# Searching in Complex Networks.

23 November 2016

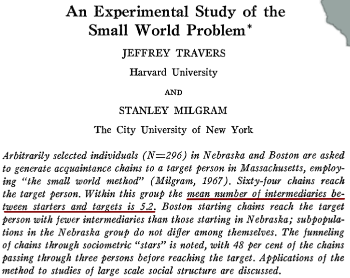
**Lecturer:** Dimitra Maoutsa

**Course:**

Network Dynamics & Complex Systems
Theoretical and Computational Tools
Georg-August-Universität Göttingen

## Six Degrees of Separation.

## Milgram's Small World Experiment.

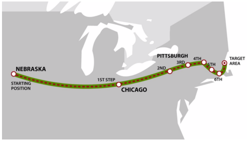


### An Experimental Study of the Small World Problem*

JEFFREY TRAVERS
Harvard University
AND
STANLEY MILGRAM
The City University of New York

*Arbitrarily selected individuals (N=296) in Nebraska and Boston are asked to generate acquaintance chains to a target person in Massachusetts, employing "the small world method" (Milgram, 1967). Sixty-four chains reach the target person. Within this group the mean number of intermediaries between starters and targets is 5.2. Boston starting chains reach the target person with fewer intermediaries than those starting in Nebraska; subpopulations in the Nebraska group do not differ among themselves. The funneling of chains through sociometric "stars" is noted, with 48 per cent of the chains passing through three persons before reaching the target. Applications of the method to studies of large scale social structure are discussed.*

Washington
Oregon
Montana
Idaho
North Dakota
Minnesota
South Dakota
Wisconsin
Michigan
Vermont
New Hampshire
Maine
New York
Massachusetts
Rhode Island
Connecticut
Wyoming
Nevada
California
Utah
Colorado
Nebraska
Iowa
Illinois
Pennsylvania
New Jersey
Ohio
Indiana
West Virginia
Delaware
Maryland
Virginia
Kansas
Missouri
Kentucky
North Carolina
Tennessee
Arizona
New Mexico
Oklahoma
Arkansas
South Carolina
Georgia
Mississippi
Alabama
Texas
Louisiana
Florida



## Modeling the Phenomenon.

- Why should there *exist* short chains of acquaintances linking together arbitrary pairs of strangers?

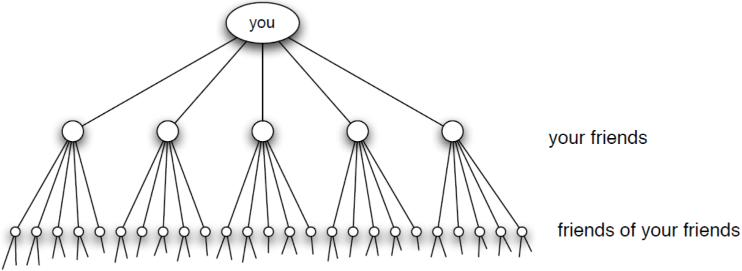
## Modeling the Phenomenon.

- Why should there *exist* short chains of acquaintances linking together arbitrary pairs of strangers?

- How could arbitrary pairs of strangers be able to *find* short chains of acquaintances that link them together?

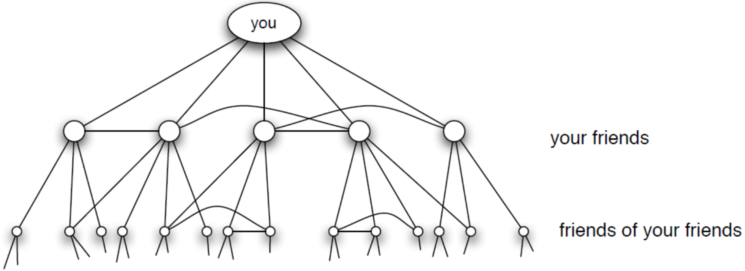
# Network Search.

