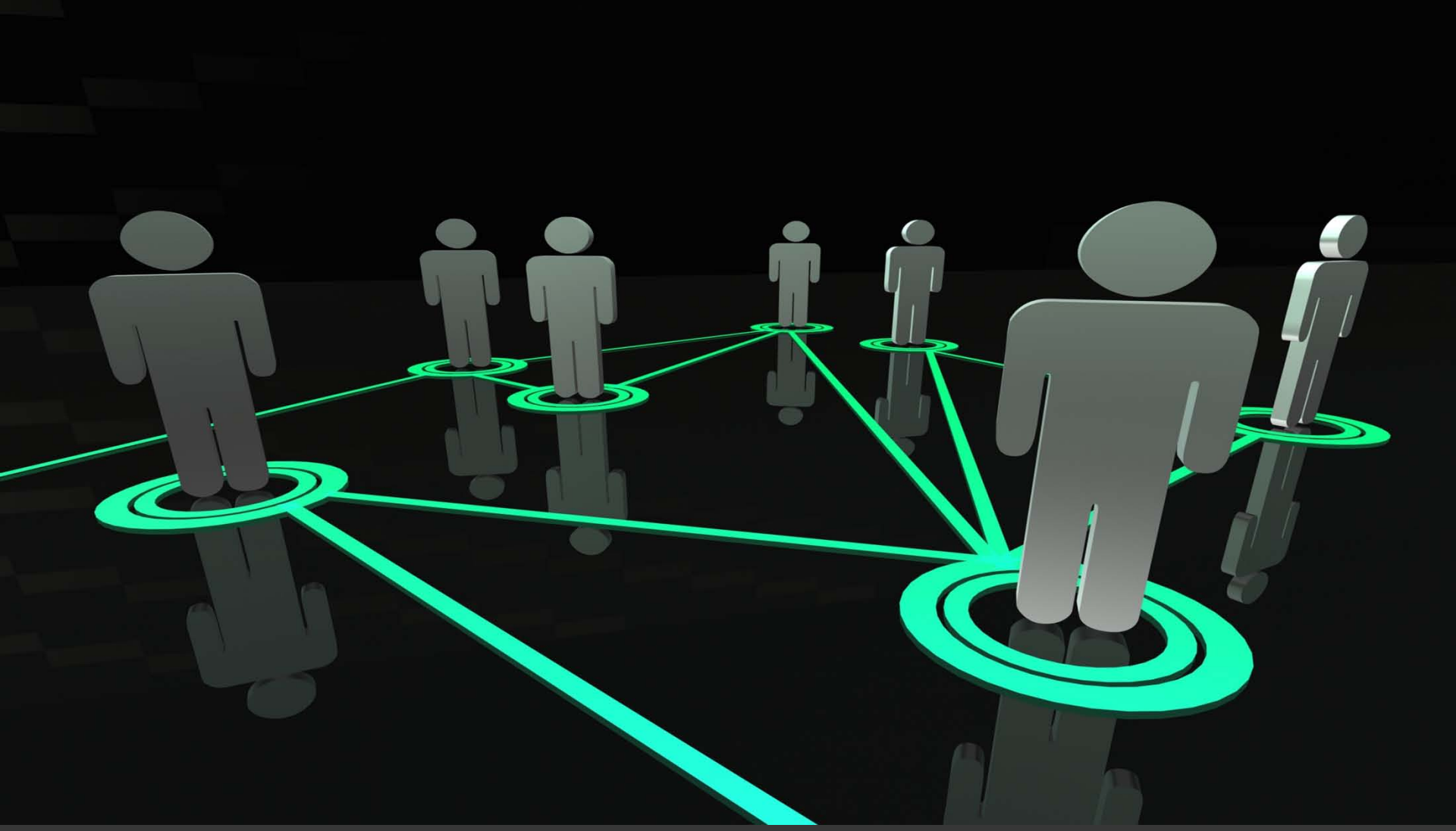
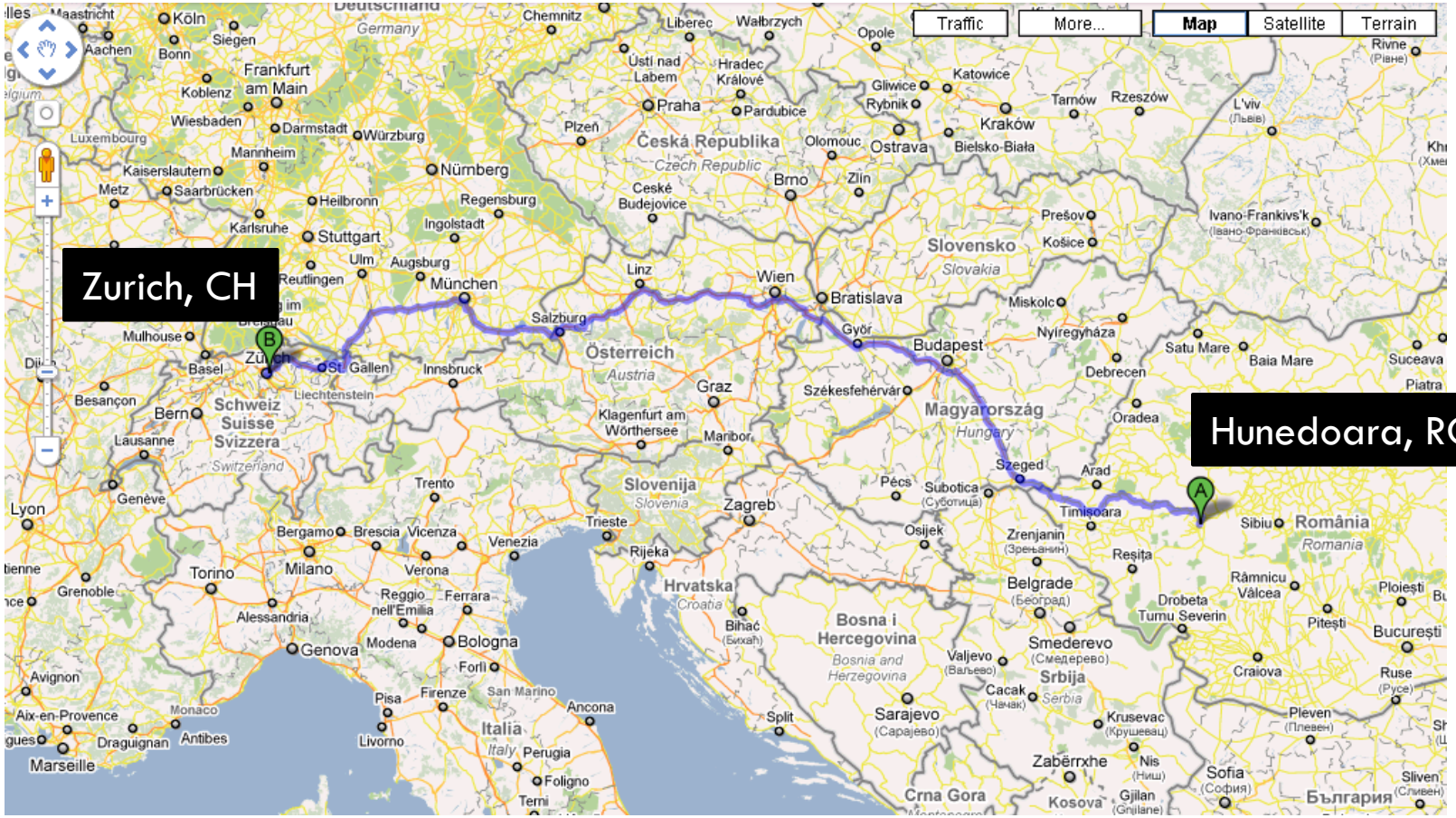


