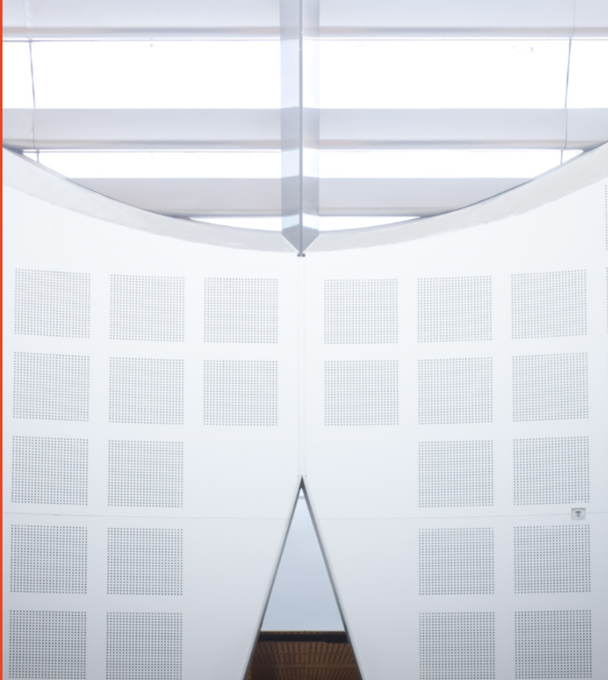
COMP5313/COMP4313 - Large Scale Networks

Week 10b: Structural Model for Small World

Lijun Chang

May 8, 2025

THE UNIVERSITY OF SYDNEY



Introduction

- As we saw at the beginning of the course
  - People are connected by very short paths through the social network
- The Small-World Phenomenon derives from this richness in short paths
  - Also called six degrees of separation.
- We now back to this phenomenon and study the way to collaboratively search through a social network

Outline

Recap of Six Degrees of Separation

The Watts and Strogatz Model

Generalized Model for Decentralized Search

Empirical Study of the Model

Six Degrees of Separation

In the 1960s, Stanley Milgram conducted the first significant empirical study of the small world phenomenon. [1]

- He asked 296 randomly chosen "starter" individuals to each try forwarding a letter to a designated "target" person living in Sharon, MA, USA.
- He provided name, address, occupation, and personal information of the target person.
- The participants could not mail the letter directly to the target
- The participants could only forward the letter to an acquaintance that (s)he knew, with the goal of reaching the target as fast as possible

[1] S. Milgram. The small world problem. Psychology today, vol. 2, no. 1, 1967.

Six Degrees of Separation

- 1/5 of the letters eventually arrived at the target
- The median of the number of steps was 6
- This style of experiments were reproduced by other people

- The experiment showed:
  - There are many short paths
  - People are effective at collaboratively finding these short paths

Six Degrees of Separation

- The fact that so many letters reached their destination by so short paths was striking
  - It is possible that, short paths are there but a letter might simply wander from one acquaintance to another, lost in the maze of social connections
  - For example, "forward this letter to user 482285204 using only people you know on a first name basis" is hopeless

- The real global network contains enough information about how people fit together, both geographic and social, to allow the process of search to focus on distant targets
  - Killworth and Bernard found that in Milgram's experiment, people use a mixture of geographic and occupational features, with different features being favored depending on the characteristics of the target in relation to the sender. [2]

[2] P. D. Killworth, H. R. Bernard. The reversal small-world experiment. Social networks, vol. 1, no. 2, 1978.

In the following, we focus on developing a model for networks such that there are a lot of triangles and short paths, and moreover a short path is easy to be found.

Outline

Recap of Six Degrees of Separation

The Watts and Strogatz Model

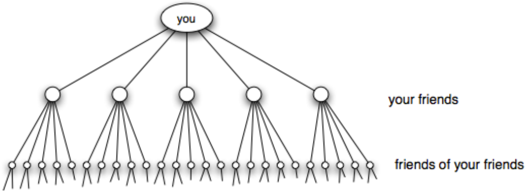
Generalized Model for Decentralized Search

Empirical Study of the Model

A Simple Model

Is it surprising that the distance between any two persons is so short?