- **Why should there *exist* short chains of acquaintances linking together arbitrary pairs of strangers?**



*Pure exponential growth produces a small world*

- **Why should there *exist* short chains of acquaintances linking together arbitrary pairs of strangers?**

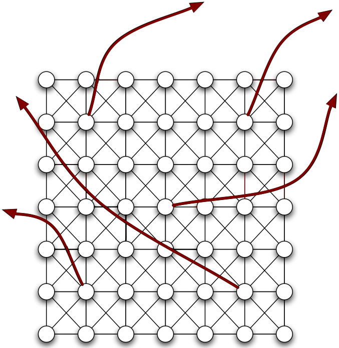


*Triadic closure reduces the growth rate*

## Watts Strogatz model.

Watts-Strogatz model captures naturally two basic social-network features:

- **Homophily.** [regular lattice]
  The principle that we connect to others who are like ourselves.
- **Weak Ties.** [shortcuts]
  The links to persons that connect us to parts of the network that would otherwise be far away.

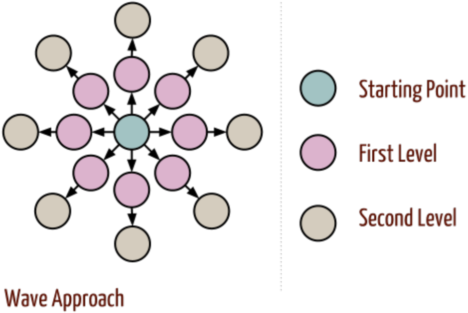


**Two dimensional Watts-Strogatz model.** Each model connects to all other nodes that lie within radius of up to $p$ grid steps away, while projecting also $q$ links to nodes selected uniformly at random from the entire grid.

## Collective Search.

- **How could arbitrary pairs of strangers be able to *find* short chains of acquaintances that link them together?**

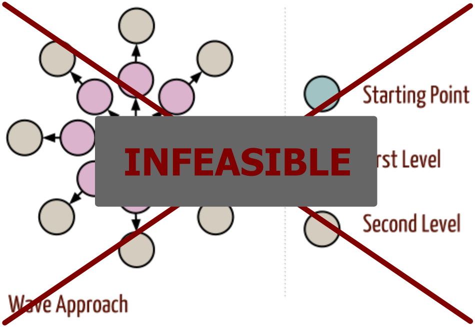
**Breadth-first Search.**



# Network Search.

- **How could arbitrary pairs of strangers be able to *find* short chains of acquaintances that link them together?**

**Breadth-first Search.**



**INFEASIBLE**

# Network Search.

- **How could arbitrary pairs of strangers be able to *find* short chains of acquaintances that link them together?**

  **Decentralised Search.**[1]
  Starting node $s$ and target node $t$.
  Each individual node has knowledge of:

  » its short- and long-range connections

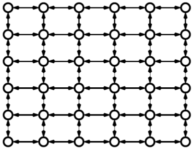
  » set of short-range connections of all nodes (i.e. the underlying grid structure)

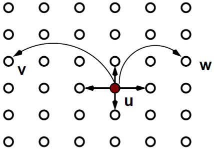
  » the location, on the lattice, of the target $t$.

  No information about the long-range ties of the other nodes.
  **Expected Delivery Time** $T$:expected number of steps from $s$ to $t$.

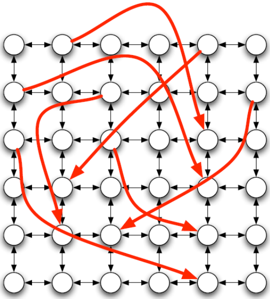
## A Model for Decentralised Search. (Kleinberg)

