Small World Phenomenon

Advanced Social Computing

Department of Computer Science
University of Massachusetts, Lowell
Fall 2020

Hadi Amiri
hadi@cs.uml.edu

Announcement

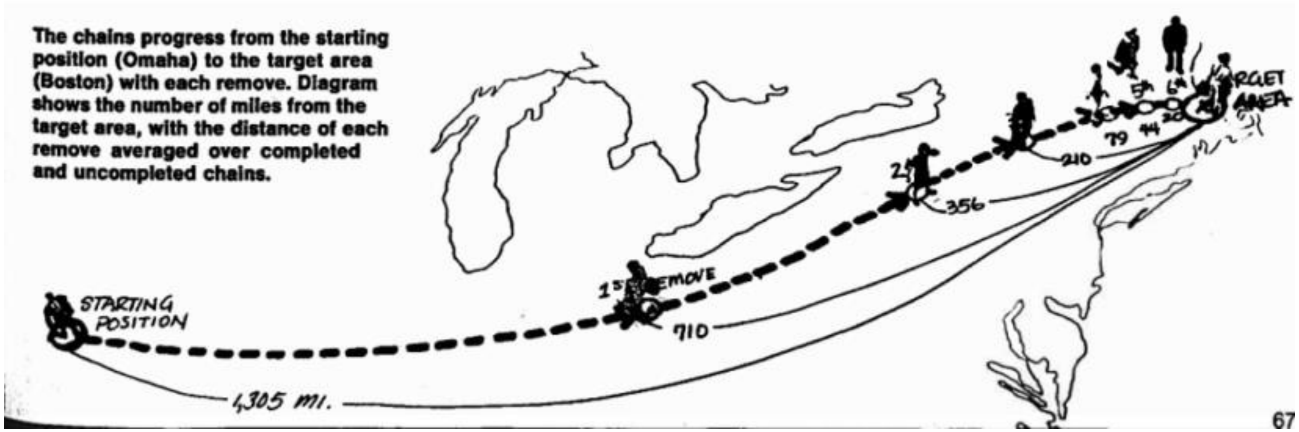
- **HW4 out**
  - Due Date: 10/14, 3:30 PM
- **Midterm Exam**
  - Date: 10/14, 3:30 - 6:20 PM
  - Open book
- **Course Survey**

Lecture Topics

- **Six Degree of Separation**
- The Watts-Strogatz Model
- Decentralized Search

Six Degree of Separation

- Small World Phenomenon
  - Stanley Milgram (1967):
    - People were asked to forward a letter to a designated target person in Sharon, MA!
      - They have target's name, address, occupation, etc.
    - Senders could only advance the letter by forwarding to contacts they knew on first-name basis
    - 33% of the letters arrived, **Median: 6 steps**



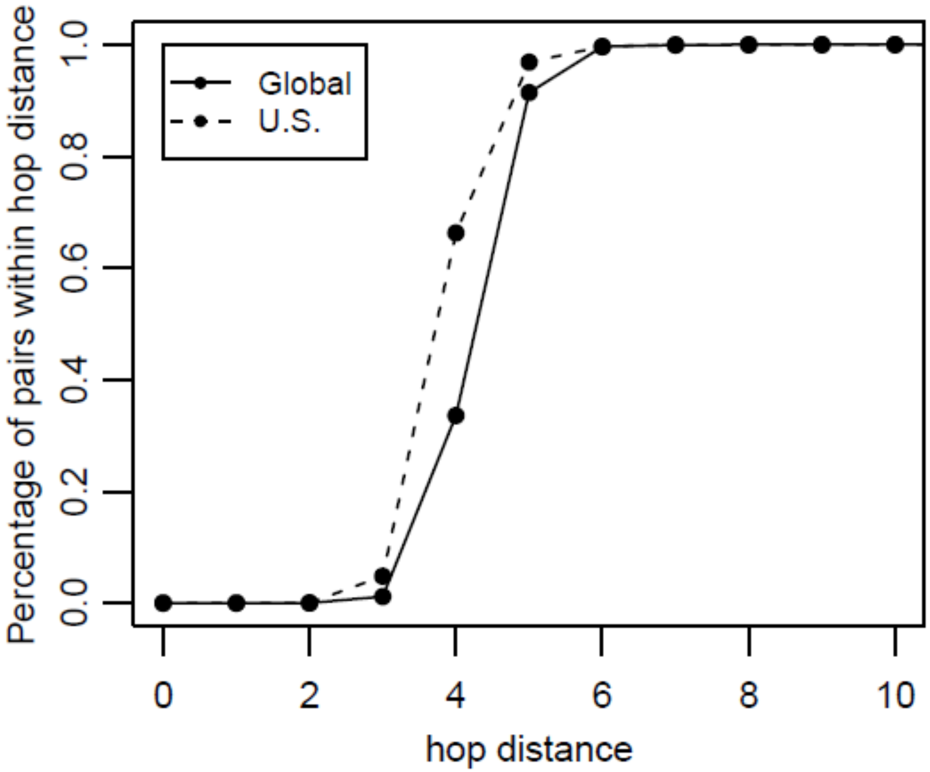
Six Degree of Separation

- Small World Phenomenon
  - Stanley Milgram (1967):
    - People were asked to forward a letter to a designated target person in Sharon, MA!
      - They have target's name, address, occupation, etc.
    - Senders could only advance the letter by forwarding to contacts they knew on first-name basis
    - 33% of the letters arrived, **Median: 6 steps**

- Focused search
- More targeted than information cascades and diffusion.

Six Degree of Separation- Cnt.

- Small World Phenomenon



Global
92.0%: within 5 degrees,
99.6%: within six degrees.

U.S. only
96.0%: within 5 degrees,
99.7%: within six degrees.

