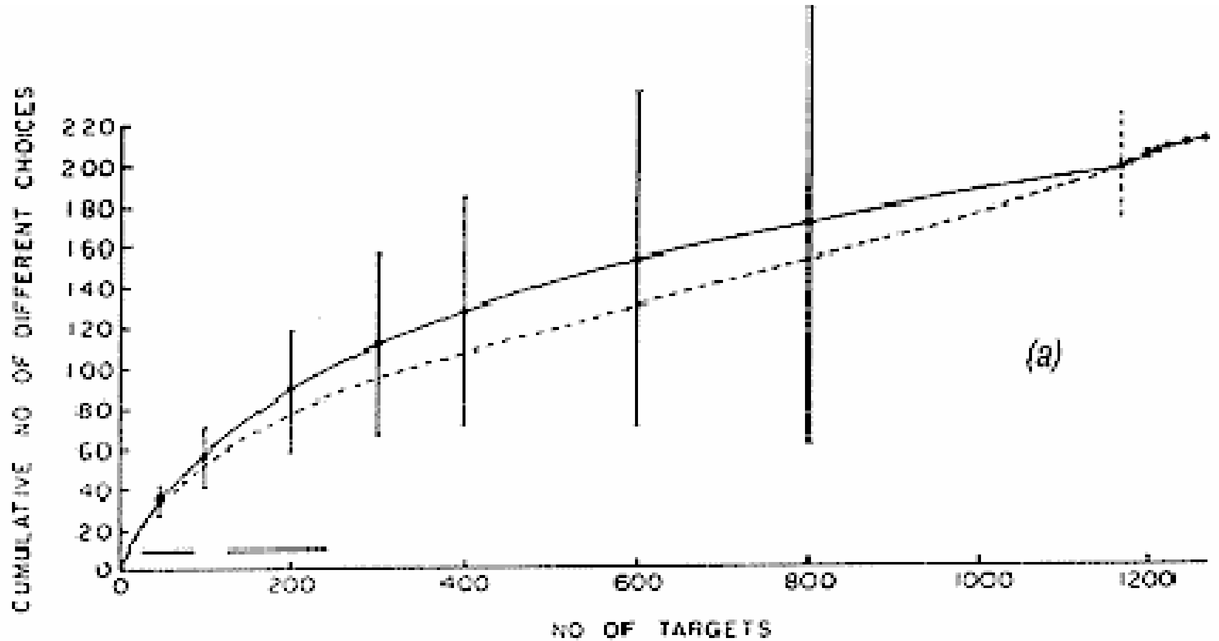
Models

geography
Kleinberg (2000)

hierarchical groups
Watts, Dodds, Newman (2001), Kleinberg(2001)

high degree nodes
Adamic, Puniyani, Lukose, Huberman (2001), Newman(2003)

Reverse small world experiment



Killworth & Bernard (1978):

Given hypothetical targets (name, occupation, location, hobbies, religion…) participants choose an acquaintance for each target

Acquaintance chosen based on

(most often) occupation, geography

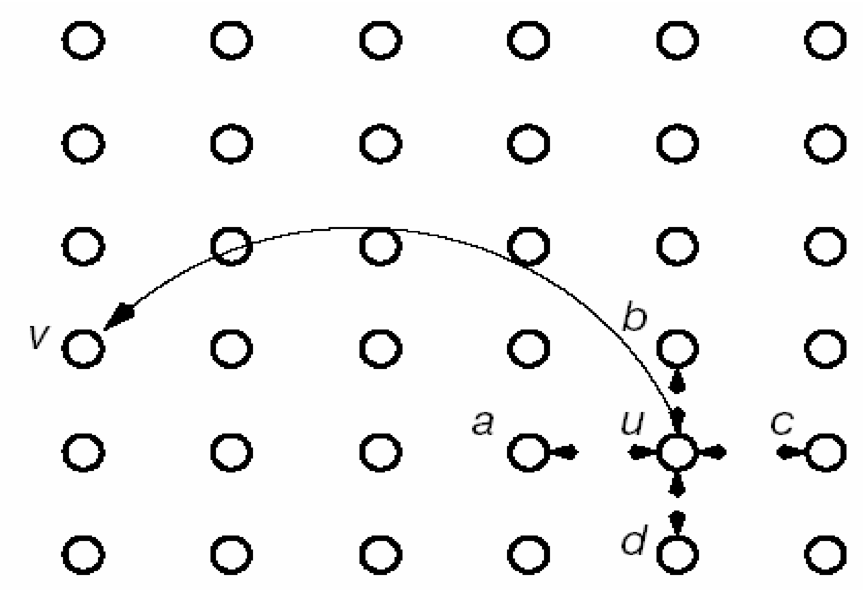
only 7% because they "know a lot of people"

Simple greedy algorithm: most similar acquaintance

two-step strategy rare

Spatial search

<u>Kleinberg, ‘The Small World Phenomenon, An Algorithmic Perspective’</u>
Proc. 32nd ACM Symposium on Theory of Computing, 2000.
(Nature 2000)



“The geographic movement of the [message] from Nebraska to
Massachusetts is striking. There is a progressive closing in on the target
area as each new person is added to the chain”
S.Milgram ‘The small world problem’, Psychology Today 1,61,1967

nodes are placed on a lattice and connect to nearest neighbors

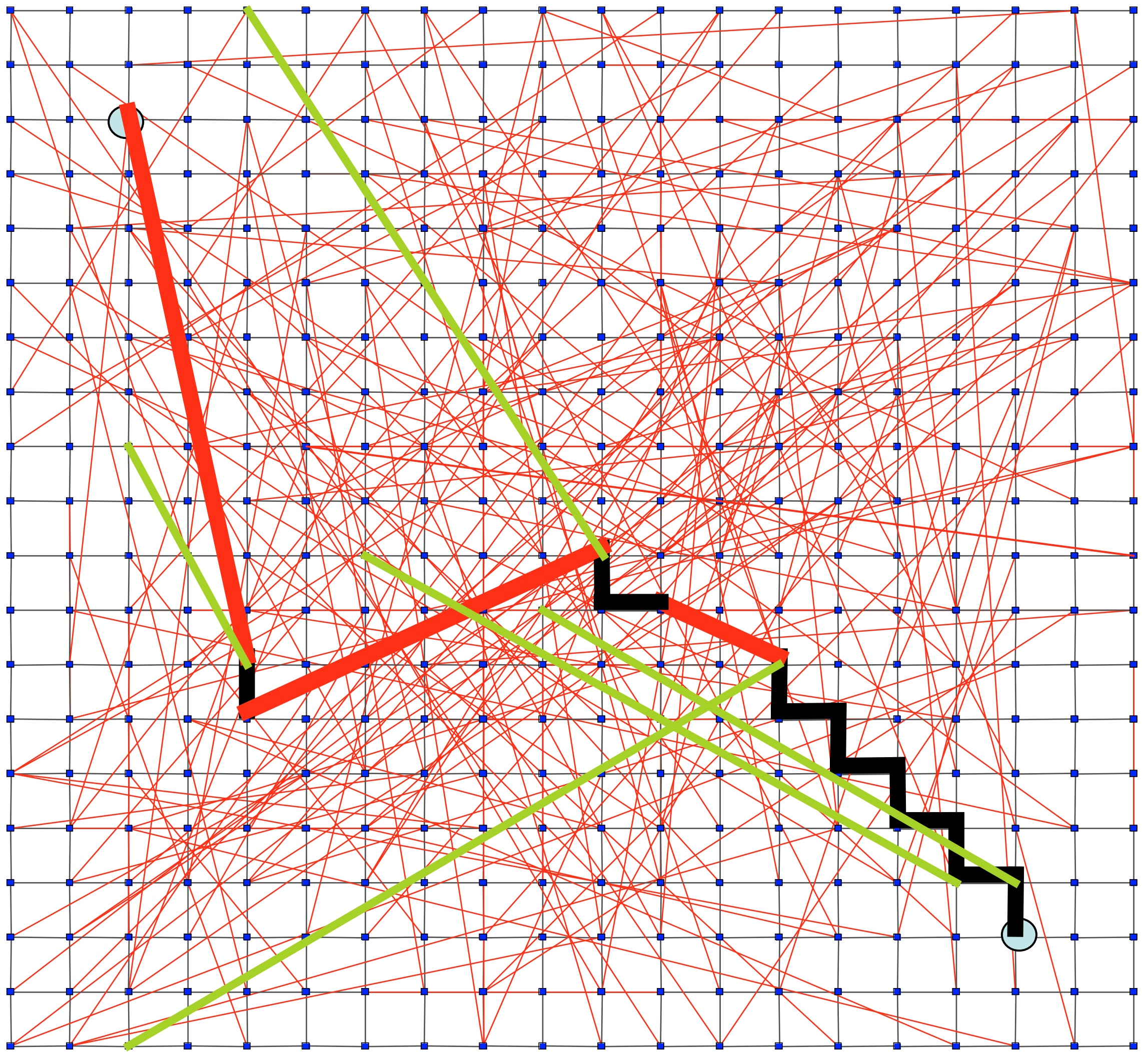
additional links placed with $p_{uv} \sim \quad d_{uv}^{-r}$