*Lattice points* in $n \times n$ square, $\{(i,j) : i \in \{1,2,\cdots,n\}, j \in \{1,2,\cdots,n\}\}$
*Lattice distance:* $d\left((i,j),(k,l)\right) = |k-i| + |l-j|$
*Local contacts:* node $u$ has directed edge to every other node within lattice distance $p \geq 1$.
*Long range contacts:* $u$ forms links to $q \geq 0$ nodes; each edge from $u$ has endpoint $v$ with prob. $\propto [d(u,v)]^{-r}$

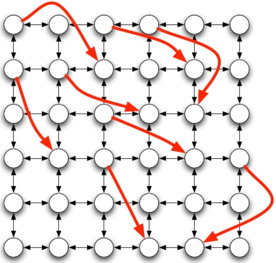




# Network Search.



(a) *A small clustering exponent*



(b) *A large clustering exponent*

**Clustering exponent $r$ determines the range of long range connections.**

## Network Search.

- $r = 0$: long-range end points are selected uniformly among the network nodes (as in WS).
- $r \gg 0$: long range links become more clustered in the vicinity of each node in the grid, not random enough for small-world.

**Is there an optimal operating point for rapid decentralised search?**

# Network Search.

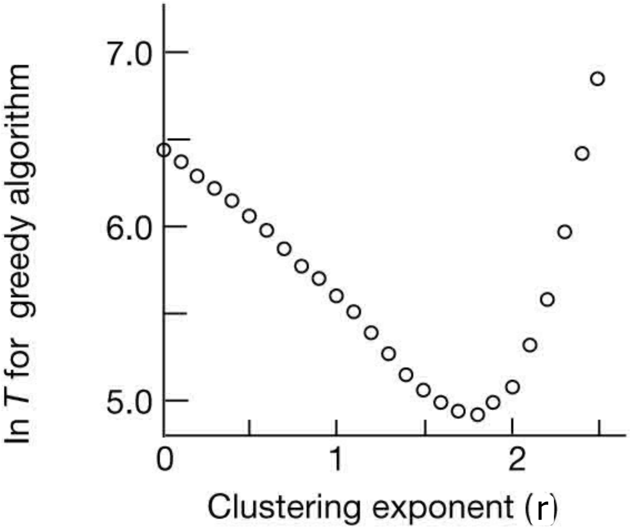
- $r = 0$: long-range end points are selected uniformly among the network nodes (as in WS).
- $r \gg 0$: long range links become more clustered in the vicinity of each node in the grid, not random enough for small-world.

**Is there an optimal operating point for rapid decentralised search?**

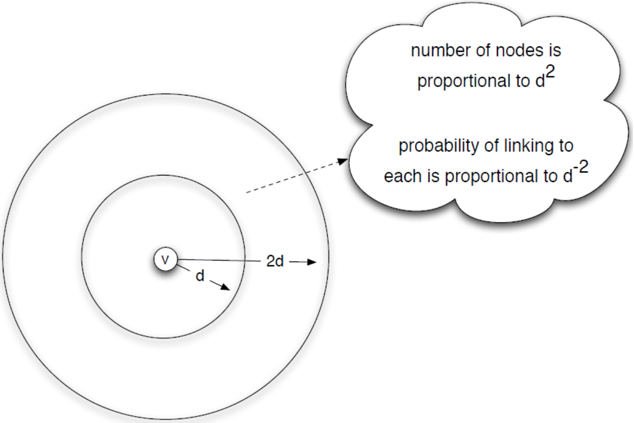
$$r = 2, \text{ for 2-dimensional grids} \tag{3}$$

$$r = D, \text{ for } D\text{-dimensional grids} \tag{4}$$

$$T \sim O(log^2 n)$$


7.0
6.0
5.0
ln T for greedy algorithm
0
1
2
Clustering exponent (r)

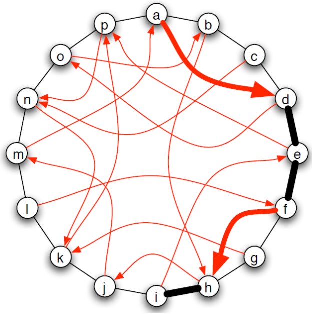
# Network Search.