Figure 2. Diameter. The neighborhood function $N(h)$ showing the percentage of user pairs that are within h hops of each other. The average distance between users on Facebook in May 2011 was 4.7, while the average distance within the U.S. at the same time was 4.3.

Six Degree of Separation- Cnt.

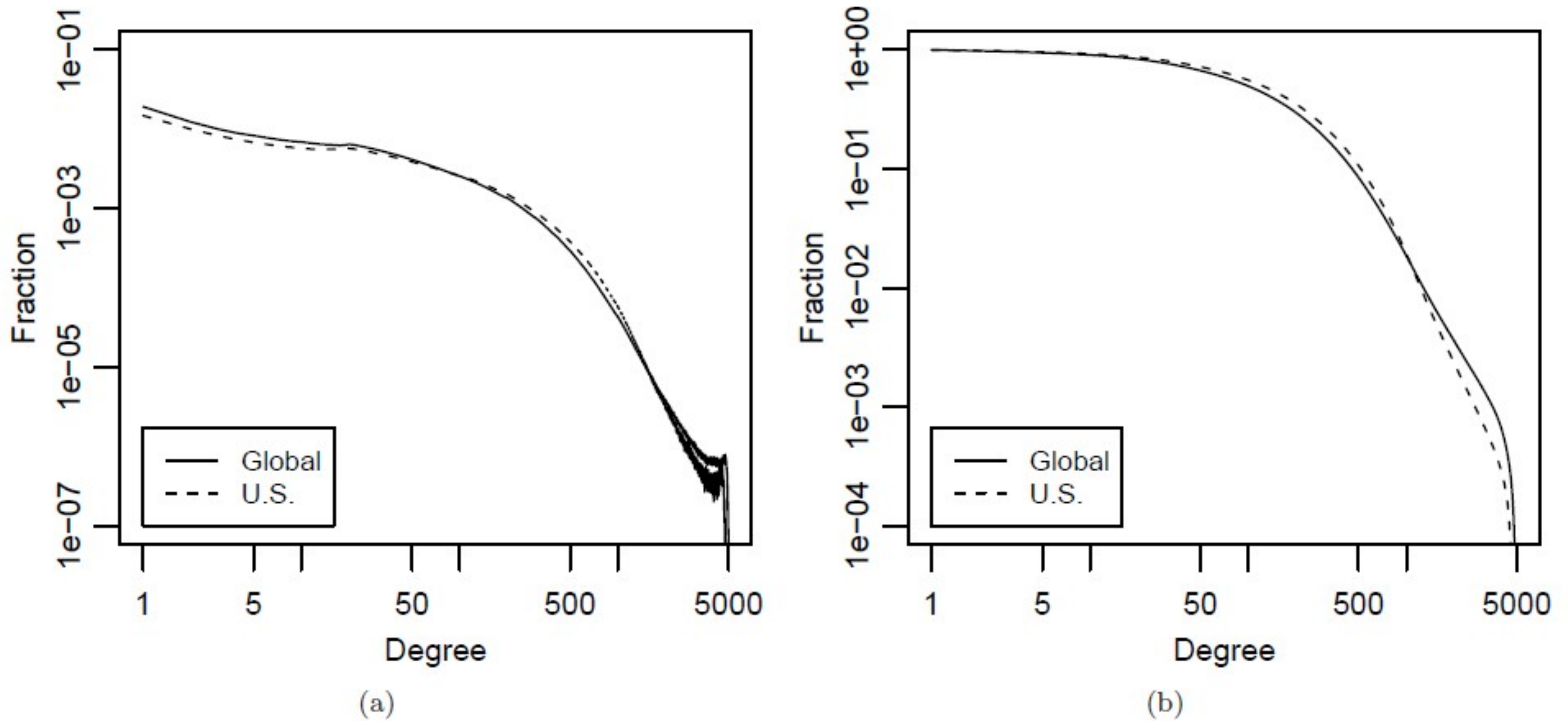


Figure 1. Degree distribution p_k. (a) The fraction of users with degree k for both the global and U.S. population of Facebook users. (b) The complementary cumulative distribution function (CCDF). The CCDF at degree k measures the fraction of users who have degree k or greater and in terms of the degree distribution is $\sum_{k' \geq k} p_{k'}$. For the U.S., the degree measures the number of friends also from the United States.

Six Degree of Separation- Cnt.

