Networks, Search, and The Small-World Problem

Duncan Watts
Columbia University

Peter Dodds
Roby Muhamad
Mark Newman
Steven Strogatz

Outline

I The Small World Problem

- Milgram's experiment
- Why is it surprising?

II Small World Networks

- The modern approach

III Small World Search

- The algorithmic problem
- A sociological approach

I: The Small World Problem

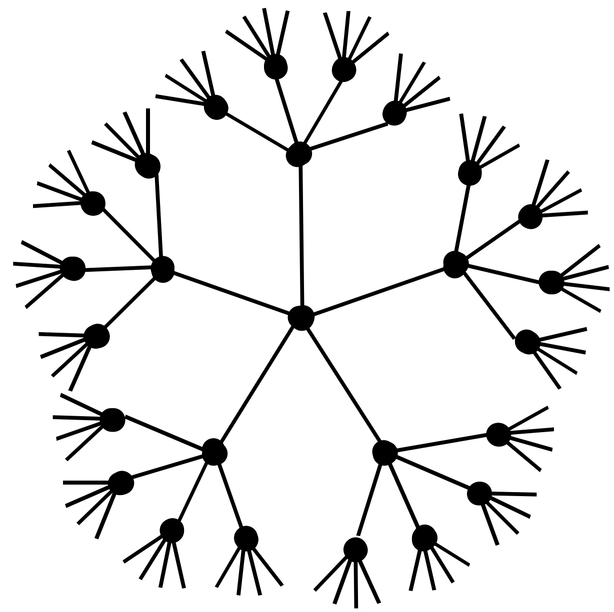
- In the late 1960^s, Travers and Milgram invented "small-world method" (Milgram 1967, Travers and Milgram 1969)
- For a single ***target*** in Boston (a stockbroker), chose 296 initial ***senders***
  - 100 in Boston, 196 in Nebraska
  - Each ***sender*** forwards letter to friend who is "**closer**" to target than themselves
  - Conditions repeat for successive senders, yielding ***message chains***
  - message chains either reached target (20%) or terminated

“Six Degrees of Separation”

- Milgram’s surprising result:
  - Average length of the completed chains was about 6
- Led to the famous phrase (Guare 1990).

A back of the envelope "explanation"?

- Ego 1
- Ego's friends 100
- Their friends $100^2 = 10K$



100^5 = 10 billion > Earth's Population!

Critical Property: When number of friends small compared to population, and social ties created at random probability of Ego's friends being friends of each other is negligible

Why was Milgram's result surprising?

- Random ties, however, are ***not*** realistic
- In reality, social networks exhibit
  - Homophily (Merton and Lazarzfeld, 1954)
  - Triadic closure (Rapoport, 1957)
- Hence Clustering/redundancy/group structure

Interesting Small World Problem is therefore:

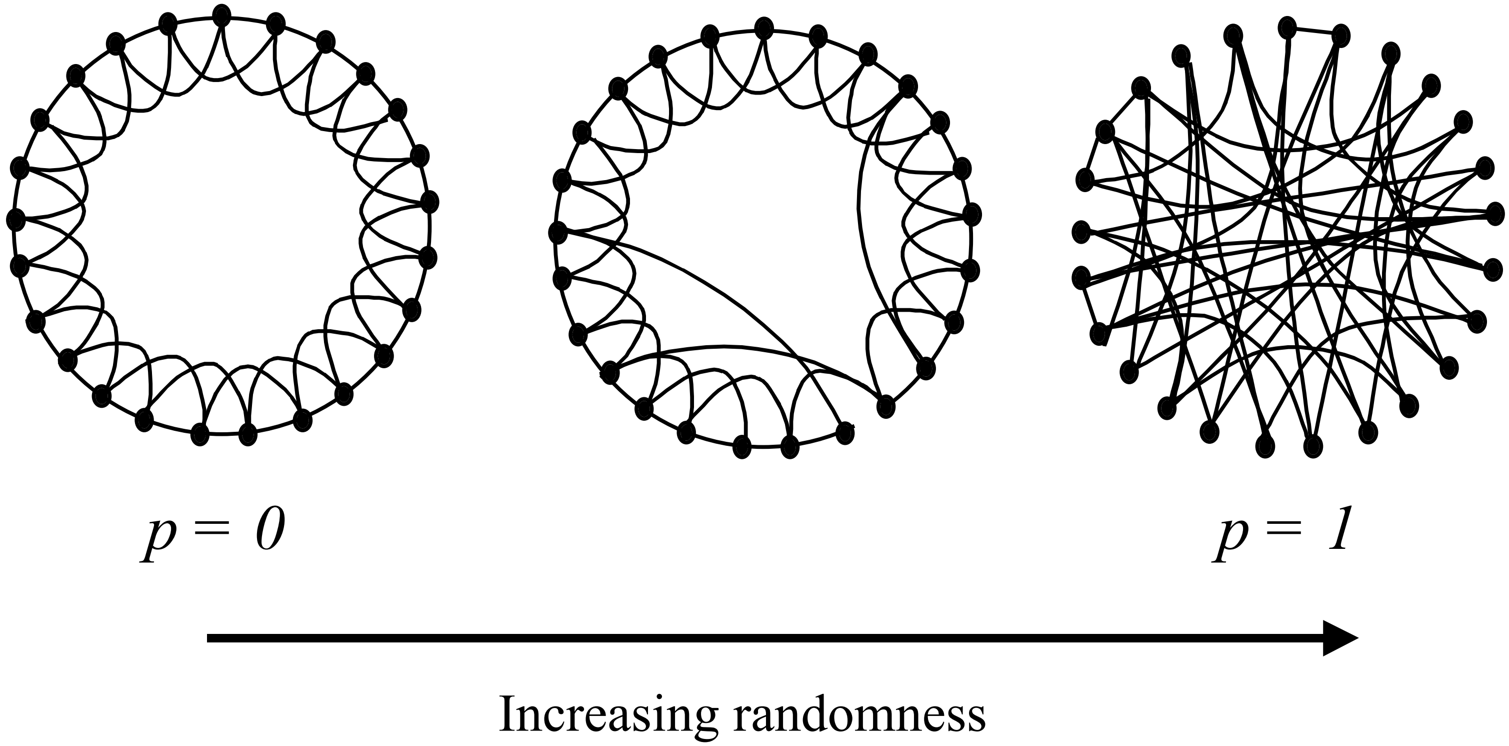
- **How is it possible for Social Networks to be:**
  - **Very highly ordered/clustered *locally* (like social groups), and**
  - **Still be "small" *globally*? (like random networks)**
- **Problem is that *Clustering* makes Analysis Hard**
  - **It was theoretical difficulty that led to Milgram's experimental approach in the first place**