SMALL-WORLD NAVIGABILITY

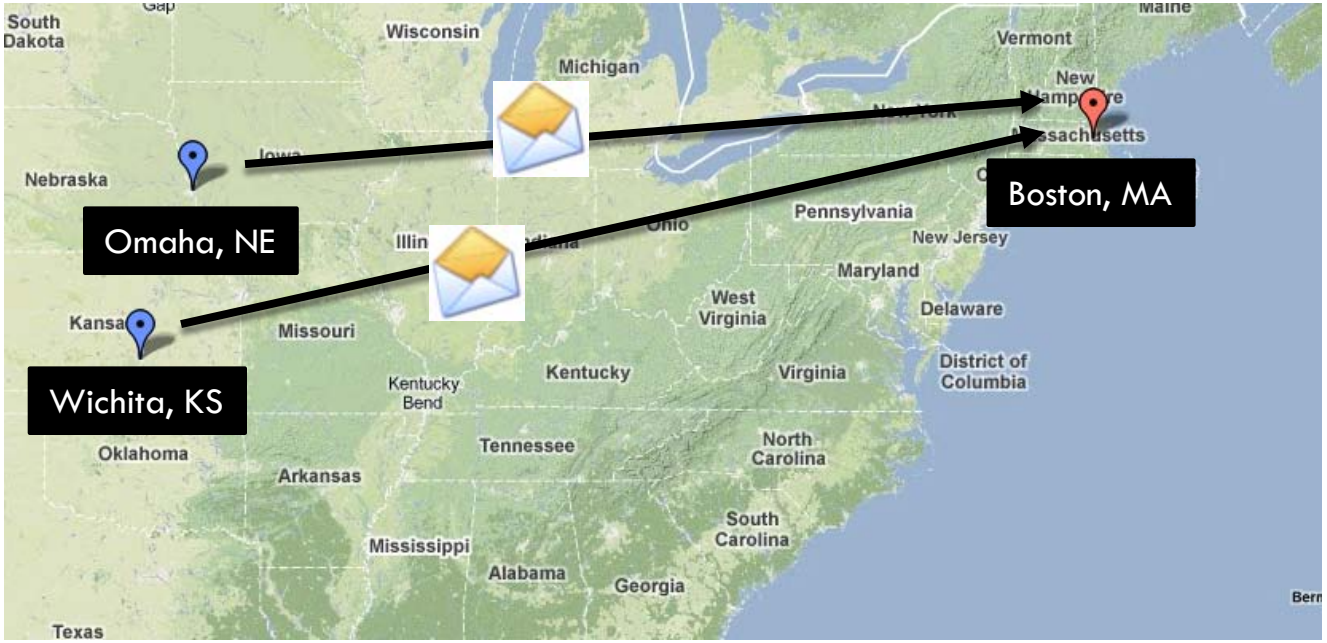
Alexandru Moga @ Seminar in Distributed Computing

Talk about a small world…



From cliché to social networks

Milgram's Experiment and The Small World Hypothesis



Human society is a **small-world** type network characterized by **short length paths**

From social networks to CS

- Models and Algorithms
- Experimental studies
- Impact in Computer Science?

Small-world phenomenon

Six degrees of separation

- "*We are all linked by short chains of acquaintance*"

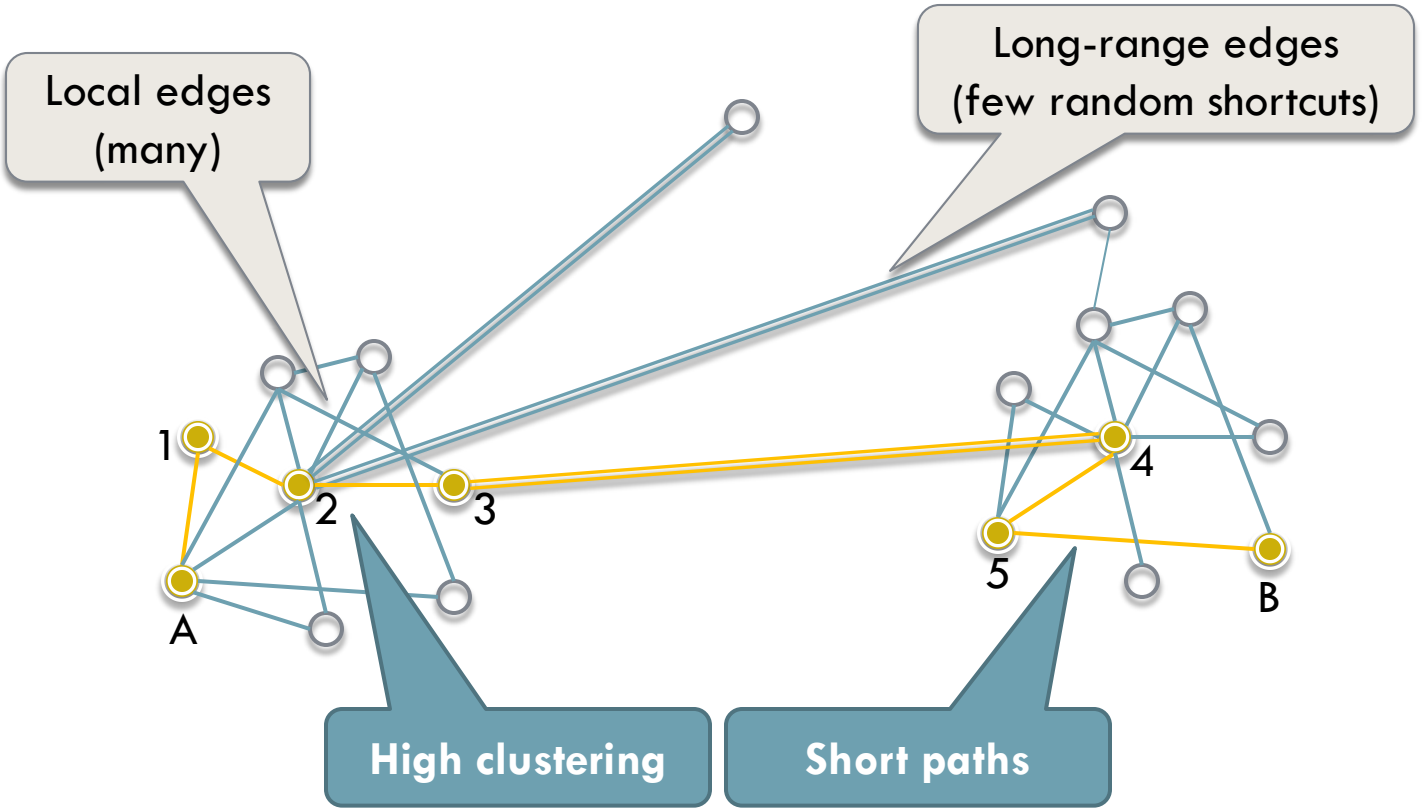
Watts-Strogatz model

- Pervasive in networks arising in nature and technology
- Fundamental factor in the evolution of WWW

