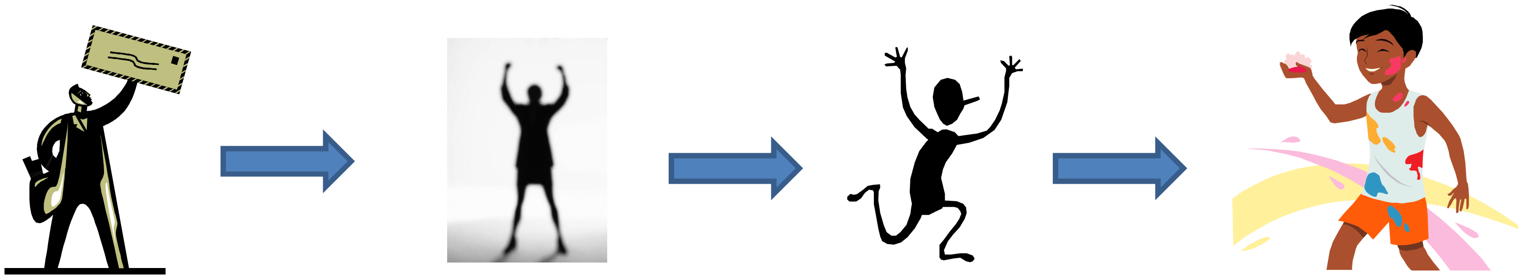
Small World Phenomenon

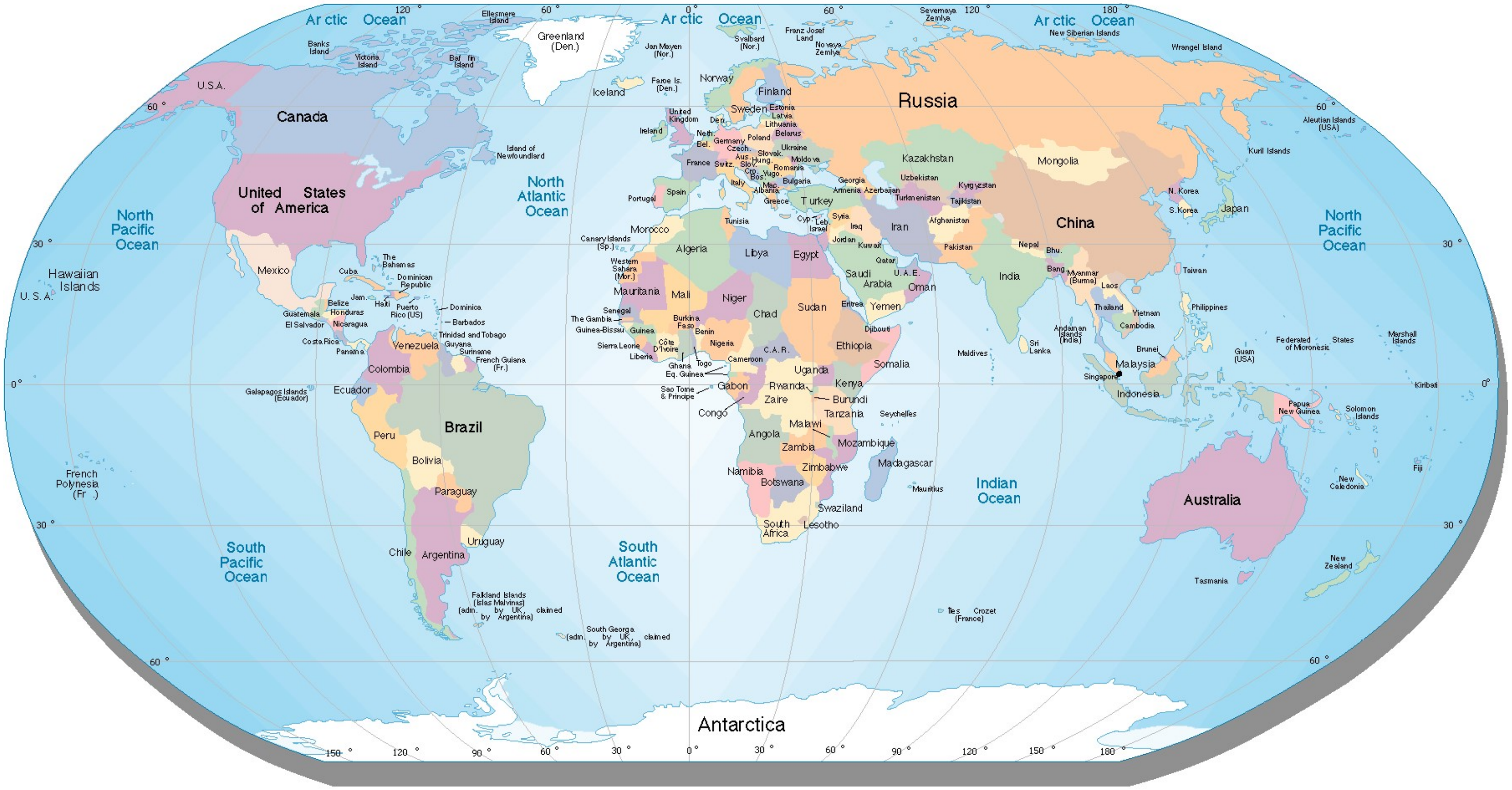
Six Degrees of Separation



- Milgram asked randomly chosen “starter” individuals to each try forwarding a letter to a designated “target” person.
- The target’s name, address, occupation and some personal information are provide.
- Each participant could only advance the letter by forwarding it to a single acquaintance that he or she knew.
- A third of the letters arrived at the target.
- In a median of six steps.