no locality

When **r=0**, links are randomly distributed, ASP ~ **log(n)**, n size of grid

When **r=0**, any decentralized algorithm is at least $\mathbf{a_0 n^{2/3}}$

$p \sim p_0$

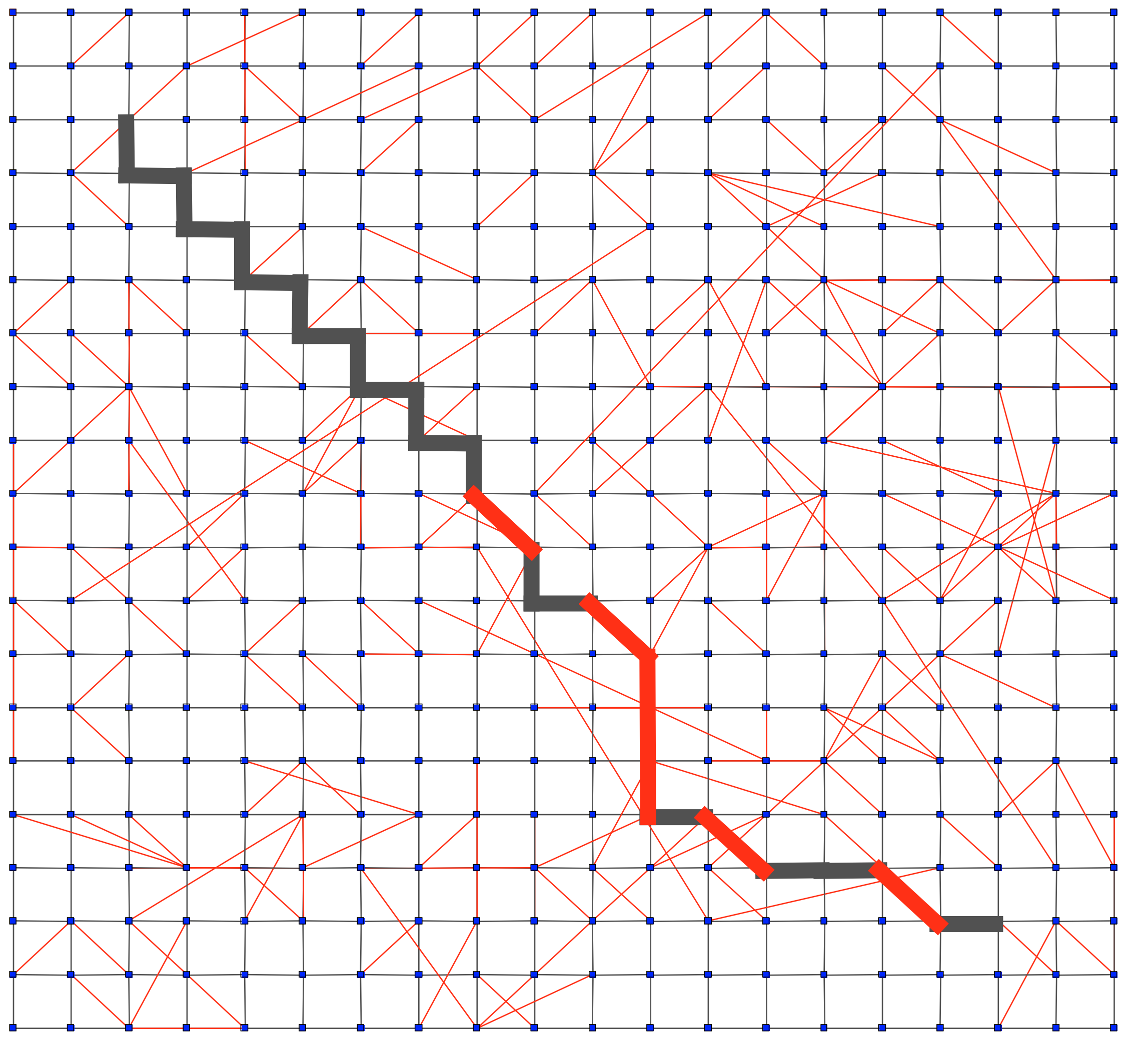


When **r<2**, expected time at least $\alpha_r n^{(2-r)/3}$

Overly localized links on a lattice

When **r>2** expected search time ~ $\mathbf{N^{(r-2)/(r-1)}}$

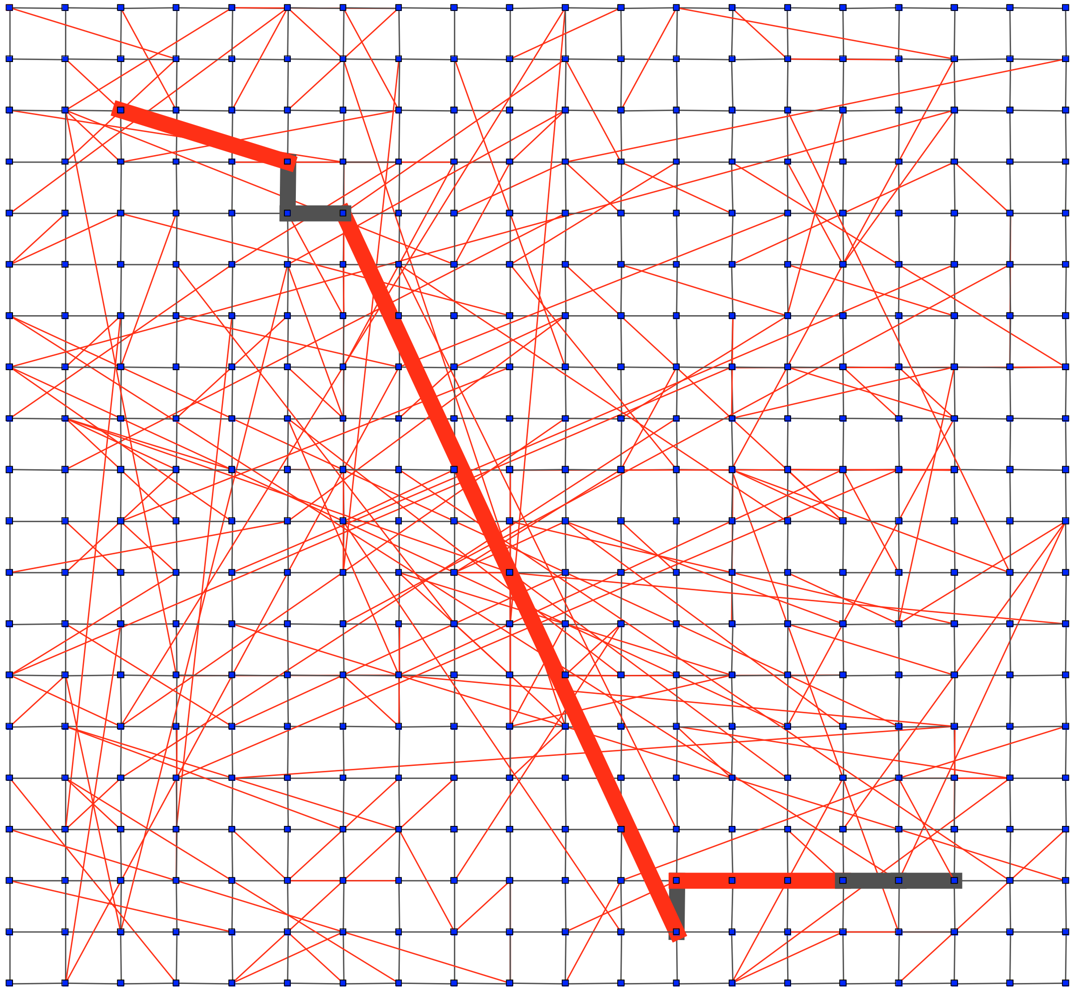
$$p \sim \frac{1}{d^4}$$



Links balanced between long and short range

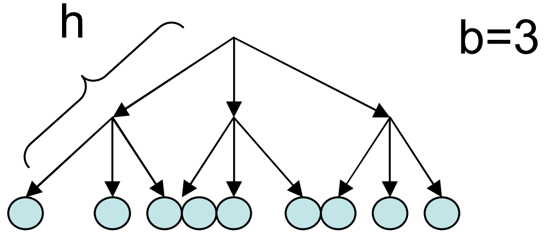
When **r=2,** expected time of a DA is at most **C (log N)²**

$$p \sim \frac{1}{d^2}$$



Kleinberg, ‘Small-World Phenomena and the Dynamics of Information’
NIPS 14, 2001

Hierarchical network models:



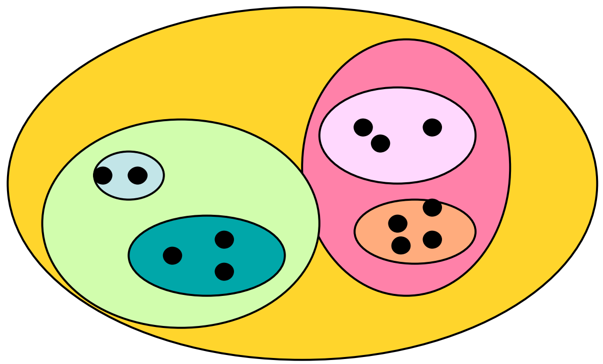
Individuals classified into a hierarchy,
h_{ij} = height of the least common ancestor.

$$p_{ij} \quad b^{-\alpha h_{ij}}$$

e.g. state-county-city-neighborhood
industry-corporation-division-group

Theorem: If α = 1 and outdegree is polylogarithmic, can
s ~ O(log n)

Group structure models:



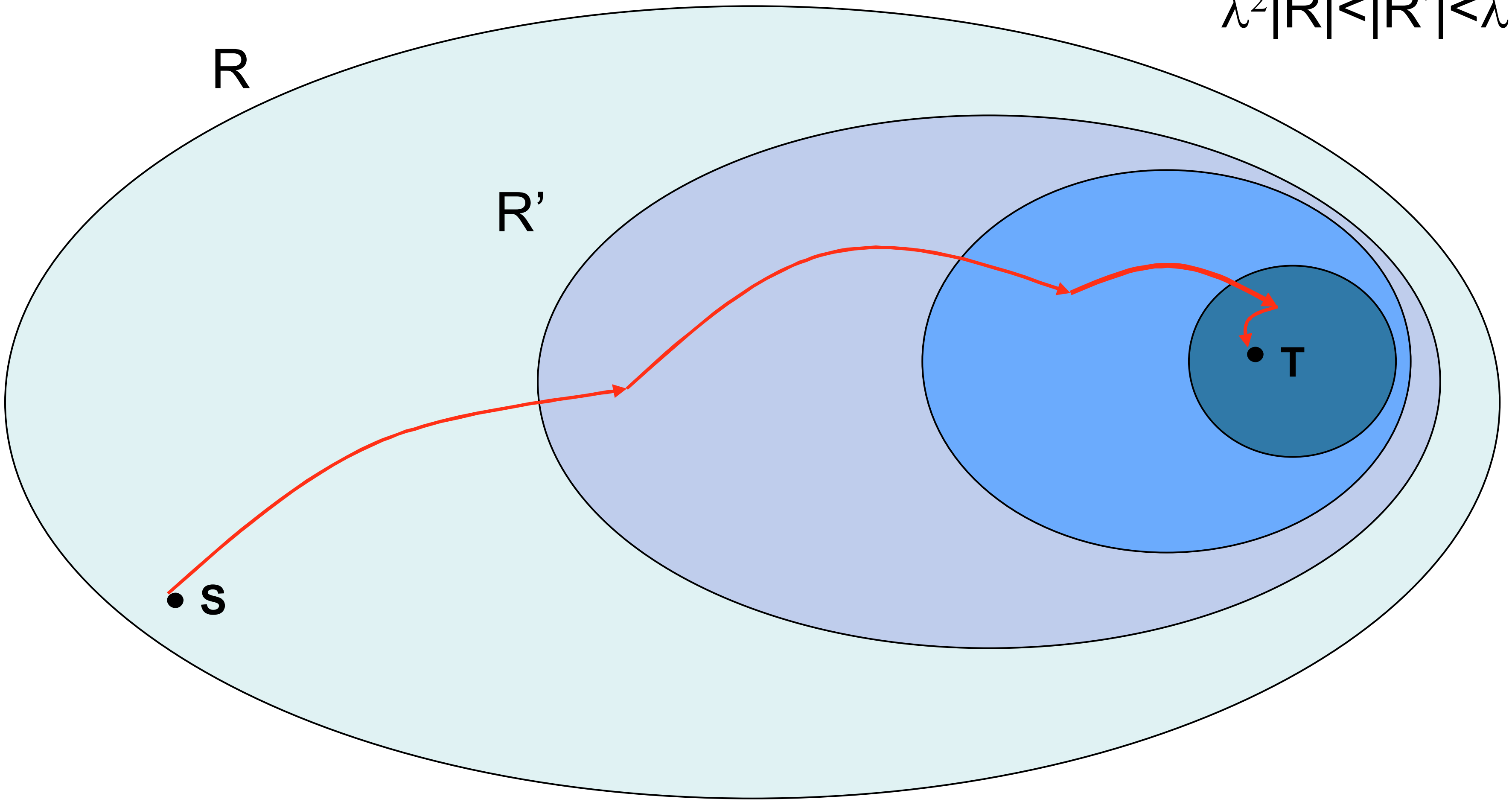
Individuals belong to nested groups
q = size of smallest group that v,w belong to

$$f(q) \sim q^{-\alpha}$$

Theorem: If α = 1 and outdegree is polylogarithmic, can
s ~ O(log n)

Sketch of proof

$\lambda^2|R|<|R'|<\lambda|R|$



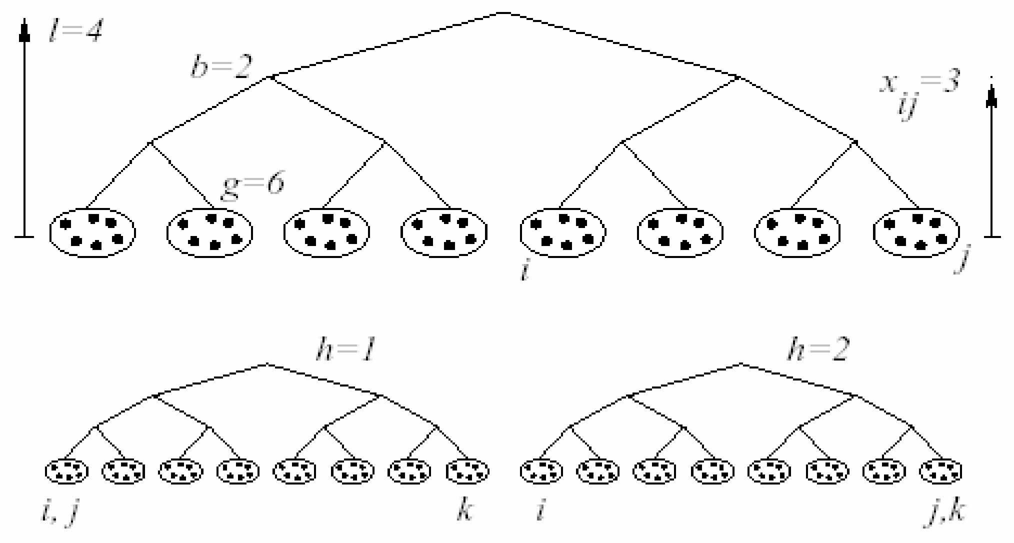
$k = c \log^2 n$

calculate probability that s fails to have a link in R'

Identity and search in social networks

Watts, Dodds, Newman (Science,2001)

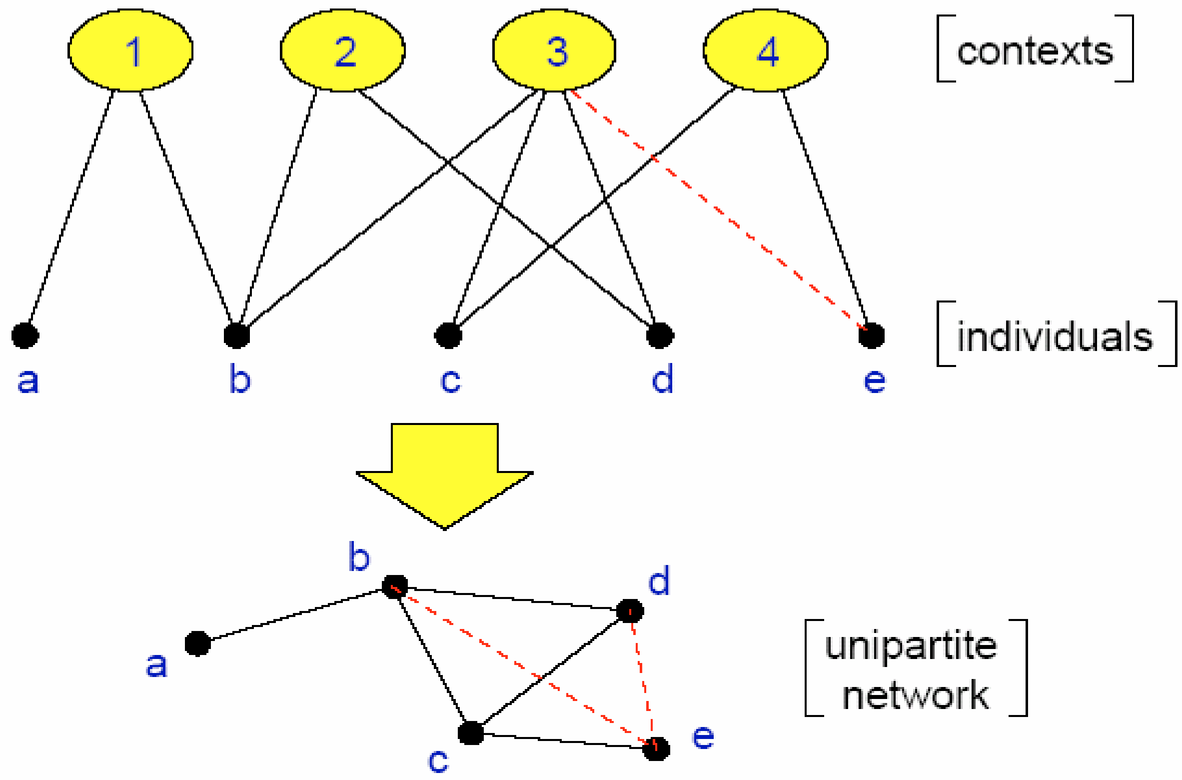
individuals belong to hierarchically nested groups



$p_{ij} \sim \exp(-\alpha\ x)$

multiple independent hierarchies h=1,2,..,H coexist corresponding to occupation, geography, hobbies, religion…

Social distance—Bipartite networks:

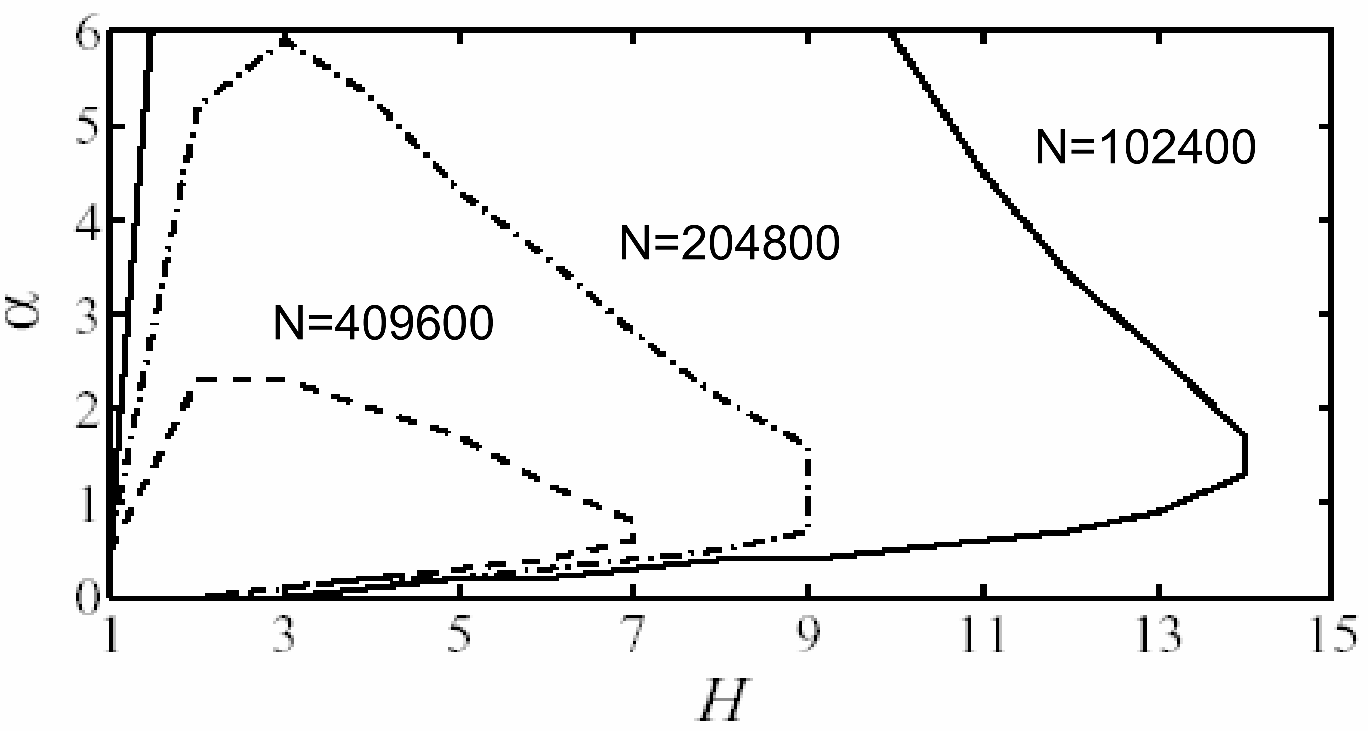


Identity and search in social networks

Watts, Dodds, Newman (2001)

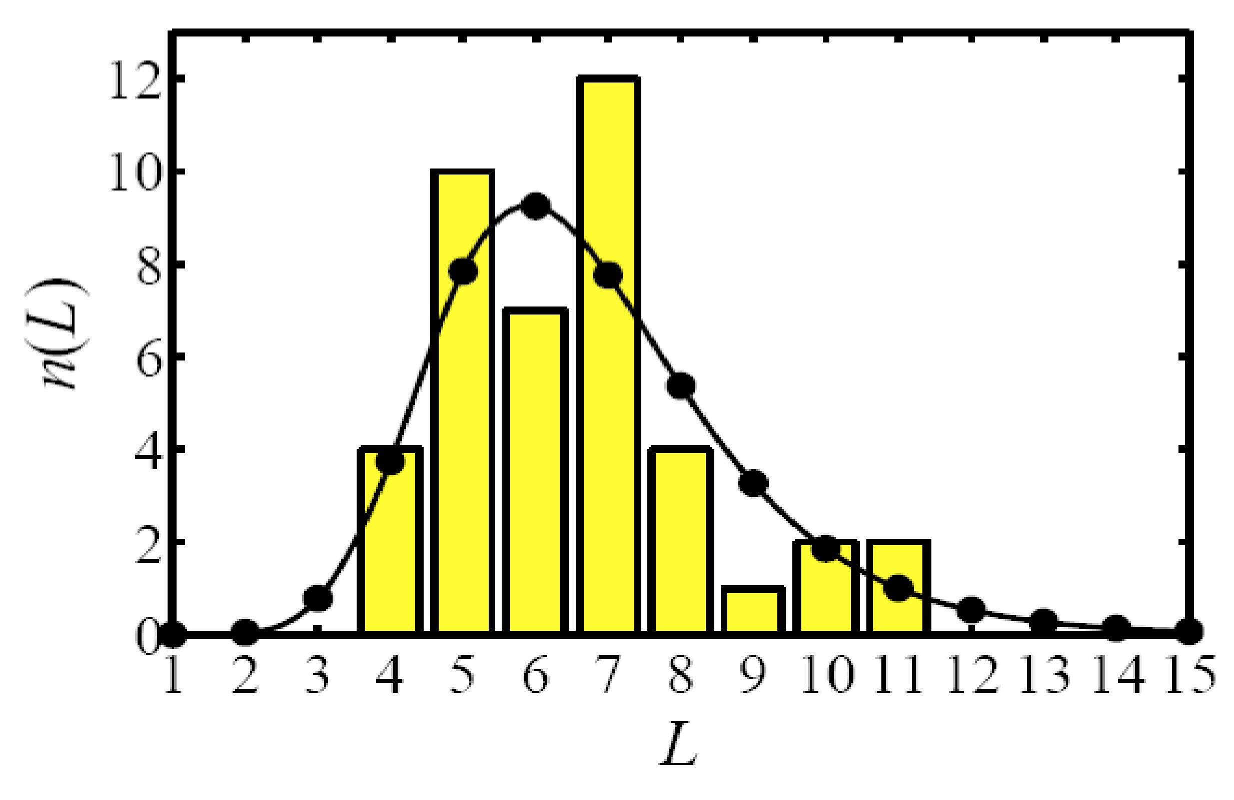
Message chains fail at each node with probability p
Network is 'searchable' if a fraction r of messages reach the target

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



Small World Model, Watts et al.

Fits Milgram's data well



Model parameters:
N = 10^8
z = 300
g = 100
b = 10
α= 1, H = 2

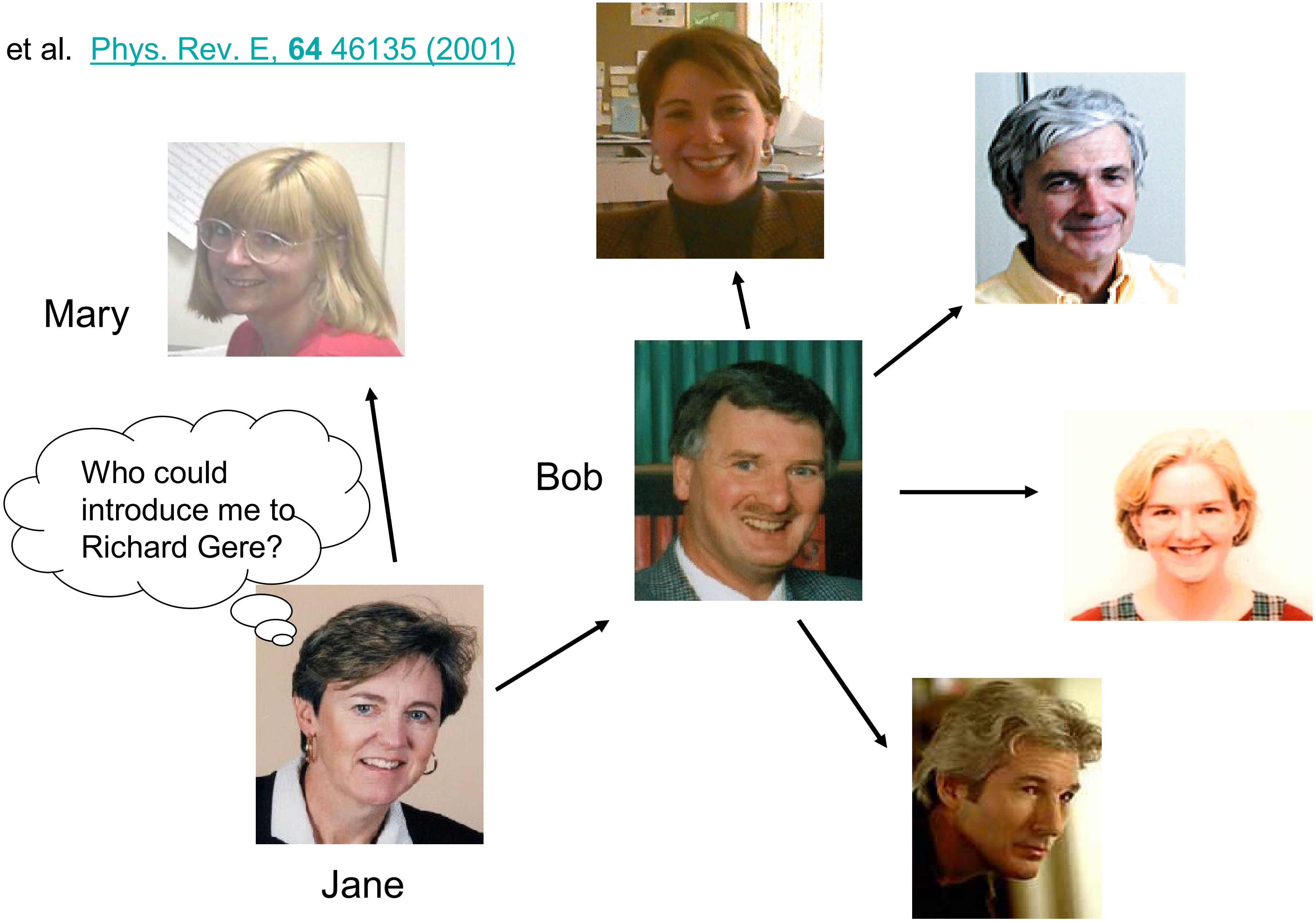
L_{model}= 6.7
L_{data} = 6.5

more slides on this:

http://www.aladdin.cs.cmu.edu/workshops/wsa/papers/dodds-2004-04-10search.pdf

High degree search

Adamic et al. Phys. Rev. E, **64** 46135 (2001)



Small world experiments so far

Classic small world experiment:

Given a target individual, forward to one of your acquaintances

Observe chains but not the rest of the social network

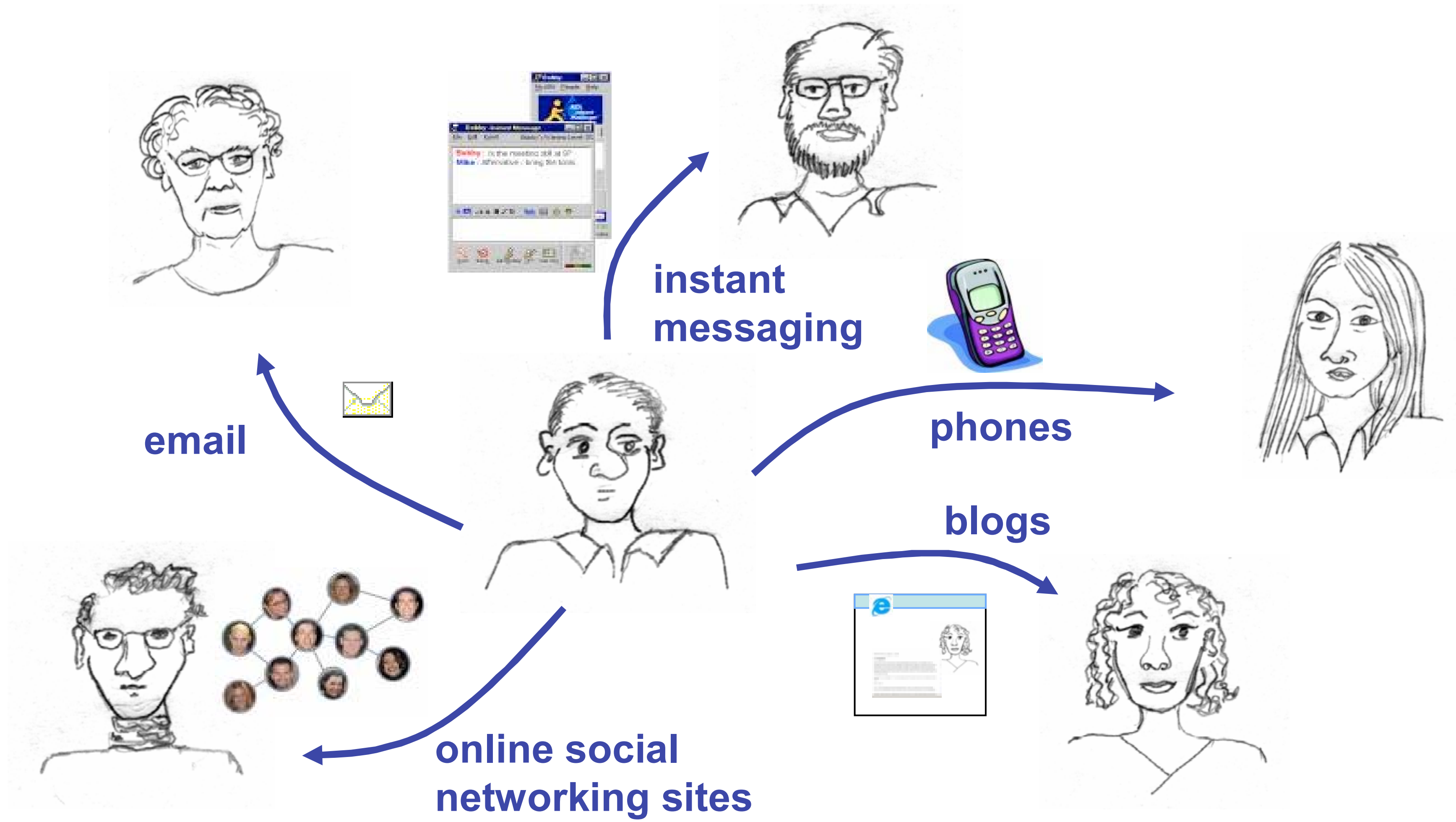
Reverse small world experiment (Killworth & Bernard)

Given a hypothetical individual,
which of your acquaintances would you choose

Observe individual's social network and possible choices,
but not resulting chains or complete social network

New data that's available

More and more social network information is available as a side-effect of people leading digital lives



Testing search models on social networks

advantage: have access to entire communication network
and to individual's attributes