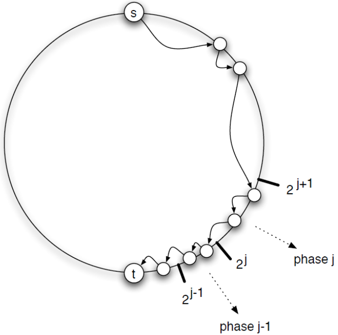

number of nodes is proportional to $d^2$
probability of linking to each is proportional to $d^{-2}$
v
d
2d

A node with several random shortcuts spanning different distance scales.

# Analysis of Decentralised Search.

## Optimal Exponent in One Dimension.





# Network Search.

To put a bound on $T$, we'll track how long it takes for the message to reduce its distance by factors of $2$ as it approaches the target $t$. Divide the vertices into phases according to their distance to $t$:

- Phase $0$: target $t$.
- Phase $1$: nodes directly adjacent to $t$.
- Phase $j$: distance to target is between $2^j$ and $2^{j+1}$.

Total search time may be written as a sum of times spent in each phase:

$$T = T_1 + T_2 + ... + T_{logn} \rightarrow E[T] = E[T_1 + ... + T_{logn}] = E[T_1] + ... + E[T_{logn}] \tag{5}$$

We'll show that $E[T_j]$ for each $j$ is at most $log\, n$.

- **Calculating the Normalizing Constant.**

$$P(v \to w) = \frac{d(v,w)^{-1}}{\sum_{v \neq w} d(v,w)^{-1}} = \frac{d(v,w)^{-1}}{Z}. \tag{6}$$

From $v$ we have $2$ nodes of distance $1$, $2$ of distance $2$ and so on, and assuming $n$ even, $1$ in distance $n/2$.
Therefore,

$$Z \leq 2\left(1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + ... + + \frac{1}{n/2}\right) \leq 2\left(1 + ln(\frac{n}{2})\right) \tag{7}$$

Considering $ln(x) \leq log_2(x)$,

$$Z \leq 2 + 2log_2(\frac{n}{2}) = 2log_2(n). \tag{8}$$

# Network Search.

Thus,

$$P(v \to w) \geq \frac{d(v,w)^{-1}}{2\,log_2(n)}. \tag{9}$$

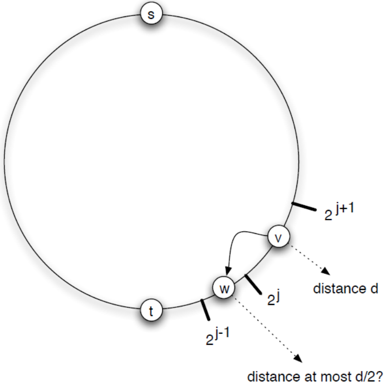
- **Calculating The Time Spent in One Phase**.

  Let us assume that the search is at some node $v$ which is at distance $d$ from target $t$ and $d \in [2^j, 2^{j+1}]$.
  The phase will finish at the point when the distance to $t$ falls under $2^j$.

a. Node $v$ has a long range connection at distance $\leq \frac{d}{2}$ from $t$. Let's calculate the probability of that: The set $I$ comprises the nodes with distance $\leq \frac{d}{2}$ from $t$. $|I| = d + 1$.
These points are at most at distance $\frac{3d}{2}$ from current node $v$, and thus their probability to be connected to $v$ (as long-range contact):

$$P(v \to w) \geq \frac{d(v,w)^{-1}}{2\,log_2(n)} \geq \frac{1}{2\,log_2(n)} \frac{1}{3d/2} = \frac{1}{3d\,log_2(n)}, w \in I \tag{10}$$

# Network Search.



Currently the search is at node $v$ at distance $d$ from target $t$ and $d \in [2^j, 2^{j+1}]$. The points within distance $\leq \frac{d}{2}$ from $t$, have mostly distance $\frac{3d}{2}$ from current node $v$.

