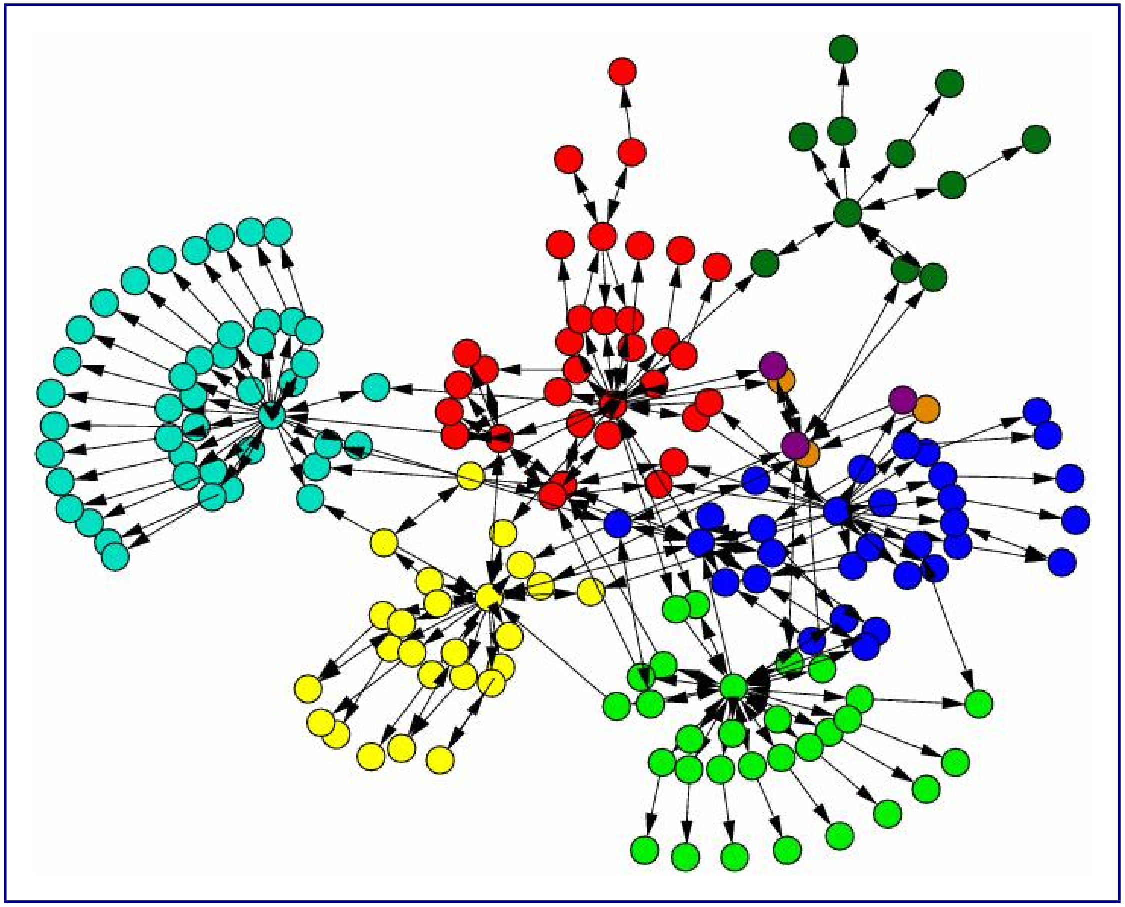
ECS 253 / MAE 253, Lecture 8

April 21, 2016



“Web search and decentralized search on small-world networks”

Search for information

Assume some resource of interest is stored at the vertices of a network:

- Web pages
- Files in a file-sharing network

Would like to determine rapidly where in the network a particular item of interest can be found.

To warehouse data or search on demand?

- Centralized : Catalogue data in one central place.

  – Makes most sense when high cost to search network in real time.

  – Requires resources for learning the data and storing it.

- Decentralized : Data is spread out in a distributed data base.