Kleinberg: People can find short paths very effectively

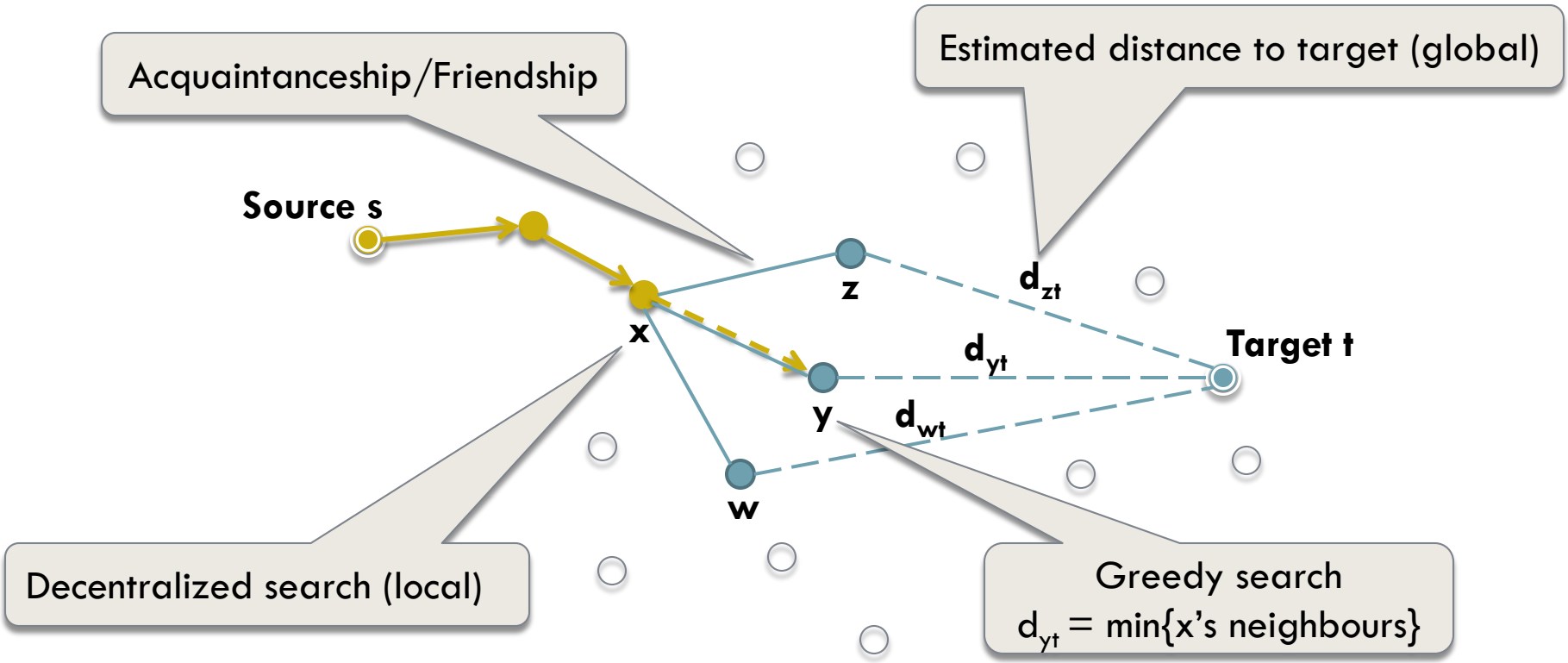
- Can we put an algorithmic price on that?

Small world characteristics



What is a good network model that exhibits such characteristics?

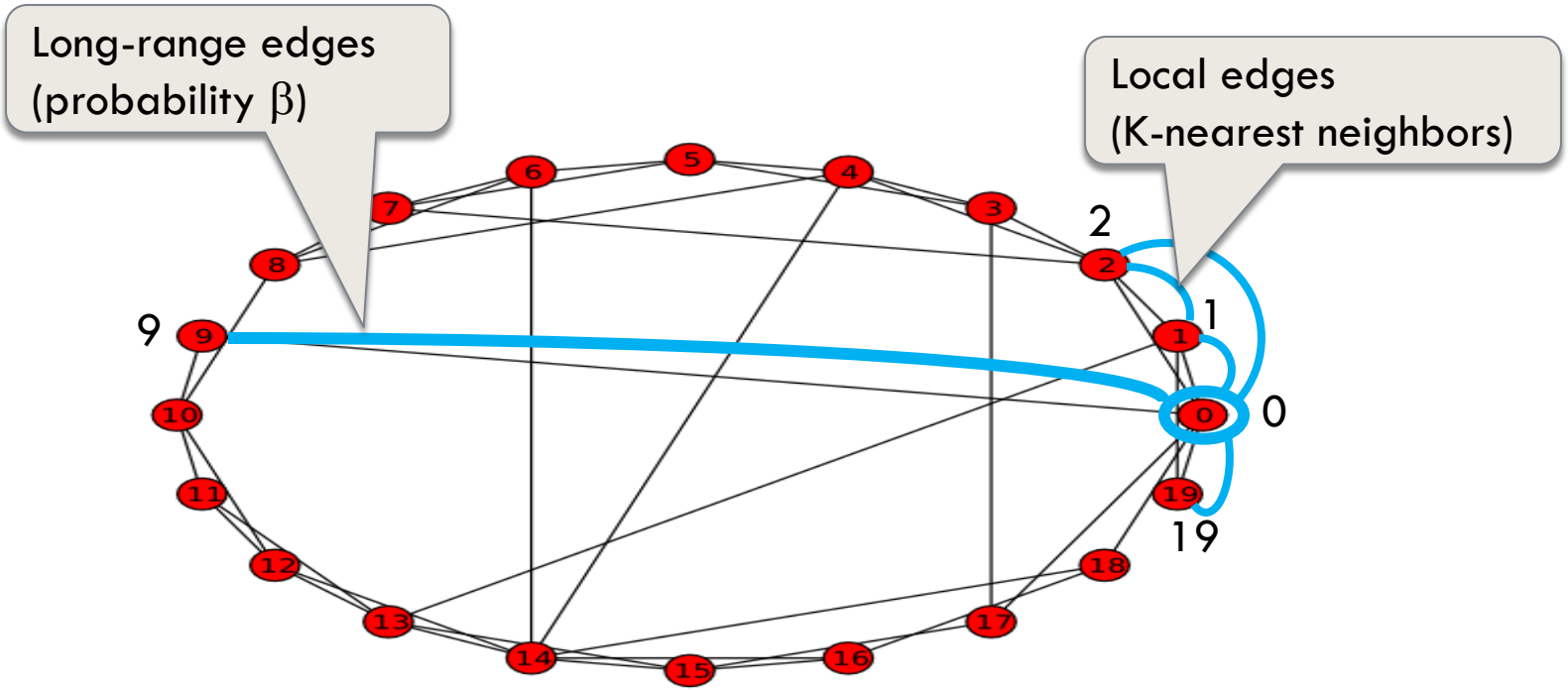
Navigation



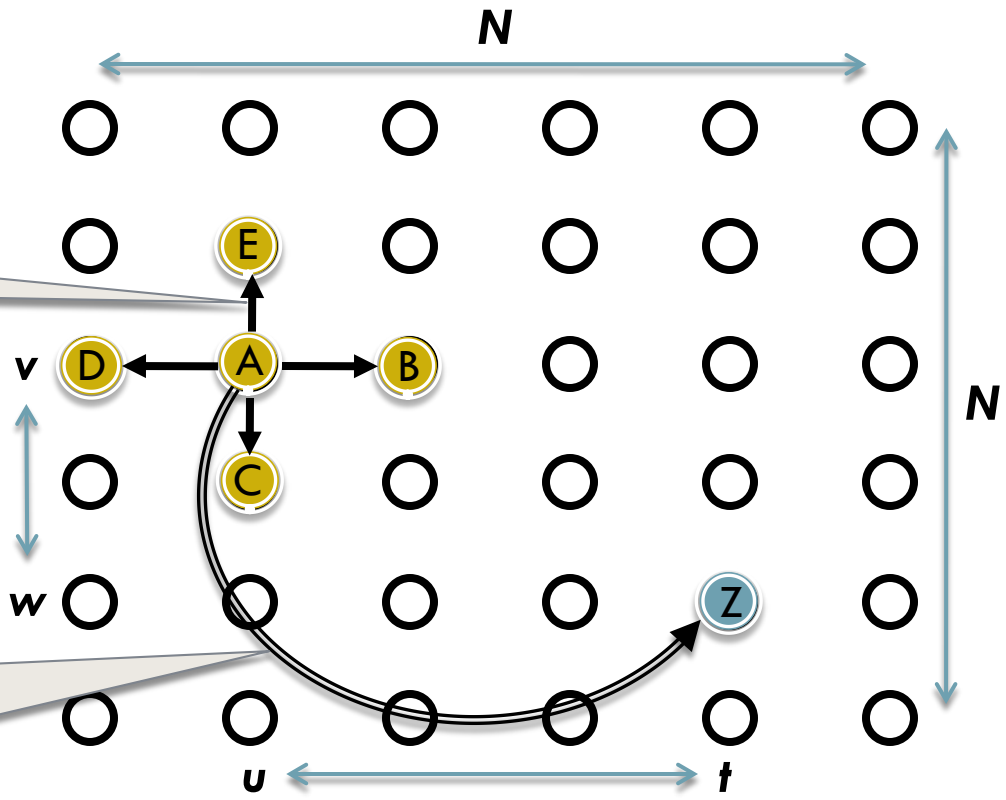
Can we effectively navigate from s to t given a network model?