II: Small World Networks

- After Milgram, not much done for 30 years
  - Experiments are hard to perform
  - Large-scale network data are hard to collect

- Arrival of modern computers enabled new theory
  - What are the conditions under which ***any*** network can be clustered and still “small”?
  - Interpolation between ordered and random networks (Watts and Strogatz 1998)

Rewiring networks from Order to Randomness



At the Extremes:

- p=0 (Ordered)

$$L \propto \frac{n}{k}$$

- "Large"

$$C \approx \frac{3}{4}$$

- "High"

- p=1 (Random)

$$L \propto \frac{\ln n}{\ln k}$$

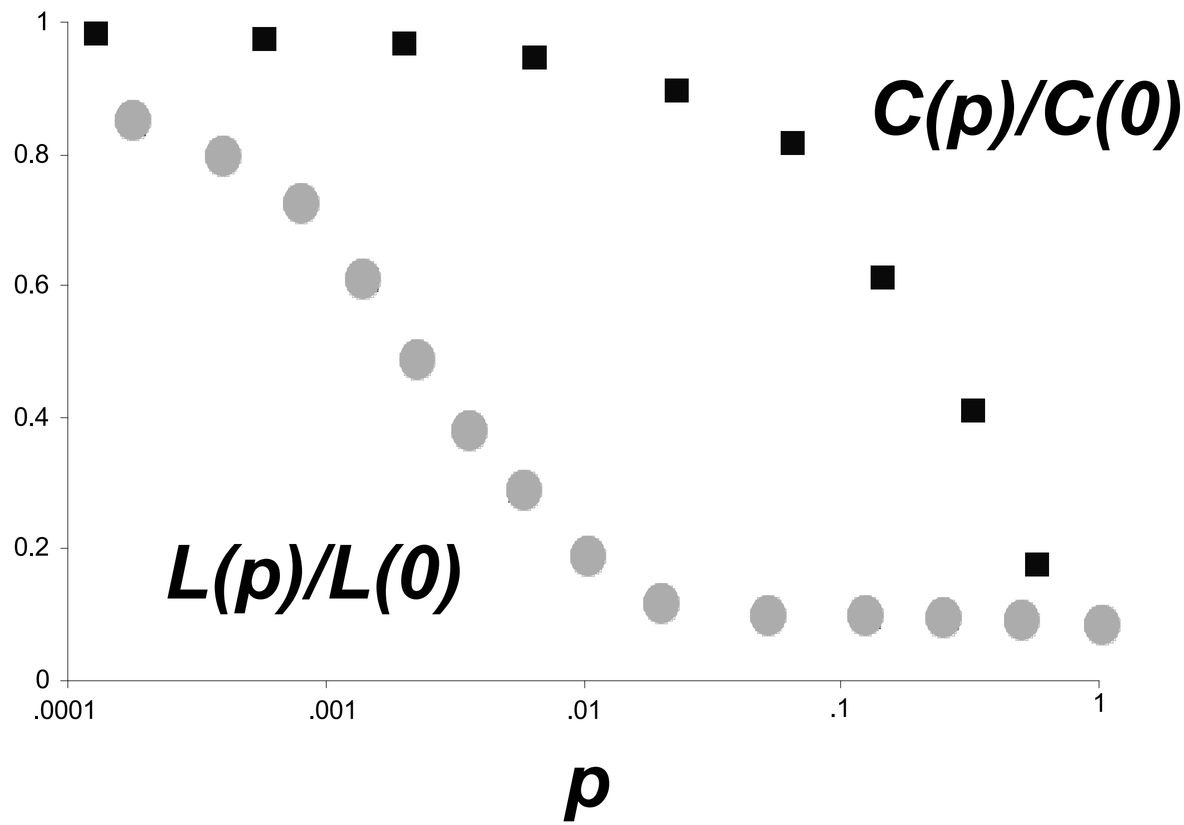
- "Small"

$$C \approx \frac{k}{n} \to 0$$

- "Low"

Intuition: the world can be *either*
"large and highly clustered",
or "small and poorly clustered",
but *not* "small and highly clustered"

Path Length and Clustering vs. Random Rewiring



Origin of Small-World Networks

- *L* is governed by **<u>Number</u>** (***p****N*)of random shortcuts
  - Surprising fact: roughly 5 shortcuts reduce average path length by factor of 1/2, *regardless* of *N*

But

- C is governed by **<u>Fraction</u>** (**p**) of random shortcuts.