- Killworth and Bernard studied the strategies that people employ for choosing how to forward a message toward a target.
  - a mixture of primarily geographic and occupational features being used, with different features being favored depending on the characteristics of the target in relation to the sender.

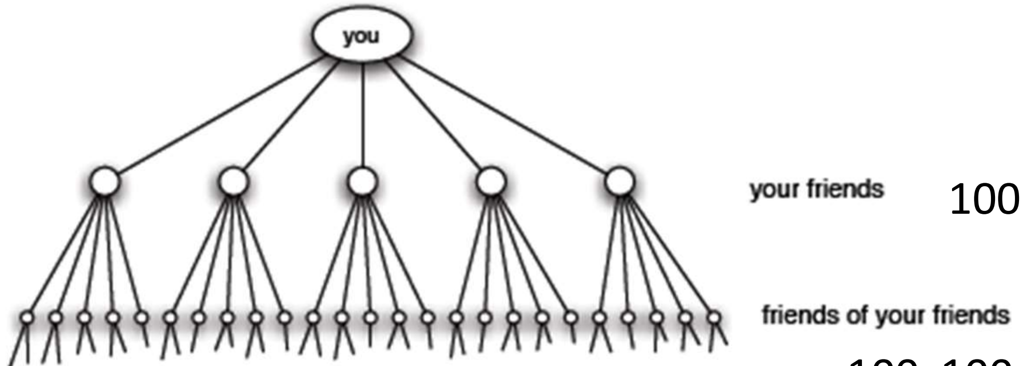
Existence of short paths



- Two striking facts
  - Short paths are there in abundance.
  - People, acting without any sort of global "map" of the network, are effective at collectively finding these short paths.
- Do you expect the same happen on social networks ?