The Watts-Strogatz model

- Re-wired ring lattice



Kleinberg's model



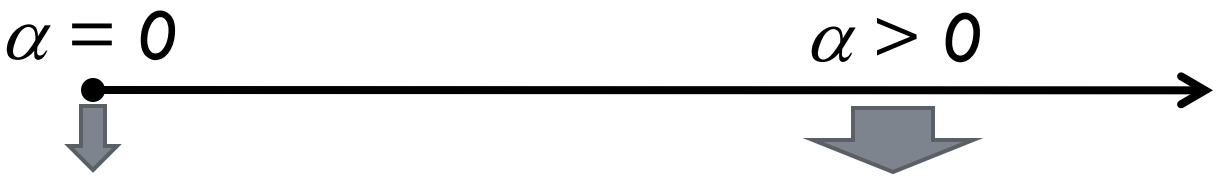
Local edges(p)
A→E := d(A,E) ≤ p

Long-range edges(q)
Pr(A→Z) ~ $1/[d(A,Z)]^{\alpha}$
Inverse α^{th}-power distribution

Lattice distance
$d(A,Z) = |t-u| + |w-v|$

Clustering exponent α

- Family of network models with parameter α



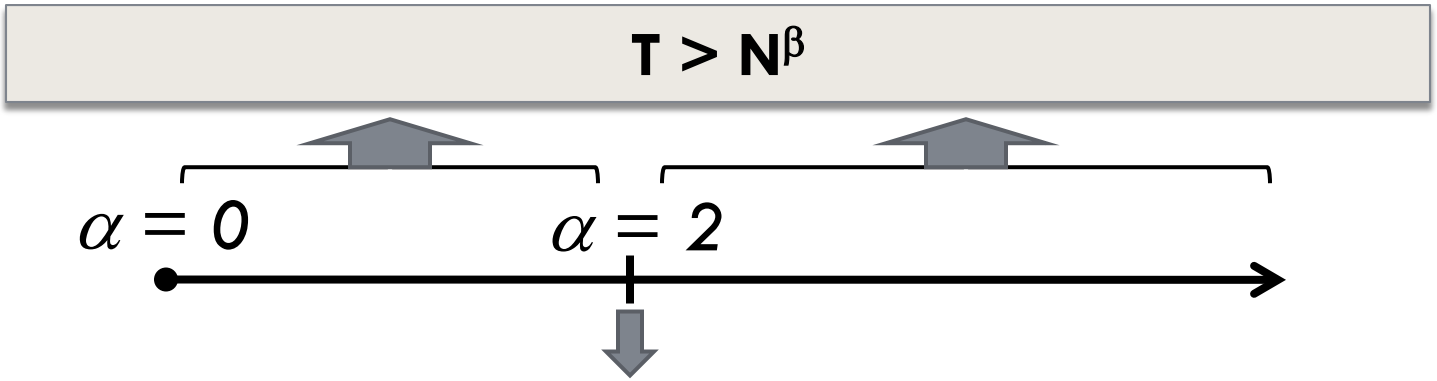
Long-range contacts chosen independently of their position (~Watts-Strogatz model)

Long-range contacts tend to cluster in the nodes' vecinity

Which α *yields an* ***effectively*** *navigable network?*

Expected delivery time T
- Expected number of steps to reach the destination
- Shortness (small T) of paths is defined as ***polylogarithmic***

Navigability in Kleinberg's model

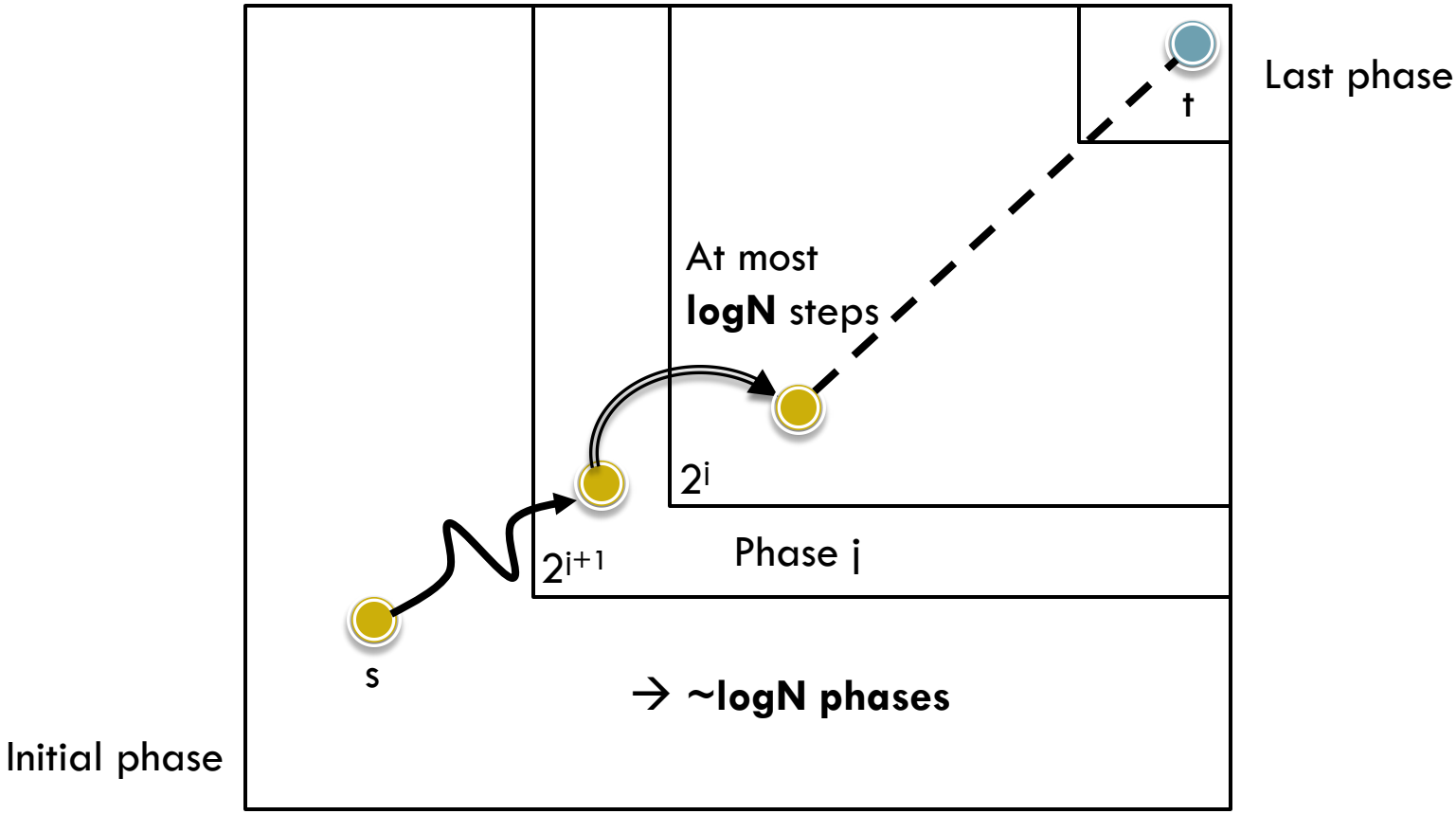


Inverse-square distribution ($1/d^2$) is the **unique** distribution that allows **polylogarythmic $T < \log^2 N$**

