Network Science: Small world networks and the navigation problem

Ozalp Babaoglu
Dipartimento di Informatica — Scienza e Ingegneria
Università di Bologna
www.cs.unibo.it/babaoglu/

Small-world networks

- An experimental study of the small world problem, Travers and Milgram, Sociometry 1969
- Abstract: *Arbitrarily selected individuals (N=296) in Nebraska and Boston are asked to generate acquaintance chains to a target person in Massachusetts, employing "the small world method" (Milgram, 1967). Sixty-four chains reach the target person. Within this group the mean number of intermediaries between starters and targets is 5.2. Boston starting chains reach the target person with fewer intermediaries than those starting in Nebraska; subpopulations in the Nebraska group do not differ among themselves. The funneling of chains through sociometric "stars" is noted, with 48 per cent of the chains passing through three persons before reaching the target. Applications of the method to studies of large scale social structure are discussed.*
- One of the earliest instances of "crowdsourcing"

Small-world networks Travers-Milgram

- Structural:

given two individuals selected randomly from the population, what is the probability that the minimum number of intermediaries required to link them is 0, 1, 2, . . . *k*?

- Algorithmic:

Perhaps the most direct way of attacking the small world problem is to trace a number of real acquaintance chains in a large population. This is the technique of the study reported in this paper.

Small-world networks Travers-Milgram — methodology

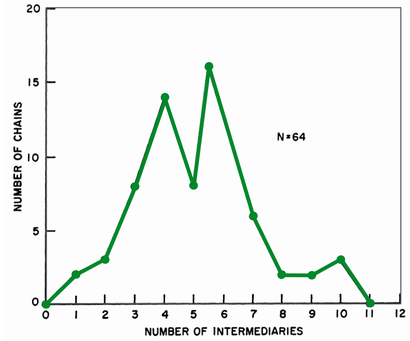
- Arbitrary "target" person and a group of "starters" selected
- Each starter given a document and asked to start moving it by mail towards the target
- The document described the experiment, named the target and asked the recipient to participate by forwarding it
- Document could be forwarded only to a first-name-based acquaintance of the sender
- Sender urged to choose recipient to advance progress of document towards target along an acquaintance chain
- Chain would end by either by reaching the target or when someone along the way declined to participate
- Information about the target (stockbroker in Boston) given to guide the choice of next recipient

Small-world networks
Travers-Milgram — methodology

- Starters: 296 volunteers total, 196 were residents of Nebraska, while 100 were recruited from the Boston area

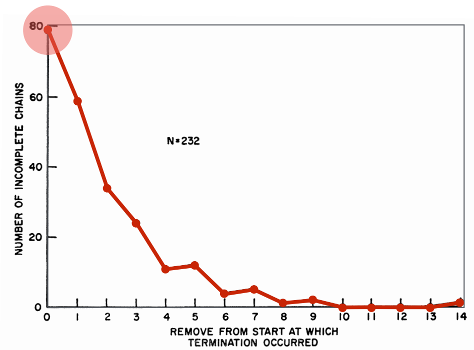
Small-world networks
Travers-Milgram — results

- Distribution of lengths of completed chains
- Only 64 out of the 296 initial chains completed



Small-world networks
Travers-Milgram — results

- Number of chains that die after making some progress



Small-world networks
Travers-Milgram — results

- Many of the completed chains passed through a very small number of penultimate individuals — "funnels"
- A certain Mr. G. responsible for forwarding 16 (out of 64) chains to the target
- Mr. D. and Mr. P. responsible for 10 and 5 chains, respectively
- "Connectors" or "hubs" with high degree often exist in social networks
- Target need not be a "connector" for small-world phenomenon to exist
- Like "hub" airports in air traffic

Small-world networks
Columbia Small Worlds Project

- *An Experimental Study of Search in Global Social Networks*, Dodds et al., Science 2003
- Modern incarnation of Travers-Milgram
- Web-based, email tracking
- 18 targets from 13 countries
- On-line registration of participants, electronic tracking
- 99K persons registered, 24K initiated chains, only 384 reached targets

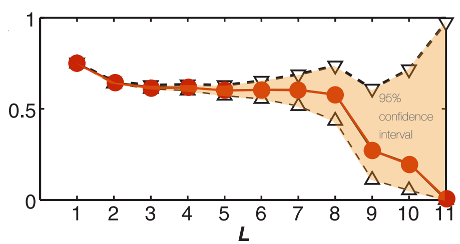
Small-world networks
Columbia Small Worlds Project