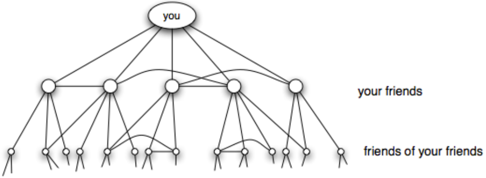
- Suppose each of us knows more than 100 other people on first-name basis
  - If your friends know 100 other people, then you are two steps away from $10,000$ people
  - $1,000,000$ are three steps away
  - 10^{10} after 5 steps?



- This intuition suggests short paths

A Simple Model

- However, real social networks are not like this simple model. Real social networks abound in triangles
  - Many of your 100 friends know each other
  - So many of the nodes you can reach from your friends are your friends



$\Longrightarrow$ The effect of triadic closure in social networks limits the number of people you can reach by following short paths

The Watts and Strogatz Model

- Can we simultaneously model the two observations?
  - Many closed triads
  - Short paths between people

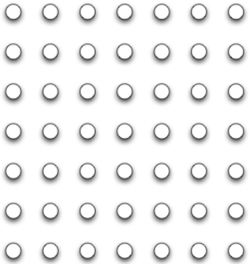
- The Watts and Strogatz model. [3]
  1. **Homophily:** the principle that we connect to others that are like ourselves
  2. **Weak ties:** the links to acquaintances that connect us to parts of the network that would otherwise be far away

- Homophily creates many triangles, while weak ties help reaching many nodes in a few steps

[3] D. J. Watts, S. H. Strogatz. Collective dynamics of “small-world” networks. Nature, vol. 393, no. 6684, 1998.

The Watts and Strogatz Model

- Let's suppose that everyone lives on a 2-dimensional grid
- Let's consider the distance as social proximity that impacts link creation

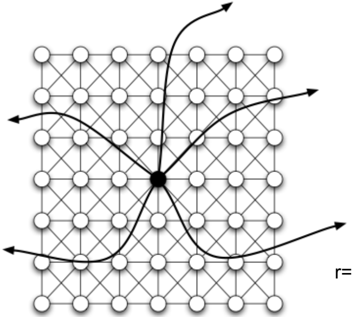


- Two nodes are one grid step apart if they are directly adjacent to each other in either the horizontal or vertical direction

The Watts and Strogatz Model

We create a network by giving nodes with two kinds of links

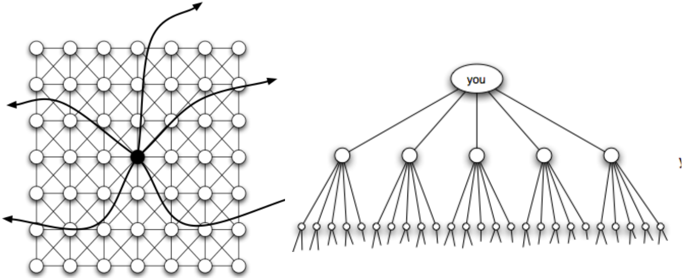
- **Homophily:** link between each node and all other nodes within a radius of r grid steps away (for some constant r)
- **Weak ties:** link between each node and k other nodes selected uniformly at random from the grid



r=1, k=5

The Watts and Strogatz Model

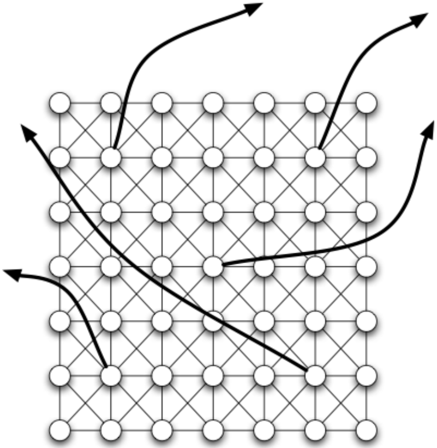
- The network has many triangles
  - Two neighboring nodes have many common friends, i.e., the intersection of the two radius-r neighborhoods.
- With high probability, there are many short paths connecting every pair of nodes
  - Consider only the k random links (weak ties) of each node



A Few Randomness is Enough

A surprisingly small amount of random links is enough to achieve the same qualitative effect.