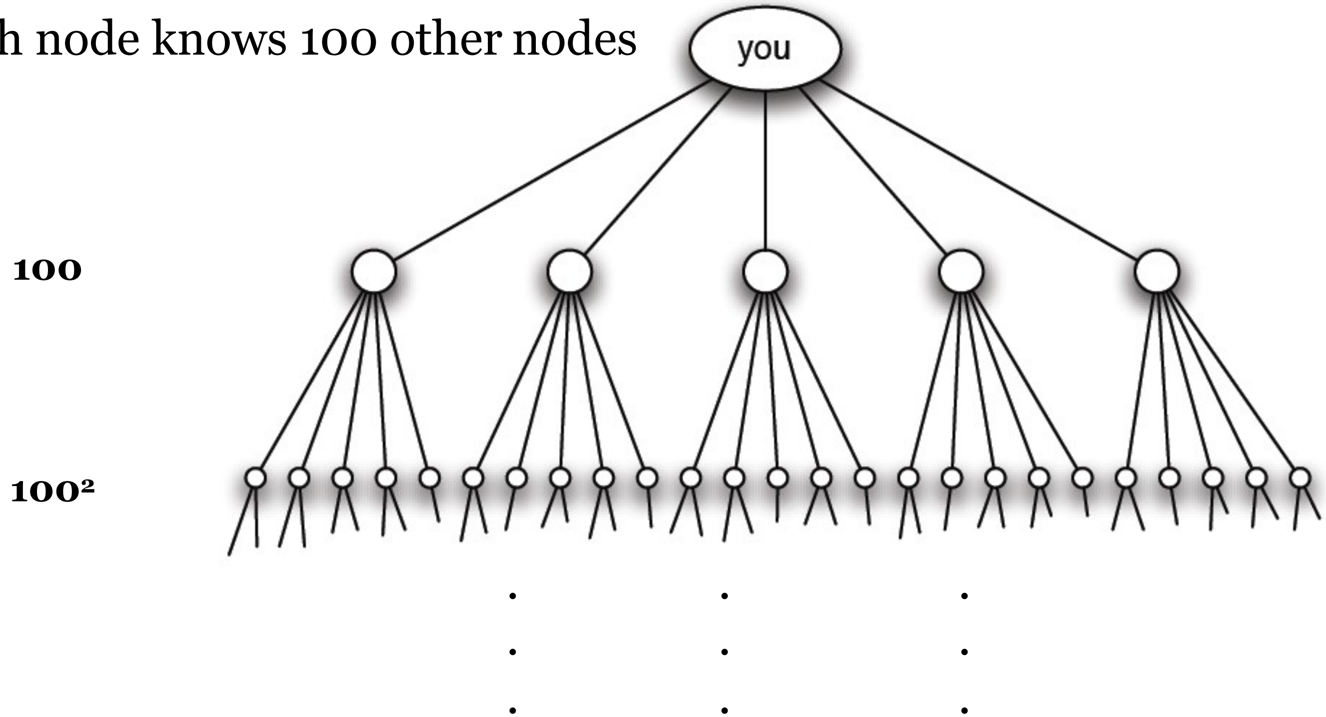
- Two facts about social networks:
  1. Short paths exist in abundance, and
  2. People are effective at collectively finding these short paths.

- We discuss models for the above principles!

Six Degree of Separation- Cnt.

- Existence of Shortest Paths / Small World

Each node knows 100 other nodes

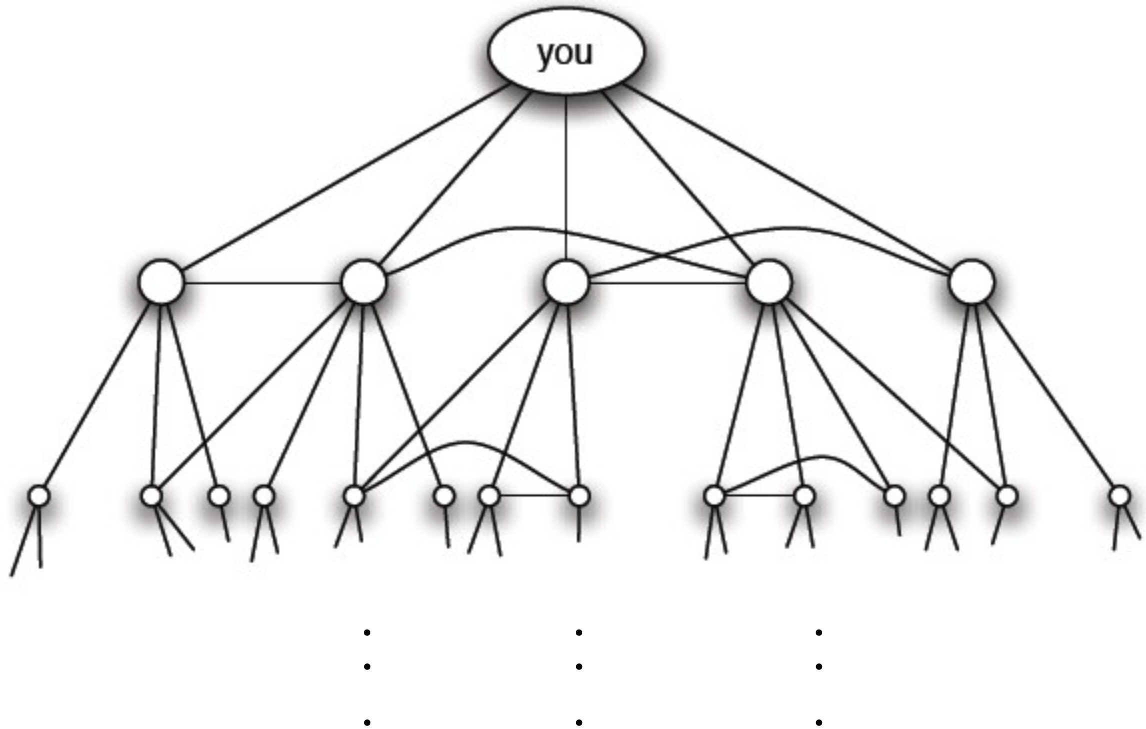


100 million after four steps,

10 billion after five steps

Six Degree of Separation- Cnt.

- Existence of Shortest Paths / Small World



- Triadic closure: social networks abound with triangles, many common friends
- Triadic closure reduces the growth rate and limits the number of people you can reach by following short paths

Lecture Topics

- Six Degree of Separation
- **The Watts-Strogatz Model**
- Decentralized Search

Watts-Strogatz Model

- A simple model that makes the word small by exhibiting the following features:
  - Many closed triads,
  - Many short paths.

Watts-Strogatz Model- Cnt.

- Watts-Strogatz (1998)
  - A model that combines two basic network ideas:
    - **Homophily**
      - The principle that we connect to others who are like ourselves.
      - Homophily creates many triangles / closed triads
    - **Weak Ties**
      - Links to acquaintances that connect us to parts of the network that would otherwise be far away.
      - Weak ties produce many short paths and let us reach many nodes in a few steps.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world'networks. nature, 393(6684), 440.

Watts-Strogatz Model- Cnt.