- Highlights of results:
  - Less than 5% of chains went through the same penultimate person (no “funneling”)
  - “Large degree” rarely a reason for forwarding choice (less than 10%)
  - Interesting “algorithmic” choices as a function of chain length (“geographic” early on, “work” later)
- Reason for choosing next recipient as a function of completed steps

L	*N*	Location	Travel	Family	Work	Education	Friends	Cooperative	Other
1	19,718	33	16	11	16	3	9	9	3
2	7,414	40	11	11	19	4	6	7	2
3	2,834	37	8	10	26	6	6	4	3
4	1,014	33	6	7	31	8	5	5	5
5	349	27	3	6	38	12	6	3	5
6	117	21	3	5	42	15	4	5	5
7	37	16	3	3	46	19	8	5	0

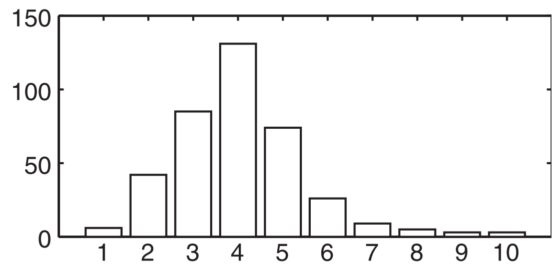
Small-world networks
Columbia Small Worlds Project

- Average attrition rates as a function of chain length



Small-world networks
Columbia Small Worlds Project

- Distribution of completed chain lengths (mean 4.05)

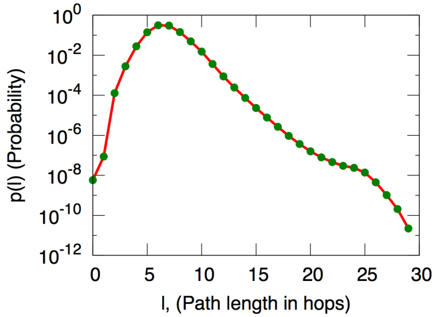


Small-world networks
Microsoft Instant Messenger

- *Worldwide buzz: Planetary-scale views on an instant-messaging network*, Leskovec and Horvitz, 2008
- "Structural" study based on 240M Microsoft IM user accounts active in 2008
- Two users considered "connected" if they communicated at least once during a month-long observation period
- No need for "tracers" since the full social graph is known
- Shortest paths computed on the graph using "breadth-first search"

Small-world networks
Microsoft Instant Messenger

- Single giant component
- Average shortest path distance 6.6, median 7
- Shortest path distribution for (only) 1000 users:

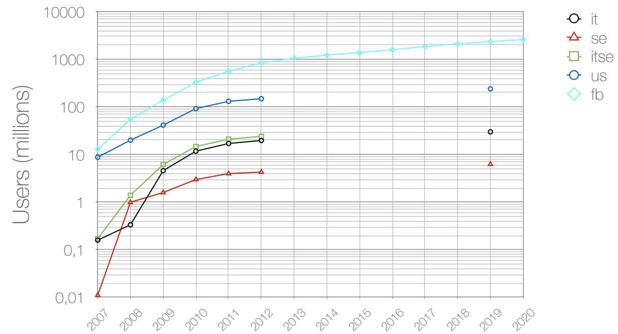


Small-world networks
Facebook study

- *Four degrees of separation*, L. Backstrom et al., 2012
- "Structural" study based on 721M active Facebook users with 69B friendship links
- Again, not a random sample from general population but by 2012, Facebook much more representative than IM in 2008
- Repeated in 2016 with 1.59B Facebook users
- The biggest technical feat of this study is the ability to process huge datasets algorithmically

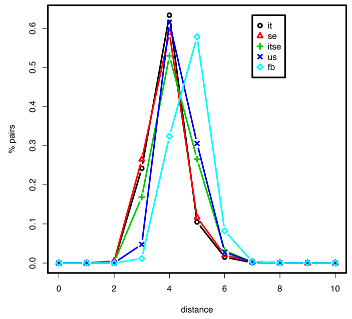
Small-world networks
Facebook study — results