- Suppose we only allow one out of every k nodes to have 1 random friend/weak tie
  - This corresponds to technologically earlier time where:
    - Most people knew their near neighbors
    - Few people knew someone far away
- Even this network has short paths between every pair of nodes



A Few Randomness is Enough

Intuition:

- Suppose we group $k \times k$ subsquares of the grid into "towns"
- Consider the small-world phenomenon at the level of towns
  - Each town contains k^2 people
    - Roughly, k people have random friends
  - So each town has collectively k links to other towns taken u.a.r.
- Like the previous model except that towns play the role of individual nodes, so we can find short paths between any pair of towns
- To find a short path between any two people:
  1. We find a short path between the two towns they inhabit
  2. We use the proximity-based edges to turn this into a path between individuals

$\implies$ Crux of Watts Strogatz model: tiny amount of randomness, in the form of long-range weak ties, is enough to make the world small

The watts and strogatz model ensures a lot of triangles and short paths, but decentralized search in this model is not easy.

Outline

Recap of Six Degrees of Separation

The Watts and Strogatz Model

Generalized Model for Decentralized Search

Empirical Study of the Model

Decentralized Search

- Let's consider the second aspect of Milgram's experiment
  - People could collectively find short paths to the target

- To really find the shortest path, one would have to instruct the starter to forward a letter to all of his or her friends, who in turn should have forwarded to all of his/her friends, and so forth
  - This flooding of the network would have reached the target as rapidly as possible
    - This is essentially the breadth first search
  - This was not possible and Milgram was forced to do a much more interesting experiment by tunneling through the network

- The success of Milgram's experiment raises two fundamental questions on collective/decentralized search
  1. Can we construct a random network in which decentralized routing succeeds?
  2. And, if so, what are the qualitative properties that are crucial for success?

Decentralized Search in the Watts and Strogatz Model

- Decentralized search for a path from s to t in the Watts and Strogatz model
  - Every node only knows the random edges out of itself, not for other nodes
  - Every node knows the location of t in the grid

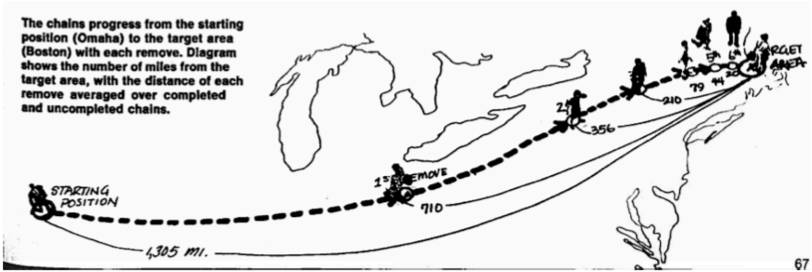
- Unfortunately, in this model, a decentralized search requires many steps to reach the target (much larger than the length of the shortest path). [4]

- This is due to that the weak ties are “too random” in this model
  - Since they are unrelated to the similarity among nodes, they are hard to be utilized reliably

[4] J. M. Kleinberg. Navigation in a small world. Nature, vol. 406, no. 6798, 2000.

Intuitions

- To reach a far away target, long range weak ties must be used properly
  - Distance to the target must constantly reduce (cf. Milgram's fig.[5])



- For successful decentralized search, weak ties cannot only span long ranges
- They have to span all the intermediary ranges as well

[5] S. Milgram. The small world problem. Psychology today, vol. 2, no. 1, 1967.

Generalized Model for Decentralized Search

- Can we adapt the Watts and Strogatz model?
  - Yes! Let's construct random edges in a structured way, and the probability of generating long edges decays with distance
  - For nodes u and v, let $d(u, v)$ be the number of grid steps between them
  - When generating an out-edge of u, we generate a link between u and v with probability proportional to:

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

    - q is the *clustering exponent*

Generalized Model for Decentralized Search

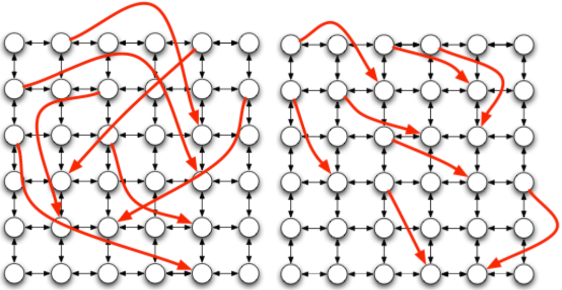
Introduce an extra quantity that controls the scale spanned by the long-range weak ties

- Each node has edges to all other nodes within r grid steps
- Each of its k random edges is generated in a way that decays with the distance controlled by a clustering exponent q:
  - For nodes u, v, we have an edge (u, v) with probability proportional to $d(u, v)^{-q}$

- Thus, we have a model for each value of q
  - $q = 0$ corresponds to the original Watts and Strogatz model
  - Varying q results in tuning how uniform the links are

