Lecture 16: Small world…?

INFO 2950:
Mathematical Methods for
Information Science

Small world

What do we mean when we say "Small world"?

- Ex. in Hawaii - see someone you know
  - Talking to some one, have a friend in common

Fred Jones of Peoria, sitting in a sidewalk cafe in Tunis, and needing a light for his cigarette, asks the man at the next table for a match. They fall into conversation; the stranger is an Englishman who, it turns out, spent several months in Detroit studying the operation of an interchangeable-bottlecap-factory. "I know it's a foolish question," says Jones, "but did you ever by any chance run into a fellow named Ben Arkadian? He's an old friend of mine, manages a chain of supermarkets in Detroit . . ."

"Arkadian, Arkadian," the Englishman mutters. "Why, upon my soul, I believe I do! Small chap, very energetic, raised merry hell with the factory over a shipment of defective bottlecaps."

"No kidding!" Jones exclaims in amazement.

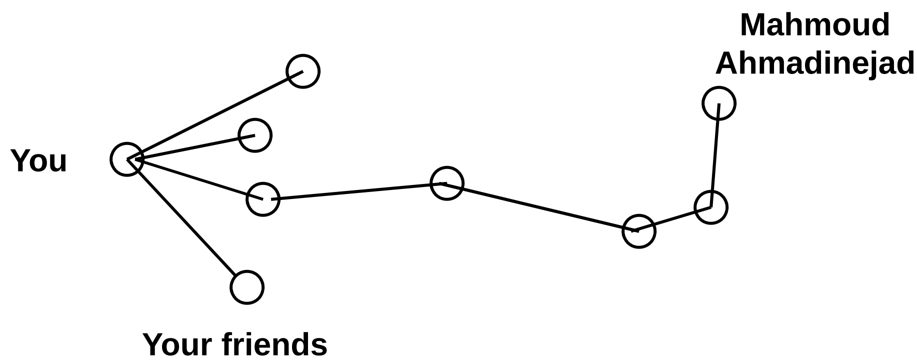
"Good lord, it's a small world, isn't it?"

What do we mean by a "small world phenomenon"? Or by "six degrees of separation"?

And what does this have to do with graph theory?

Six degrees

In the social network, there is a path of at most six edges.



Origins

From a famous paper of Stanley Milgram (1967), trying to find a basis for the “small world” experience.

Idea: Pick some “targets” and a large number of “starters”; try to get starters to forward a message to the targets.

Details

Targets:
1. Wife of divinity school student in Cambridge, MA
2. Stockbroker working in Boston, living in Sharon, MA

Starters:
1. From Wichita, KS
2. From Omaha, NE

Details

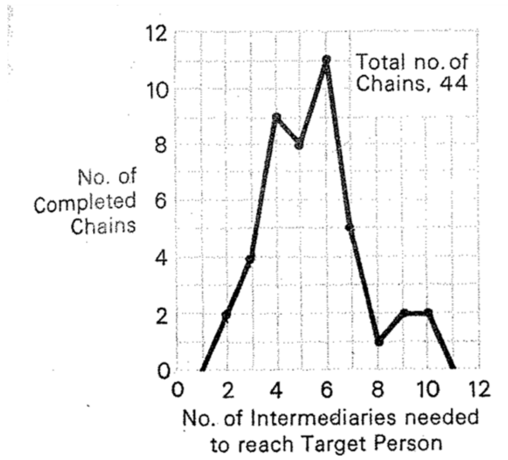
Directions: “Mail this folder to a personal acquaintance more likely than you to know the target person”.

Folder contains information about who the target person is, what he/she does, where he/she lives, etc.

Folder keeps track of who has forwarded it.

Each person forwarding sends a postcard to Milgram.

Average chain length



Average a bit bigger than 5; hence “6 degrees”.

But is it really six?

Suppose after folder is forwarded, then next person fails to forward with probability p.

Then do chains look longer or shorter than they are in reality?

Ideal • → • → • → • → Target

Real • → • → • ?

• → • → • → • Target

• → • → • → • → • p

Chains will look shorter in reality

More problems (Kleinfeld 2002)

How many chains actually completed?

1. 50 started, 3 completed, 8 intermediaries.
2. 217 started, 64 completed
   - 100 “blue chip” stock owners
   - 96 “random” (newspaper ad)
   - 100 from Boston

Very typical!