Use a well defined network:
HP Labs email correspondence over 3.5 months

Edges are between individuals who sent
at least 6 email messages each way

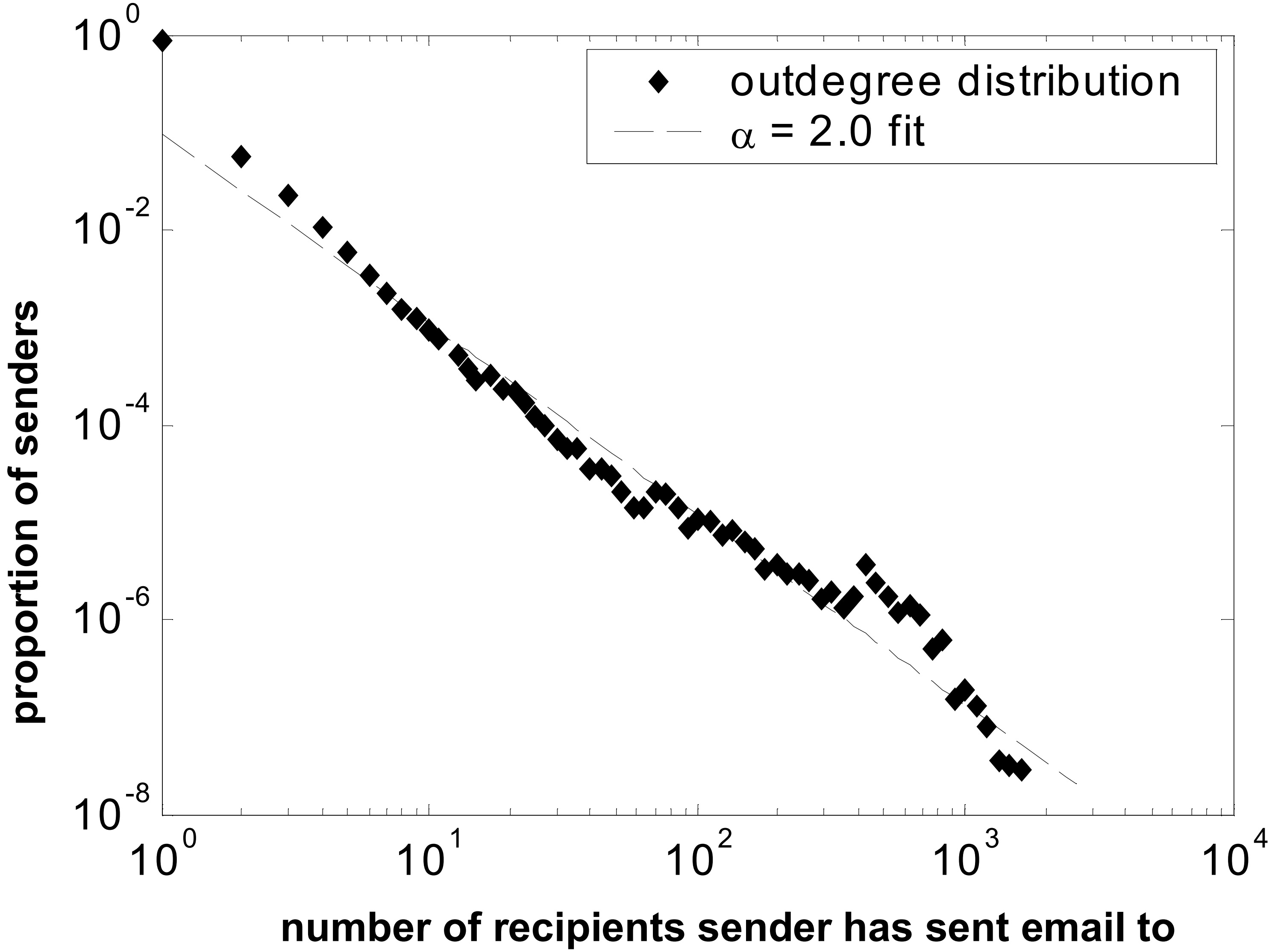
450 users
median degree = 10, mean degree = 13
average shortest path = 3

Node properties specified:
degree
geographical location
position in organizational hierarchy

Can greedy strategies work?

Strategy 1: High degree search

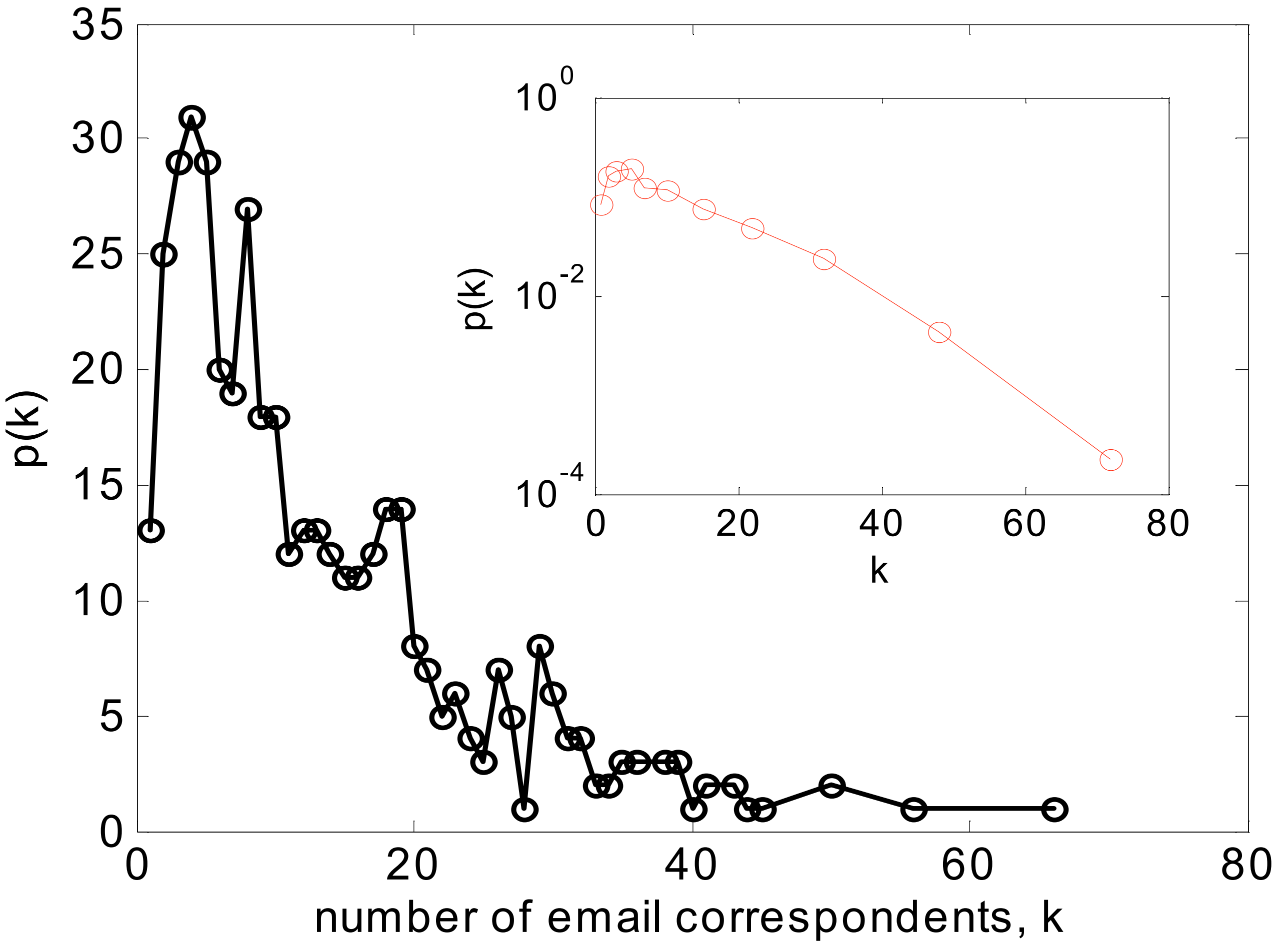
Power-law degree distribution of all senders of email passing through HP labs



Filtered network

(at least 6 messages sent each way)

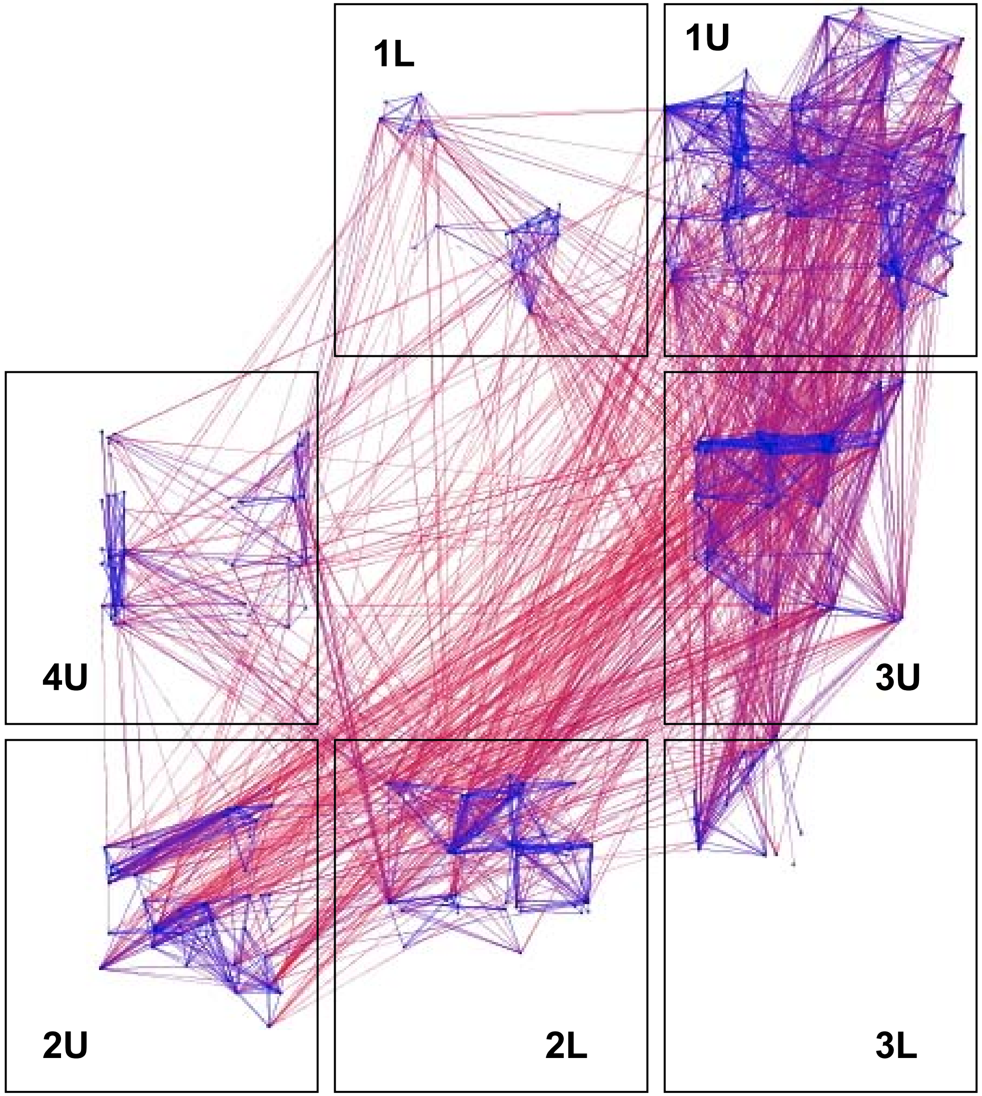
Degree distribution no longer power-law, but Poisson



It would take 40 steps on average (median of 16) to reach a target!