Generalization

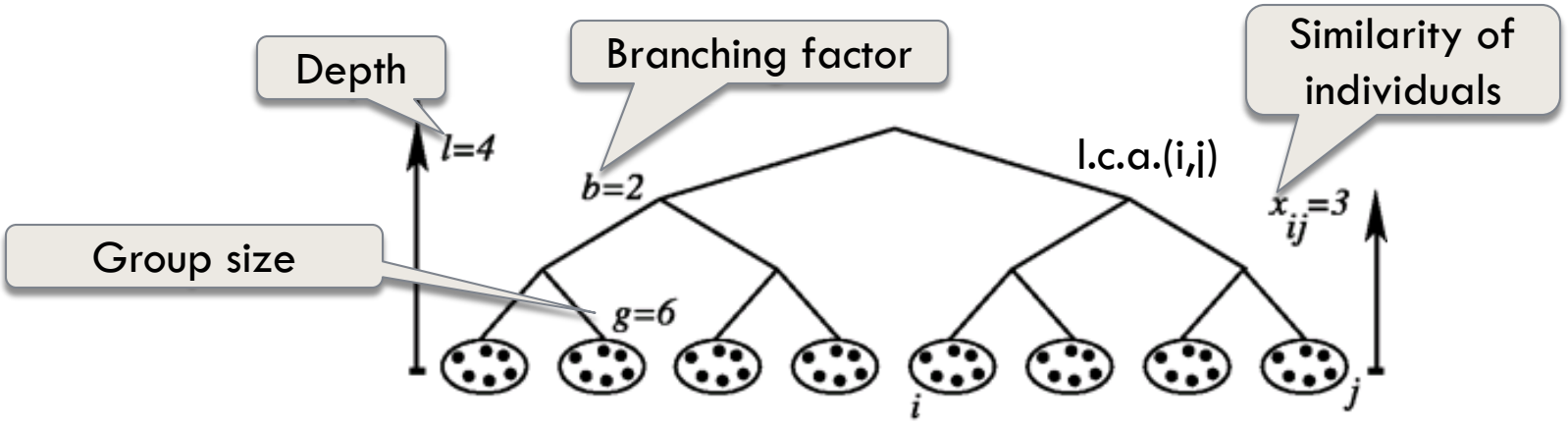
For a k-dimensional lattice, paths are polylogarithmic iff $\alpha = k$

Inverse-square distribution



Plausible social structures (*Watts et al.*)

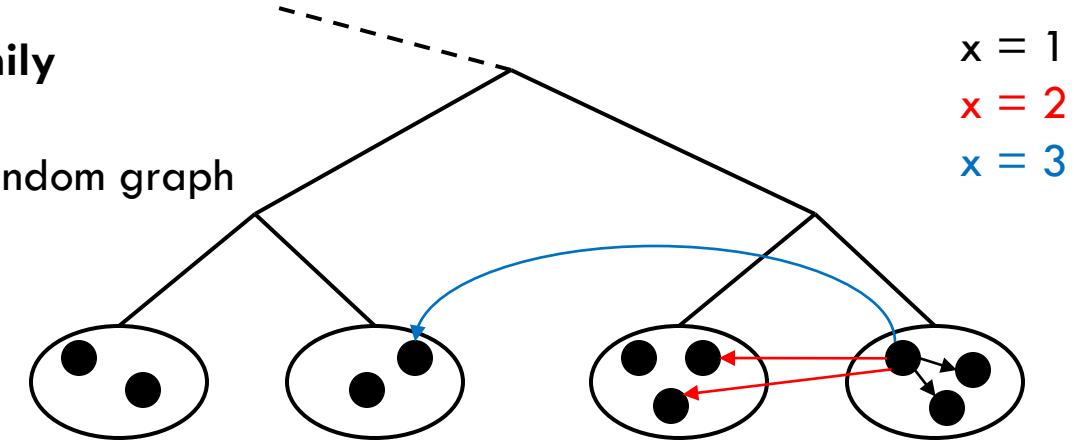
1. Individuals have identities
2. World is partitioned hierarchically (cognitively)
   - Group management is easier (typically 100 individuals)



Plausible social structures

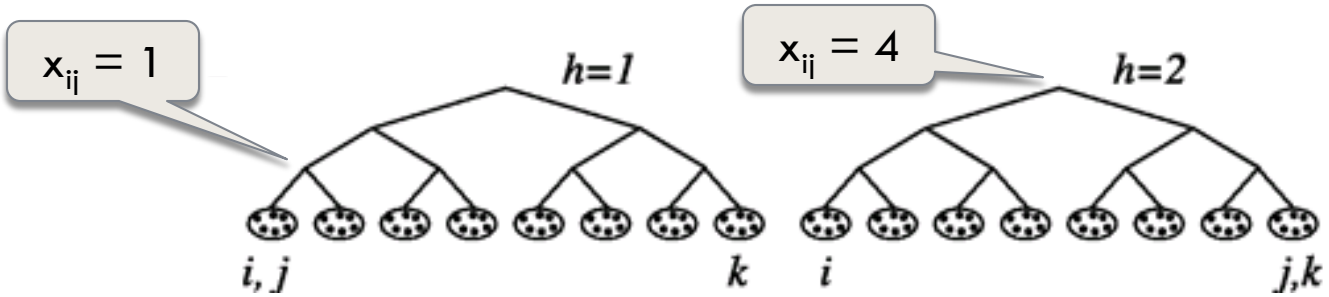
3. Network structure
   - Pr(acquaintance) decreases with decreasing similarity
   - Choose i and a link distance with $Pr(x) = ce^{-\alpha x}$
   - Choose j that is in distance x from i
   - Continue until individuals have an average of z friends

α - shows **homophily**
$e^{-\alpha} << 1$: cliques
$e^{-\alpha} = b$: uniform random graph



Plausible social structures

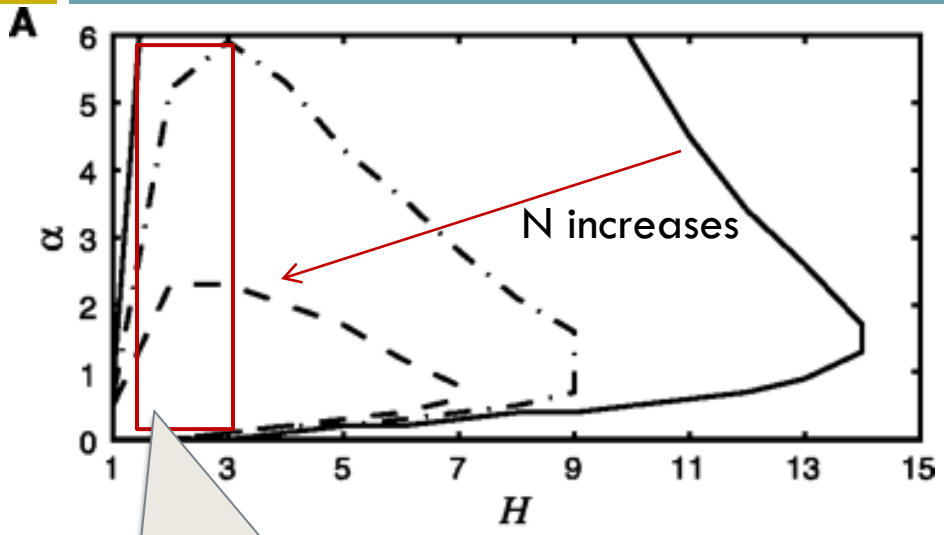
4. Social world is multi-dimensional (H)
   - Each dimension corresponds to an independent hierarchical division (e.g. geography, occupation)
   - Node identity: H-dimensional vector



$y_{ij} = 1$
$y_{jk} = 1$
$y_{ik} = 4$
$y_{ij} + y_{jk} < y_{ik}$!!!