Origin of Small-World Networks

- Main result:
  - For large ***N***, a small **fraction** (***p***) of shortcuts will contract ***L***, but leave ***C*** unchanged.
- Conclusions:
  - Small-World Networks are generic
  - Should be widespread
  - Not confined to social networks

Examples of Small-World Networks

	L_{Actual}	L_{Random}	C_{Actual}	C_{Random}
Movie Actors	3.65	2.99	0.79	0.00027
Power Grid	18.7	12.4	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

Examples of Small-World Networks

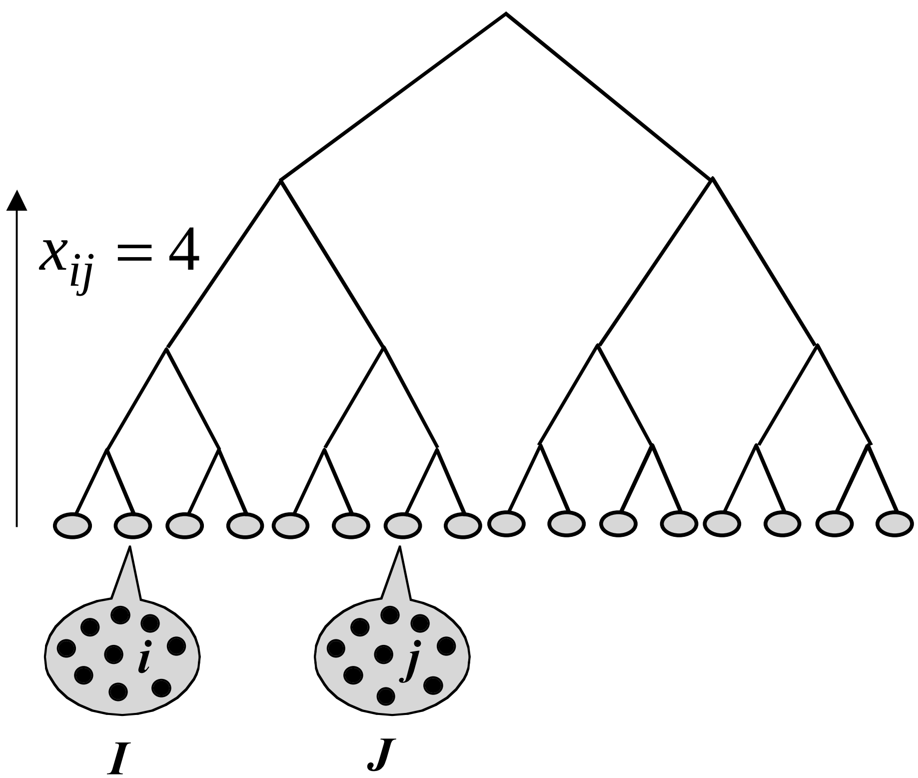
- Movie actors
- Power grid of Western United States
- Neural network of C. elegans
- World Wide Web
- Ownership network of German firms
- Metabolic network of E. coli
- Collaboration networks of scientists
- Boards of directors of Fortune 1000 Companies

III: Small-World Search

- Travers and Milgram showed not only that
  - short paths exist between randomly-selected pairs

  but
  - individuals could actually ***find*** these paths using only:
    - Local information about the network
    - Simple heuristic strategies
- Jon Kleinberg (1999, 2001) identified this "Algorithmic Small-World Problem"

Sociology Important!



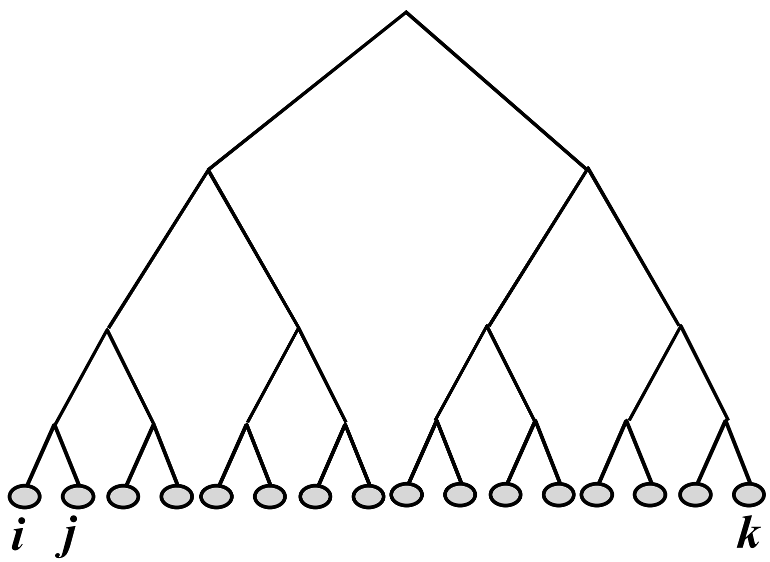
1. Individuals ***i,j*** belong to groups ***I,J***
2. Group membership equivalent to ***social identity***
3. Individuals ***partition*** the world hierarchically
4. Distance between groups measured on hierarchy

Social Identity:

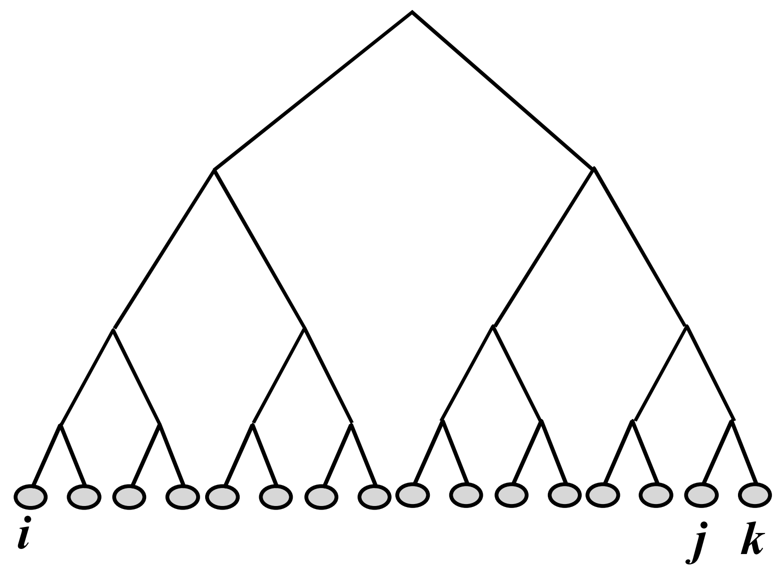
- Hierarchy is a cognitive device that defines similarity and difference between individuals.
- But it isn't actually the network.
- Network is generated as function of social distance x: $p_{ij} = c \exp(-\alpha x_{ij})$
- α is homophily parameter

Multiple Dimensions

- Crucial feature: individuals cluster the world in ***<u>multiple</u>*** ways
- Leads to the notion of ***Social Identity***



Geography

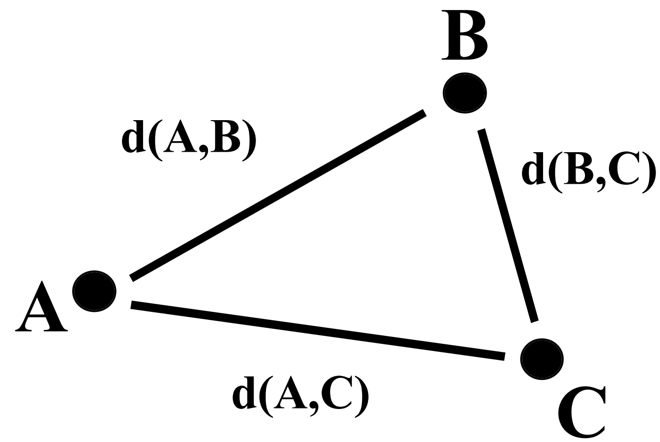


Occupation

Social Distance

❖ Social distance is **minimum** distance across all dimensions

- ❖ Minimal "metric" violates "***triangle inequality***"
- ❖ Individuals have 2 levels of information
  - ◆Social "distance" (Global)
  - ◆Local knowledge of network
- ❖ Neither of these – on its own – is adequate
  - ◆Social "distance" not a true distance
  - ◆Network "distance" only locally known
- ❖ But ***together***, they resolve the search problem via a simple greedy algorithm



$$d(A,B) \le d(A,C) + d(B,C)$$

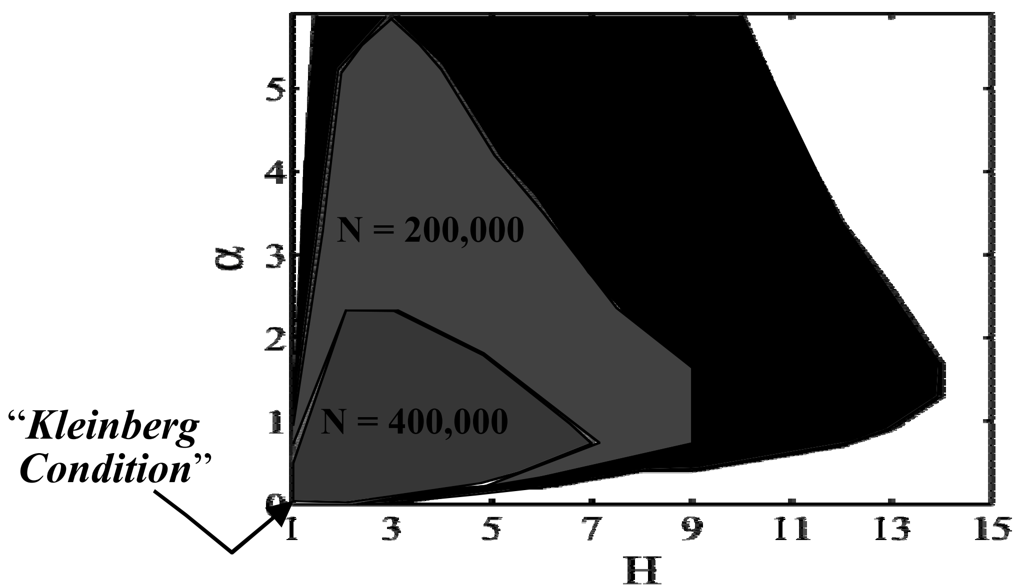
Local Search Algorithm

- Each node has the following information
  - Coordinates (“Identity”) of target (t)
  - Coordinates of self
  - Coordinates of immediate neighbors
- Node i passes message to its neighbor j, that has the smallest social “distance” y(j,t).
- In effect, the same algorithm used by Milgram’s subjects

What is “Small”?

1. Assume: Message failure probability = 25%
2. Require: 5% of chains complete

→ _small ≤ 11 steps



Parameter regions in which networks are searchable

<u>Main Result:</u>

Searchable Networks are Generic

Some Consequences

- In a world of one social dimension – "Kleinberg condition" is required for searchability