- Watts-Strogatz (1998)
  - A model that combines two basic network ideas:
    - **Homophily**
      - Th

Local Structure

    - **Weak Ties**
      - Link … f the
      - no

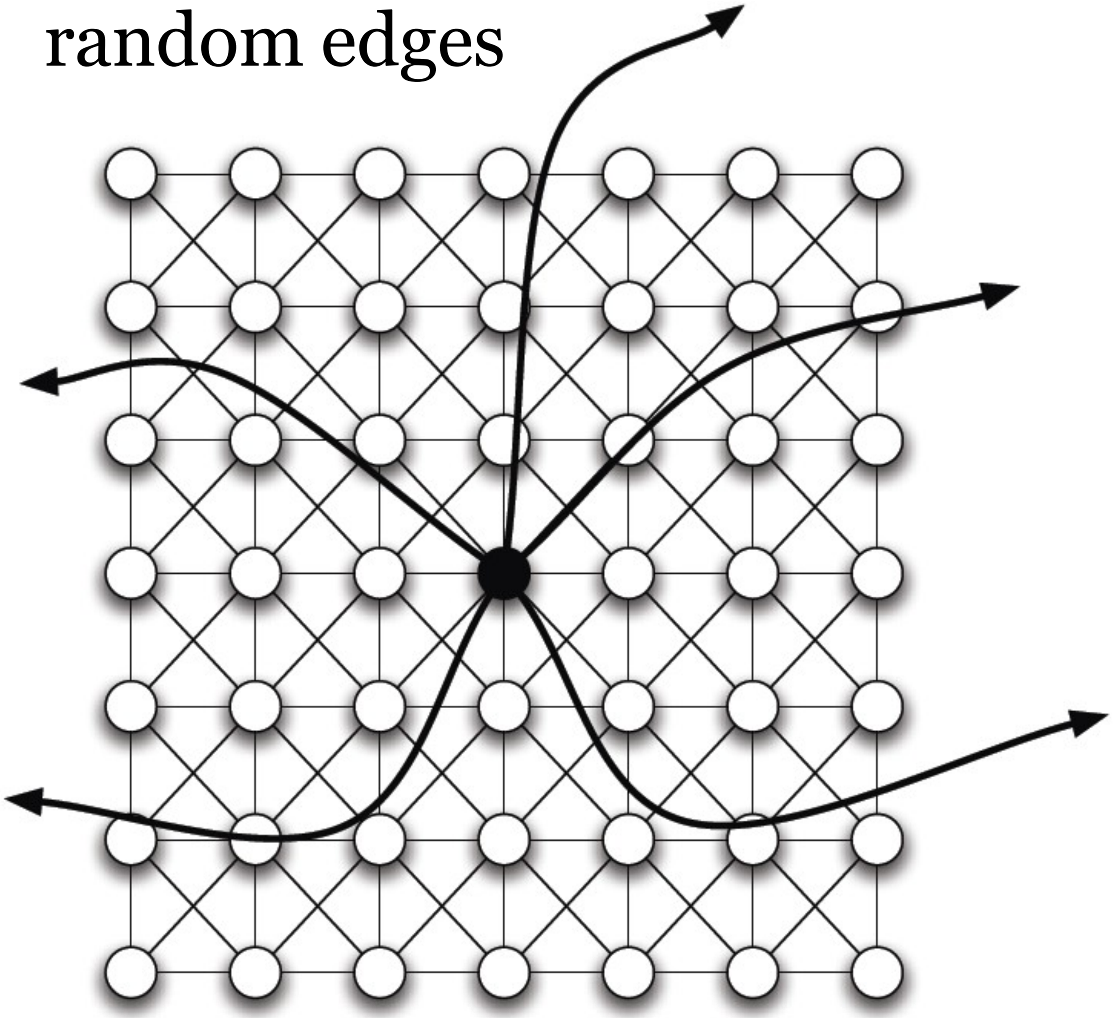
Branching Structure

Watts-Strogatz Model- Cnt.

- Suppose nodes live on a 2-dimensional grid, and create two kinds of edges:
  - Each node links to all other nodes that are *r* grid-steps away from it, *r* is a constant
    - these are the links to similar people (Homophily)!
  - Each node also forms a link to *k* other nodes selected *uniformly at random* from the grid
    - Connecting nodes who lie very far apart on the grid (Weak Ties).

Watts-Strogatz Model- Cnt.