Structure and Randomness

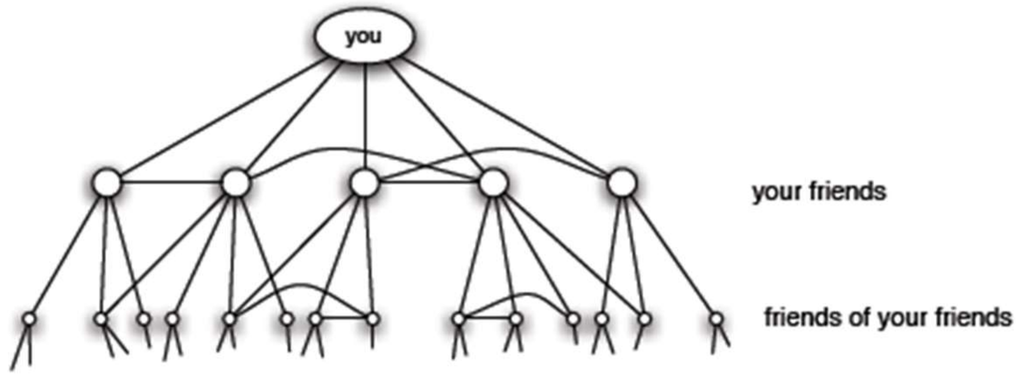
Each friend is linked to 100 new people



100 friends

100x100 friends

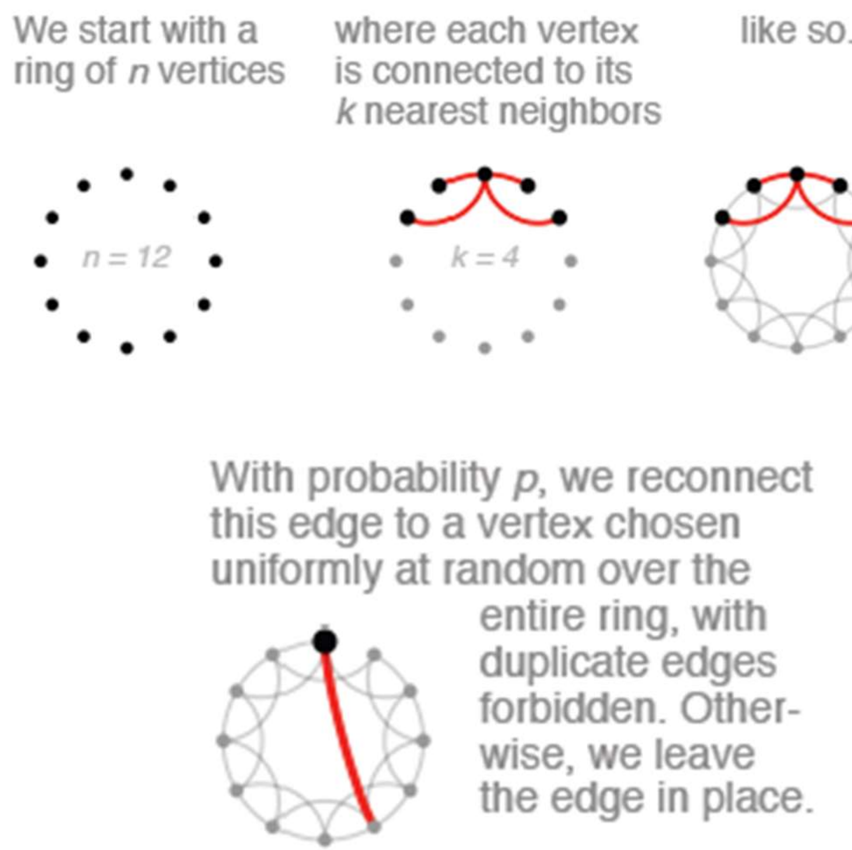
Triadic closure



Highly clustered individuals, so why do shortest paths exist ?

Watts-Strogatz model (Ring)

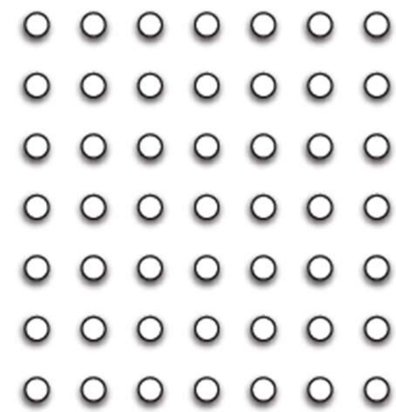
- Two types of links
  - Homophily
    - We connect to others who are like ourselves.
    - Many triangles
  - Weak ties
    - The links to acquaintances that connect us to parts of the network that would otherwise be far apart.



Watts-Strogatz model (Grid)

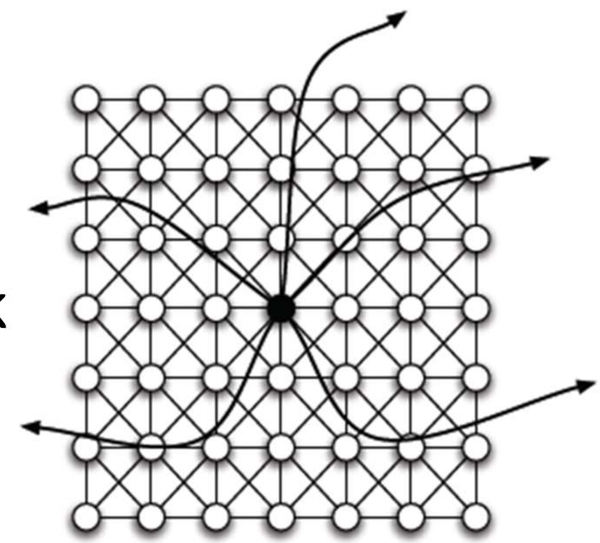
Everyone lives on a two-dimensional grid

- Two nodes are one grid-step apart if they are directly adjacent to each other.

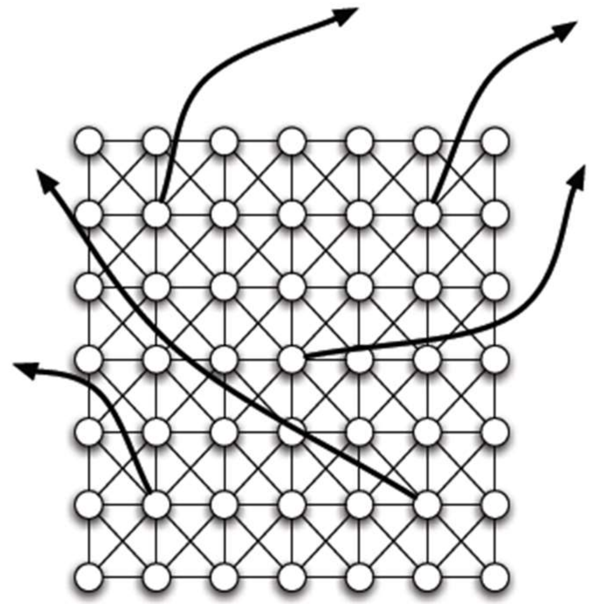


Two types of links

- Homophily : each node form a link to all other nodes that lie within a radius of up to r grid steps away.
- Weak ties : each node also forms a link to k other nodes selected uniformly at random from the grid (very far apart).