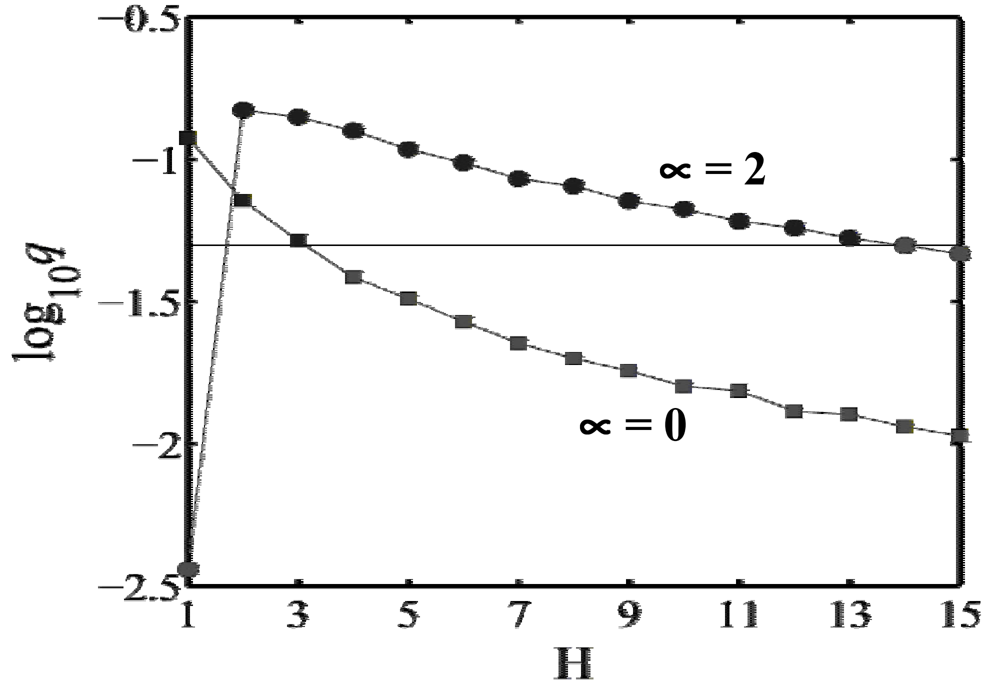
  But,

- in a world of multiple social dimensions – homophilous networks work better

  or

- in a homophilous world, multiple social dimensions are essential for searchability

Some Consequences

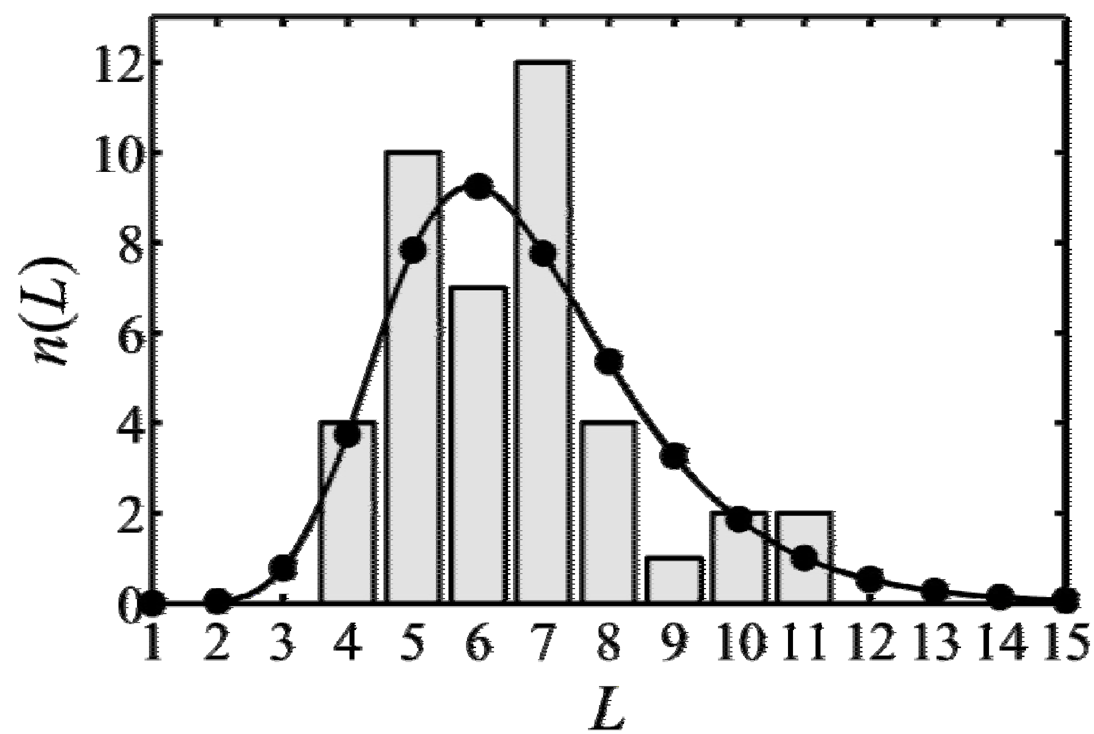


∝ = 0   corresponds to Kleinberg condition

∝ = 2   corresponds to homophilous network

The Model – Results

❖Milgram's Nebraska-Boston data



Key Notion

- ***Social identity*** governs both
  - The creation of the network
  - Successful search strategies on the network
- Identity makes search possible
  - ***Network structure is not enough***

The New Small-World Experiment
("bigger, faster, and less expensive")

- Columbia Small-World Research Project
- Very similar to Milgram's Experiment, but web-based
  - smallworld.sociology.columbia.edu
- Initial results (Dodds, Muhamad, and Watts, 2002)
  - 60,000 senders
  - 19 targets
  - 171 countries
- 380 chains complete (worse attrition than Milgram)
- Median chain length ranges from 5 (same country) to 7 (different country)

Who Cares Anyway?

- Small world problem is a particularly clean example of ***social search*** (locate remote target using local ties)
- Social search critical aspect of ***problem solving*** when
  - Environment is uncertain/ambiguous
  - Central database/directory is absent
- Technological example: ***peer-to-peer networks***
- But human organizations already have efficient peer-to-peer networks.
- By extracting essence of social search, may be able to design better protocols and "smarter" networks.