- The resulting net built from local structure and random edges

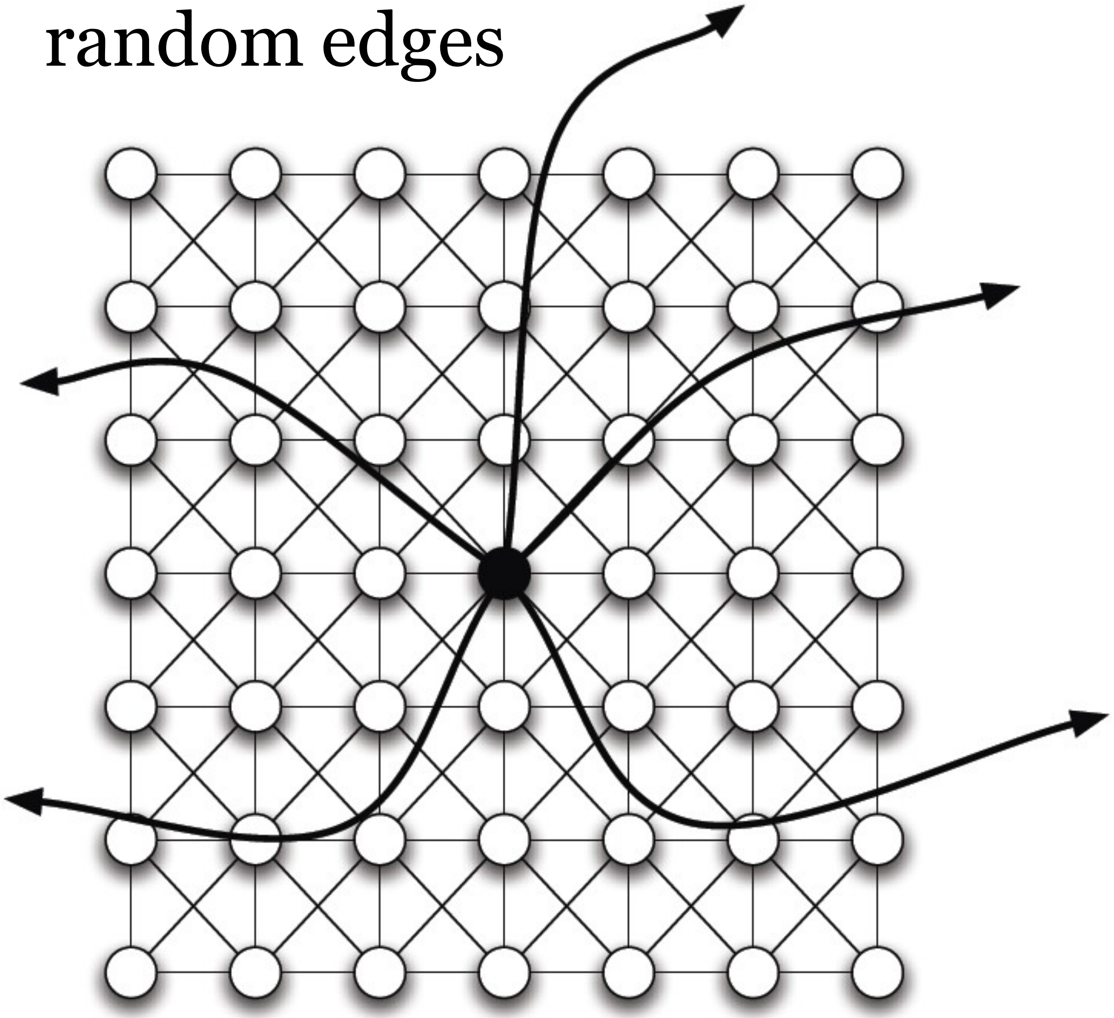


The network has many triangles

Two neighboring nodes (or nearby nodes) will have many common friends!

Watts-Strogatz Model- Cnt.

- The resulting net built from local structure and random edges

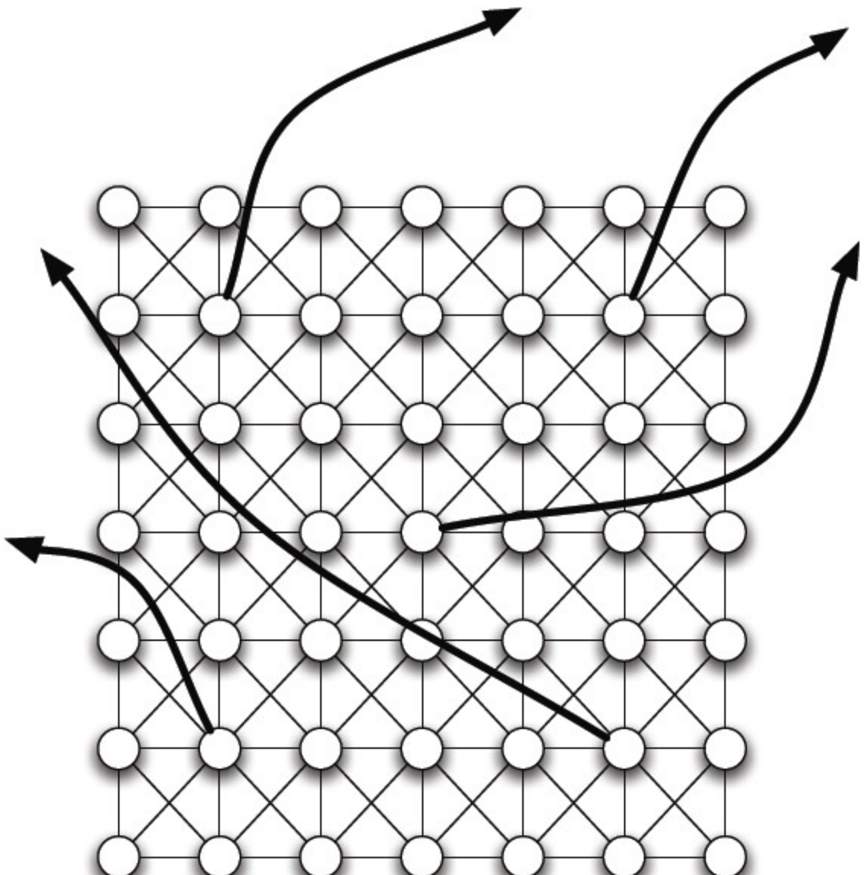


There are many short paths connecting pairs of nodes in the net!

Trace paths outward from a node using its weak ties. It's very unlikely to see a node twice in the first few steps. Reach huge number of nodes!

Watts-Strogatz Model- Cnt.

- It was shown that a small amount of randomness is just needed to achieve the same qualitative effect.