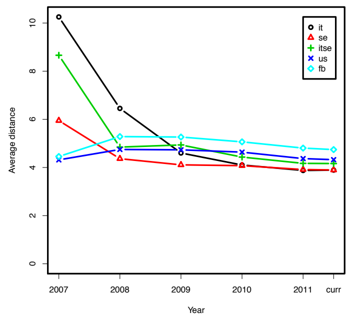
- Growth of active Facebook users



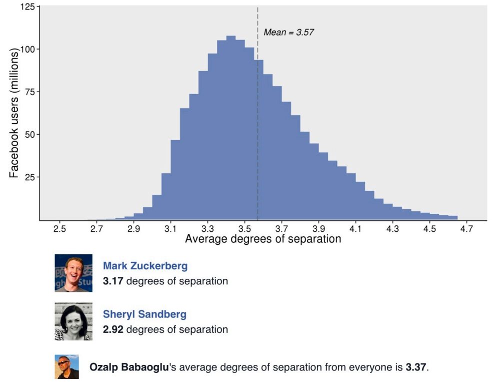
Small-world networks Facebook study — 2012 results

- Shortest path length: current distribution and averages over the years
- Overall average path length: 4.74





Small-world networks Facebook study — 2016 results



https://research.facebook.com/blog/three-and-a-half-degrees-of-separation/

The navigation problem

- Suppose you are a node in a very large social network
- You want to find a short path to another node in the network
- You do not have a global view of the network
- You only know who your immediate neighbors are
- You can ask your neighbors to make introductions
- Relevant not only for social networks but also in many technological contexts — Internet packet routing, peer-to-peer file sharing

The navigation problem

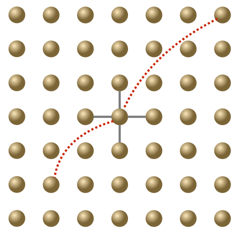
- Two aspects for solving the navigation (search) problem:
  - Verify the existence of short paths in the network — structural
  - Allow people to actually find these short paths using only distributed, local information — algorithmic
- Algorithmic constraints
  - Only know your immediate neighbors
  - Limited information about the target
  - Simple heuristic strategies

Small-world networks
Kleinberg's model

- Recall that to find short paths in networks
  - Short paths must exist (structural property — small diameter)
  - Must be able to find these short paths using only *local* forwarding information (algorithmic property)
- Kleinberg's model: abstract formulation of the navigation problem in a small-world network to study the *structural* and *algorithmic* constraints
- *Navigation in a small world*, J. Kleinberg, Nature 2000

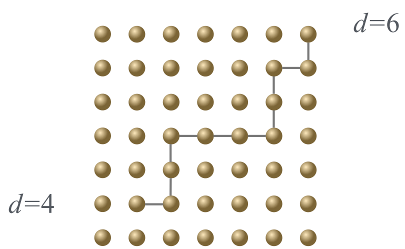
Small-world networks
Kleinberg's model — definition

- Start with a $k \times k$ regular grid of nodes ($n=k^2$)
- Each node connected to its 4 compass neighbors
- Each node gets one additional random "long-distance" edge



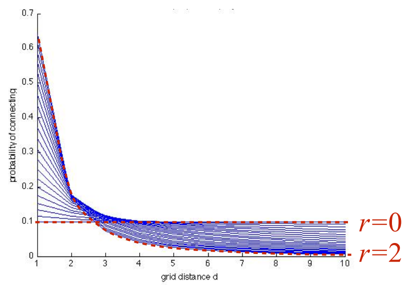
Small-world networks
Kleinberg's model — definition

- Let d denote the "grid (Manhattan) distance" between two nodes



Small-world networks
Kleinberg's model — definition

- Let the probability of the random edge connecting to a node at grid distance d be proportional to d^{-r} for some $r \geq 0$
  - Smaller r — most "long-distance" edges are uniform random
  - Larger r — most "long-distance" edges are actually "local"



Small-world networks
Kleinberg's model — constraints

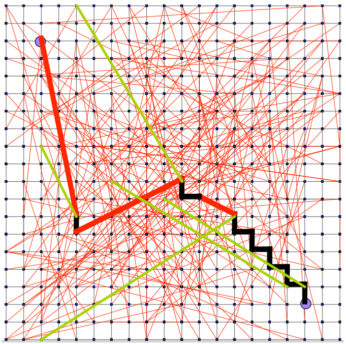
- Which values of r permit efficient navigation?
- "Efficient navigation" — the number of hops is bounded by a function $\log^a(n)$
- Choice of r constrains the problem *structurally*
- What are the *algorithmic* constraints?
  - Nodes know the coordinates of their neighbors
  - Nodes know the coordinate of the target
  - Nodes always forward to neighbors closest to target in grid distance ("greedy" strategy excludes "backwards" hops even though they may lead to shorter paths)
  - Forwarding based on *local geometric* information only (with global knowledge, the solution becomes trivial)