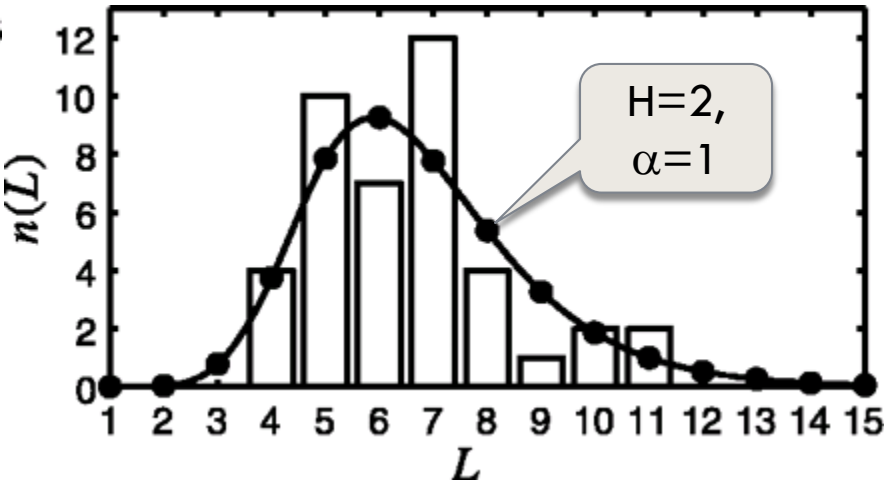
5. Perceived similarity yields "social distance"
   - Minimum similarity across all dimensions

Searchability with social distance



Searchable networks in the H-α space

Comparison to original Milgram experiment



L~6.5 (Milgram) vs. L~6.7

- Individuals are basically homophilous
- Similarity is judged along more than 1 dimenations (2-3)

Experimental studies

- Real-world social networks
- Large-scale
- Geography and occupation are crucial
- Network structure alone may not be sufficient

Geography in small-world networks (*Nowell et al.*)

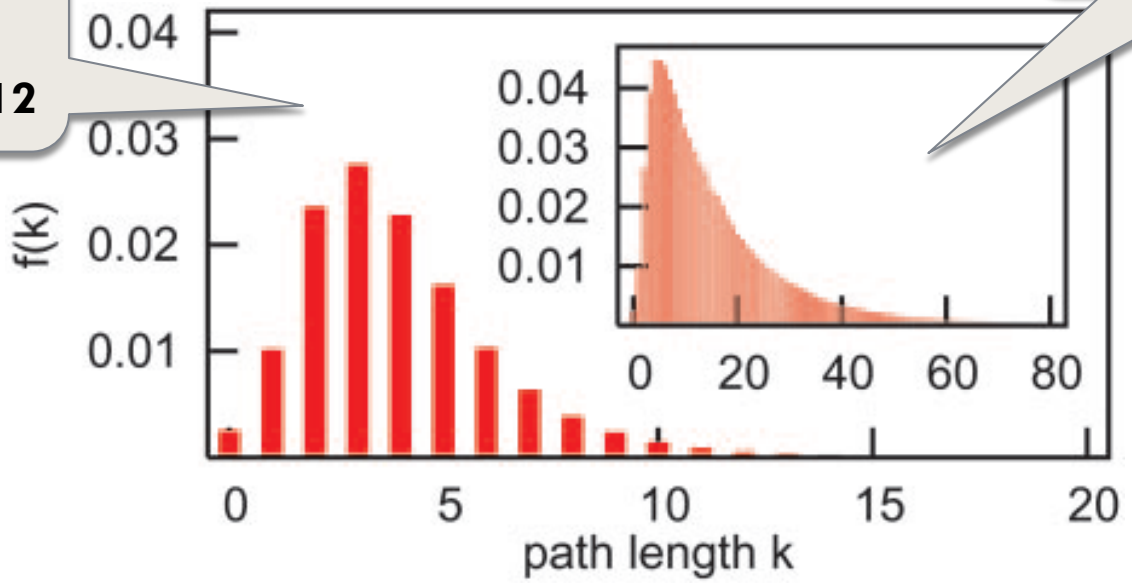
What is the importance of geography in navigation?

- LiveJournal online community
  - ~500.000 bloggers located in US
  - Friendship-based network
  - **Global routing** with GEOGREEDY

GEOGREEDY simulation

13% of chains completed with avg. length of **4.12**

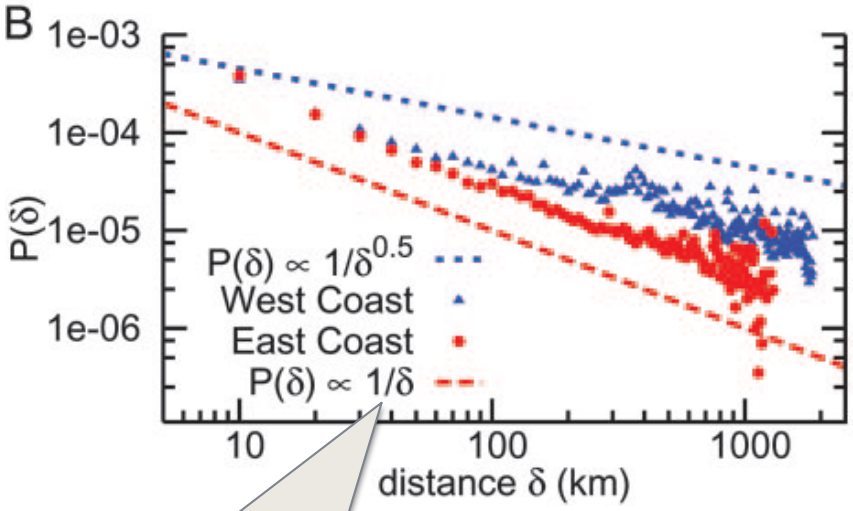
80% of chains completed with avg. length of **16.74**



What is the relation between geography and friendship?