Korte-Milgram (1970)

- 18 targets in NYC: 9 white, 9 black
- 18 x 30 white starters in LA
  - 88 white target chains complete (33%)
  - 35 black target chains complete (13%)

More recently

Dodds, Muhamad, Watts (2003)

Email variant: 18 targets from 13 countries

- 98,000 people register to participate
- 24,000 chains are started
- 384 complete
- Average length: 4.05

Back to graph theory

Supposing “six degrees” is true, what does that mean that the social network is like?

- Connected
- Edges dense - needed?
- Shortest path between any two nodes ≤ 6

$\Rightarrow$ diameter of $G \leq 6$.

What kinds of random graphs might look like this?

Erdős-Rényi graph: each edge present with probability p for some p.

What is expected number of edges? Degree? Clustering coefficient?

Expected number of edges

$X =$ # of edges

$X_{(u,v)} = \begin{cases} 1 & \text{if edge is present} \\ 0 & \text{if not} \end{cases}$

$E[X_{(u,v)}] = p \cdot 1 + (1-p) \cdot 0 = p$

$E[X] = \sum_{(u,v) \in} E[X_{(u,v)}] = \binom{n}{2} p$

$p = \frac{1}{2} \Rightarrow$ graph is very dense

Expected degree

$X =$ # of neighbors of a node v

$E[X] = p(n-1)$

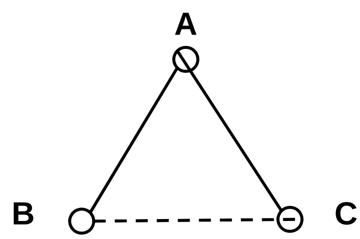
For social network, what should p be?

$p = \frac{1}{2} \Rightarrow$ know half the network, not realistic

$p = \frac{100}{n-1} \Rightarrow$ expect to know 100 people

Triangles?

Clustering coefficient: Given a random node A and two random edges (A,B) and (A,C), the probability that the edge (B,C) exists.



Expected clustering coefficient

Erdös-Renyi graphs: (B,C) exists w/ prob p.

If $p = \frac{60}{(n-1)}$, then prob. your friends know each other very low.

Some real networks

From Newman 2002:

Network	*n*	Mean degree *z*	Cc	Cc for random graph
Internet (AS level)	6,374	3.8	0.24	0.00060
WWW (sites)	153,127	35.2	0.11	0.00023
Power grid	4,941	2.7	0.080	0.00054
Biology collaborations	1,520,251	15.5	0.081	0.000010
Mathematics collaborations	253,339	3.9	0.15	0.000015
Film actor collaborations	449,913	113.4	0.20	0.00025
Company directors	7,673	14.4	0.59	0.0019
Word co-occurrence	460,902	70.1	0.44	0.00015
Neural network	282	14.0	0.28	0.049
Metabolic network	315	28.3	0.59	0.090
Food web	134	8.7	0.22	0.065

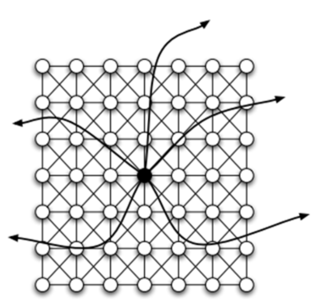
21

Slide credit: Dragomir Radev

Another model

Watts and Strogatz (1998):

People in a grid, and they know all their neighbors, plus a k random people far away.



Properties

Expected degree: $\delta+k$

Expected number of edges: $\frac{1}{2}n(\delta+k)$

Clustering coefficient: high

Paths

Can also show that there are likely to be short paths in the network between any pair of nodes.

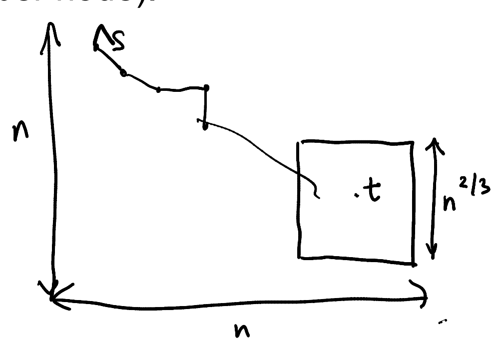
But

Can anyone find the short path?

How would you try to forward a message to someone in Sharon, MA?

- Same profession
- Same area
  or as close as you can get.

Spse we try to get from a vertex s to a vertex t by trying to get into a square around t as soon as possible. Assume k=1 (only one random link per node).