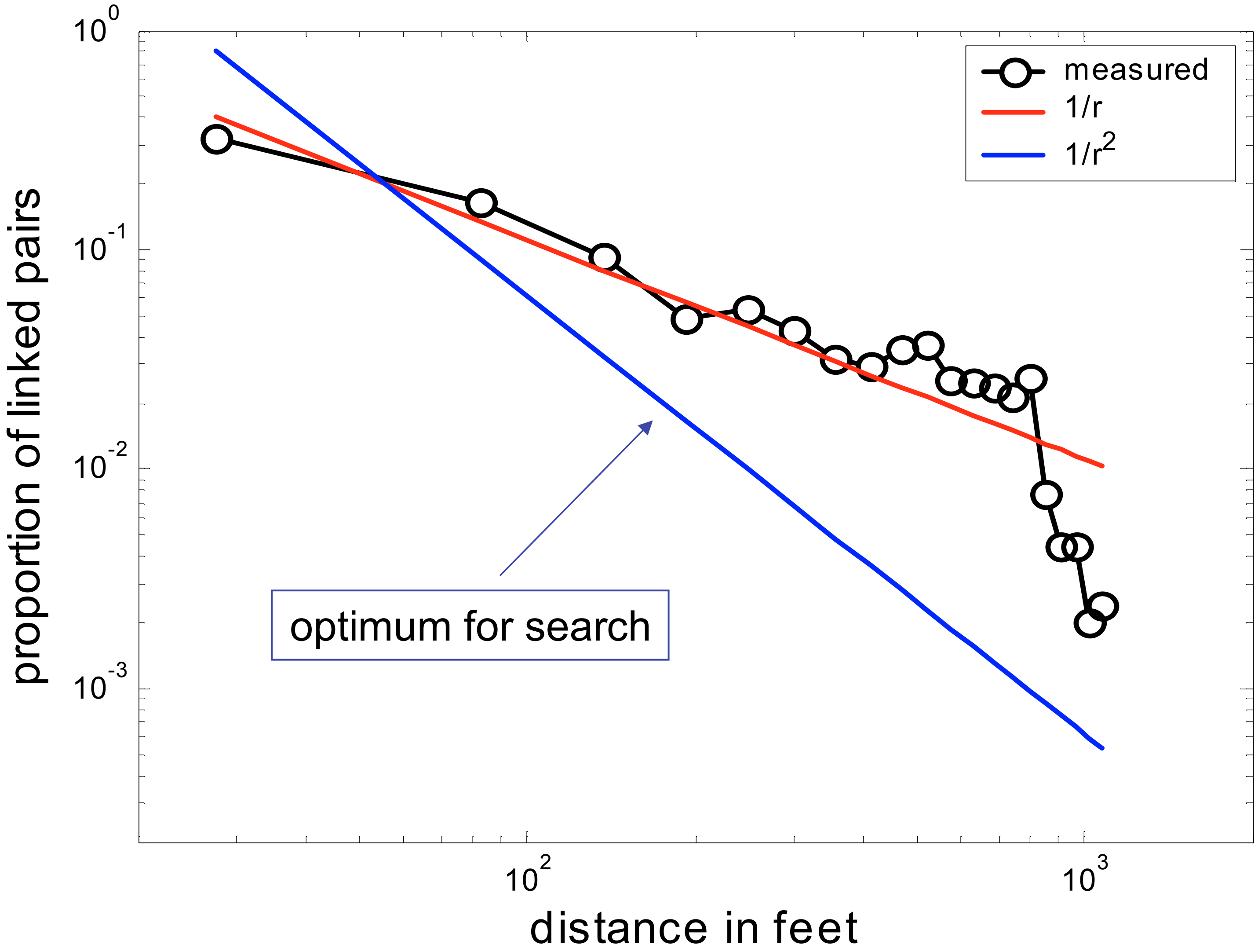
Strategy 2:
Geography

Communication across corporate geography

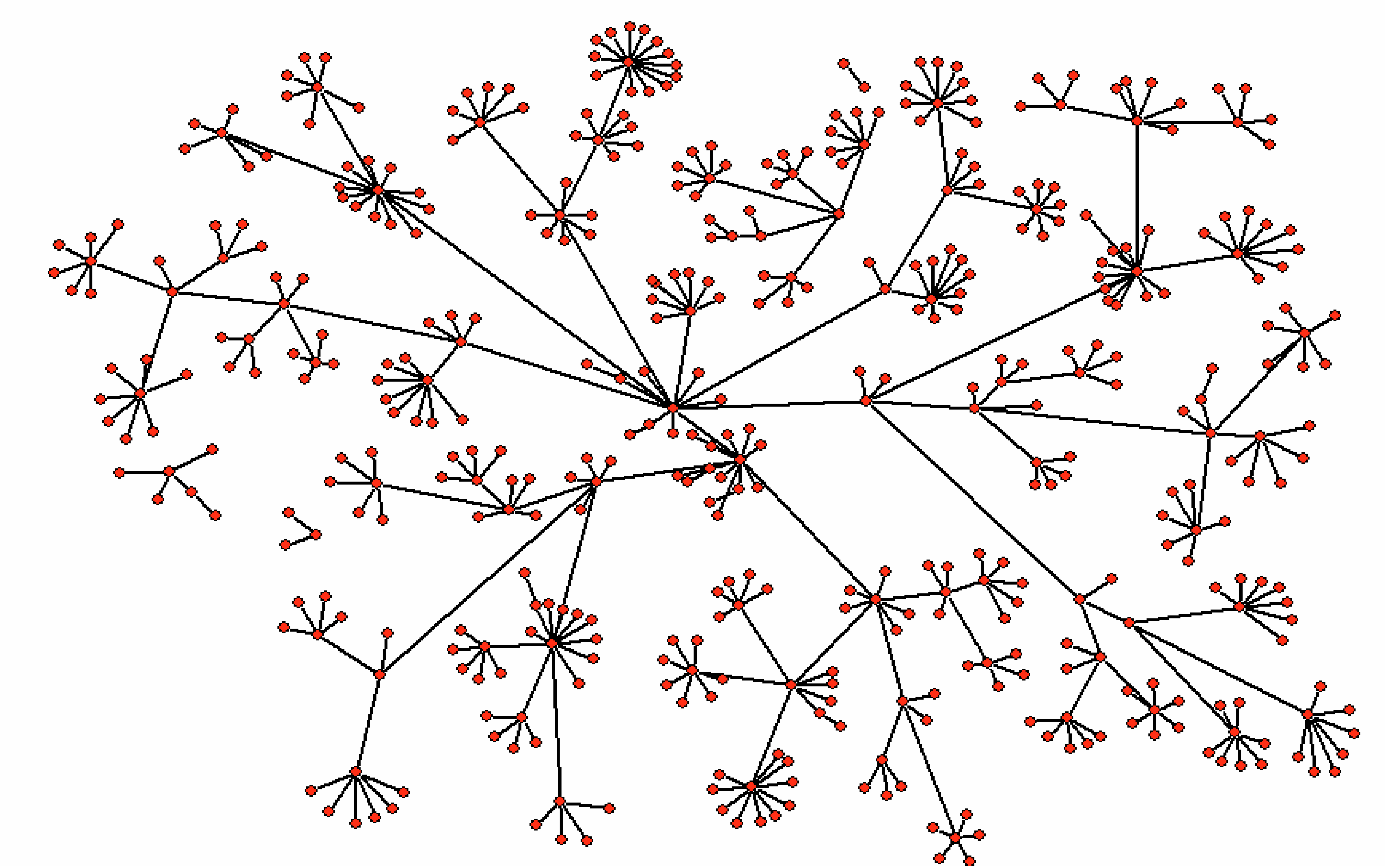


87 % of the 4000 links are between individuals on the same floor

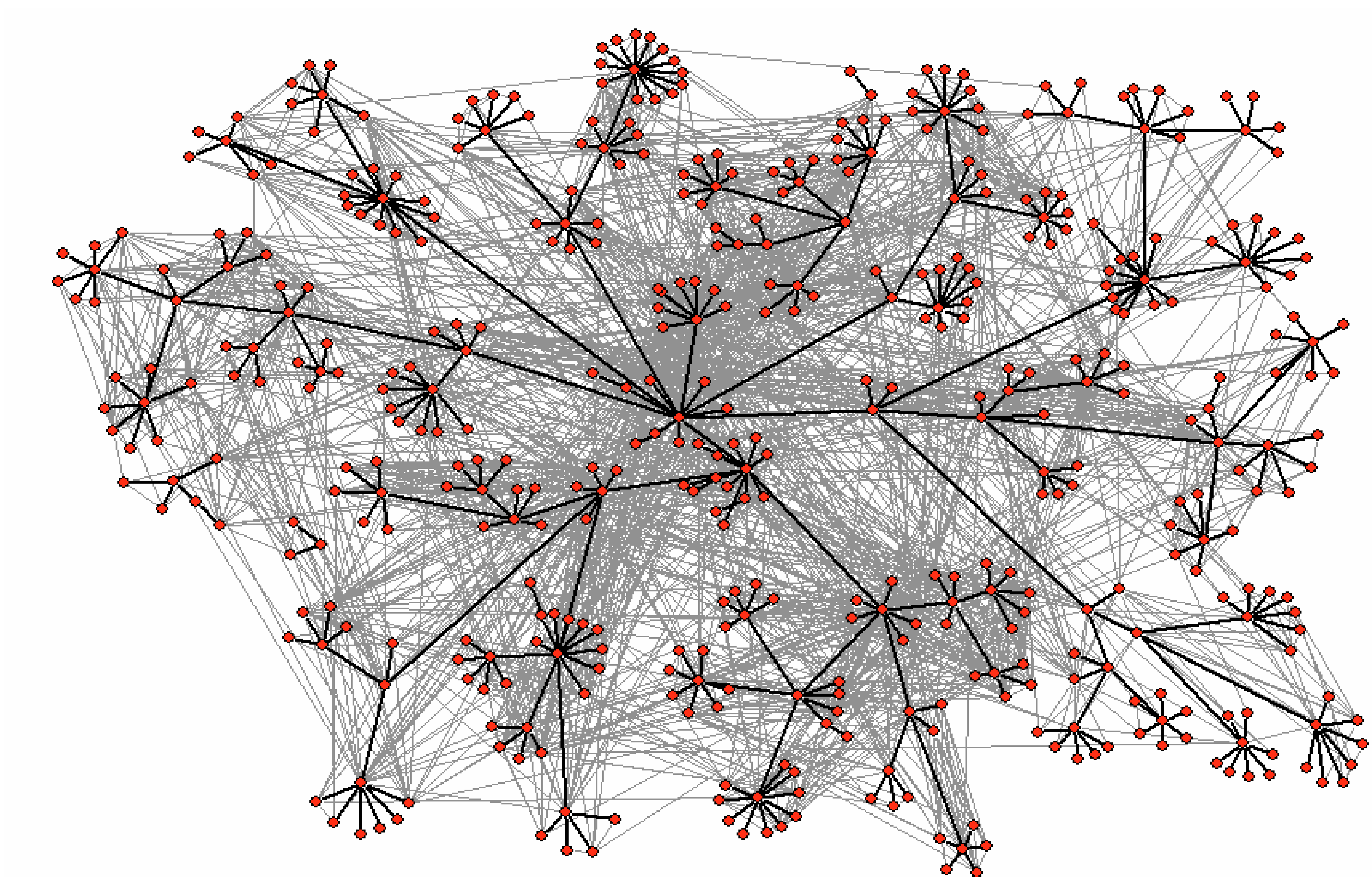
Cubicle distance vs. probability of being linked