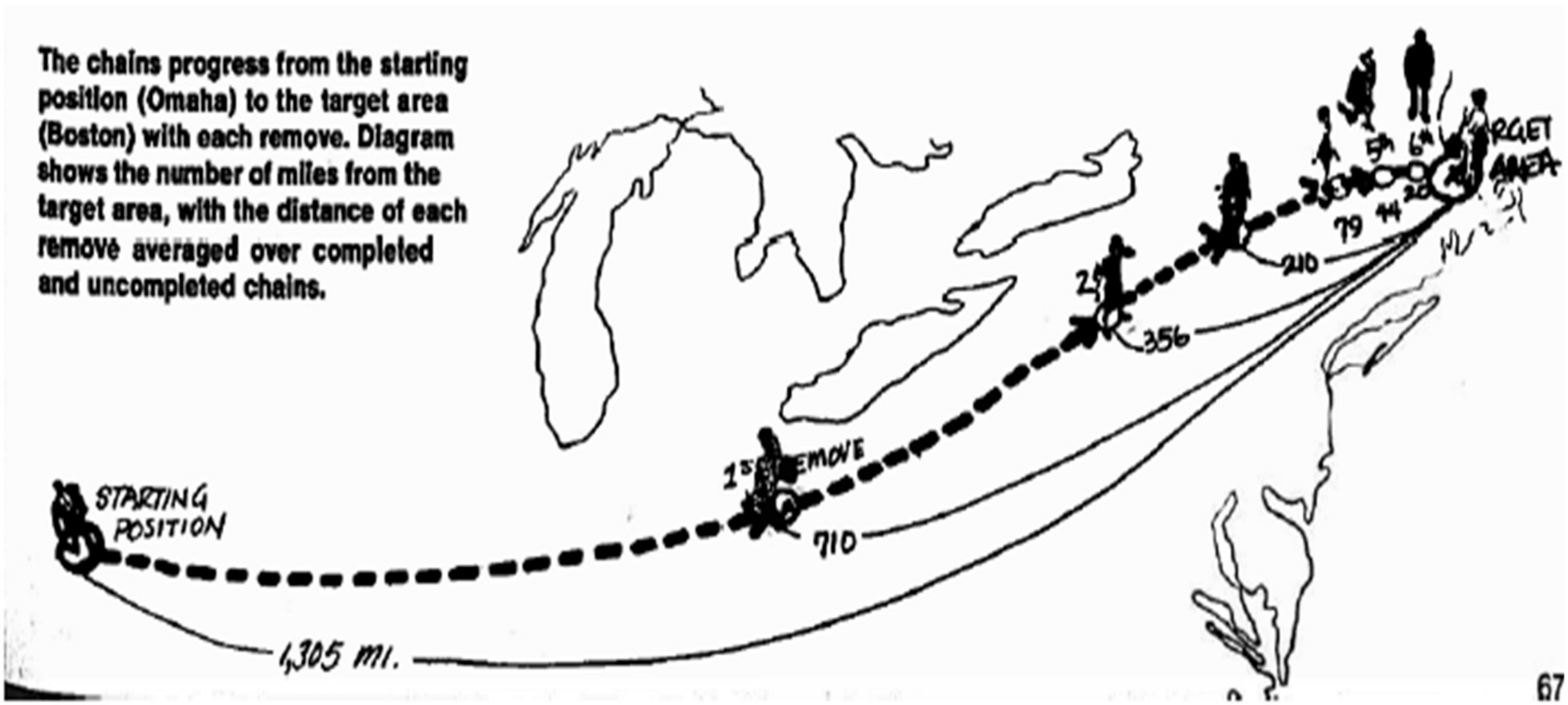


- The combinations
  - weak ties
    - reach many people in a small no. steps
  - mainly homophilous links and few weak ties
    - many triangles
    - If only one out of every k nodes is allowed to have a single random friend, this is enough to make the world small.



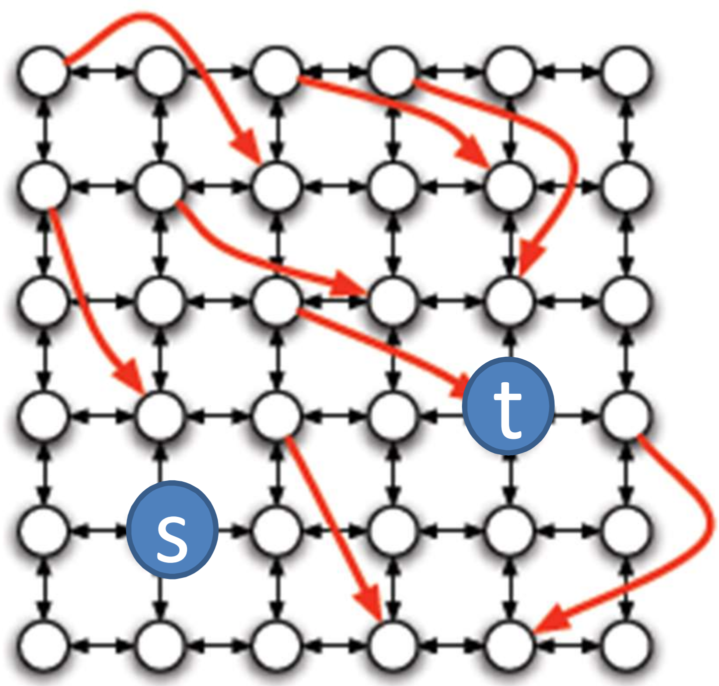
Decentralized Search

- Milgram small-world experiment



A model for decentralized search

- Starting with the grid-based model of Watts and Strogatz
- Node s has to forward a message to node t passing it along edges of the network.
- s knows the location of t but s does not know the random edges out of any nodes other than itself.
- Each intermediate nodes has this partial information as well.