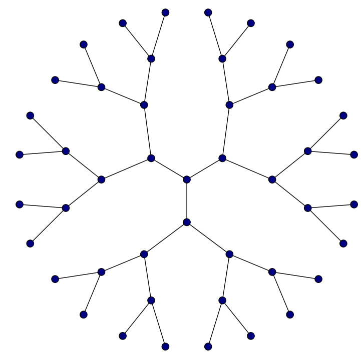
  – Can be a very slow process to search.

  – But dependent on network topology may be able to devise "quick" algorithms.

Distributed search

Some resource of interest is stored at the vertices of a network: i.e., Files in a file-sharing network

Search on arbitrary networks: $O(N)$ (search every node)



- Depth-first
- Breadth-first

Search on power-law random graphs

- Breadth-first passing always to highest-degree node possible. [Adamic, R. M. Lukose, A. R. Puniyani, and B. A. Huberman, "Search in power-law networks", Phys. Rev. E, 64 2001], find between $O(N^{2/3})$ and $O(N^{1/2})$.

Decentralized search on small-world networks

Consider a network with small diameter.

Will this help with decentralized search? (i.e. Find a local algorithm with performance less than $O(N)$?)

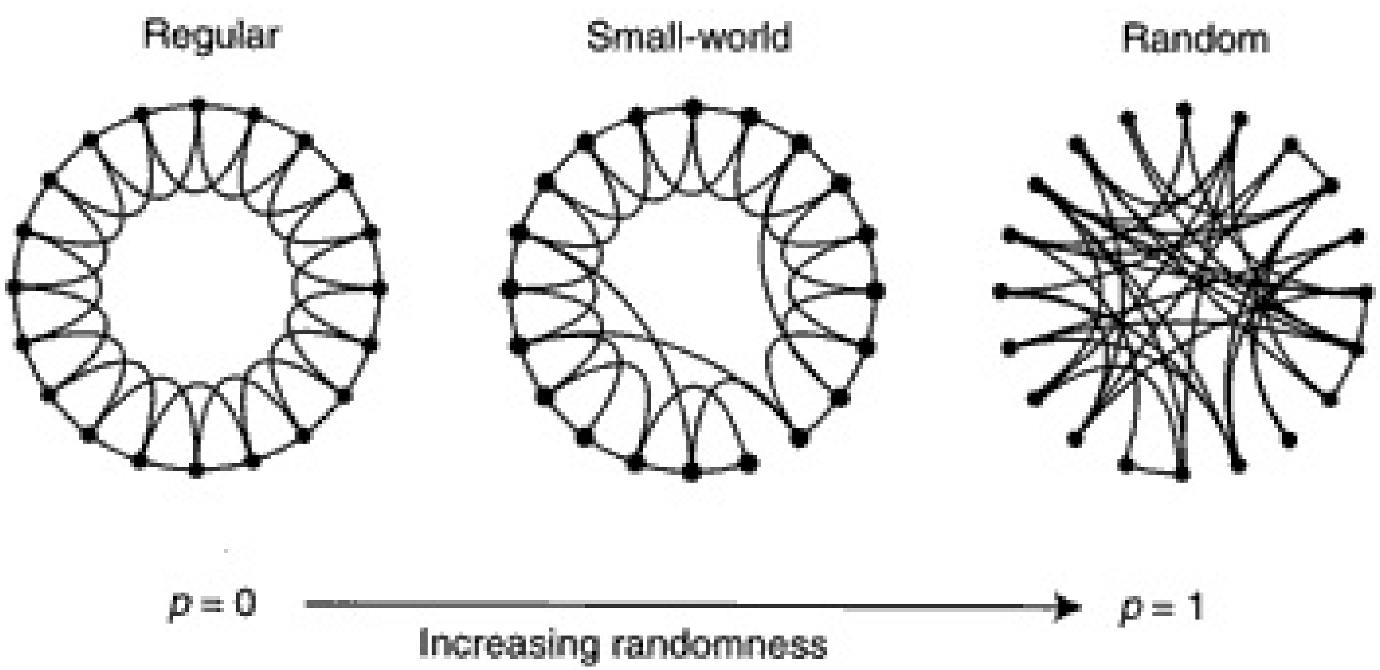
What is a small world?