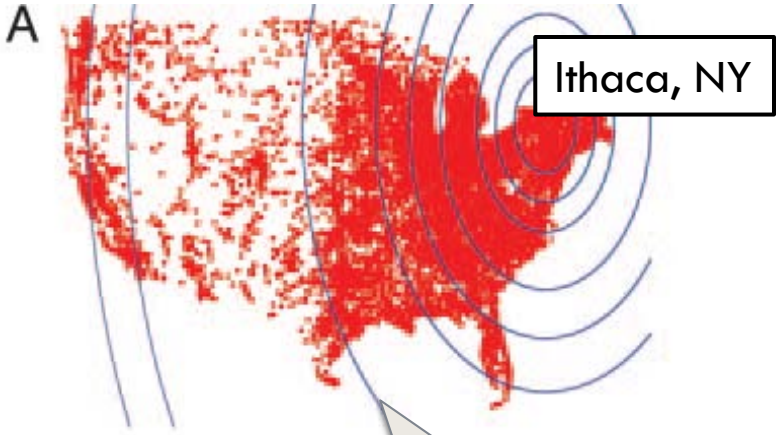
Geographic friendship probability

$Pr^{Kleinberg}(\delta) \sim 1/\delta^2$



$Pr^{LiveJournal}(\delta) \sim 1/\delta^{\alpha}$, $\alpha \sim 1$

+50.000 people



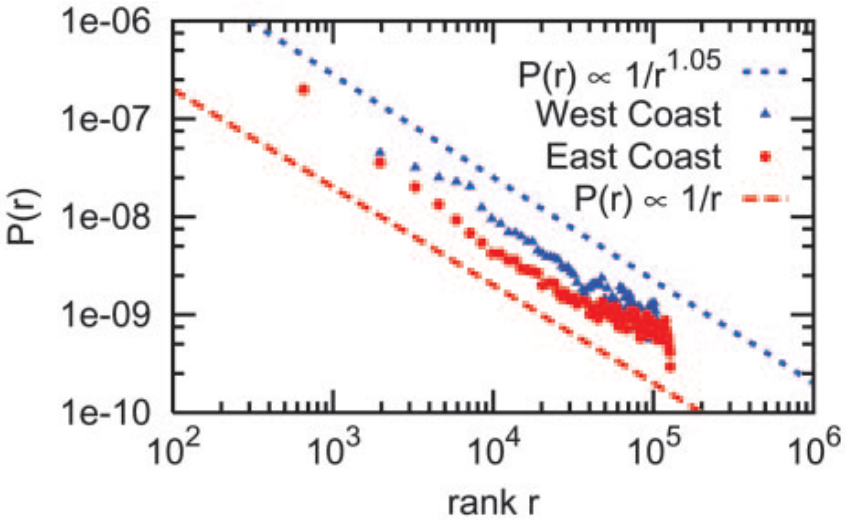
LiveJournal network exhibits large variance in population density

What is a good interpretation of geographic friendship?

Rank-based friendship

Rural Iowa

Manhattan



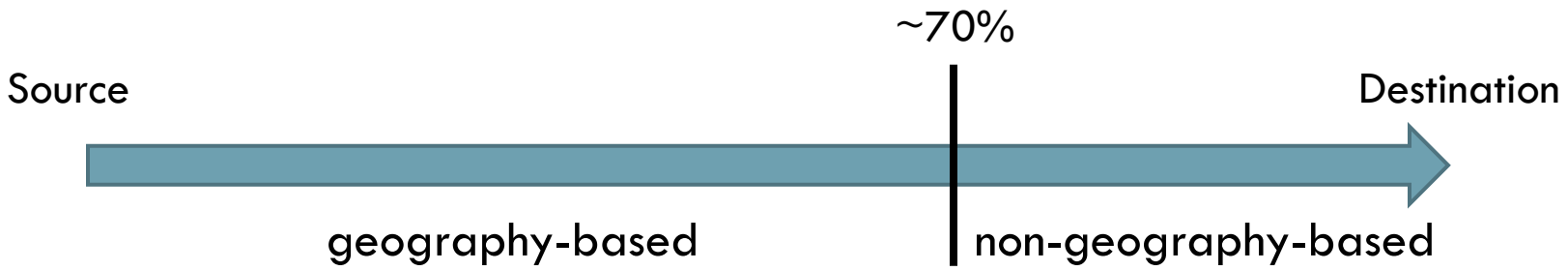
$$\text{rank}_u(v) := |\{w : d(u,w) < d\{u,v\}\}|$$

$$\Pr[u \to v] \sim 1/\text{rank}_u(v)$$

*In a network formed by **rank-based friendship**, GEOGREEDY can find short paths (**polylogarithmic**)*

Navigability in global social networks (*Dodds et al.*)

- Routing in the LiveJournal community



- Geography and occupation are the most important factors in establishing short chains

E-mail replication experiment

- Human participants (not simulated)
- ~100k individuals, 18 targets in 13 countries

Type of relationship	%	Origin of relationship	%	Strength of relationship	%
Friend	67	Work	25	Extremely close	18
Relatives	10	School/university	22	Very close	23
Co-worker	9	Family/relation	19	Fairly close	33
Sibling	5	Mutual friend	9	Casual	22
Significant other	3	Internet	6	Not close	4

Geography vs. occupation

Geography matters more in the early stages of the chain (3 steps)

L	*N*	Location	Travel	Family	Work	Education	Friends	Cooperative	Other
1	19,718	33	16	11	16	3	9	9	3
2	7,414	40	11	11	19	4	6	7	2
3	2,834	37	8	10	26	6	6	4	3
4	1,014	33	6	7	31	8	5	5	5
5	349	27	3	6	38	12	6	3	5
6	117	21	3	5	42	15	4	5	5
7	37	16	3	3	46	19	8	5	0

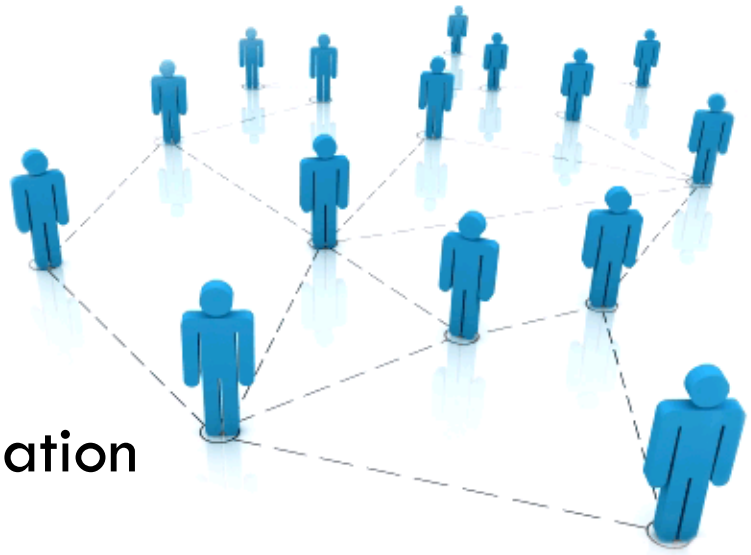
Occupation clearly takes over in the later stages

Results of the study

- Without enough incentives, the small-world hypothesis may not hold
  - E.g. Target 5 (university prof.) accounted for 44% of the completed chains → good reachability
- Network structure alone is not enough