Small-world networks
Kleinberg's model — intuition

- If r is too *small* (no local bias), we can get close to the target quickly but then need to use grid edges to conclude
- If r is too *large* (strong local bias), then "long-distance" edges are actually local and do not help much — short paths may not even exist
- "Efficient" navigation requires a delicate mix of local and long-distance edges
- SmallWorldSearch NetLogo demo

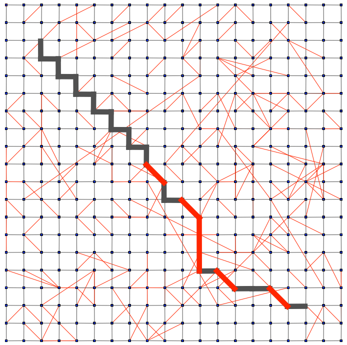
Small-world networks
Kleinberg's model — intuition

- r=0



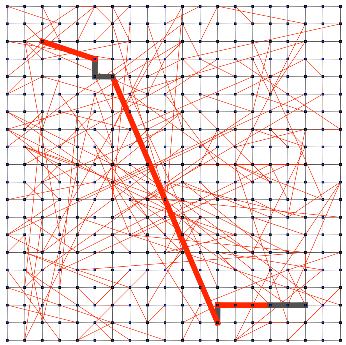
Small-world networks
Kleinberg's model — intuition

- r=4



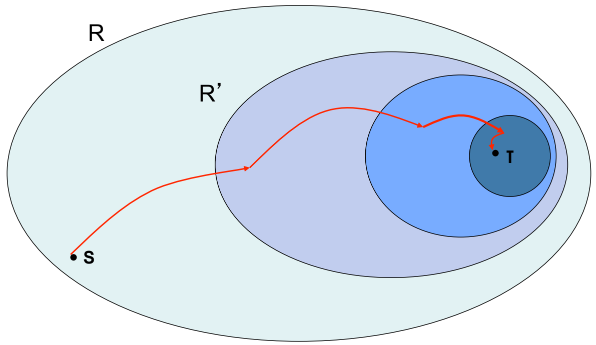
Small-world networks
Kleinberg's model — intuition

- $r=2$



Small-world networks
Kleinberg's model — intuition

- Navigability requires networks to be ***multiscale***

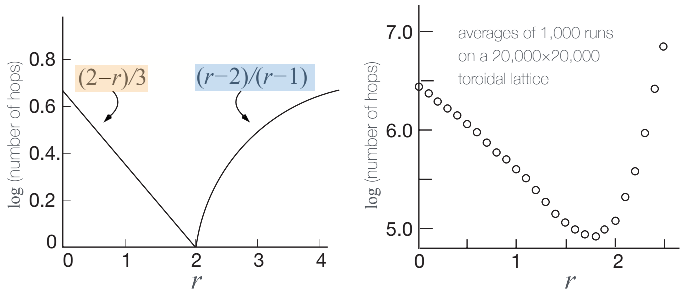


Small-world networks
Kleinberg's model — results

- As n becomes large, for *any* decentralized navigation algorithm, the expected number of hops is bounded by a function proportional to:
  - $n^{(2-r)/3}$ if $r<2$
  - $n^{(r-2)/(r-1)}$ if $r>2$
  - $\log^2 n$ if $r=2$
- Results can be generalized to d-dimensional lattices for any value of $d\geq1$
- The critical value becomes $r=d$

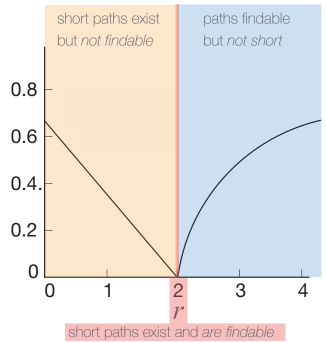
Small-world networks
Kleinberg's model — results

- Expected number of hops bounded by:
  - $n^{(2-r)/3}$ if $r<2$
  - $n^{(r-2)/(r-1)}$ if $r>2$



