Overview

Social Goal. Explain why information and disease spread so quickly in social networks.

Mathematical Approach. Model social networks as random graphs and argue that they are likely to have low diameter.

Definition: The clustering coefficient of a node \( v \) is the fraction of pairs of \( v \)'s friends that are connected to each other by edges.

Clustering Coefficient = \( 1/2 \)

The higher the clustering coefficient of a node, the more strongly triadic closure is acting on it.

Erdos-Renyi Graph models

- Randomly choose
- How well does the E-R model characterize real networks?
  - high school friendship networks
  - road networks
  - peer-to-peer filesharing networks
  - product co-purchase networks (people who bought X also bought Y)
  - other?

Random graph diameter

Technique. Grow trees to bound path lengths.

1. Show when trees are small enough (< \( \sqrt{n} \)), # of leaves doubles.
2. Grow small trees (< \( \sqrt{n} \)) around a pair of nodes.
3. Use birthday paradox to argue trees prob. intersect.

Conclude diameter is 2 log \( \sqrt{n} = \log n \).
Small world phenomenon

Milgram’s experiment (1960s).

Ask someone to pass a letter to another person via friends knowing only the name, address, and occupation of the target.

Last time: short paths exist.
(argument: flood the network)

This time: and people can find short paths!
(without flooding the network)
The Milgram experiment

Experiment:

We choose a random “source” and “target”.

Your goal: pass ball from “source” to “target” by throwing it to people you know on a first-name basis.

How did you find short paths?

Social network models

What information did you use to choose the next hop?

Random rolodex model

Random wiring model
Random wiring model

People have a predictable structure of local links (e.g., neighbors, colleagues)

And a few random long-range links (e.g., someone you meet on a trip)

Local links have grid-like structure (you know the person on your left/right and front/back)

Local links represent homophily, the idea that we know people similar to us

Long-range links are random – each node chooses one long-range link
Random wiring model

Long-range links represent weak ties, the links to acquaintances that would otherwise be far away.

What do you expect the diameter to be? Do you expect people to find short paths? If so, how short?

Short paths? Problem: navigational clues lost in long-range links. What if long-range links are uniformly random?

Problem: increases path-length. How long-range links are very localized?
Decentralized search

Idea: Suppose long-range links are just slightly more likely to be to close nodes.

Result: Then decentralized search finds short paths.

Optimal tradeoff

Suppose links are proportional to \((1/\text{distance})^2\), i.e., inverse square.

Inverse square?? WTF?
The Science of Networks

**Scales of resolution**

- **Street**: location within 2 miles
- **City**: location within 4 miles
- **County**: location within 8 miles

Each new scale doubles distance from the center.
Scales of resolution

Long-range links equally likely to connect to each different scale of resolution!

(allows people to make progress towards destination no matter how far away they are)
How many people do you know?
How well did this work?

For me:
- Neighborhood (Trinity Heights)
- Region (RTP)
- State (NC)
- Southeast
- Eastern US
- United States
- World

Next topic

decentralized search

How to route

(sub) Problem. How can I get this message from me to the far-away target?

Solution. Pass message to a friend.

closer

Scales of resolution

Each new scale doubles distance from the center.
Long-range links

Suppose each person has a long-range friend in each scale of resolution.

How to route

Algorithm. Pass the message to your farthest friend that is to the left of the target.

Trace of route

Analysis

old dist. 1 2 4 2^i 2^{i+1} new dist.
Distance is cut in half every step!

**Analysis**

1. Original distance is at most $n$.
2. Distance is cut in half every step (at least).
3. Number of steps is at most $\log n$.

**Finding the Short Paths**

- Milgram’s experiment, Columbia Small Worlds, E-R, $\alpha$-model...
  - all emphasize existence of short paths between pairs
- How do individuals find short paths?
  - in an incremental, next-step fashion
  - using purely local information about the NW and location of target
  - note: shortest path might require taking steps “away” from the target!
- This is not (only) a structural question, but an algorithmic one
  - statics vs. dynamics
- Navigability may impose additional restrictions on formation model!
- Briefly investigate two alternatives:
  - a local/long-distance mixture model [Kleinberg]
  - a “social identity” model [Watts, Dodd, Newman]
### Kleinberg’s Model
- Start with an \( n \) by \( n \) grid of vertices (so \( N = n^2 \))
- Add some long-distance connections to each vertex:
  - \( k \) additional connections
  - Probability of connection to grid distance \( d \): \( \sim (1/d)^r \)
- So full model given by choice of \( k \) and \( r \)
- Large \( r \): Heavy bias towards “more local” long-distance connections
- Small \( r \): Approach uniformly random
- Kleinberg’s question:
  - What value of \( r \) permits effective navigation?
  - \# hops \( \ll N \), e.g. \( \log(N) \)
- Assume parties know only:
  - Grid address of target
  - Addresses of their own immediate neighbors
- Algorithm: Pass message to neighbor closest to target in grid

### Kleinberg’s Result
- Intuition:
  - If \( r \) is too large (strong local bias), then “long-distance” connections never help much; short paths may not even exist (remember, grid has large diameter, \( \sim \sqrt{N} \))
  - If \( r \) is too small (no local bias), we may quickly get close to the target; but then we’ll have to use grid links to finish
    - Think of a transport system with only long-haul jets or donkey carts
  - Effective search requires a delicate mixture of link distances
- The result (informally):
  - \( r = 2 \) is the only value that permits rapid navigation (~\( \log(N) \) steps)
  - Any other value of \( r \) will result in time \( \sim N^c \) for \( 0 < c \leq 1 \)
    - \( N^c > > \log(N) \) for large \( N \)
  - A critical value phenomenon or “knife’s edge”; very sensitive
- Note: Locality of information crucial to this argument
  - Centralized, “birds-eye” algorithm can still compute short paths at \( r < 2! \)
  - Can recognize when “backwards” steps are beneficial

### Navigation via Identity
- Watts et al.:
  - We don’t navigate social networks by purely “geographic” information
  - We don’t use any single criterion; recall Dodds et al. on Columbia SW
  - Different criteria used at different points in the chain
- Represent individuals by a vector of attributes
  - Profession, religion, hobbies, education, background, etc...
  - Attribute values have distances between them (tree-structured)
  - Distance between individuals: minimum distance in any attribute
  - Only need one thing in common to be close!
- Algorithm:
  - Given attribute vector of target
  - Forward message to neighbor closest to target
- Permits fast navigation under broad conditions
  - Not as sensitive as Kleinberg’s model