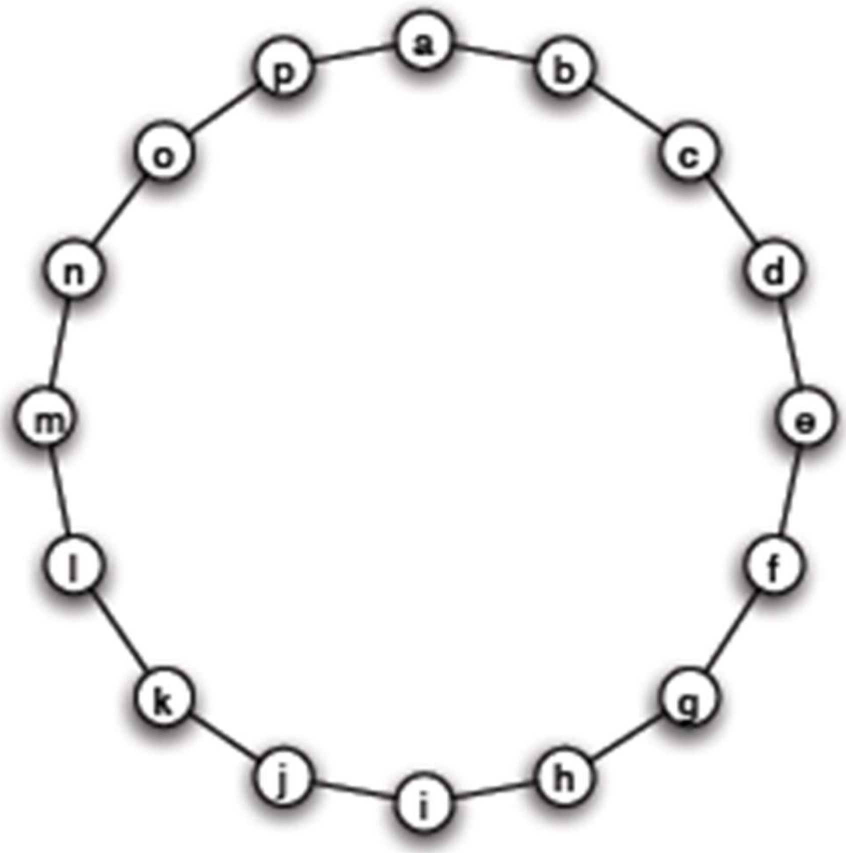


A model for decentralized search

- We will evaluate different search procedures according to their delivery time.
- Require a large number of steps to reach a target — much larger than the true length of the shortest path.

