Some problems for people thinking about people?:

How are social networks structured?
▶ How do we define connections?
▶ How do we measure connections?
▶ (remote sensing, self-reporting)

What about the dynamics of social networks?
▶ How do social networks evolve?
▶ How do social movements begin?
▶ How does collective problem solving work?
▶ How is information transmitted through social networks?

Social Search

A small slice of the pie:
▶ Q. Can people pass messages between distant individuals using only their existing social connections?
▶ A. Apparently yes...

Handles:
▶ The Small World Phenomenon
▶ or “Six Degrees of Separation.”
The problem

Stanley Milgram et al., late 1960’s:
- Target person worked in Boston as a stockbroker.
- 296 senders from Boston and Omaha.
- 20% of senders reached target.
- average chain length $\simeq 6.5$.

Two features characterize a social ‘Small World’:

1. Short paths exist and
2. People are good at finding them.

60,000+ participants in 166 countries
18 targets in 13 countries including
- a professor at an Ivy League university,
- an archival inspector in Estonia,
- a technology consultant in India,
- a policeman in Australia,
  and
- a veterinarian in the Norwegian army.
24,000+ chains

Milgram's participation rate was roughly 75%
Email version: Approximately 37% participation rate.
Probability of a chain of length 10 getting through:
\[
0.37^{10} \approx 5 \times 10^{-5}
\]
⇒ 384 completed chains (1.6% of all chains).

Motivation/Incentives/Perception matter.
If target seems reachable
⇒ participation more likely.
Small changes in attrition rates
⇒ large changes in completion rates
e.g., ↓ 15% in attrition rate
⇒ ↑ 800% in completion rate

Successful chains disproportionately used
- weak ties (Granovetter)
- professional ties (34% vs. 13%)
- ties originating at work/college
- target's work (65% vs. 40%)

... and disproportionately avoided
- hubs (8% vs. 1%) (+ no evidence of funnels)
- family/friendship ties (60% vs. 83%)

Geography → Work
Social search—the Columbia experiment

Senders of successful messages showed little absolute dependency on
- age, gender
- country of residence
- income
- religion
- relationship to recipient

Range of completion rates for subpopulations: 30% to 40%

Nevertheless, some weak discrepancies do exist...

An above average connector:
Norwegian, secular male, aged 30-39, earning over $100K, with graduate level education working in mass media or science, who uses relatively weak ties to people they met in college or at work.

A below average connector:
Italian, Islamic or Christian female earning less than $2K, with elementary school education and retired, who uses strong ties to family members.

Mildly bad for continuing chain: choosing recipients because “they have lots of friends” or because they will “likely continue the chain.”

Why:
- Specificity important
- Successful links used relevant information.
  (e.g. connecting to someone who shares same profession as target.)

Basic results:
- $\langle L \rangle = 4.05$ for all completed chains
- $L_\star = \text{Estimated ‘true’ median chain length (zero attrition)}$
- Intra-country chains: $L_\star = 5$
- Inter-country chains: $L_\star = 7$
- All chains: $L_\star = 7$
- Milgram: $L_\star \approx 9$
Previous work—short paths

- Connected random networks have short average path lengths:
  \[ \langle d_{AB} \rangle \sim \log(N) \]
  
  \( N \) = population size, 
  \( d_{AB} \) = distance between nodes A and B.

- But: social networks aren’t random...

Non-randomness gives clustering

\[ d_{AB} = 10 \rightarrow \text{too many long paths.} \]

Randomness + regularity

\[ \langle d \rangle \text{ decreases overall} \]

Need “clustering” (your friends are likely to know each other):
Small-world networks

Introduced by Watts and Strogatz (Nature, 1998) \[6\]
“Collective dynamics of ‘small-world’ networks.”

Small-world networks were found everywhere:

- neural network of C. elegans,
- semantic networks of languages,
- actor collaboration graph,
- food webs,
- social networks of comic book characters,...

Very weak requirements:

- local regularity + random short cuts

The structural small-world property

previous work—finding short paths

But are these short cuts findable?

No.

Nodes cannot find each other quickly with any local search method.
Previous work—finding short paths

- What can a local search method reasonably use?
- How to find things without a map?
- Need some measure of distance between friends and the target.