Six Degrees:
The Science of A Connected Age
(W. W. Norton, 2003)

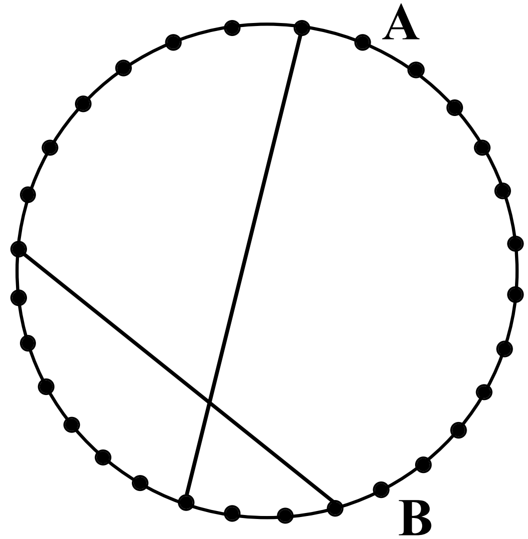
Home Page

http://www.sociology.columbia.edu/people/index.html

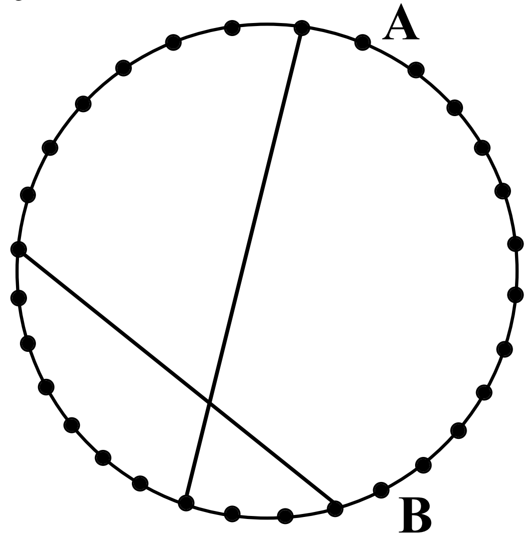
Small World Project

http://smallworld.sociology.columbia.edu

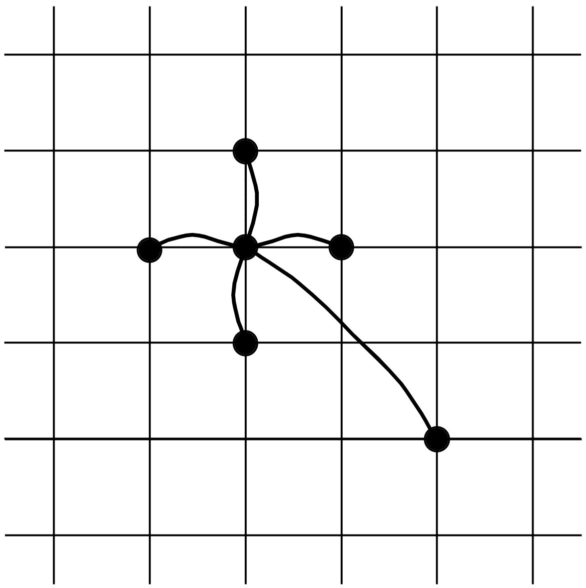
- First, Kleinberg proved that when random edges are added with uniform probability (with respect to lattice distance), individuals cannot find short paths.



- Reason: uniform edges are not correlated with underlying “social distance”; hence, having used one shortcut to get closer to target, additional shortcuts are equally likely to move message far away.



Kleinberg's Model



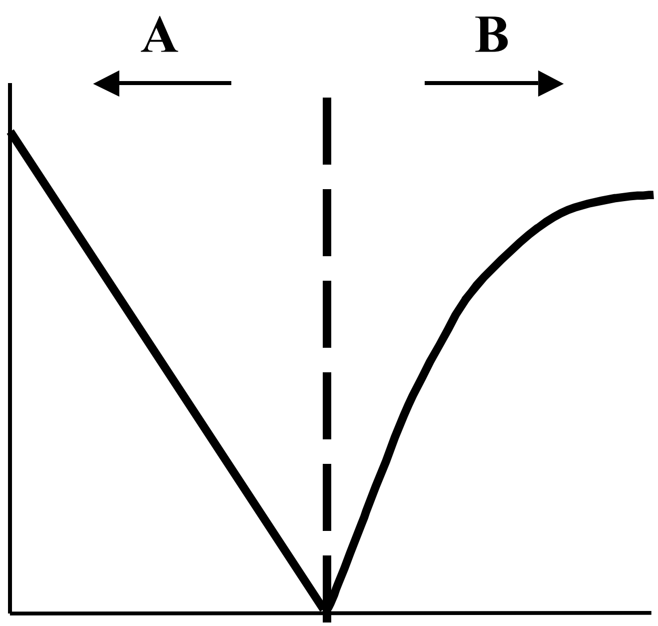
- ❖Local contacts (lattice)
- ❖Random contacts $p(r) = cr^{-\gamma}$

When $\gamma = 0$ get uniform random edges

When $\gamma >> 1$ all contacts are local

What happens for intermediate values of γ?