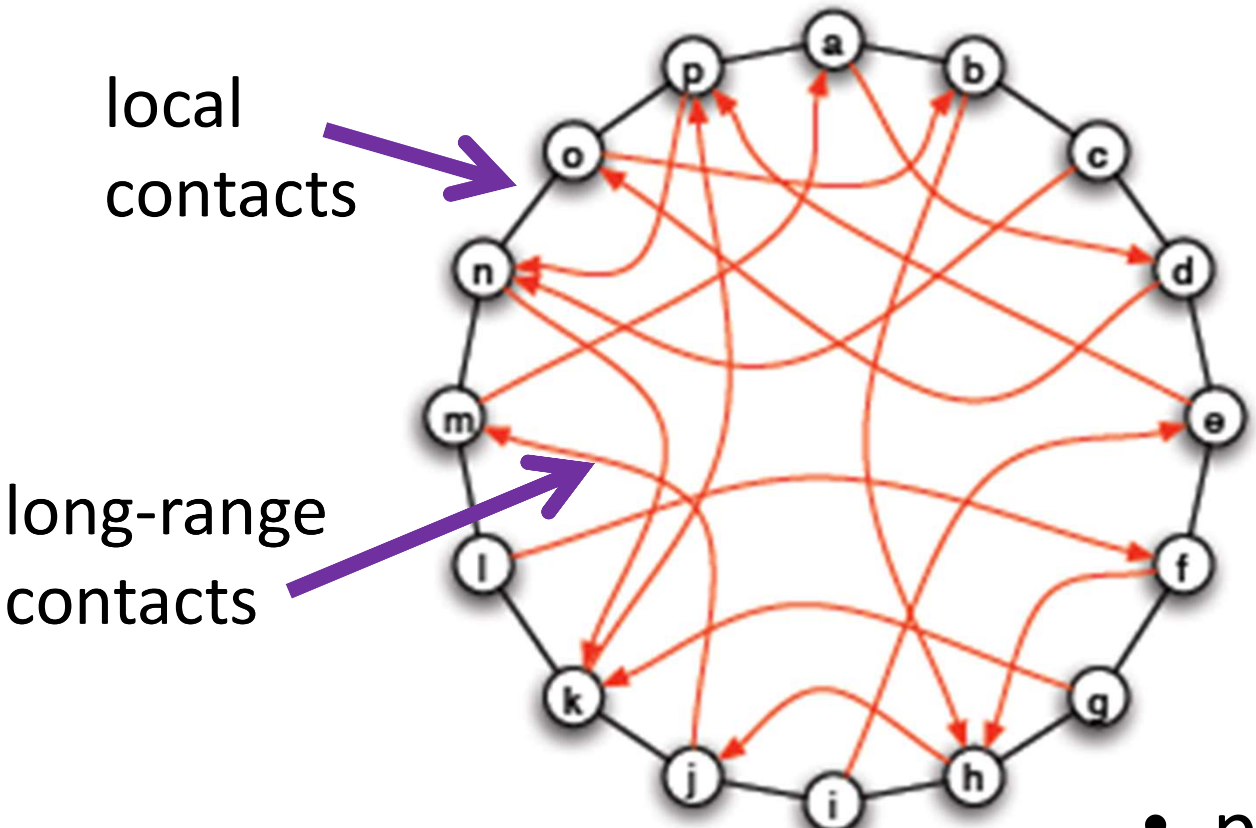
Analysis of Decentralized Search

- A set of n nodes arranging on a 1D ring



- directed edges to two adjacent nodes

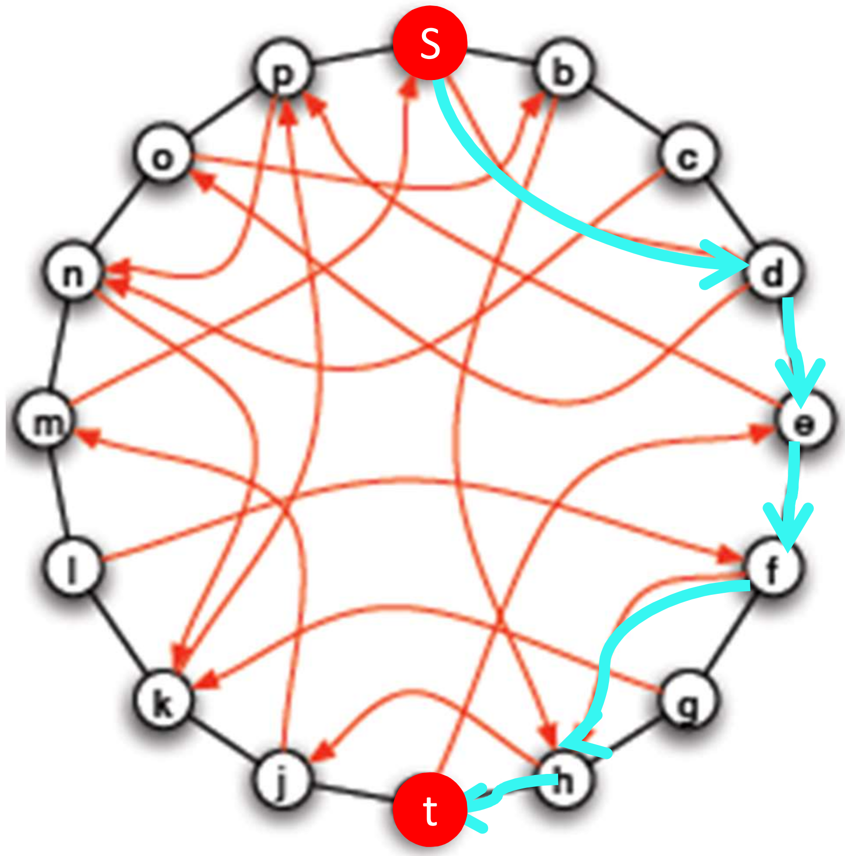
- A set of n nodes arranging on a 1D ring



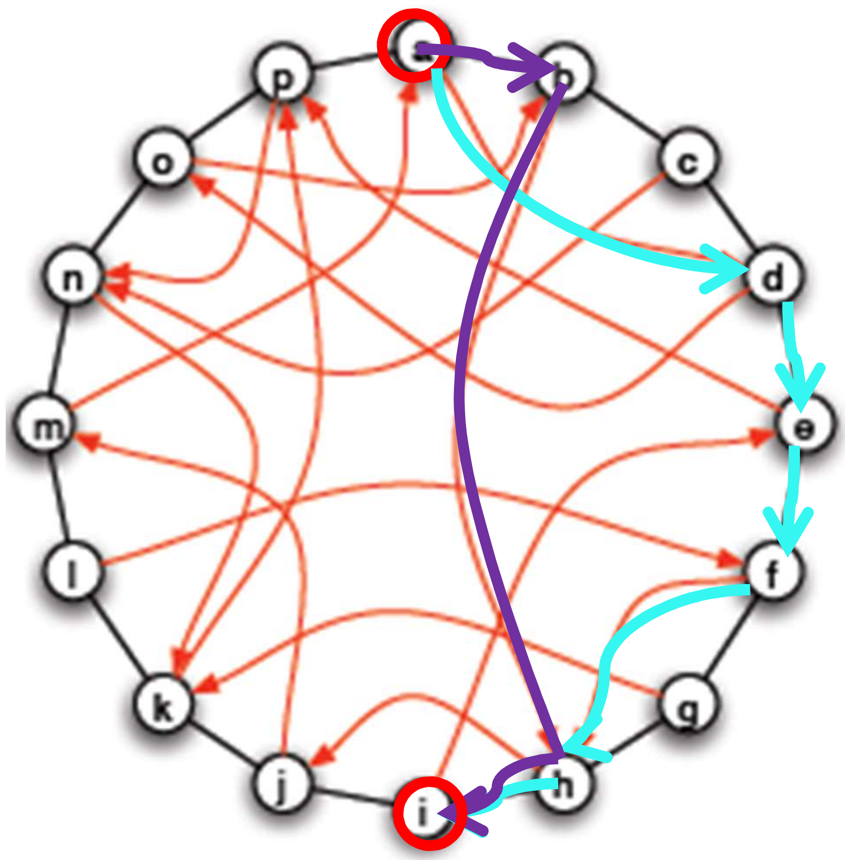
- directed edges to two adjacent nodes
- a single directed edge to some other node
- prob $\propto d(v,w)^{-1}$