Some possible knowledge:
- Target's identity
- Friends' popularity
- Friends' identities
- Where message has been

Previous work—finding short paths

Kleinberg's Network:
1. Start with regular d-dimensional cubic lattice.
2. Add local links so nodes know all nodes within a distance $q$.
3. Add $m$ short cuts per node.
4. Connect $i$ to $j$ with probability
   \[ p_{ij} \propto d_{ij}^{-\alpha}. \]

- $\alpha = 0$: random connections.
- $\alpha$ large: reinforce local connections.
- $\alpha = d$: same number of connections at all scales.

Previous work—finding short paths


Allowed to vary:
1. local search algorithm
2. network structure.

Previous work—finding short paths

Theoretical optimal search:
- “Greedy” algorithm.
- Same number of connections at all scales: $\alpha = d$.

Search time grows slowly with system size (like $\log^2 N$).

But: social networks aren't lattices plus links.
Previous work—finding short paths

- If networks have hubs can also search well: Adamic et al. (2001) \[^{[1]}\]
  \[ P(k_i) \propto k_i^{-\gamma} \]
  where \( k = \) degree of node \( i \) (number of friends).
- Basic idea: get to hubs first (airline networks).
- But: hubs in social networks are limited.

The problem

If there are no hubs and no underlying lattice, how can search be efficient?

- Which friend of a is closest to the target b?
- What does ‘closest’ mean?
- What is ‘social distance’?

The model

One approach: incorporate identity. (See “Identity and Search in Social Networks.” Science, 2002, Watts, Dodds, and Newman \[^{[2]}\])

Identity is formed from attributes such as:

- Geographic location
- Type of employment
- Religious beliefs
- Recreational activities.

Groups are formed by people with at least one similar attribute.

Attributes ↔ Contexts ↔ Interactions ↔ Networks.

Social distance—Bipartite affiliation networks
The model

- Individuals are more likely to know each other the closer they are within a hierarchy.
- Construct \( z \) connections for each node using

\[
p_{ij} = c \exp\{ -\alpha x_{ij} \}.
\]

- \( \alpha = 0 \): random connections.
- \( \alpha \) large: local connections.

Distance between two individuals \( x_{ij} \) is the height of lowest common ancestor.

\[
x_{ij} = 3, \ x_{ik} = 1, \ x_{iv} = 4.
\]

Social distance—Generalized context space

(Blau & Schwartz, Simmel, Breiger)
The model

\[
\vec{v}_i = [1, 1, 1]^T, \quad \vec{v}_j = [8, 4, 1]^T
\]

\[
x_{ij}^1 = 4, \quad x_{ij}^2 = 3, \quad x_{ij}^3 = 1.
\]

Social distance:

\[
y_{ij} = \min_h x_{ij}^h.
\]

Triangle inequality doesn't hold:

\[
y_{jk} = 4 > y_{ij} + y_{jk} = 1 + 1 = 2.
\]

The model

▶ Individuals know the identity vectors of
  1. themselves,
  2. their friends,
  and
  3. the target.
▶ Individuals can estimate the social distance between
  their friends and the target.
▶ Use a greedy algorithm + allow searches to fail
  randomly.

The model-results—searchable networks

\[
\alpha = 0 \text{ versus } \alpha = 2 \text{ for } N \sim 10^5:
\]

\[
q \geq r \quad q < r \quad r = 0.05
\]

q = probability an arbitrary message chain reaches a
  target.
▶ A few dimensions help.
▶ Searchability decreases as population increases.
▶ Precise form of hierarchy largely doesn’t matter.
The Small-World Phenomenon

History
An online experiment
Previous theoretical work
An improved model
References

Social search—Data

Adamic and Adar (2003)

- For HP Labs, found probability of connection as function of organization distance well fit by exponential distribution.
- Probability of connection as function of real distance $\propto 1/r$.

Social Search—Real world uses

- Tags create identities for objects
- Website tagging: http://www.del.icio.us
- (e.g., Wikipedia)
- Photo tagging: http://www.flickr.com
- Dynamic creation of metadata plus links between information objects.
- Folksonomy: collaborative creation of metadata

Recommender systems:

- Amazon uses people's actions to build effective connections between books.
- Conflict between 'expert judgments' and tagging of the hoi polloi.
Conclusions

- Bare networks are typically unsearchable.
- Paths are findable if nodes understand how network is formed.
- Importance of identity (interaction contexts).
- Improved social network models.
- Construction of peer-to-peer networks.
- Construction of searchable information databases.

References


