Small world networks

Social and Technological Networks

Rik Sarkar

University of Edinburgh, 2018.
Milgram’s experiment

- Take people from random locations in USA
- Ask them to deliver a letter to a random person in Massachusetts
- A person can only forward the letter to someone you know
- Question: How many hops do the letters take to get to destination?
Results

• Out of 296 letters, only 64 completed
• Number of hops varied between 2 and 10
• Mean number of hops 6
• There were a few people that were the last hop in most cases
Discussion of experiment
Discussion of experiment

• Short paths exist between pairs (small diameter)
• More surprisingly, people find these short paths
• Without knowing the entire network
• *Decentralized search*
• Analogous to routing without a routing table
• People use a “greedy” strategy
• Forward to the friend nearest to the destination
Recent results

• Milgrams results reproduced on better data
• Use online data (Livejournal, facebook)
• Containing approximate locations
• Simulate the process of forwarding letters
• Results similar to original experiment
• Relatively short diameter, successful decentralized search
In popular culture

• Erdos distance
• Kevin bacon distance
Definition of small worlds

• Small diameter

• Large clustering coefficient
  – Related to homophily — similar people connect to each-other
  – “Similar”: close in some coordinate value (or metric)

• Supports decentralized search
  – People find short paths without knowing the entire network

• (Usually) High expansion
Model 1: Watts and Strogatz

Nature 1998

- Parameters $n, k, p$  $n > k > \ln n$
  - Often $k$ is taken to be a constant in practice with the idea that people cannot have infinitely large friend-circles
- Put nodes in a ring of size $n$
- Connect each to $k/2$ neighbors on each side
Model 1: Watts and Strogatz

- Parameters: $n, k, p$, such that $n > k > \ln n$
  - Often $k$ is taken to be a constant in practice with the idea that people cannot have infinitely large friend-circles

- Put nodes in a ring of size $n$

- Connect each to $k/2$ neighbors on each side

- What is the diameter and CC?
Model 1: Watts and Strogatz

Nature 1998

- Parameters: \( n, k, p \) \( n > k > \ln n \)
  - Often \( k \) is taken to be a constant in practice with the idea that people cannot have infinitely large friend-circles
- Put nodes in a ring of size \( n \)
- Connect each to \( k/2 \) neighbors on each side
- With probability \( p \) rewire each edge of a vertex to a random vertex
Small world

• In between random and structured
Small world
Properties

• Average clustering coefficient per vertex bounded away from zero
  – In other words: at least a constant

• Connected: sufficient random edges + regular edges

• Short diameter
Watts-strogatz model does not explain milgram’s experiment
Watts-strogatz model does not explain milgram’s experiment

• Milgram’s experiment was on 2D plane
• Watts strogatz does not support decentralized search
Watts-strogatz model does not explain milgram’s experiment

- Milgram’s experiment was on 2D plane
- Watts-strogatz does not support decentralized search \((\text{poly} (\log n))\) steps to destination
Outline

• A good small world model should support \(\text{poly}(\log n)\) steps to destination
• We will show that watts strogatz needs \(\text{poly}(n)\)

• Kleinberg’s model supports \((\log(n))^2\) step route
  – Show that the route can be divided into \(O(\log n)\) phases
  – Each phase takes \(O(\log n)\)
Decentralized search in random link networks

• Decentralized search does not work to produce short paths

• Let us consider 2D (n x n grid):
  – We want to show that if every node works only on its local information (edges it has)

• Then there is no algorithm that delivers the message in less than poly(n) messages.
Decentralized search in random link networks

- Consider $s$ and $t$ separated by $\Omega(n)$ hops
- Take ball $B$ of extrinsic radius around $n^{2/3}$
  - There are $O(n^{2/3})^2$ nodes in $B$
- When we are already at distance $n^{2/3}$ (on the edge of $B$)
  - A long link can help only if it falls inside $B$
- Otherwise we take a step along a short link
- What is the probability that a random link from $s$ hits $B$?
- This is $\sim O((n^{2/3})^2/n^2) = O(n^{-2/3})$
- The expected number of steps before getting a useful long link is: $\Omega(n^{2/3})$
Decentralized search

- Therefore long links are not really useful in reaching $t$
- The number of steps is $\text{poly}(n)$.
Model 2: Kleinberg’s model

• Idea: Long links are not helping much
  – Getting closer to the destination does not increase the chances of getting a long link close to destination.

• Make the probability of a long link sensitive to the distance
  – Nearby nodes are more likely to have a long link

Model 2: Kleinberg’s model

• Suppose $d(u, v)$ is the extrinsic distance between nodes $u$ and $v$ in the plane.
• Then $u$ connects its long link to $v$ with probability $\propto \frac{1}{d(u, v)^\alpha}$. 
Kleinberg’s model

• Links to nearby nodes are more likely
  – A node knows more people locally
  – With increasing distance, it knows fewer and fewer people
  – At the largest scale it knows only a handful
  – More representative of how people have their contacts spread

• We want to show that the model permits short paths to be found
The proportionality constant

\[ \text{Pr}[(u, v)] = \frac{1}{\gamma} \frac{1}{d(u,v)^\alpha} \]

\[ \alpha = 2 \Rightarrow \gamma = \Theta(\ln n) \]

• Sketch of proof: Take rings of thickness 1 at distances 1, 2, 3...

• The number of nodes at distance \( d \sim \Theta(d) \)

• Thus from any node:

\[ \frac{1}{\gamma} \sum_{d=1}^{n} d^{-2} \Theta(d) = 1 \]

\[ \frac{1}{\gamma} \Theta \left( \sum_{d=1}^{n} \frac{1}{d} \right) = 1 \]

\[ \Rightarrow \frac{1}{\gamma} \Theta(\ln n) = 1 \]
Theorem

\[ \alpha = 2 \]

- Permits finding \( O(\log^2 n) \) intrinsic length paths
- Using local routing: Always move to the neighbor nearest to the destination
Proof

- Main idea:
- In $O(\log n)$ steps, the extrinsic distance is halved
  - Let us call this one phase
- In $O(\log n)$ phases, the distance will be 1
- So, we need to show the first claim: one phase lasts $O(\log n)$ steps
One Phase lasts $\log n$ steps

- Suppose distance from $s$ to $t$ is $d$
- take ball $B$ of radius $d/2$ around $t$
- There are about $\Theta(d^2)$ nodes in this area
- The probability that a long link hits $B$ is

$$\frac{1}{\Theta(n \log n)} \sum_{v \in B} d(s,v)^{-2} \geq \Theta \left( \frac{1}{\log n} d^2 d^{-2} \right) = \Theta \left( \frac{1}{\log n} \right)$$
One phase lasts $\log n$ steps

• Thus, the expected number of steps before we find a link into B is $\log n$.
• And there are $\log n$ such phases
• Therefore, this method finds a path of $\log^2 n$ steps
Other exponents

• $< 2$ : more like uniform random
• $> 2$ : Shorter links, almost same as basic grid..
Generality

• Search is a very general problem
• Search for an item, search for a path, search for a set, search for a configuration
• Decentralized: Operation under small amount of information. (local, easy to distribute)
Small worlds in other networks

- Brain neuron networks
- Telephone call graphs
- Voter network
- Social influence networks ...

- Applications:
- Peer to peer networks
- Mechanisms for fast spread of information in social networks
- Routing table construction