Myopic Search

- start node s → target node t



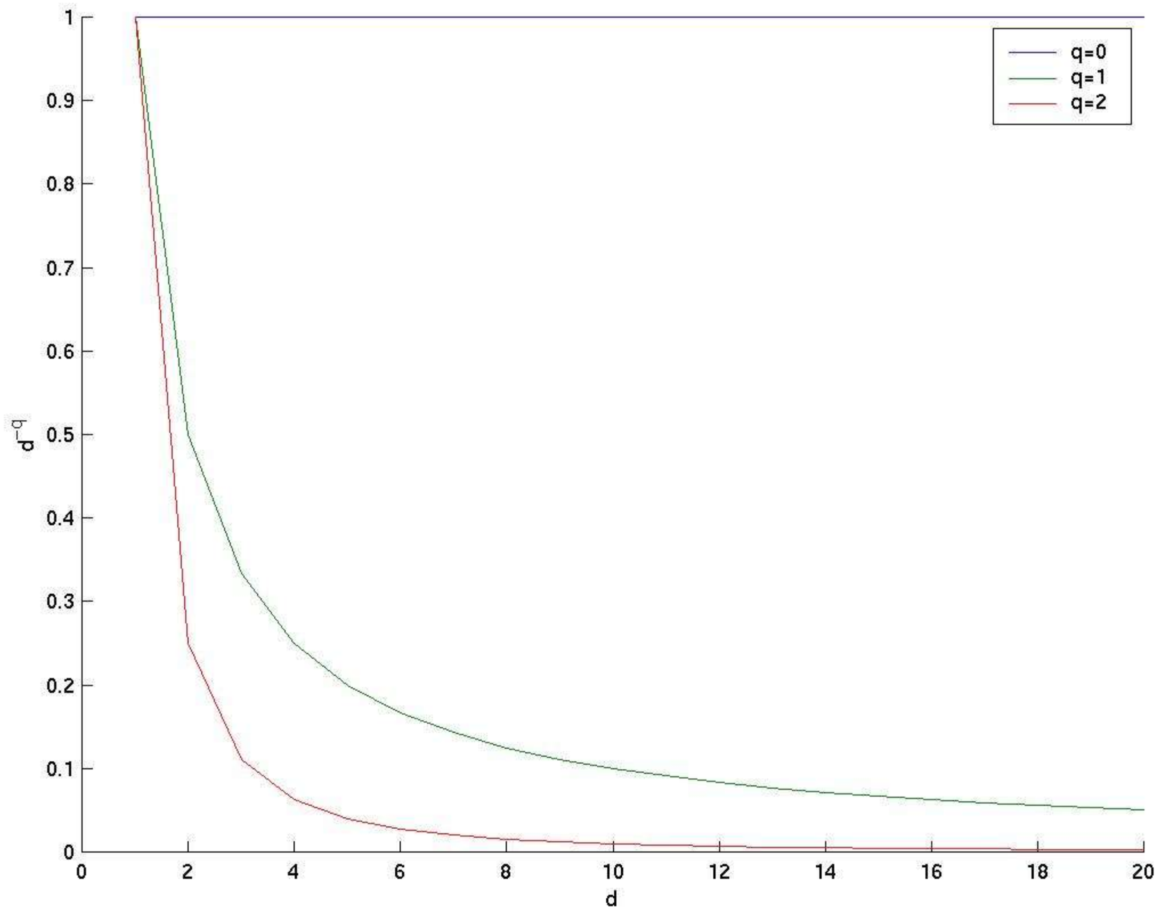
the shortest path from a to i



Generalizing the network model

- Nodes on a grid and each node has edges to each other node within r grid steps.
- Each of its k random edges is generated in a way that decays with distance
  - Let d(v, w) denote the number of grid steps between nodes v and w.
  - Probability of an edge is proportional to $d(v, w)^{-q}$

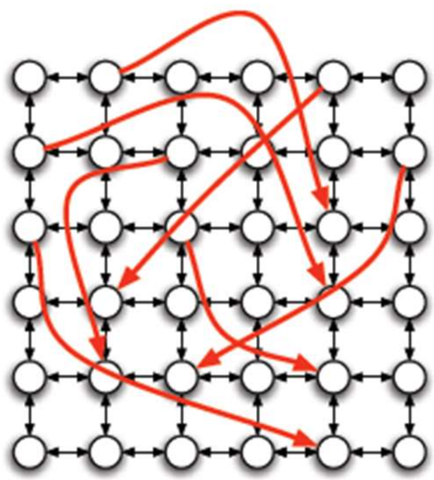
Graph of d^{-q}

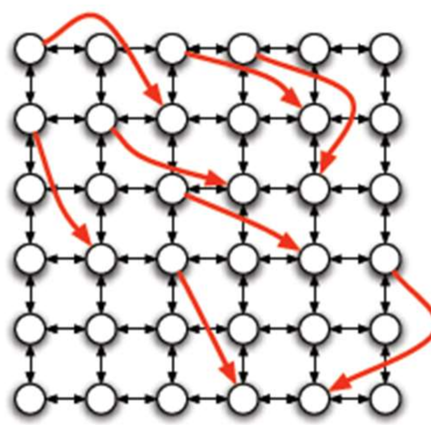


Generalizing the network model

- q=0, the links are chosen uniformly at random
- Small q, the long-range links are "too random"
- Large q, the long-range links are "not random enough,"