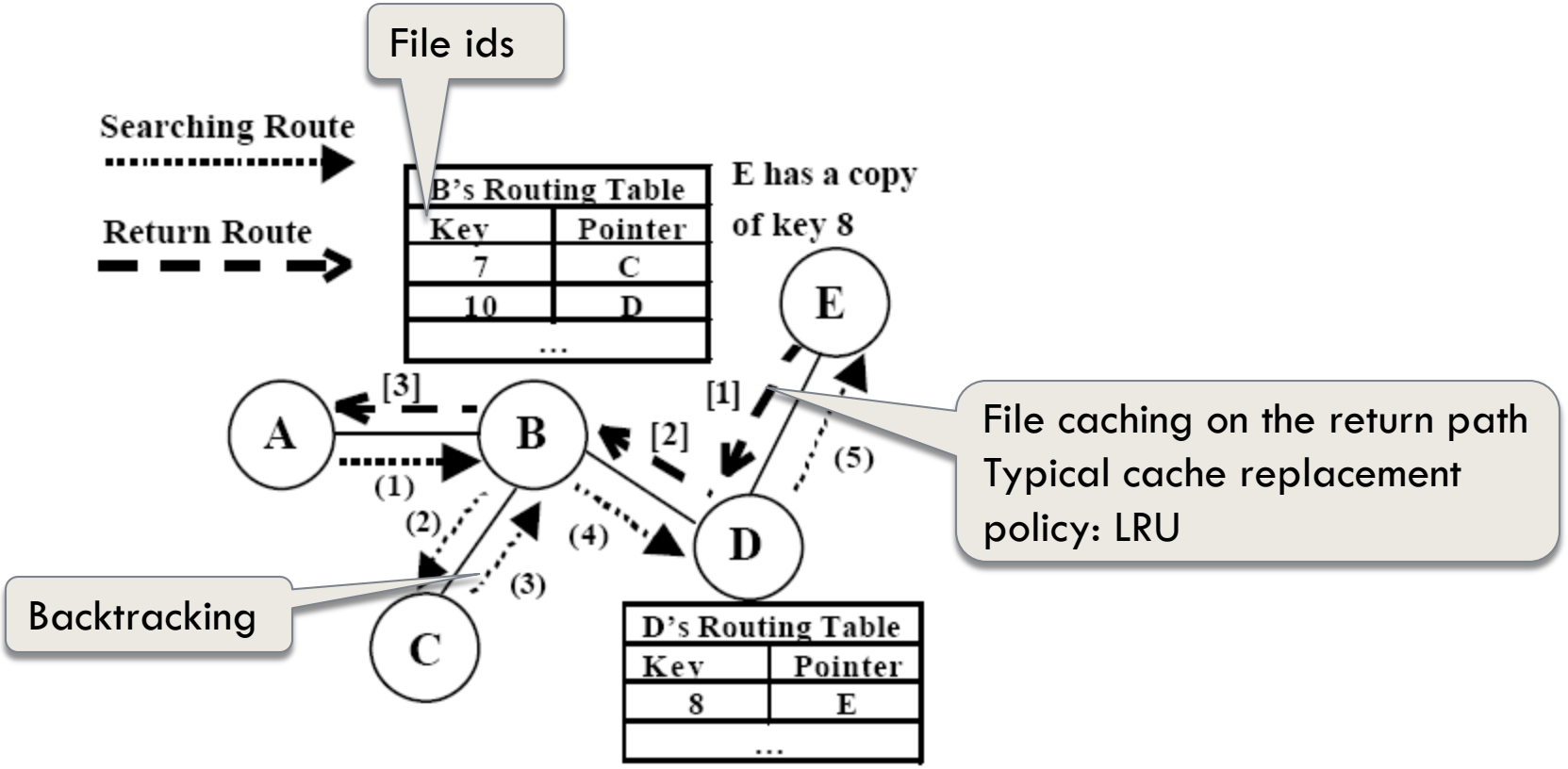
Case study: Freenet

- P2P system
  - Collaborating group of Internet nodes
  - Overlay special-purpose network
  - Application-level routing



- Freenet
  - Distributed anonymous information storage and retrieval
  - Unstructured system

Case study: Freenet

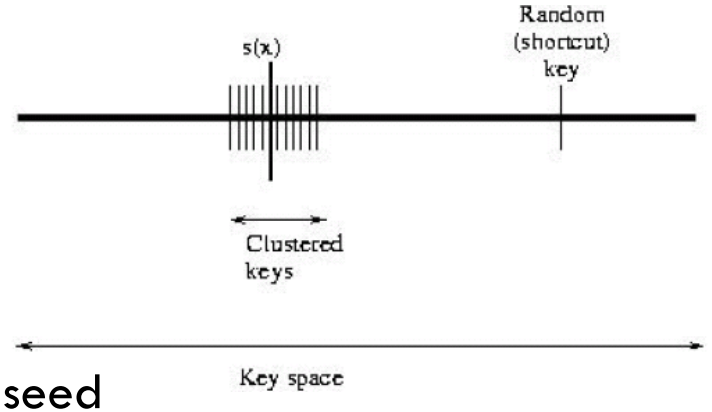


Case study: Freenet

- At low load:
  - Freenet network shown to evolve into a "small-world" (high clustering + logarithmic paths)
- At high load:
  - Frequent local caching actions
  - Clusters may break → small-world hypothesis might not hold

Case study: Freenet

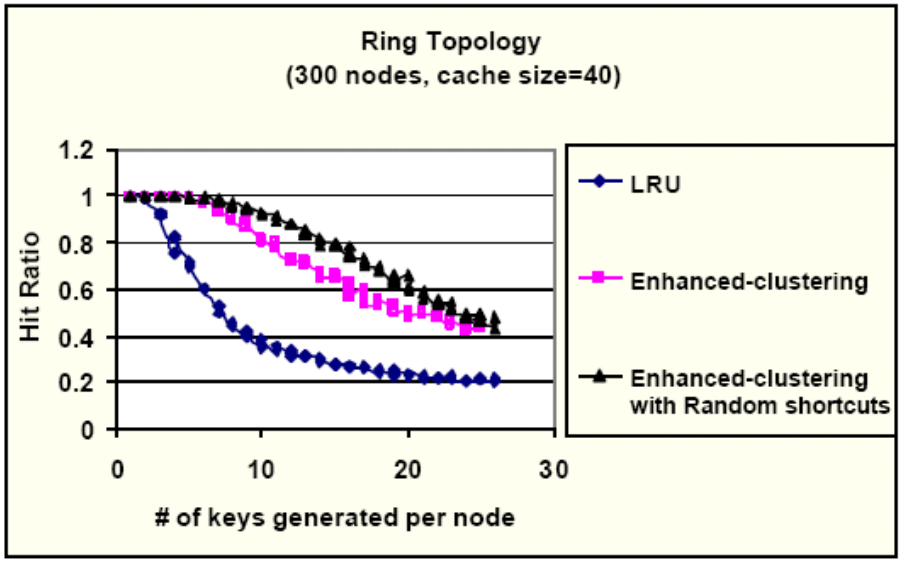
- Enhanced-clustering cache replacement policy
  - Preserve key clustering in the cache
  - Each node chooses a seed s(x) randomly from the key space



  - At node x (datastore full)
    - key u arrives
    - choose v which is farthest from the seed
      - *Distance(u, seed)* ≤ *Distance(v, seed)*: cache u, evict v, create entry for u
      - *Distance(u, seed)* > *Distance(v, seed)*: cache u, evict v, create entry for u with probability p (randomness)

Case study: Freenet

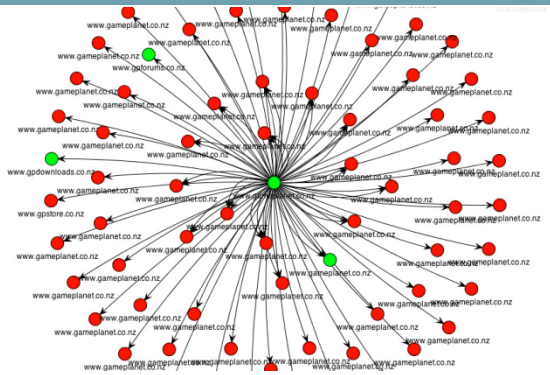
- Empirical results



- Analytically
  - $f(d(x,y)) \sim 1/d(x,y) = 1/|s_x - s_y|$
  - Expected delivery time: $O(\log^2 n)$

Other applications

- Crawling the WWW



- On-line search in the unknown

- Supercomputing

Conclusion

A small-world network is characterized by:
High clustering of nodes
"Short" paths

Small-world phenomenon has two sides
Existential and Algorithmic

Unsupervised networks are generally small-worlds

References

- **Identity and Search in Social Networks**
  Watts, D.J. and Dodds, P.S. and Newman, MEJ In Science 2002.
- **Small-World Phenomena and the Dynamics of Information**, Kleinberg, J. In NIPS 2002.
- **Small-world Phenomena: An algorithmic perspective**, Kleinberg, 2000
- **Navigation in a Small World**, Kleinberg, Nature 406 (2000)
- **Geographic routing in social networks**
  Liben-Nowell, D. and Novak, J. and Kumar, R. and Raghavan, P. and Tomkins, A. PNAS 2005.
- **An Experimental Study of Search in Global Social Networks**
  Dodds, P.S. and Muhamad, R. and Watts, D.J. In Science 2003.
- **The Small World Web**, L. Adamic, 1999
- **Growing and Navigating the Small World Web by Local Content**, F. Menczer, 2002
- **Using the Small-World Model to Improve Freenet Performance**, Zhang et al.