The probability that *one* of the $(d+1)$ nodes in $I$ contacts $v$ is at least:

$$d\frac{1}{3d\,log_2(n)} = \frac{1}{3\,log_2(n)}. \tag{11}$$

Therefore at each step the phase $j$ has probability at least $\frac{1}{3\,log_2(n)}$ to terminate.

b. Node $v$ does node have a long-range connection below $\frac{d}{2}$ from $t$.

The search will continue within the same phase.

The probability that phase $j$ will run at least $i$ times (fail $i-1$) is:

$$Pr[T_j \geq i] = \left(1 - \frac{1}{3\,log_2(n)}\right)^{i-1}. \tag{12}$$

We may write for $T_j$

$$\begin{aligned} E[T_j] &= 1 \cdot Pr[T_j = 1] + 2 \cdot Pr[T_j = 2] + 3 \cdot Pr[T_j = 3] + ... \\ &= Pr[T_j \geq 1] + Pr[T_j \geq 2] + Pr[T_j \geq 3] + ... \\ &\leq 1 + \left(1 - \frac{1}{3\,log_2(n)}\right) + \left(1 - \frac{1}{3\,log_2(n)}\right)^2 + ... \end{aligned} \tag{13}$$

The above is a geometric sum with multiplier $\left(1 - \frac{1}{3\,log\,n}\right)$ and converges to

$$\frac{1}{1 - \left(1 - \frac{1}{3\,logn}\right)} = 3\,log\,n \tag{14}$$

Therefore $E[T_j] \leq 3\,log\,n \Longrightarrow E[T] \leq 3\,(log\,n)^2$.

# Network Search.

If there is no underlying lattice, how can search be efficient?

- What does *closest* mean?

- We need some sort of *distance*.

Incorporate **Identity**.

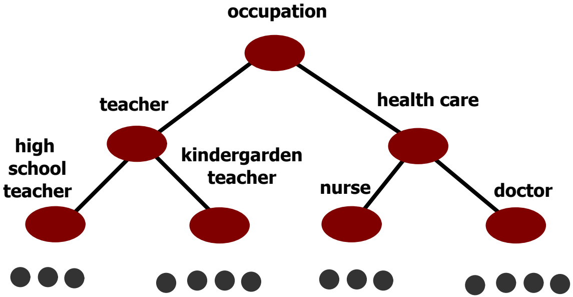
(Watts, Dodds and Newman [2])

Identity is formed from attributes such as:

- georgraphic location
- occupation
- religious beliefs
- recreational activities

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

# Network Search.

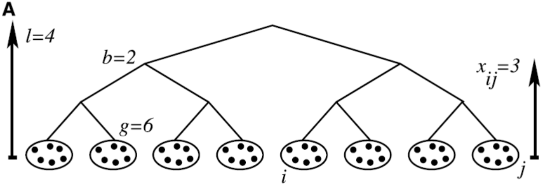


Individuals partition the world (i.e. identities of others) hierarchically.

Group hierarchies captured by branching trees of $l$ layers and branching ratio $b$.

# Network Search.

Distance between two individuals $i$ and $j$, $x_{i,j}$, defined as the lowest common ancestor.



# Network Search.

- Hierarchies reflect the social distance!They do not represent the actual network connections.

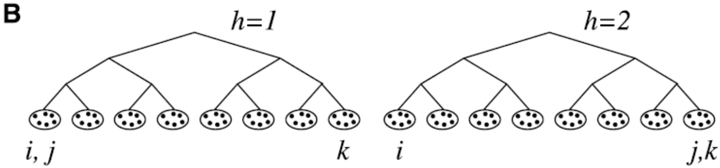
- However, individuals are more likely to know each other the closer they are in the hierarchy.