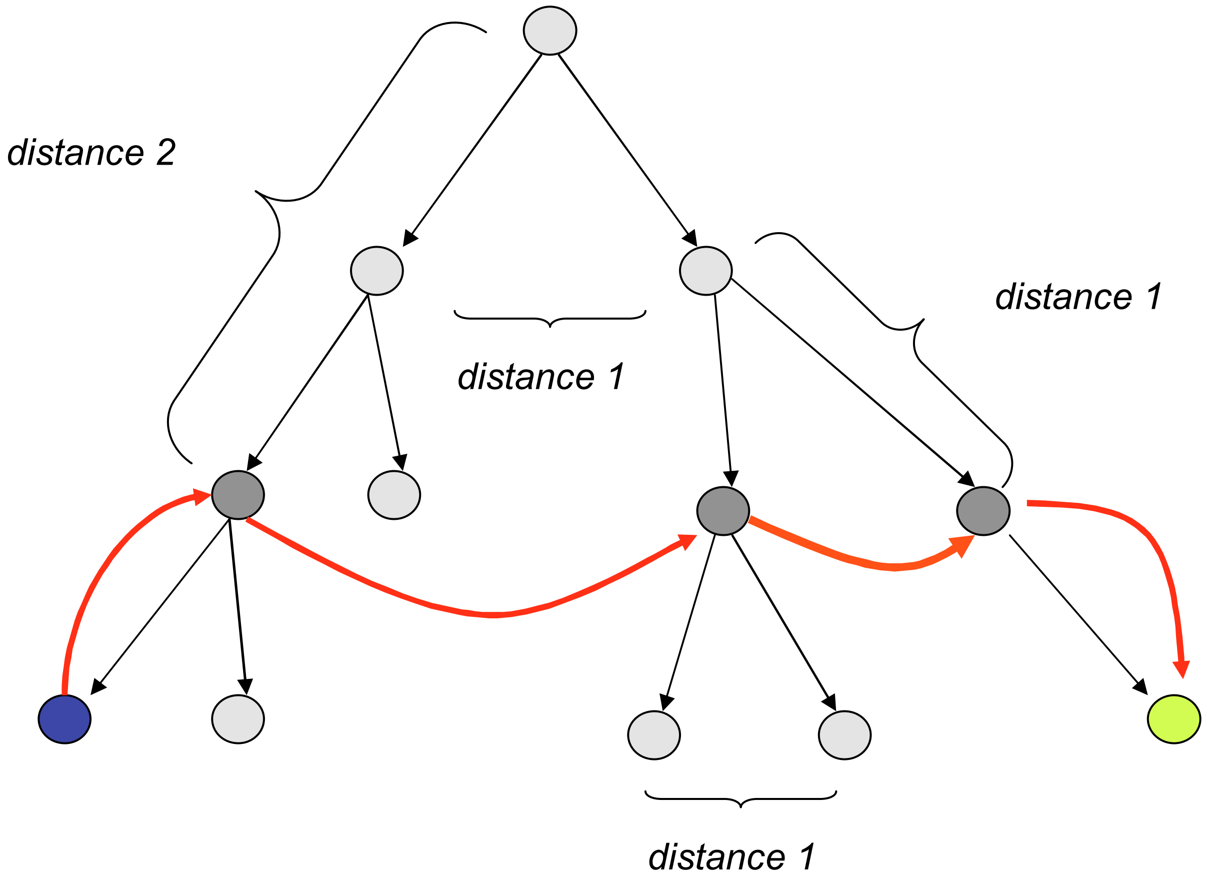
Strategy 3: Organizational hierarchy



Email correspondence superimposed on the organizational hierarchy

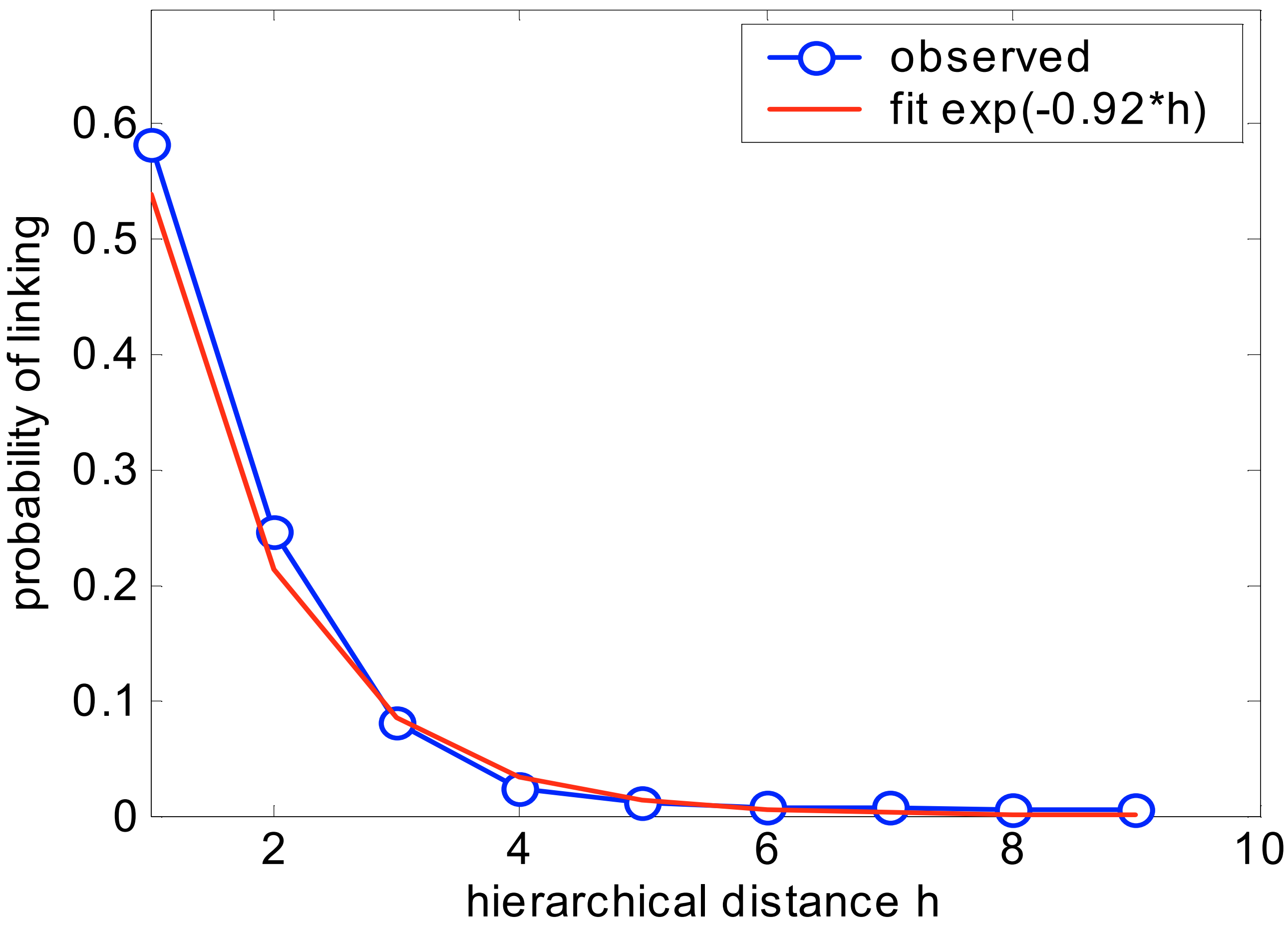


Example of search path



hierarchical distance = 5
search path distance = 4

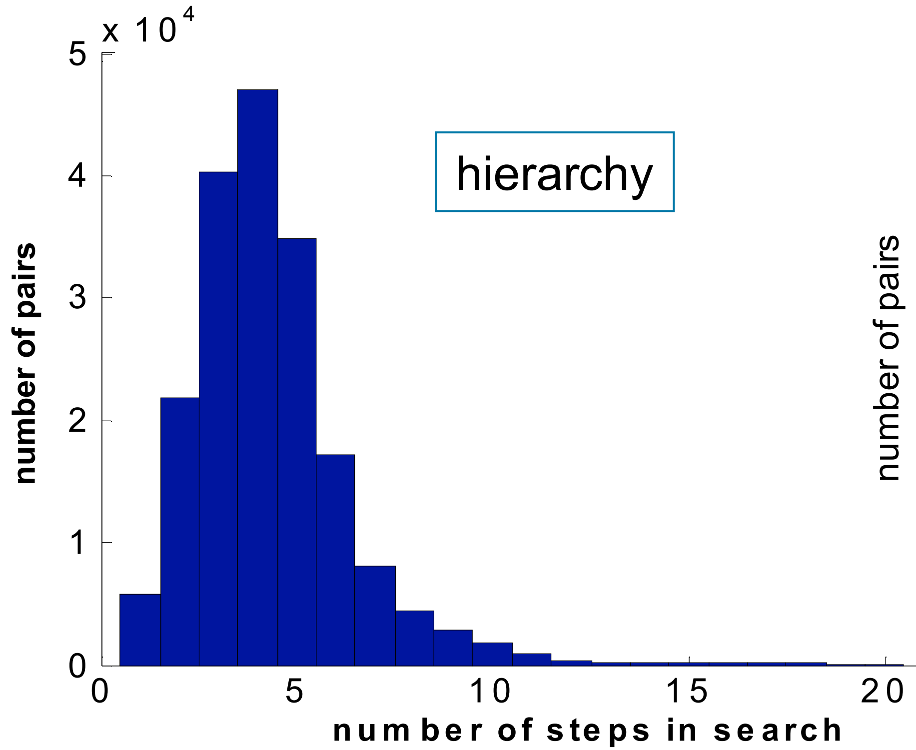
Probability of linking vs. distance in hierarchy

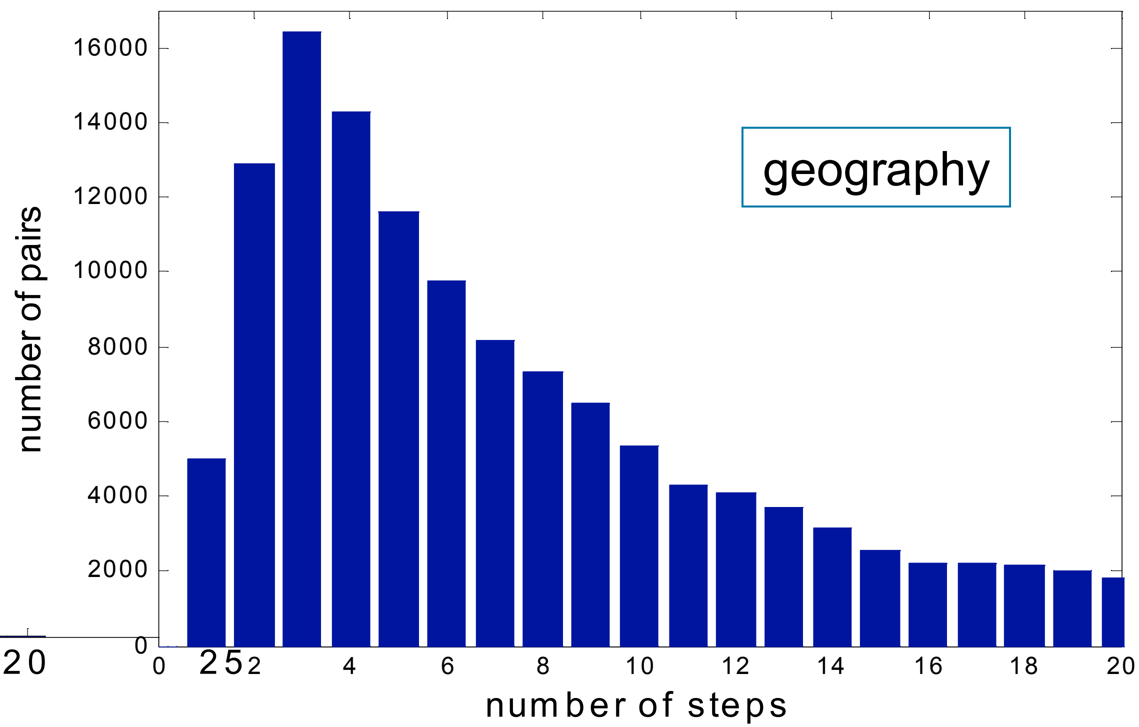


in the 'searchable' regime: $0 < \alpha < 2$ (Watts, Dodds, Newman 2001)

Results

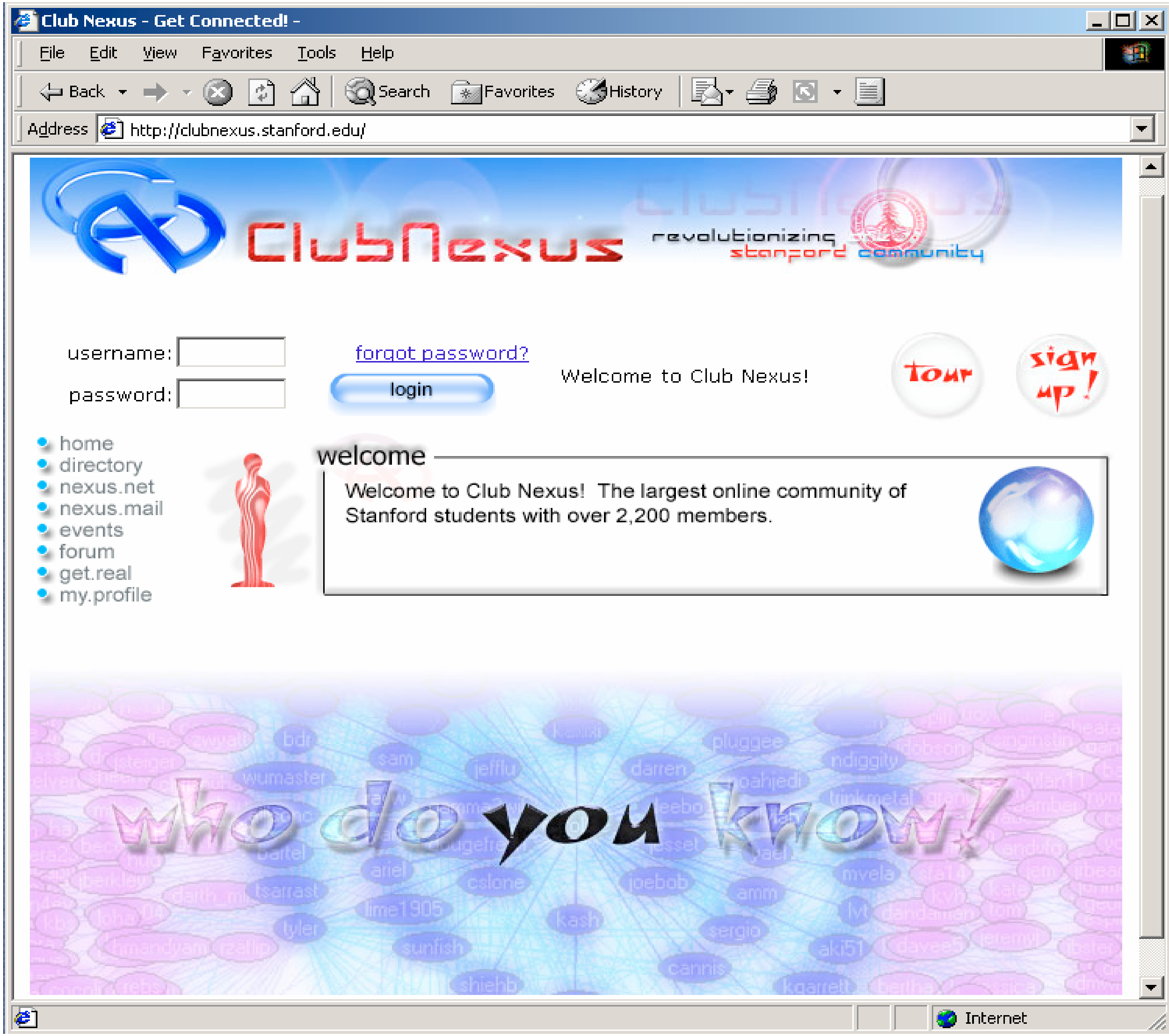
distance	hierarchy	geography	geodesic	org	random
median	4	7	3	6	28
mean	5.7 (4.7)	12	3.1	6.1	57.4