- Dates back at least to 1929, Frigyes Karinthy, “Everything is different”
  – The modern world was “shrinking”
  – Due to technological advances in communications and travel, friendship networks could grow larger and span greater distances
- Michael Gurevich, 1961 MIT PhD dissertation
- Manfred Kochen, 1978: For the US size population, “it is practically certain that any two individuals can contact one another by means of at most two intermediaries”
- Stanley Milgram, “The Small World Problem” in *Psychology Today*, 1967.
- John Guare, “Six Degrees of Separation” a play in 1990 and film in 1993.
- Watts and Strogatz, *Nature* 1998

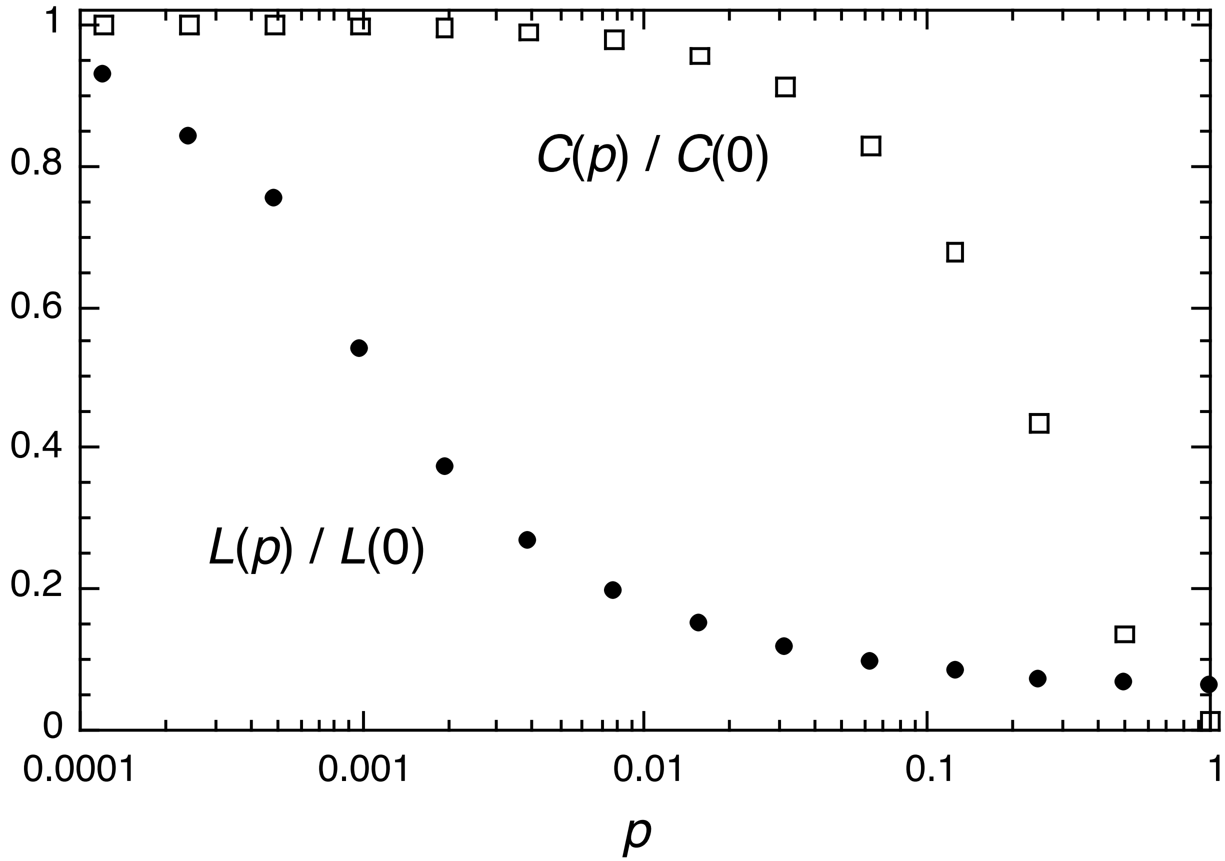
Mathematically defining a small-world

[Watts and Strogatz, "Collective dynamics of 'small-world' networks", *Nature*, 393 (1998)]

- Start with regular 1D lattice, with each node connected to it's k nearest neighbors.
- Randomly re-wire each link independently with probability p.