Generalized Model for Decentralized Search

- With a small clustering exponent q, the random edges can span long distances
  - As q increases, the random edges become shorter

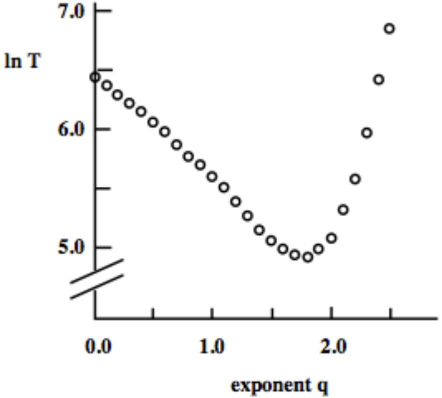


- Is there an optimal operating point for the network, where the distribution of long-range links is sufficiently balanced between these extremes to allow for rapid decentralized search?
  - Yes, decentralized search is most efficient when $q = 2$
  - $d(u, v)^{-2}$ is called the inverse-square distribution

Empirical Results on q for Efficient Decentralized Search

- Performance of a basic decentralized search method with different values of q, for a network of several hundred million nodes

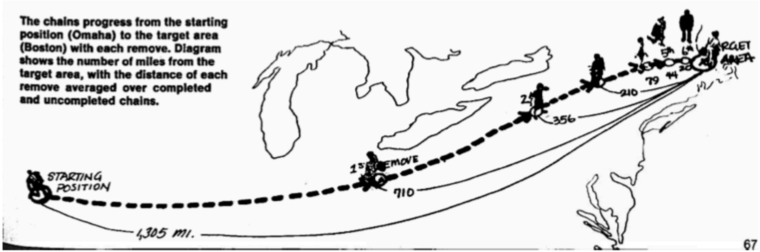
- As the network size increases (and goes to infinity), the best performance occurs at exponents q closer and closer to 2



Why Choose $q = 2$?

Different Scales of Resolution

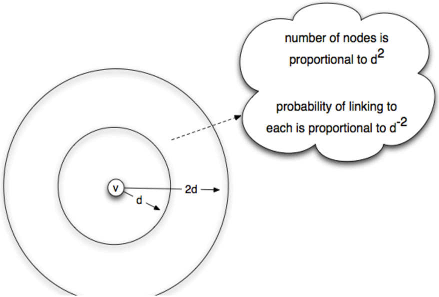
- In Milgram's experiment, we mentally organize distances into different "scales of resolution"
  - Down the block, across town, across the state, across the country, around the world
- Effective decentralized search "funnels inward" through these different scales of resolution, as we see from the way the letter reduces its distance to the target by approximately a factor of two with each step



Why Choose $q = 2$?

How does $q = 2$ interact with the different scales of resolution?

- Intuition behind the scale of resolution is to consider the groups of all nodes at increasingly large ranges of distance from a node:
  - 2-4, 4-8, 8-16 ...
- Consider node v in the network and nodes between d and $2d$ distance from v
- Since the area grows like the square of the radius, the total number of nodes is proportional to d^2
- The probability that v connects a node in this area is proportional to d^{-2}
- These two values approximately cancel out
- The probability that a random edge links into some node in this ring is approximately independent of d.



Why Choose q= 2?

How does $q = 2$ interact with the different scales of resolution?

- Intuitively, long-range weak ties are being formed in a way that's spread roughly uniformly over all different scales of resolution.
- Thus, the people, forwarding the message, can consistently find ways of reducing their distance to the target, no matter how near or far they are from it.

Outline

Recap of Six Degrees of Separation

The Watts and Strogatz Model

Generalized Model for Decentralized Search

Empirical Study of the Model

LiveJournal

- To answer the question "Does the link formation in real social networks follows this model?"
- Liben-Nowell et al. used the blogging site LiveJournal, and analyzed roughly 500,000 users who provided a U.S. ZIP code for their home address, as well as links to their friends on the system. [6]

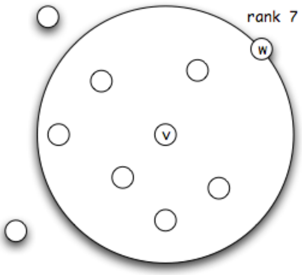
[6] D. Liben-Nowell, J. Novak, R. Kumar, P. Raghavan, A. Tomkins. Geographic routing in social networks. Proceedings of the National Academy of Sciences, vol. 102, no. 33, 2005.