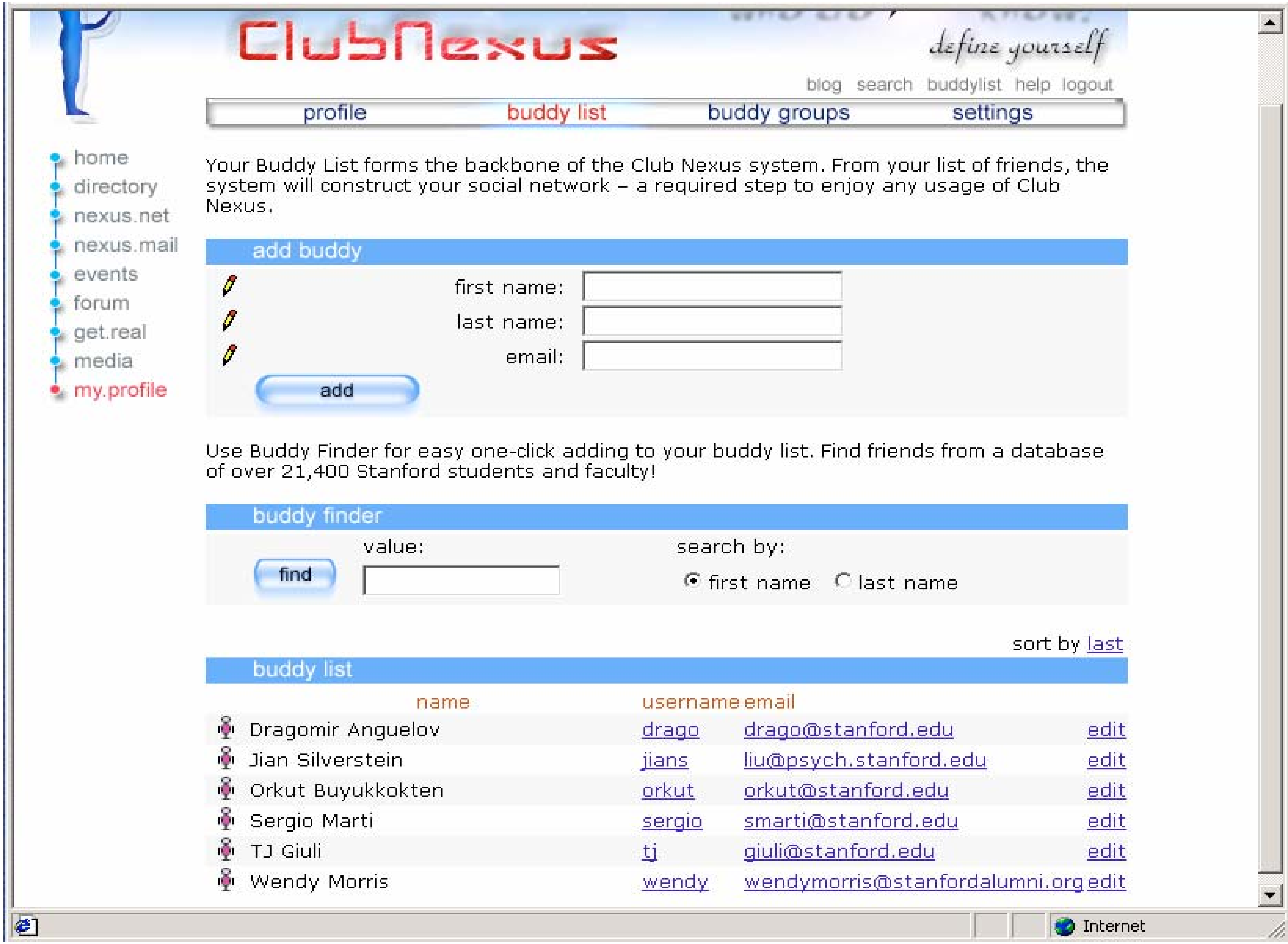


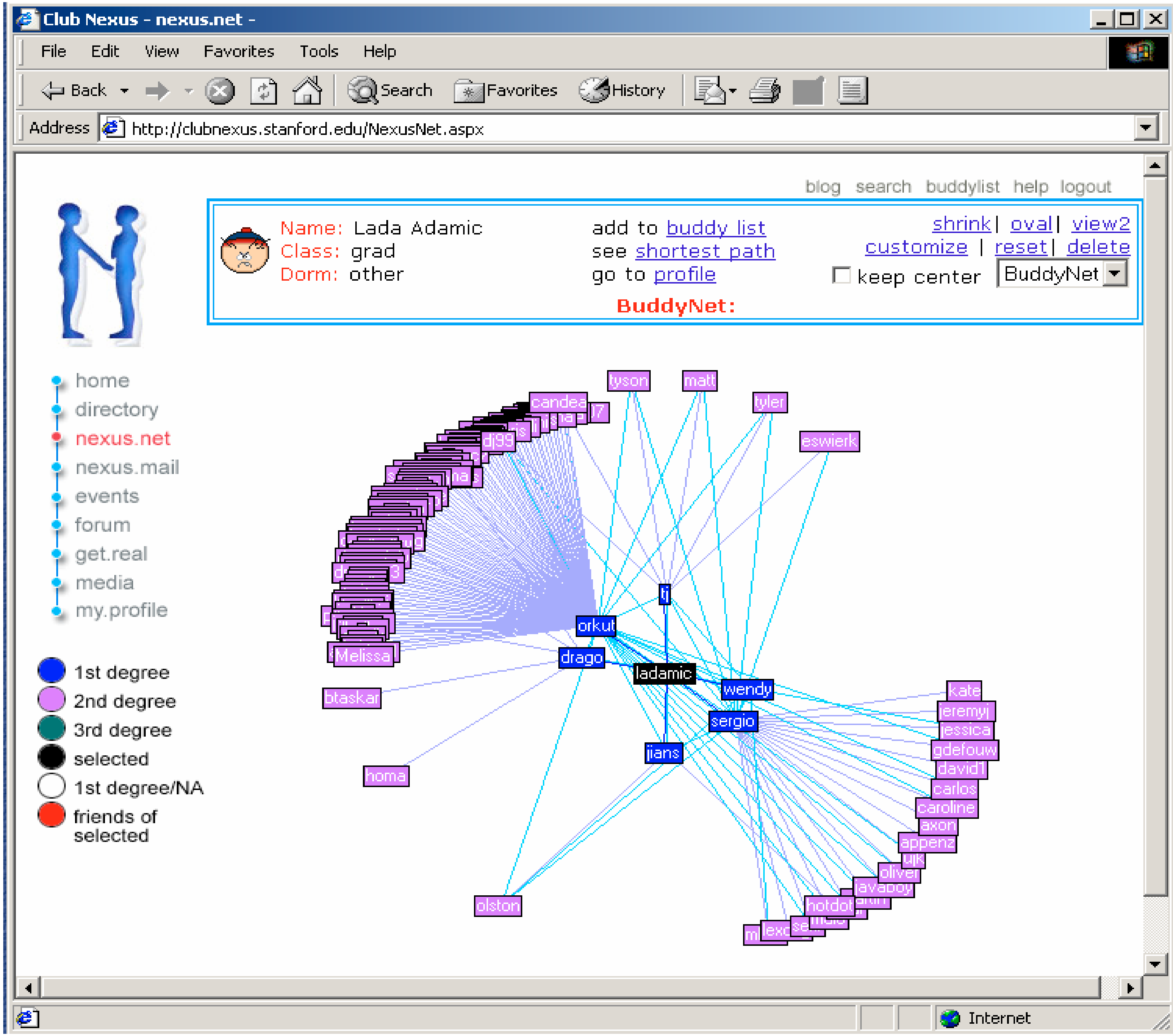


Expt 2

Searching a social networking website




ClubNexus
define yourself
blog search buddylist help logout
profile
buddy list
buddy groups
settings
home
directory
nexus.net
nexus.mail
events
forum
get.real
media
my.profile
Your Buddy List forms the backbone of the Club Nexus system. From your list of friends, the system will construct your social network – a required step to enjoy any usage of Club Nexus.
add buddy
first name:
last name:
email:
add
Use Buddy Finder for easy one-click adding to your buddy list. Find friends from a database of over 21,400 Stanford students and faculty!
buddy finder
value:
search by:
find
first name
last name
sort by last
buddy list
name
username
email
Dragomir Anguelov drago drago@stanford.edu edit
Jian Silverstein jians liu@psych.stanford.edu edit
Orkut Buyukkokten orkut orkut@stanford.edu edit
Sergio Marti sergio smarti@stanford.edu edit
TJ Giuli tj giuli@stanford.edu edit
Wendy Morris wendy wendymorris@stanfordalumni.org edit
Internet


Club Nexus - nexus.net -
File Edit View Favorites Tools Help
Back
Search
Favorites
History
Address
http://clubnexus.stanford.edu/NexusNet.aspx
blog search buddylist help logout
Name: Lada Adamic
Class: grad
Dorm: other
add to buddy list
see shortest path
go to profile
shrink | oval | view2
customize | reset | delete
keep center
BuddyNet
BuddyNet:
home
directory
nexus.net
nexus.mail
events
forum
get.real
media
my.profile
1st degree
2nd degree
3rd degree
selected
1st degree/NA
friends of selected
tyson
matt
tyler
eswierk
orkut
drago
ladamic
wendy
sergio
jians
Melissa
btaskar
homa
olston
kate
jeremyj
jessica
gdefouw
david1
carlos
caroline
axon
appenz
oliver
hotdot
Internet

Profiles:

status (UG or G)
year
major or **department**
residence
gender

Personality	(choose 3 exactly):
you	funny, kind, weird, …
friendship	honesty/trust, common interests, commitment, …
romance	- “ -
freetime	socializing, getting outside, reading, …
support	unconditional accepters, comic-relief givers, eternal optimists

Interests	(choose as many as apply)
books	mystery & thriller, science fiction, romance, …
movies	western, biography, horror, …
music	folk, jazz, techno, …
social activities	ballroom dancing, barbecuing, bar-hopping, …
land sports	soccer, tennis, golf, …
water sports	sailing, kayaking, swimming, …
other sports	ski diving, weightlifting, billiards, …

Finding correlations between user attributes

Are people who consider themselves funny also more likely to enjoy comedies?

518 funny users
74 % of users overall like comedies
416 (80% of) funny users like comedies,

this is 3.4 standard deviations (=10) above expected (383)

Z score = 3.4

Z scores with absolute value > 2 are significant at the p = 0.05 level
3.4 is significant at the 0.0003 level

→ small differences (10%) can be significant.

Personality and tastes (just a few examples)

creative	book	art & photography, philosophy, fiction & literature, classics
	music	folk, bluegrass/rural, jazz
	movie	art, documentary, independent

successful	book	business
	landsport	tennis
	other	weightlifting
	social	barbecuing
	watersport	boating, jet skiing, water skiing
	free time	fulfilling commitments, catching up on chores and things

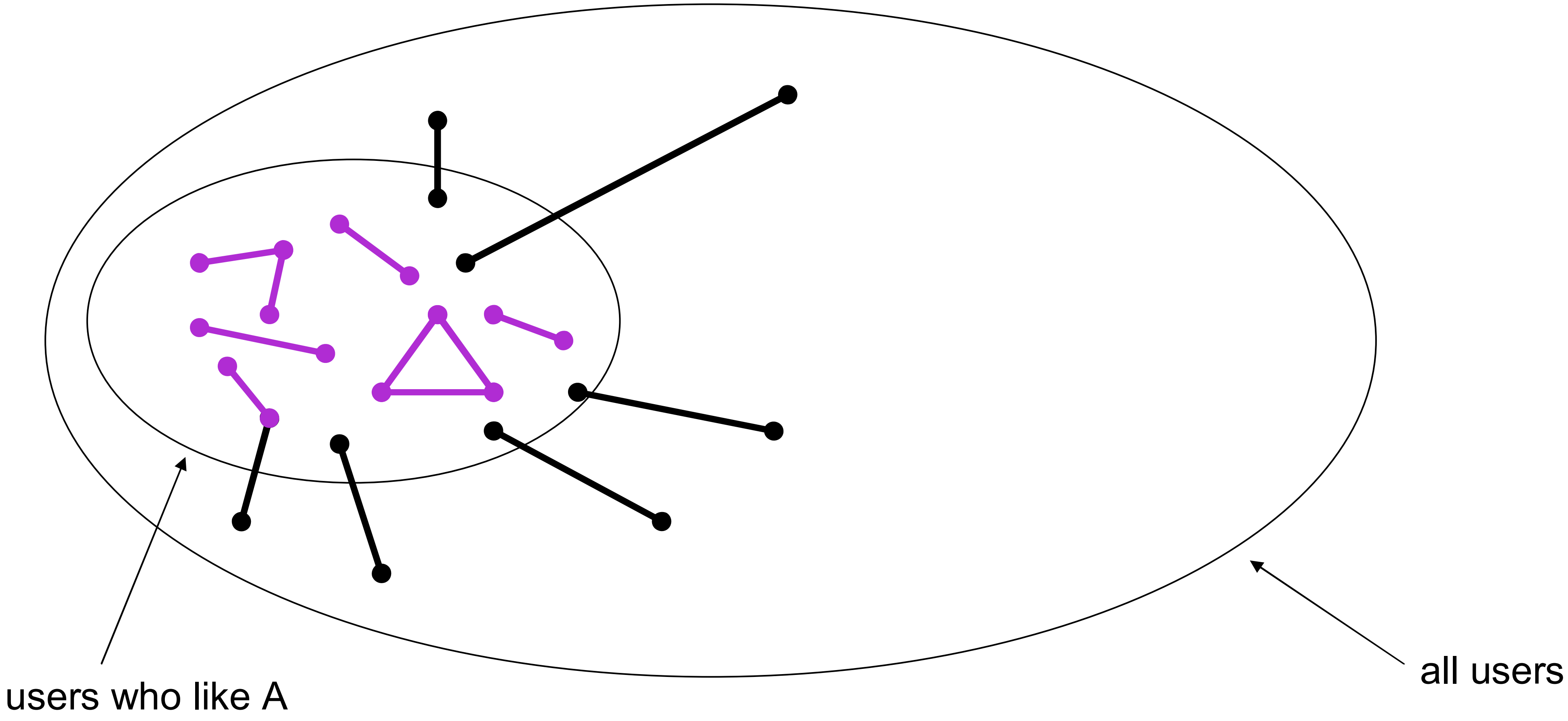
not responsible	book	sex
	movie	erotic & softcore, gay & lesbian, independent
	music	funk, jungle, reggae, trance
	other	skateboarding
	social	raving

Major and personality

personality (% of total)	major
free time: learning (17%)	Physics (46%), Philosophy (37%), Math (31%), EE (26%), CS (24%)
free time: reading (26%)	English (55%)
free time: staying at home (8%)	History (24%)
free time: doing anything exciting (52%)	undecided/undeclared (62%)
you: weird (12%)	Physics (34%), Math (28%), EE (18%)
you: intelligent (32%)	Philosophy (59%), CS (42%)
you: successful (4%)	CS (7%)
you: socially adaptable (14%)	STS (46%)
you: attractive (16%)	Political Science (29%), International Relations (25%)
you: lovable (12%)	Political Science (24%)
you: kind (25%)	Public Policy (45%)
you: funny (25%)	Philosophy (6%)
you: fun (26%)	Human Biology (38%)
you: creative (22%)	Product Design (62%), English (42%)
you: sexy (8%)	English (18%), EE (2%)

Association ratios

p = (# users who like A)/(total #users)
L = # connections A users have
m = expected number of links to other A users = L*p
r = (# links between A users)/m



Interests and association ratios

	high association	low association
book	gay & lesbian, professional & technical, computers, teen, sex, sports	history, fiction & literature, outdoor & nature
movie genres	gay & lesbian, performing arts, religion, erotic & softcore, sports	drama, mystery, documentary, comedy
music genres	gospel, jungle, bluegrass/rural, heavy metal, trance	pop, classical, rock
land sport	lacrosse, field hockey, wrestling, cricket	tennis, martial arts, bicycling, racquetball
water sport	synchronized swimming, diving, crew	swimming, fishing windsurfing
social	raving, ballroom dancing, Latin dancing	partying, camping

Nexus Karma

Rank how ‘trusty’, ‘nice’, ‘cool’, and ‘sexy’ your buddies
are on a scale of 1 to 4

446 users ranked 1735 different friends

correlations between scores given (users were ranked as
‘3,3,3,3’ more often than ‘1,4,2,3’

average scores: nice (3.37), trusty (3.22), cool (3.13), sexy(2.83)

trusty--nice and cool--sexy more highly correlated (ρ = 0.7) vs.
trusty--sexy and nice--sexy (ρ = 0.4)

no relationship	between average score received and # of friends
negative correlation	between average score *given* and # of friends

How users view themselves vs. how others view them

	trusty (3.22)	nice (3.37)	cool (3.13)	sexy (2.83)
responsible	↑ 3.36		↓ 3.02	↓ 2.67
sexy	↓ 3.10	↓ 3.23		↑ 3.03
attractive	↓ 3.09	↓ 3.25		↑ 2.93
kind	↑ 3.34	↑ 3.46		
friendly		↑ 3.44		
weird				↓ 2.67
funny		↓ 3.31		

Additional insights from Nexus Karma

Users *receiving* higher ‘nice’ scores *give* higher ‘trusty’, ‘nice’, and ‘cool’ scores (ρ = 0.14-0.17)

If one user gives another user a higher ‘trusty’ or ‘nice’ score than their other friends, that same friend is more likely to reciprocate.