Ave shortest path $L(p)$ and clustering coefficient $C(p)$

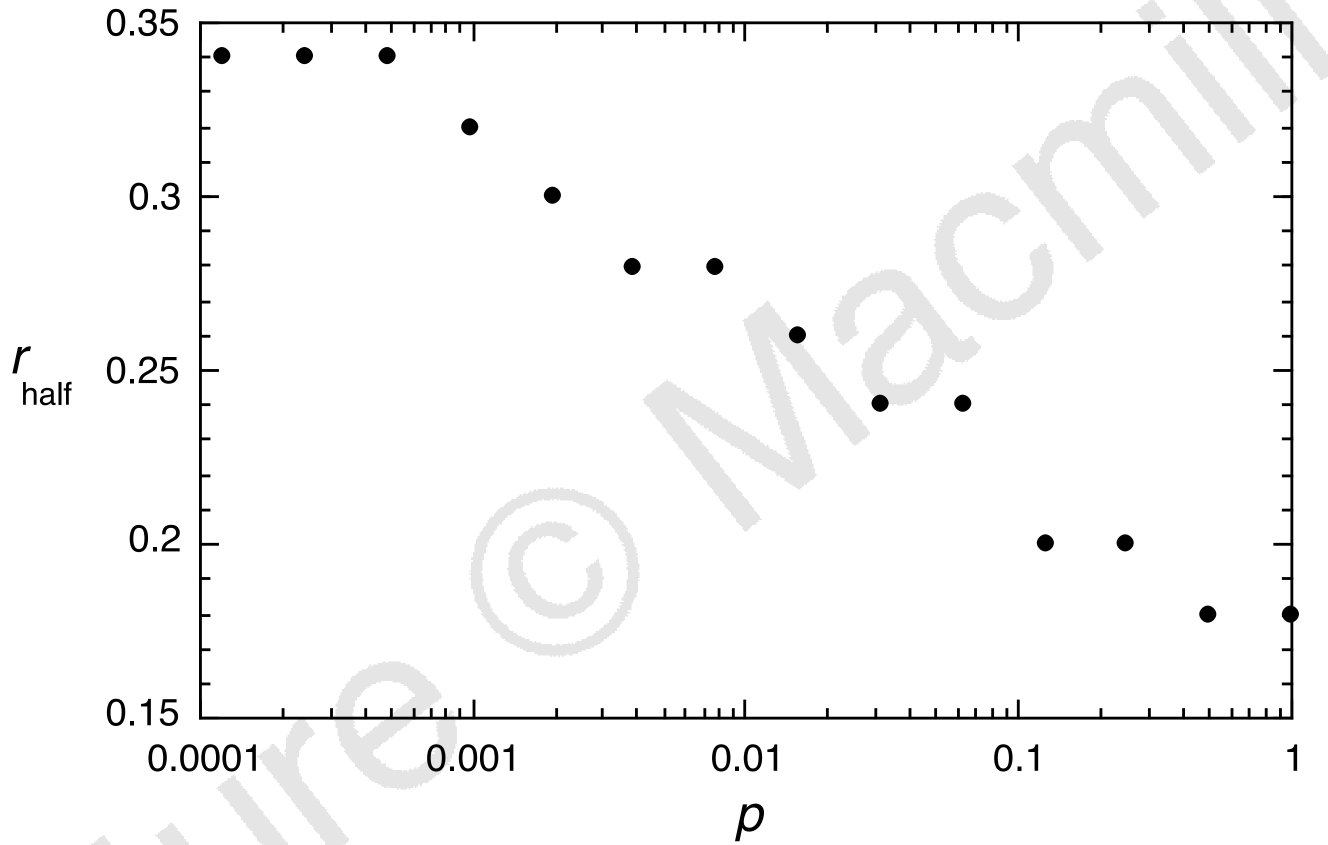


- Small-worlds have small diameter and large clustering coefficient.
- They are remarkably easy to generate (just a tiny p required).