E_j = event that on the j^{th} step there's a random edge into the square B

$$p(E_j) = \frac{\#\text{ of nodes in square } B}{n^2} = \frac{n^{2/3} \times n^{2/3}}{n^2} = \frac{n^{4/3}}{n^2} = n^{-2/3}$$

A = event that any of the first $\frac{1}{2}n^{2/3}$ steps on path has an edge into B

$$A = \bigcup_{j=1}^{\frac{1}{2}n^{2/3}} E_j \qquad p(A) \leq \sum_{j=1}^{\frac{1}{2}n^{2/3}} p(E_j) = \left(\frac{1}{2}n^{2/3}\right)\left(n^{-2/3}\right) = \frac{1}{2}$$

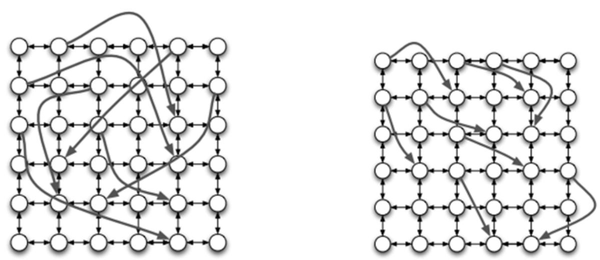
If we start outside B and event A does not occur need $\frac{1}{2}n^{2/3}$ steps

X = # of steps

$$E[X] = E[X|A]p(A) + E[X|\bar{A}]p(\bar{A})$$
$$\geq 0 + \frac{1}{2}n^{2/3} \cdot \frac{1}{2} = \frac{1}{4}n^{2/3}$$

Another model

Kleinberg (2000): Assume that random links are to nodes at distance d with probability proportional to d^{-2}.



Zeno's paradox

Basic idea: Same as Zeno's paradox.

First get to some node at distance $\frac{n}{2} \le \quad \le n$ from target

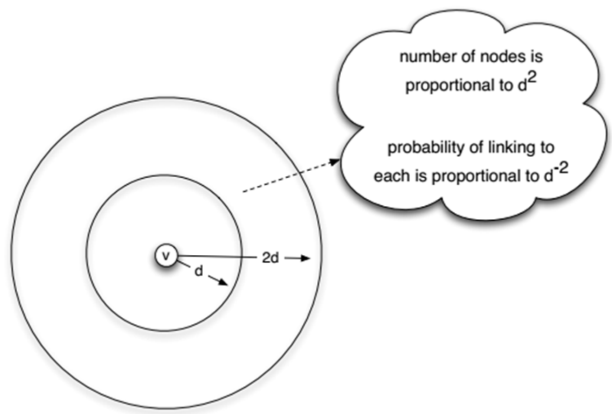
Then " " " " " " $\frac{n}{4} \le \quad \le \frac{n}{2}$

Then " $\frac{n}{8} \le \quad \le \frac{n}{4}$

$\vdots$

$1 \le \quad \le 2$

Basic idea for a step



Can show that the probability of an edge to some node at distance between d and 2d from target is c / $\log_2$ n for some constant c.

X_d = # of steps to get to nodes of distance $d \leq \;\; \leq 2d$ from target

$$E[X_d] = 1 \cdot p(X_d = 1) + 2 \cdot p(X_d = 2) + 3 \cdot p(X_d = 3) + \cdots$$

$$= p(X_d \geq 1) + p(X_d \geq 2) + p(X_d \geq 3) + \cdots$$

$$\leq 1 + \left(1 - \frac{c}{\log_2 n}\right) + \left(1 - \frac{c}{\log_2 n}\right)^2 + \cdots$$

$$= \frac{1}{1 - \left(1 - \frac{c}{\log_2 n}\right)} = \frac{\log_2 n}{c}$$

X = # of steps taken in total

$$X = X_{n/2} + X_{n/4} + X_{n/8} + \cdots + X_1 + 2$$

$$E[X] = E[X_{n/2}] + E[X_{n/4}] + E[X_{n/8}] + \cdots + E[X_1] + 2$$

$$\leq \frac{\log_2 n}{c} + \frac{\log_2 n}{c} + \frac{\log_2 n}{c} + \cdots + \frac{\log_2 n}{c} + 2$$

$$= \log_2 n \left(\frac{\log_2 n}{c}\right) + 2 = \frac{(\log_2 n)^2}{c} + 2$$

Reality

Some evidence that the online world (at least) is really like this (random links with prob. proportional to d^{-2}).

Liben-Nowell, Novak, Kumar, Raghavan, Tomkins 2005: LiveJournal

Backstrom, Sun, Marlow 2010: Facebook

Summary

- Is it a small world? A bit hard to know.
- Some evidence that the world is structured so that Milgram's experiment can work.
- A little probability and graph theory can take you a long way.

Have a great spring break!