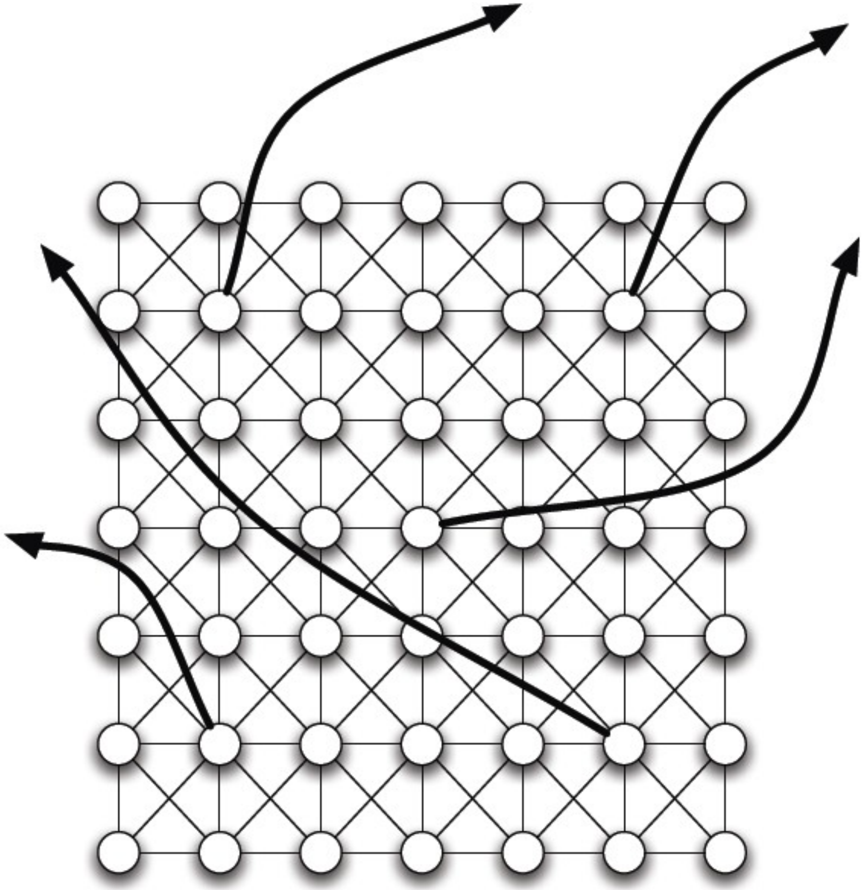
Instead of k random friends, allow one out of every k nodes to have a single random friend!

<u>Interpretation</u>: group $k \times k$ nodes into one towns. Each town has k links to other towns → Just like previous model.

To find a short path btw 2 people, first find a short path btw their towns, then use the local edges to find the actual path.

Watts-Strogatz Model- Cnt.

- It was shown that a small amount of randomness is just needed to achieve the same qualitative effect.



A small amount of randomness (weak ties) is enough to make the world "small," with short paths between every pair of nodes.

Lecture Topics

- Six Degree of Separation
- The Watts-Strogatz Model
- **Decentralized Search**

Decentralized Search

- Two striking facts about social networks:
  1. Short paths exist in abundance, and
  2. People are effective at collectively finding these short paths: **Decentralized Search**

Decentralized Search- Cnt.

- A collective procedure for finding a path from a **starting node (*s*)** to a **target node (*t*)**.
- Setting:
  - Consider grid-based model of Watts and Strogatz
  - *s* must forward a given message to *t* through net.
  - Each node only knows its neighbors and location of *t* on the grid
    - no other info about the structure of the net.
  - Each node must choose a neighbor to send the message to next.

Decentralized Search- Cnt.

- Decentralized search in the Watts-Strogatz model requires a **large number of steps** to reach a target
  - Larger than the true length of the *shortest path*
  - Why?

Decentralized Search- Cnt.

- Decentralized search in the Watts-Strogatz model requires a **large number of steps** to reach a target
  - Larger than the true length of the *shortest path*
  - Weak ties are **too random** in this model:
    - hard to use reliably as they are unrelated to the similarity btw nodes that produces homophily links.

Decentralized Search- Cnt.

- Generalizing Watts-Strogatz network model
  - Nodes on a grid and each node:
    - has edges to nodes within r grid steps,
    - has k random edges (weak ties) that are generated in a way that **decays with distance**
      - v links to w with probability proportional to $\mathbf{d(v, w)^{-q}}$
        - d(v,w): number of grid steps between nodes v and w.
        - q >= 0
- q controls how uniform the random links are!
  - We call q clustering exponent!

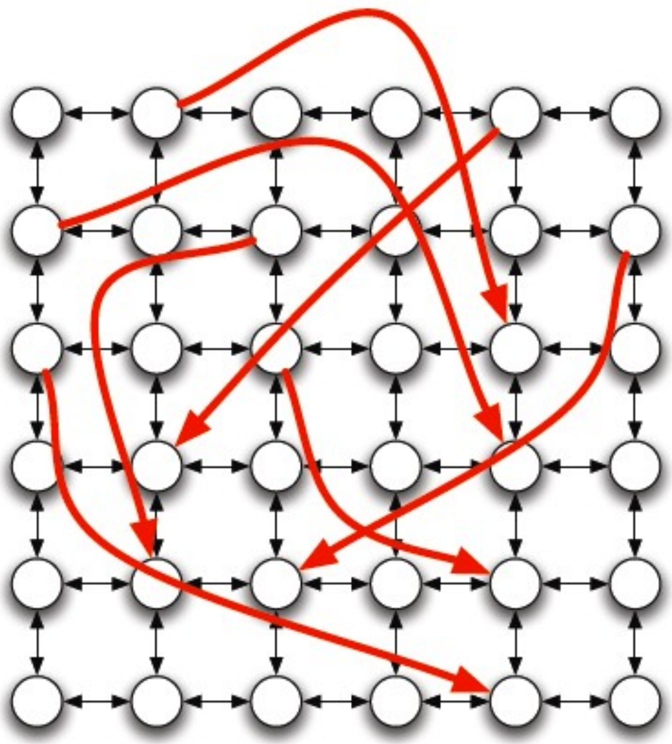
Decentralized Search- Cnt.

