Search in networks

Argimiro Arratia & Ramon Ferrer-i-Cancho

Universitat Politècnica de Catalunya

Complex and Social Networks (2023-2024)
Master in Innovation and Research in Informatics (MIRI)

Instructors

- Ramon Ferrer-i-Cancho, rferrericancho@cs.upc.edu, `http://www.cs.upc.edu/~rferrericancho/`
- Argimiro Arratia, argimiro@cs.upc.edu, `http://www.cs.upc.edu/~argimiro/`

Please go to `http://www.cs.upc.edu/~csn` for all course's material, schedule, lab work, etc.

Milgram's small-world experiment

[Milgram, 1967, Travers and Milgram, 1969]

Instructions

Given a target individual (stockbroker in Boston), pass the message to a person you correspond with who is "closest" to the target

Outcome

20% of initiated chains reached target average chain length = 6.5

Small-world experiment revisited
[Dodds et al., 2003]

We report on a global social-search experiment in which more than **60,000 e-mail users** *attempted to reach one of* **18 target persons** *in* **13 countries** *by forwarding messages to acquaintances. We find that successful social search is conducted primarily through intermediate to weak strength ties, does not require highly connected "hubs" to succeed, and, in contrast to unsuccessful social search, disproportionately relies on professional relationships. By accounting for the attrition of message chains, we estimate that social searches can reach their targets in a median of* **five to seven steps**, *depending on the separation of source and target, although small variations in chain lengths and participation rates generate large differences in target reachability. We conclude that although global social networks are, in principle, searchable, actual success depends sensitively on individual incentives.*

Reflections on Milgram's experiment

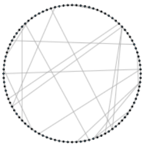
1. Short paths exist between random pairs of people ("six degrees of separation")
   - Explained by models with small diameter, e.g. Watts-Strogatz model

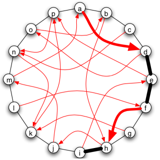
2. With little "local" information, people are **able to find them**
   - [Kleinberg, 2000b]: *The success of Milgram's experiment suggests a source of latent navigational "cues" embedded in the underlying social network, by which a message could implicitly be guided quickly from source to target. It is natural to ask what properties a social network must possess in order for it to exhibit such cues, and enable its members to find short chains through it.*

Reproducing Milgram's result: Kleinberg's model

[Kleinberg, 2000b, Kleinberg, 2000a]

Variation on Watts-Strogatz small-world model[1]





- n nodes arranged on a ring
- each node connects to immediately adjacent nodes
  - mimics "local" information
- each node has an additional long-range **shortcut**
  - $Prob(\text{shortcut from } u \text{ to } v) \propto d(u, v)^{-\alpha}$
  - α is a parameter, the "clustering exponent"
    - if $\alpha = 0$, like WS model
    - if $\alpha > 0$, preference for *closer* nodes

[1]Originally defined on a 2D grid, here explained with 1D ring for simplicity.

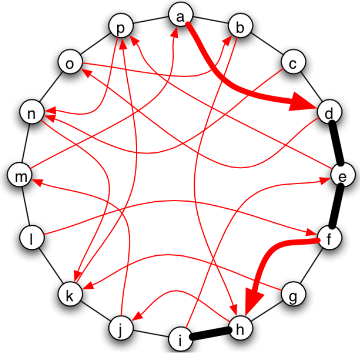
Myopic search

Goal: to show that networks are navigatable: for any i, j, there are shortest paths and to find them one do not need to know all the network, but only use *local information*.

Given a source node s and a destination node t, the decentralized algorithm works as follows:

1. Each node has a coordinate and knows its position on the ring, included the positions of s and t ("geographical" information)
2. Each node knows its neighbors and its shortcut ("local" information)
3. Each node forwards the "message" *greedily*, each time moving as close to t as possible

Myopic search
Example



- Source $s = a$; destination $t = i$
- Myopic search selects path $a - d - e - f - h - i$ (length 5)
- Shortest path is $a - b - h - i$ (length 3)

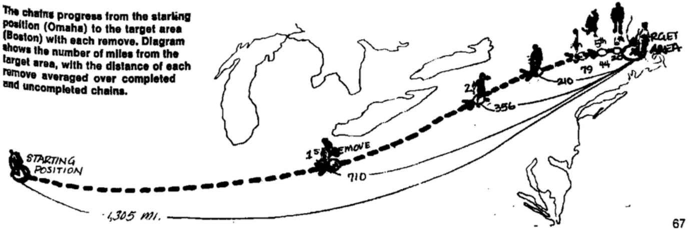
For $\lim_{n\to\infty}$, the expected number of steps needed to reach target $E[X]$, is:

$$E[X] = \begin{cases} \Omega(n^{1-\alpha}) & \alpha < 1 \\ O(\log^2 n) & \alpha = 1 \\ \Omega(n^{\alpha-1}) & \alpha > 1 \end{cases}$$