LiveJournal

- This is different from our grid-based model
  - Here, the population density is skewed
  - The inverse-square distribution is nice for finding node when nodes are uniformly spaced in 2 dimensions

Rank-based Friendship

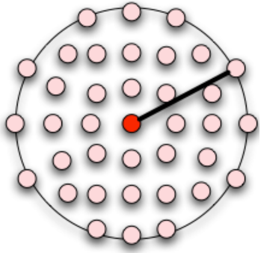
- Suppose that as a node v looks out at all other nodes, it ranks them by proximity
  - The rank of a node w, denoted rank(w), is equal to the number of other nodes that are closer to v than w is
  - rank$(w) = 7$, since seven other nodes (including v itself) are closer to v than w is.
- Now, suppose that for some exponent p, node v creates a random link as follows
  - It chooses a node w as the other end with probability proportional to rank$(w)^{-p}$.
  - We call this rank-based friendship with exponent p



- Which choice of exponent p would generalize the inverse-square distribution for uniformly-spaced nodes?

Rank-based Friendship

- In a uniformly-spaced grid, let's consider a node w that is at distance d from v
  - It lies on the circumference of a disc of radius d containing about d^2 closer nodes
  - Its rank is approximately d^2.
- Thus, linking to w with probability proportional to d^{-2} is approximately the same as linking with probability $\text{rank}(w)^{-1}$
  - This suggests that exponent $p = 1$ is the right generalization of the inverse-square distribution
- Liben-Nowell et al. were able to prove that for essentially any population density, if random links are constructed using rank-based friendship with exponent 1, the resulting network allows for efficient decentralized search

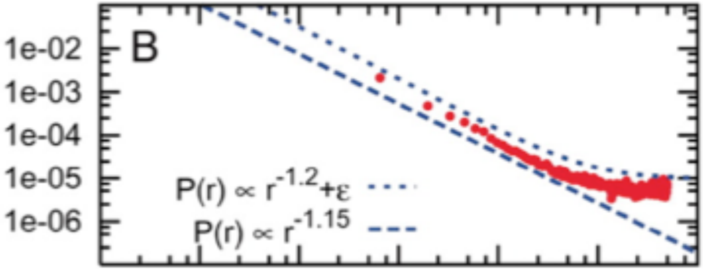


Rank-based Friendship

- How well rank-based friendship fits the distribution of actual social network links?
  - Consider pairs of nodes where one assigns the other a rank of r, and we ask what fraction f of these pairs are actually friends, as a function of r.
  - Does this fraction decrease approximately like r^{-1}?
- Since we're looking for a power-law relationship between the rank r and the fraction of edges f, we can proceed as in the previous lecture:
  - rather than plotting f as a function of r, let us plot $\log f$ as a function of $\log r$,
  - see if we find an approximately straight line and
  - estimate the exponent p as the slope of this line

Rank-based Friendship

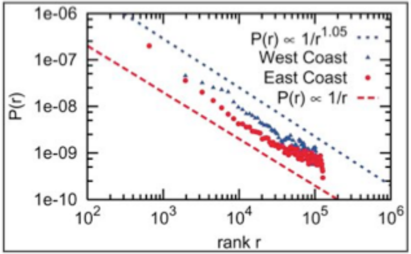
- Here is the result for the LiveJournal data



- Much of the body of the curve is approximately a straight line sandwiched between slopes of -1.15 and -1.2, and hence close to the optimal exponent of -1

Rank-based Friendship

- It is also interesting to work separately with the more structurally homogeneous subsets of the data consisting of West-Coast users and East-Coast users: the exponent becomes very close to the optimal value of -1.



- The lower dashed line is what you should see if the points followed the distribution rank^{-1}, and the upper dotted line is what you should see if the points followed the distribution $\text{rank}^{-1.05}$.

Conclusion

- To construct a network that is efficiently searchable, create a link to each node with probability that is inversely proportional to the number of closer nodes

Reading

- Reading for this week
  - Chapters 18 and 20 (excluding 20.7 Advanced Material) of the textbook

- Reading for next week
  - Hari Balakrishnan, M. Frans Kaashoek, David R. Karger, Robert Tappan Morris, Ion Stoica: Looking up data in P2P systems. Commun. ACM 46(2): 43-48 (2003) https://dl.acm.org/doi/10.1145/606272.606299