Kleinberg's Model



$\gamma = 2$

At critical point

Short paths exist

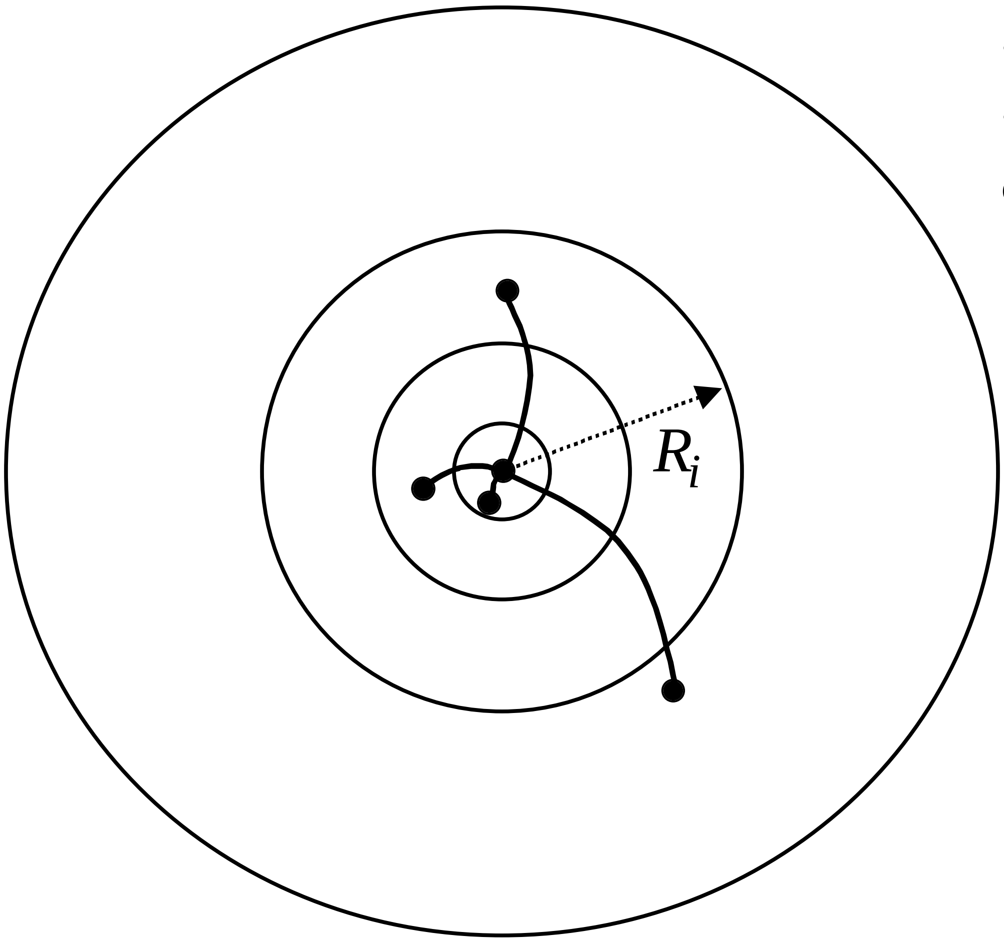
And findable

- ❖ General Idea
  - ◆ Distribution of random contacts ***encodes*** information about underlying "***social structure***"

A : short paths exist but can't be found

B : paths easy to find but not short

How Does It Work?

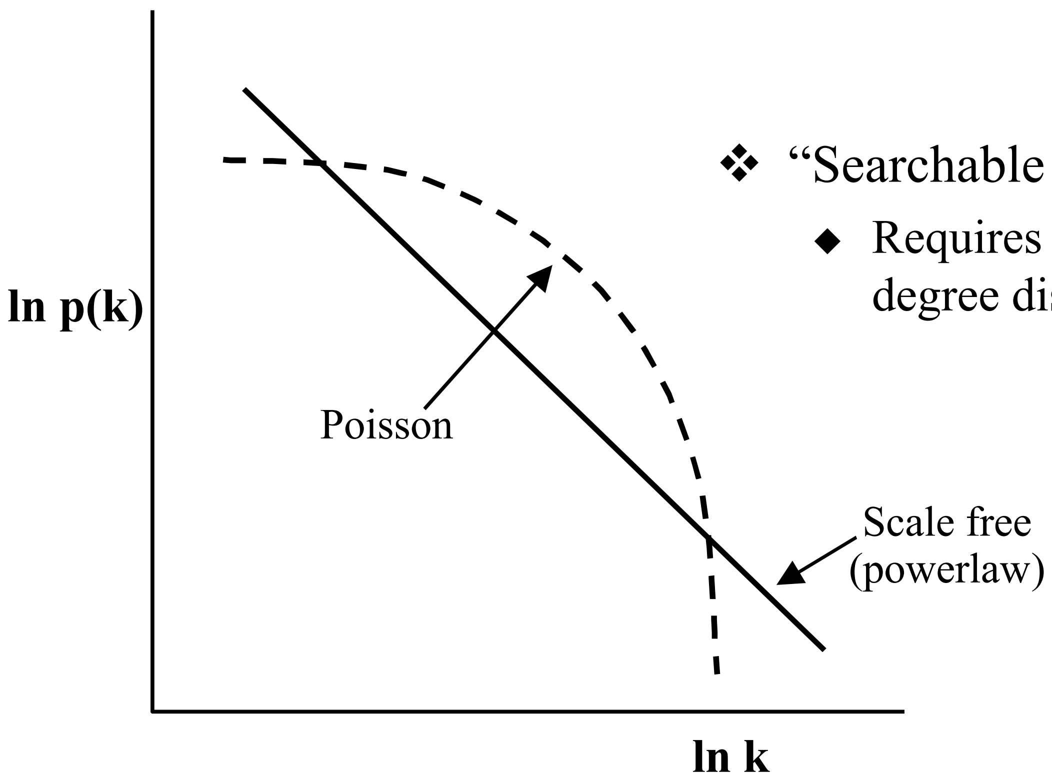


- Partition world into "phases"
- Picture as concentric rings with exponential radius: $R_i = 2^i$