- Generalizing Watts-Strogatz network model
- $d(v, w)^{-q}$:
  - Smaller q:
  - Larger q:

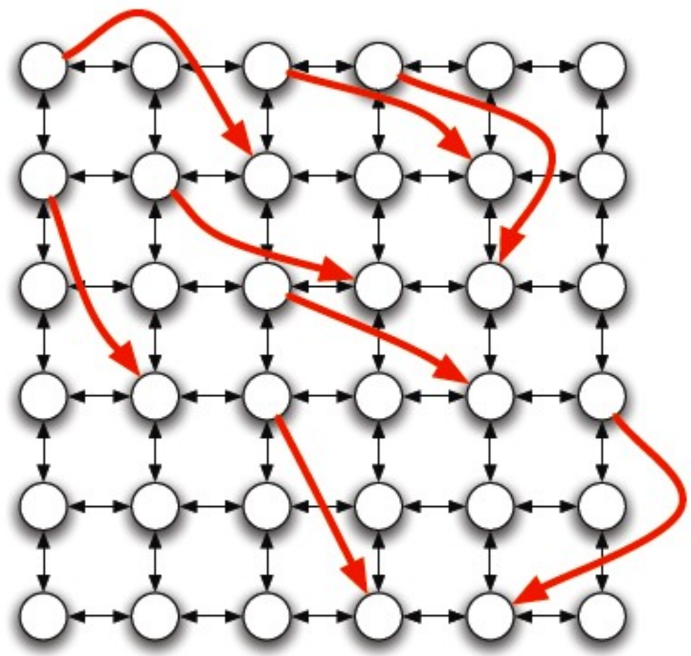
Decentralized Search- Cnt.

- Generalizing Watts-Strogatz network model
- $d(v, w)^{-q}$:
  - Smaller q:
    - Links are too random
      - $q=0$: the original model; links are generated uniformly at random.
    - Can't be used effectively for decentralized search
  - Larger q:
    - Links are not random enough
    - Not enough for long-distance jumps that are needed to create a small world.

Decentralized Search- Cnt.



(a) *A small clustering exponent*



(b) *A large clustering exponent*

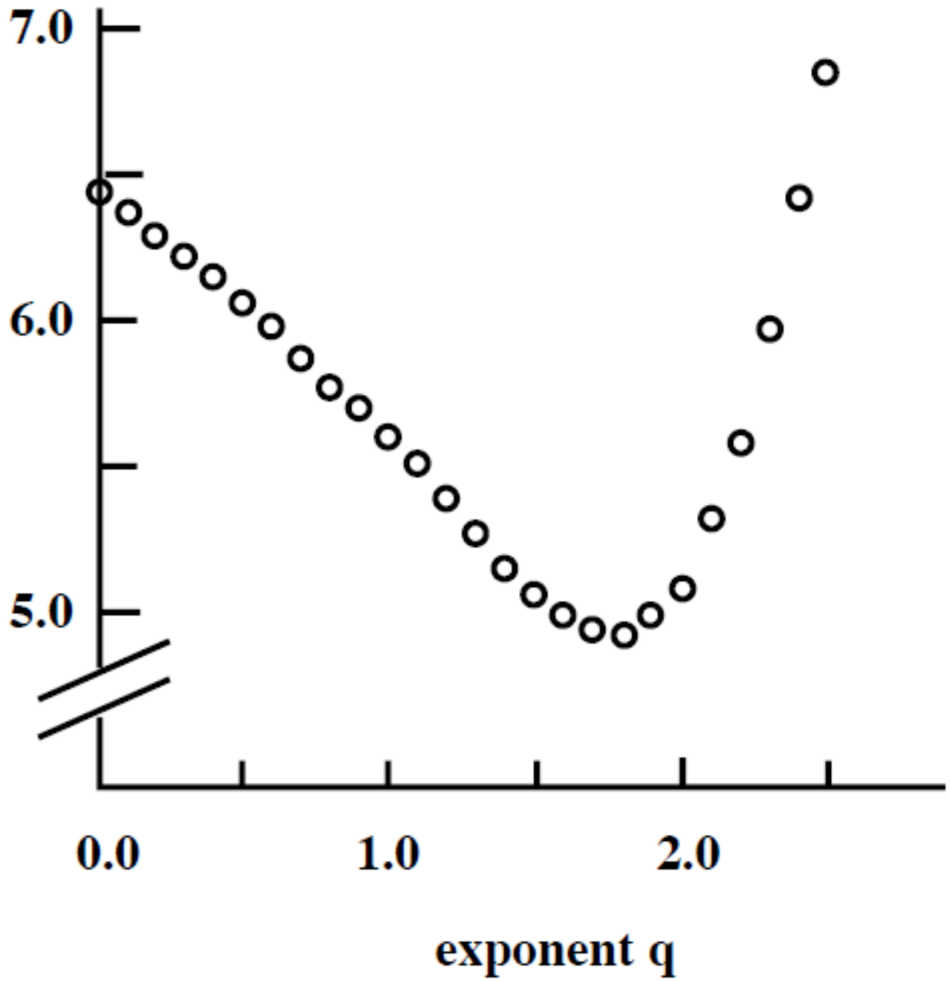
$$d(v, w)^{-q}$$

Is there an optimal *q* for the network that allows rapid decentralized search?

Decentralized Search- Cnt.

- 400M nodes
- Each point is the avg. of 1k runs
- Delivery time: exp. # of steps to reach the target.