Small-world networks
Kleinberg's model — results

- For *any* decentralized navigation algorithm, expected number of hops is proportional to:
  - $n^{(2-r)/3}$ if $r<2$
  - $n^{(r-2)/(r-1)}$ if $r>2$
  - $\log^2 n$ if $r=2$

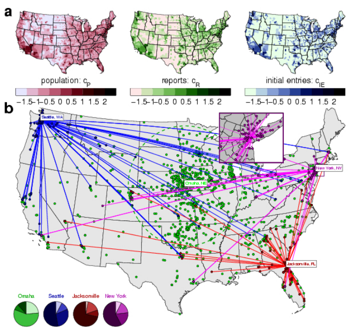


Small-world networks
Where's George

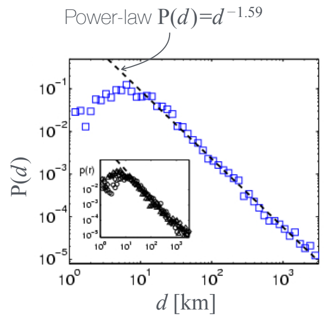
- Further confirmation of Kleinberg's results
- *The scaling laws of human travel,* Brockmann et al., Nature 2006
- Based on the "Where's George?" dataset
- Tracks movement of dollar bills
- Illustration of ***multiscale networks***
- Idea: movement of dollar bills can be a good proxy for movement of people

Small-world networks
Where's George

- https://youtu.be/kn32vavZqvg?t=28
- Movement of 4 dollar bills originating in 4 different cities



Small-world networks
Where's George



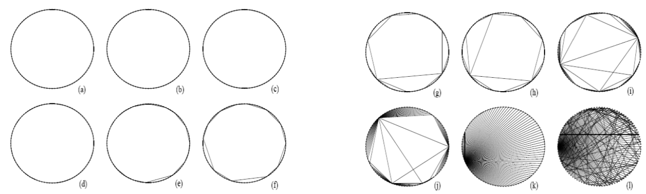
$P(d)$=probability of traversing distance d in 4 days

Small-world networks
Physical models

- Why do small-world networks form in the physical world?
- A model
- Each network has an associated "energy level" which the topology tries to minimize
- Define the energy level E as a weighted sum of two terms:
  $E = \lambda L + (1-\lambda) W$
  where L is the average shortest distance in hops, W is the average Euclidian distance (in meters) and λ is a parameter between 0 and 1

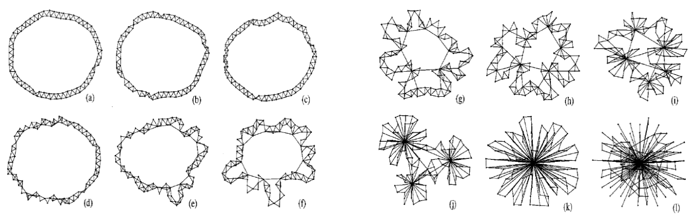
Small-world networks
Physical models

- Varying λ from 0 to 1
- Optimization through "simulated annealing"

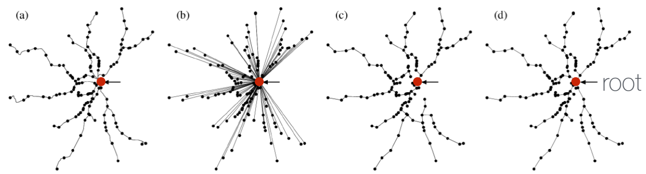


Small-world networks
Physical models

- Varying λ from 0 to 1
- Allow the nodes to move in physical space using a "spring" algorithm

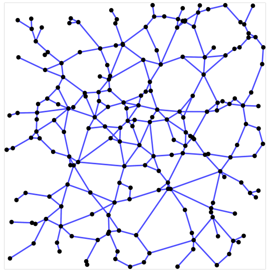


Small-world networks
Physical models

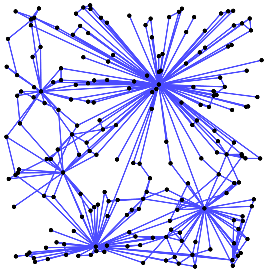


(a) Commuter rail network in the Boston area
(b) Star graph
(c) Minimum spanning tree
(d) The model applied to the same set of stations

Small-world networks
Physical models



Highways



Air routes