Small q, more long distances edges

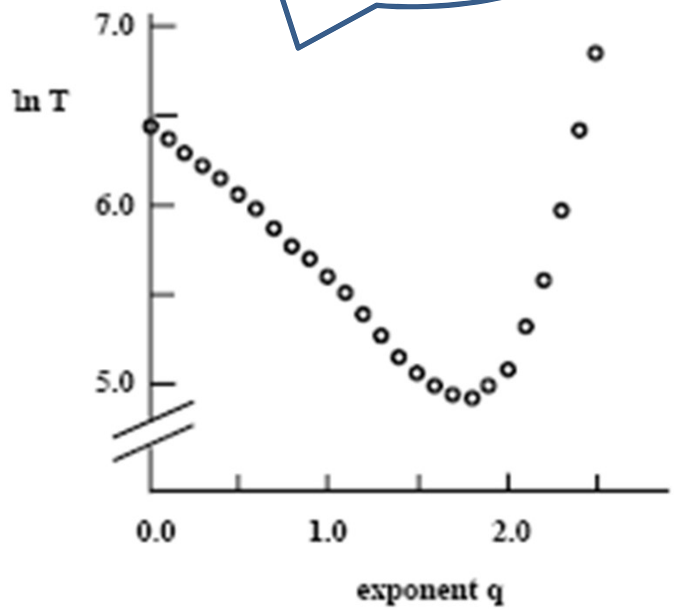




Large q, more short distances edges

- Decentralized search is most efficient when q = 2 (so that random links follow an inverse-square distribution).
- Decentralized search has about the same efficiency on networks of hundreds million nodes across all exponents q between 1.5 and 2.0.

Simulation of decentralized search with 400 million nodes



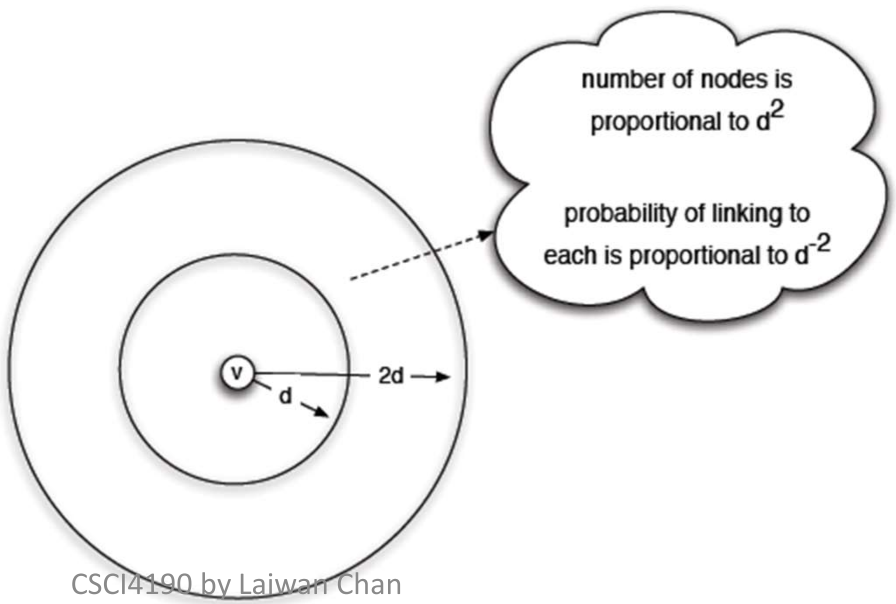
Efficiency of Decentralized Search

- Kleinberg 2000
  - nxn grid, $\alpha_0, \alpha_2, \alpha_r$ constants
  - Direct link to nodes within r grid steps away.
  - Prob. of a long range edge : $d(v, w)^{-q}$
- When q=0, the expected delivery time is at least
  $\alpha_0 n^{2/3}$
- When $0 \le q < 2$, the expected delivery time is at least
  $\alpha_r n^{(2-q)/3}$
- When q>2, the expected delivery time is at least
  $\alpha_r n^{(q-2)/(q-1)}$
- When q = 2 and r =1, the expected delivery time is at most
  $\alpha_2 (\log n)^2$

A Rough Calculation Motivating the Inverse-Square Network

- A node v in the network
- Considering the group of nodes lying at distances between d and 2d from v
- The probability that a random edge links into any node in this group is approximately independent of the value of d.

When q=2,
same probability,
regardless of the distance



Geographic Data on Friendship (Liben-Nowell et al)

- The population density of the LiveJournal network, a blogging site with 500,000 users
- Users provided links to their friends.
- The population density of the users is extremely non-uniform.
- Distance based model

  → Rank-based model