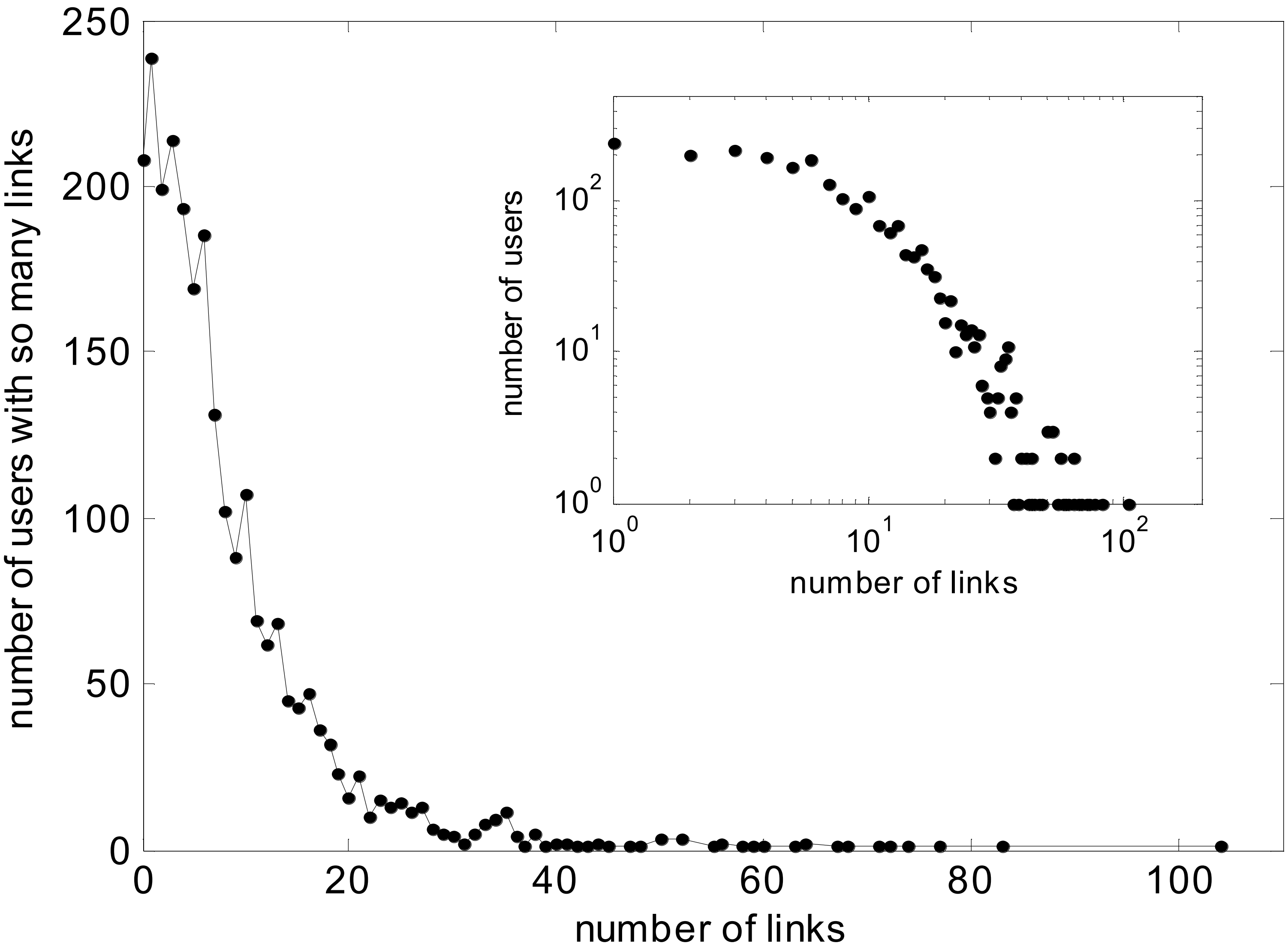
Users who share friends are more likely to give each other high scores (ρ = 0.10-0.13)

Differences between data sets

HP labs email network	Online community
• complete image of communication network	• partial information of social network
• affinity not reflected	• only friends listed

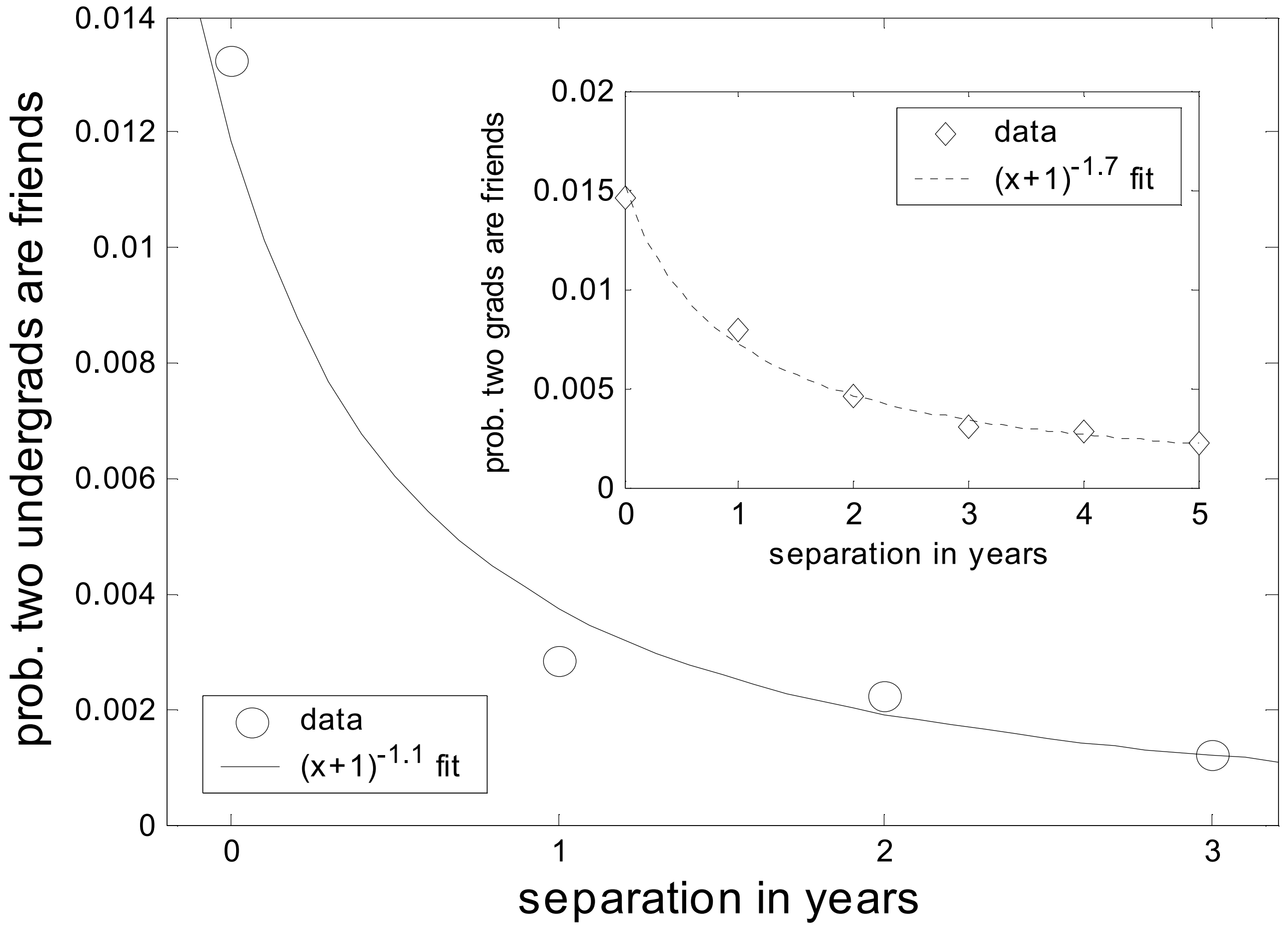
Degree Distribution for Nexus Net

2469 users, average degree 8.2



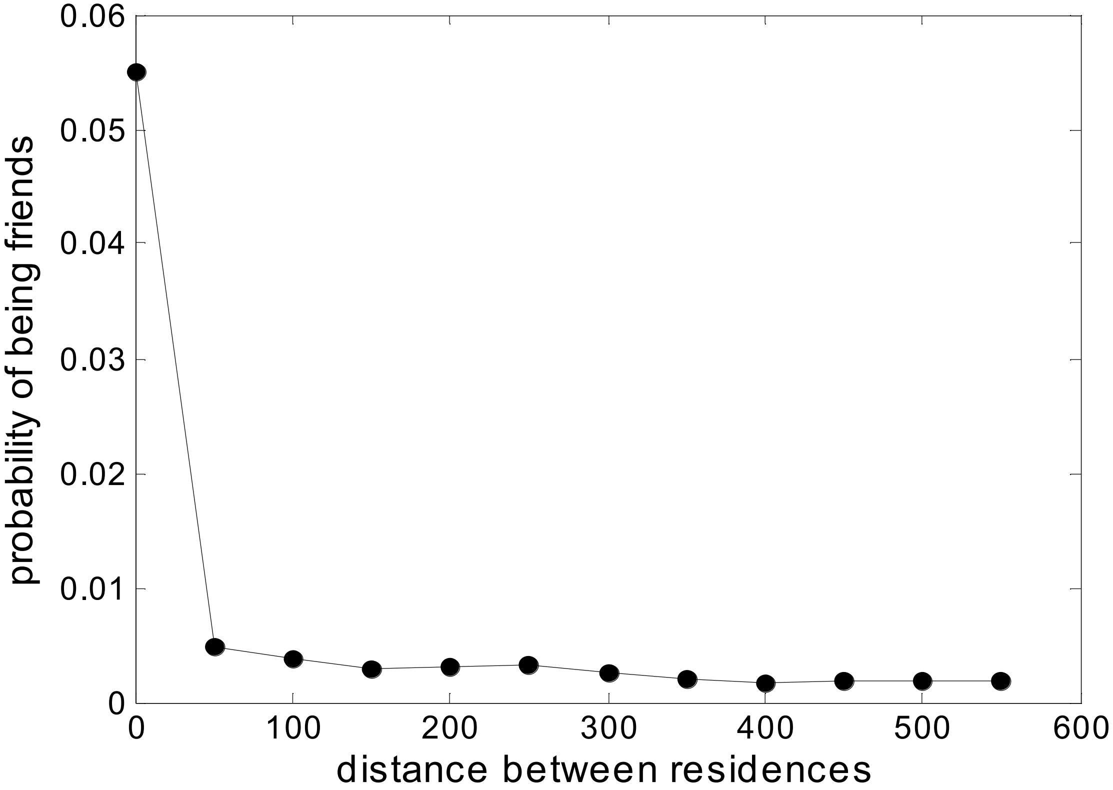
Problem: how to construct hierarchies?

Probability of linking by separation in years



Hierarchies not useful for other attributes:

Geography



Other attributes: major, sports, freetime activities, movie preferences…

Strategy using user profiles

prob. two undergrads are friends (consider simultaneously)

- both undergraduate, both graduate, or one of each
- same or different year
- both male, both female, or one of each
- same or different residences
- same or different major/department

Results

strategy	median	mean
random	133	390
high degree	39	137
profile	21	53

With an attrition rate of 25%, 5% of the messages get through at an average of 4.8 steps,
=> hence network is barely searchable

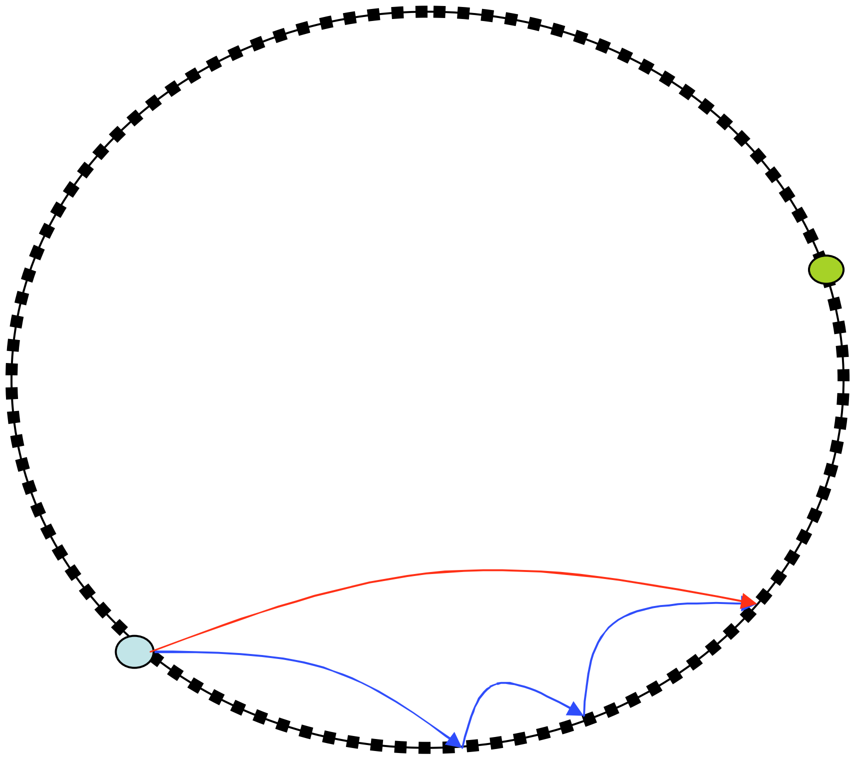
Search Conclusions

- Individuals associate on different levels into groups.
- Group structure facilitates decentralized search using social ties.
- Hierarchy search faster than geographical search
- A fraction of ‘important’ individuals are easily findable
- Humans may be more resourceful in executing search tasks:
  - making use of weak ties
  - using more sophisticated strategies

How do networks become navigable?

Aaron Clauset and Christopher Moore

arxiv.org/abs/cond-mat/0309415



In the limit N-> ∞
long range
link distribution becomes 1/r,
r = lattice distance between
nodes

Applications to peer to peer networks

Adriana Iamnitchi, Matei Ripeanu, Ian Foster
“Small-World File-Sharing Communities”, http://arxiv.org/abs/cs.DC/0307036
create localized indeces for peers with similar download patterns

Foreseer:
Proposed P2P architecture with friend & neighbor overlay
friend: has shared a file
neighbor: short ping time

Fletcher, George , Sheth, Hardik and Börner, Katy. (2004). Unstructured Peer-to-Peer Networks: Topological Properties and Search Performance. Third International Joint Conference on Autonomous Agents and MUlti-Agent Systems. W6: Agents and Peer-to-Peer Computing, Moro, Gianluca, Bergmanschi, Sonia and Aberer, Karl, Eds., New York, July 19-23, pp. 2-13.
http://ella.slis.indiana.edu/~katy/paper/04-fletcher.pdf