Table 1 Empirical examples of small-world networks

	L_{actual}	L_{random}	C_{actual}	C_{random}
Film actors	3.65	2.99	0.79	0.00027
Power grid	18.7	12.4	0.080	0.005
C. elegans	2.65	2.25	0.28	0.05

Infection rates on Small-worlds



- Start with *one* infected individual.
- r is transmission rate of disease (infect neighbors at rate r).
- $r_{\rm half}$ is the the value of r required for half the population to get the disease.

Their concluding words

"We hope that our work will stimulate further studies of small-world networks. Their distinctive combination of high clustering with short characteristic path length cannot be captured by traditional approximations such as those based on regular lattices or random graphs. Although small-world architecture has not received much attention, we suggest that it will probably turn out to be widespread in biological, social and man-made systems,often with important dynamical consequences."

“Six degrees” (i.e., a small world)

- Six Degrees of Separation — 1993 Dramatic Film
- Six Degrees: The Science of a Connected Age. — 2003 book by Duncan Watts

Fun related websites

- The Oracle of Bacon – The path to Kevin Bacon
  http://oracleofbacon.org/

- SixDegrees.org – connecting causes and celebrities
  http://www.sixdegrees.org/

- Six degrees of Kevin Garnett
  http://www.slate.com/articles/sports/slate_labs/2013/10/six_degrees_of_kevin_garnett_connect_any_two_athletes_who_ve_ever_played.html

- Six degrees of NBA seperation
  http://harvardsportsanalysis.wordpress.com/2011/03/04/six-degrees-of-nba-separation/
  (Blog post explaining use of Dijkstra's algorithm)

Watts-Strogatz small-world model

- Together with Barabasi-Albert launched the flurry of activity on networks.

- Watts and Strogatz showed that networks from both the natural and manmade world, such as the neural network of C. elegans and power grids, exhibit the small-world property.

- Originally they wanted to understand the synchronization of cricket chirps.

- Introduced the mathematical formalism, which *interpolates between lattices and networks* .

A new paradigm

”I think I’ve been contacted by someone from just about every field outside of English literature. I’ve had letters from mathematicians, physicists, biochemists, neurophysiologists, epidemiologists, economists, sociologists; from people in marketing, information systems, civil engineering, and from a business enterprise that uses the concept of the small world for networking purposes on the Internet.” – Duncan Watts

Navigation

Clearly if central coordination, can use short paths to deliver info quickly.

But, can someone living in a small world actually make use of this info and do efficient decentralized routing?

- Instead of designing search algorithms, given a local greedy algorithm, are there any topologies that enable $O(\log N)$ delivery times?

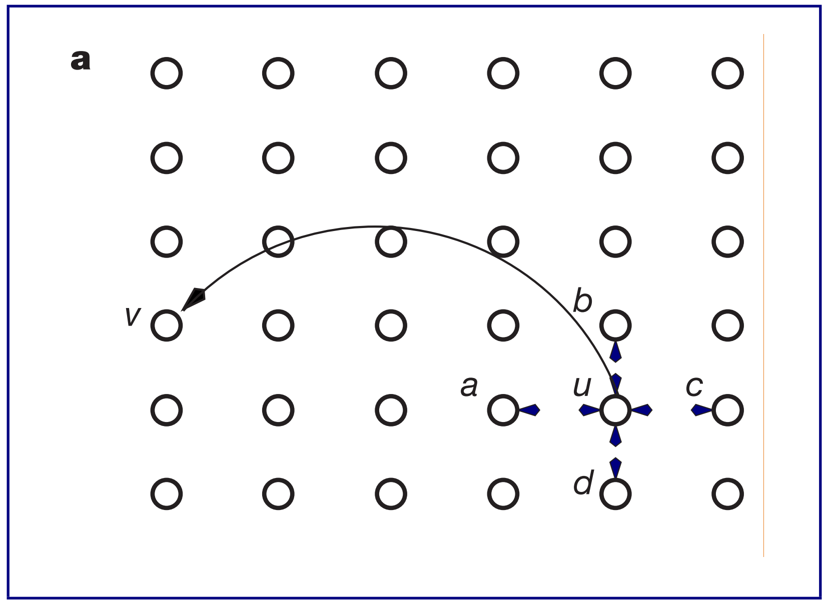
Precise topologies required

[J. M. Kleinberg, “Navigation in a small world”, *Nature*, 406 (2000)]