- Construct $z$ links for each node with prob. of connection between $i$ and $j$:

$$p_{i,j} = ce^{-ax_{i,j}} \tag{15}$$

  - a: measures homophily
    - $a = 0$ random connections - hierarchy is irrelevant.
    - $a \gg 0$ local connections.
  - c: normalisation constant

# Network Search.



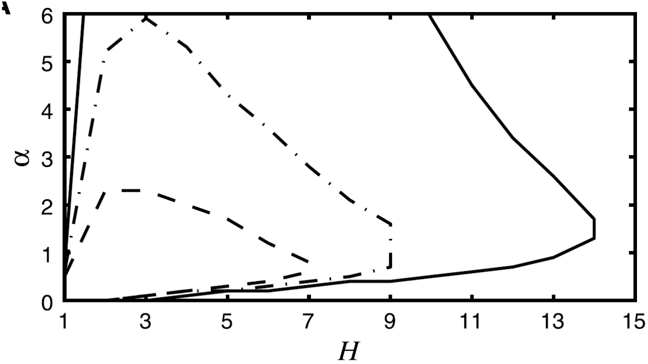
But there may be multiple independent hierarchies (e.g. occupation and geography).

Social Distance $y_{i,j} = min_H x_{i,j}$, where $H$: # of hierarchies.

## Searching in Watts-Dodds-Newmann model.

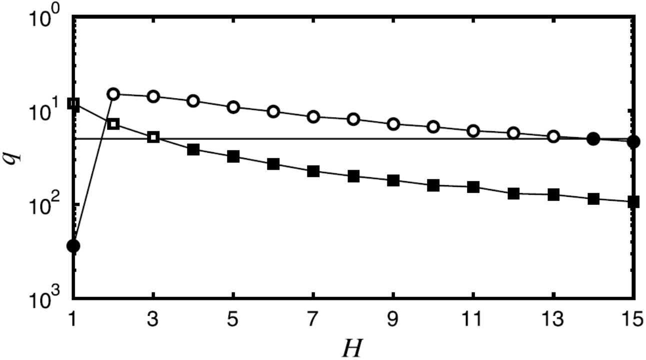
- Individuals know identities of themselves, their friends and the target.
- Individuals may estimate the social distance $y_{it}$ between their friends and the target.
- Perform a greedy search and allow searches to fail with probability $p = 0.25$ at each step [3].
- Network is searchable if a fraction $r$ of searches reach the target $q = \langle (1-p)^L \rangle \geq r$, where $L$ the average path length in the network.

# Network Search.



**Search success for different network sizes.** Boundaries of regions in the $H - \alpha$ space, where searchable networks exist [$N = 102400$ (solid), $N = 204800$ (dot-dash), and $N = 409600$ (dash)].

Network Search.



**Probability of successful search** when $\alpha = 0$ (squares) and $a = 2$ (circles). Open symbols indicate searchable networks.

# Network Search.

- Searchable networks occupy a broad range of parameter space $(\alpha, H)$ with almost all searchable networks having $\alpha > 0$ and $H > 1$.
- Increasing group dimension beyond $H = 1$ yields a dramatic increase in search success, but the improvement is lost as $H$ increases further.

## Overview.

- The existence of short paths does not guarantee that they may be identified without knowledge of the global network structure.
- Kleinberg showed that networks may be designed in such a way to allow for rapid search with greedy algorithms based on local information.
- Watts-Dodds-Newmann model extends the idea of searchable networks to socially relevant settings.
- Group structure may facilitate decentralized search using social ties.

# References

[1] Jon M Kleinberg. Navigation in a small world. *Nature*, 406(6798):845–845, 2000.

[2] Duncan J Watts, Peter Sheridan Dodds, and Mark EJ Newman. Identity and search in social networks. *science*, 296(5571):1302–1305, 2002.

[3] Jeffrey Travers and Stanley Milgram. An experimental study of the small world problem. *Sociometry*, pages 425–443, 1969.

# Questions?