- When γ is at critical value, network provides an equal ***number*** of random contacts at every scale

- "Kleinberg Condition" guarantees each phase requires only few steps
- Exponential radius ensures only few phases

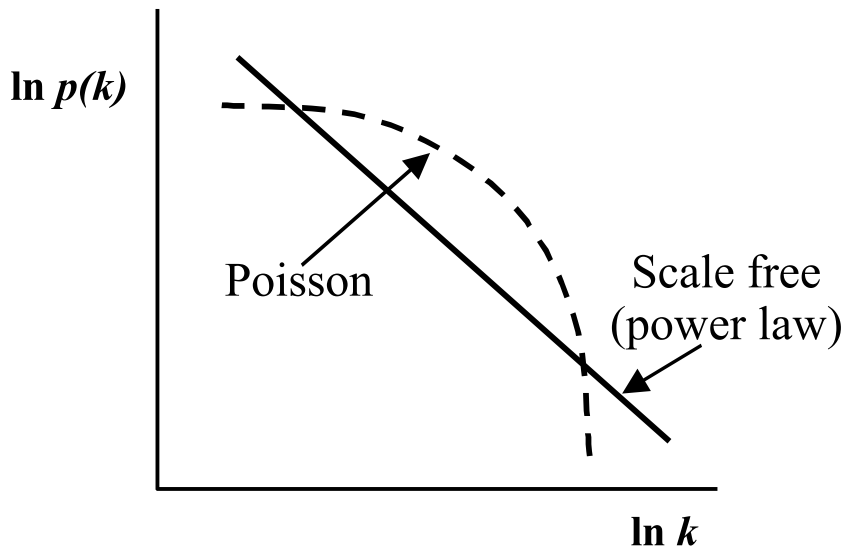
Another Attempt to Explain the "six degrees" phenomenon:



- ❖ "Searchable Small-World Networks"
  - ◆ Requires the assumption of a "scale-free degree distribution"

“Searchable Small-World Networks”

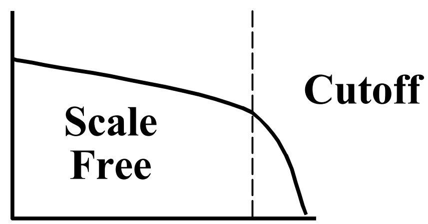
- ❖ “Scale-free degree distribution” implies the existence of a small fraction of highly connected “***hub nodes***”



A simple search algorithm
– ***direct message to your most connected neighbor*** –
quickly finds hubs and jumps around randomly until target is found.

Some Problems –

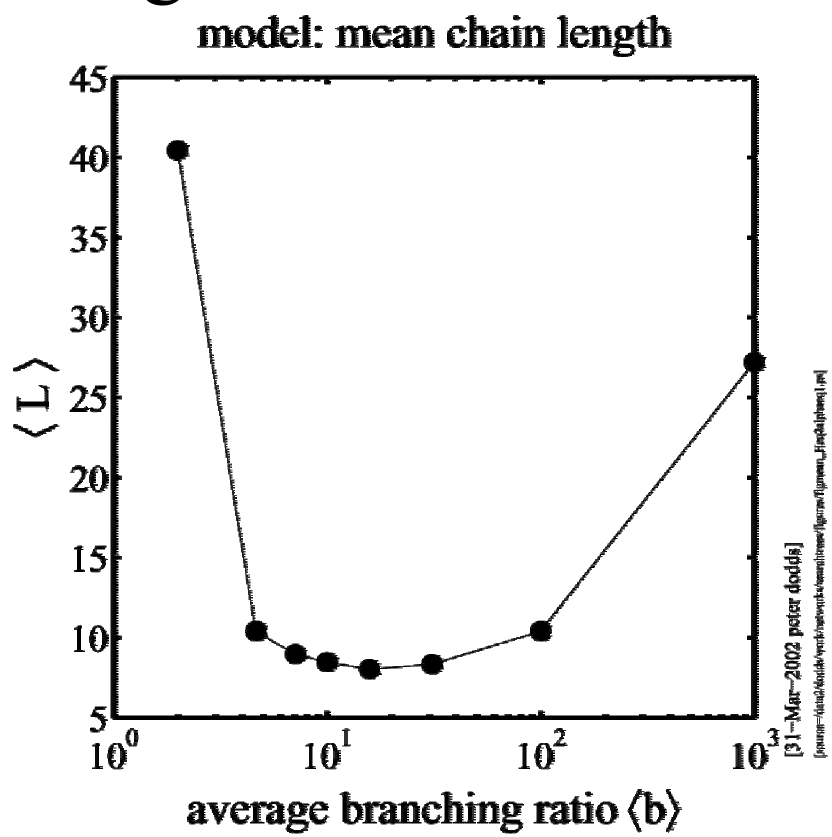
- There is no evidence that social networks are built on ***geometric lattices***
- There is no organizing mechanism to drive the parameter to the sweet spot – i.e., searchability is not generic
- No evidence that ***real*** social networks are ***scale-free*** (at the very least, they have cut-offs)



- Evidence on search algorithms shows that social characteristics like geography, occupation are important (not just degree)

The Model – Results

- ❖Mean Chain Length



Early History

- Anecdotal observation since at least 1920's (Karinthy)
- Academic Study commenced in 1950's
- Pool (political scientist) and Kochen (mathematician) became interested in mobilization of political power (Eventually published in *Social Networks I*, 1978)
- Their theoretical work attracted interest of the social psychologist, Stanley Milgram

Relevance of Small World Problem

- Role of social information in financial markets
- Efficient matching in labor markets
- Diffusion of ideas or innovations
- Robust architectures for organizations or redistribution networks (airlines, Internet)
- Efficiently searchable distributed databases