Fast myopic search with $\alpha = 1$

Intuition

From [Milgram, 1967]:



"Track how long it takes to for the message to reduce its distance by factors of 2"

- X_j is the nr. of steps taken in *phase j*
- Phase j: portion of the search in which message is at distance between 2^j and 2^{j+1}
- Will show that $E[X_j] = O(\log(n))$ for each j

$$E[X] = E[X_1] + E[X_2] + .. + E[X_{\log n}]$$

Normalizing constant for $\alpha = 1$

What is the probability distribution, exactly?

$$P[\text{shortcut from } u \text{ to } v] \propto \frac{1}{d(u, v)}$$

Need to figure out normalizing constant $Z = \sum_v \frac{1}{d(u,v)}$ for the distribution of shortcuts for node u.

Fix arbitrary node u. Then, there are 2 nodes at distance 1, 2 at distance 2, and in general 2 at each distance up to $n/2$:

$$\begin{aligned} Z &= 2\left(1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n/2}\right) \\ &\leqslant 2(1 + \ln(n/2)) \\ &\leqslant 2(1 + \log_2(n/2)) \\ &= 2\log_2(n) \end{aligned}$$

Halving the distance to destination is quick for $\alpha = 1$ I

- Assume we are at node v in phase j, somewhere at distance d from destination for $2^j \leqslant d \leqslant 2^{j+1}$
- If current node v has shortcut to node at distance at most $d/2$, then we are done (since current node takes shortcut and search leaves phase j and goes into phase $j-1$ or better)
- There are $d+1$ nodes at distance $d/2$ from destination; these nodes are at distance at most $d + d/2 = 3d/2$ from v
- Probability of v having a shortcut to any one of these $d+1$ nodes (call it w) is
  $Prob[v \text{ shortcuts to } w] = \frac{1}{Z}\frac{1}{d(v,w)} \geqslant \frac{1}{2\log(n)}\frac{1}{3d/2} = \frac{1}{3d\log(n)}$
- The probability that v shortcuts to *any* one of them is at least $\frac{1}{3\log(n)}$ (since there are $d+1$ of them)

Halving the distance to destination is quick for $\alpha = 1$ II

- After each step, the probability of leaving phase j is at least $\frac{1}{3\log(n)}$ so the probability of staying in phase j for i steps is at least $(1 - \frac{1}{3\log(n)})^{i-1}$ and so $P[X_j \geqslant i] \leqslant (1 - \frac{1}{3\log(n)})^{i-1}$
- Now,

$$\begin{aligned}
E[X_j] &= \sum_{k \geqslant 0} k \times P[X_j = k] \\
&= 1P[X_j = 1] + 2P[X_j = 2] + 3P[X_j = 3] + .. \\
&= P[X_j \geqslant 1] + P[X_j \geqslant 2] + P[X_j \geqslant 3] + .. \\
&\leqslant 1 + \left(1 - \frac{1}{3\log(n)}\right)^1 + \left(1 - \frac{1}{3\log(n)}\right)^2 + .. \\
&= 3\log(n)
\end{aligned}$$

where the last step is due to the geometric series $\frac{1}{1-x} = \sum_{n \geqslant 0} x^n$ Q.E.D.

References I

Barabasi, A.-L. and Albert, R. (1999).
Emergence of scaling in random networks.
Science, 286(5439):509–512.

Dodds, P. S., Muhamad, R., and Watts, D. J. (2003).
An experimental study of search in global social networks.
science, 301(5634):827–829.

Kleinberg, J. M. (2000a).
Navigation in a small world.
Nature, 406(6798):845–845.

Kleinberg, J. M. (2000b).
The small-world phenomenon: an algorithm perspective.
In *Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing, May 21-23, 2000, Portland, OR, USA*, pages 163–170.

References II

Milgram, S. (1967).
The small world problem.
Psychology today, 2(1):60–67.

Pastor-Satorras, R. and Vespignani, A. (2001).
Epidemic spreading in scale-free networks.
Phys. Rev. Lett., 86(14):3200–3203.

Travers, J. and Milgram, S. (1969).
An experimental study of the small world problem.
Sociometry, pages 425–443.