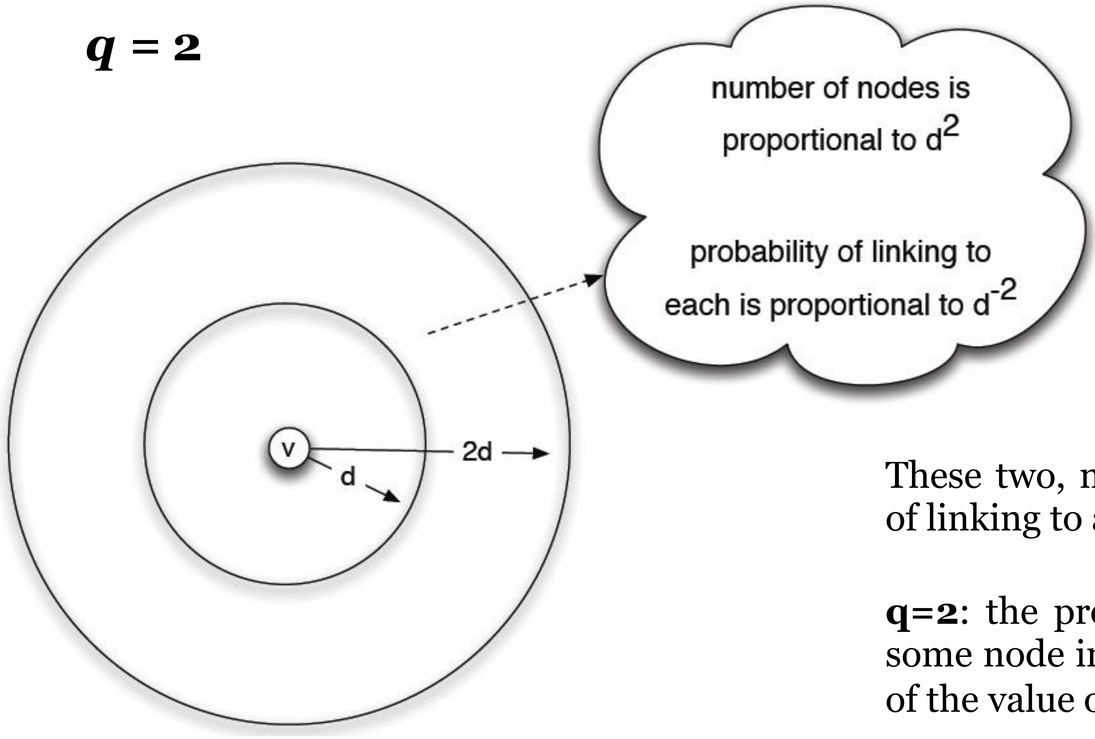
- Best delivery time q ~ 1.5 - 2
  - ~ Inverse square distribution!



Decentralized Search- Cnt.

- Consider groups of all nodes at increasingly large ranges of distance from a single node *v*
  - nodes at distance 2-4, 4-8, 8-16, etc.

***q* = 2**



The area in the plane grows like the square of radius, total number of nodes in the ring is proportional to **d^2**.

q= 2: The probability that *v* links to a node in the ring is proportional to **d^{-2}**.

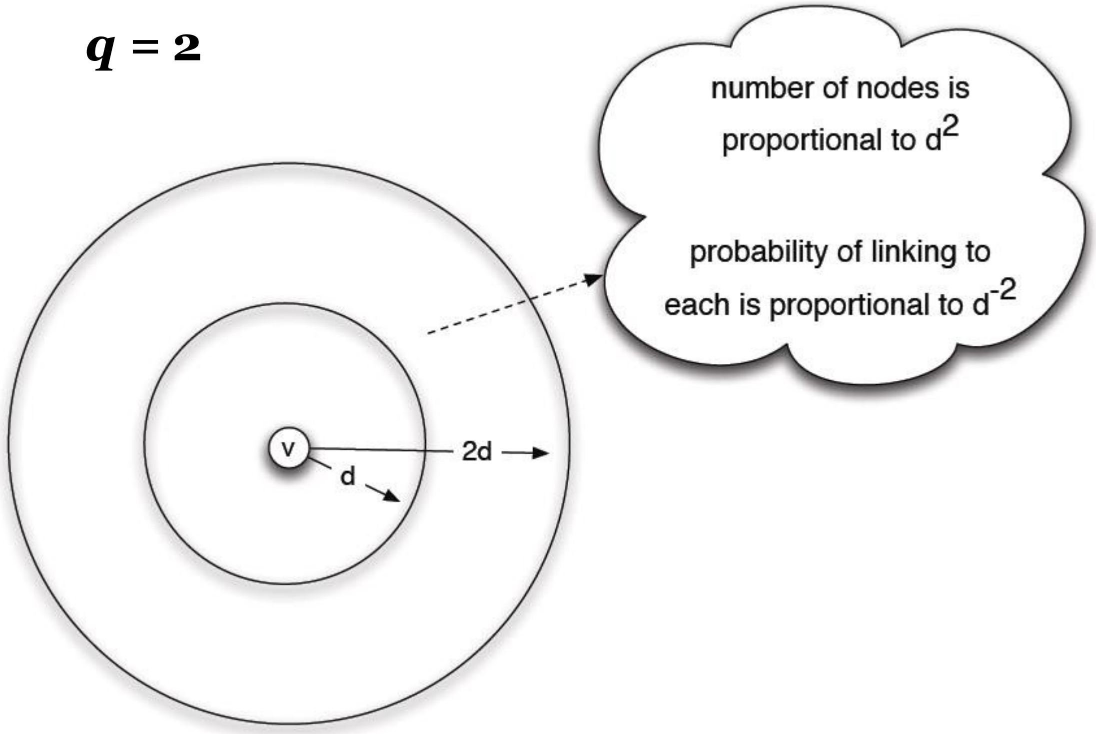
These two, number of nodes in ring and probability of linking to any of them, approximately cancel out.

q=2: the probability that a random edge links into some node in this ring is approximately independent of the value of *d*.

Decentralized Search- Cnt.

- Consider groups of all nodes at increasingly large ranges of distance from a single node v
  - nodes at distance 2-4, 4-8, 8-16, etc.

$q = 2$



Weak ties are formed in a way that is spread roughly uniformly over all different scales of resolution.

This allows nodes to consistently find ways of reducing their distance to the target, no matter how near or far they are from it.

Reading

- Ch.20 The Small-World Phenomenon [NCM]