- Start with a regular 2D square lattice (consider vertices and edges).
- Add random long links, with bias proportional to distance between two nodes:

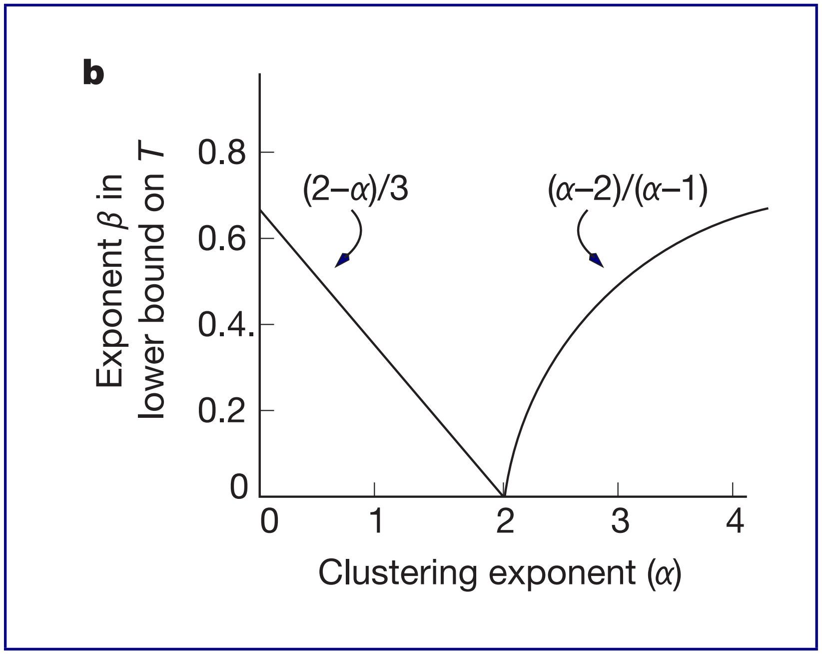
$$\boxed{p(e_{ij}) \propto 1/d_{ij}^{\alpha}}$$

- Call α the “clustering exponent”


a
v
b
a
u
c
d

- Find mean delivery time $t \sim N^{\beta}$, unless $\alpha = 2$.

- Only for $\alpha = 2$ will decentralized routing work, and packet can go from source to destination in $O(\log N)$ steps.



- For d-dimensional lattice need $\alpha = d$.

But we know greedy decentralized routing works for human networks (c.f. Milgram's experiments "six-degrees of separation" [S. Milgram, "The small world problem", *Psych. Today*, 2, 1967.]

So how do we get beyond a lattice model?

Navigating social networks

[Watts, P. S. Dodds, and M. E. J. Newman, "Identity and search in social networks", Science, 296 (2002)]

[Kleinberg, "Small world phenomena and the dynamics of information", in Proceedings of NIPS 2001].

- Premise: people navigate social networks by looking for common features between there acquaintances and the targets (occupation, city inhabited, age, ...)
- Brings in DATA!

Hierarchical “social distance” tree

- Individuals are grouped into categories along many attributes.
- One tree for each attribute.
- Trees are not the network, but complementary mental constructs believed to be at work.
- Assume likelyhood of acquaintance falls of exponentially with “social distance”.

Building P2P architectures

- “Chord A Scalable Peer-to- peer Lookup Service for Internet Applications

  I. Stoica, R. Morris, D. Karger, F. Kaashoek, H. Balakrishnan, ACM SIGCOMM, 2001.
  (Cited 4825 times)

- Gnutella

  “Peer-to-Peer Architecture Case Study: Gnutella Network”, M Ripeanu, Proceedings of International Conference on Peer-to-peer Computing, 2001.

Summary

Web search

- Centralized
- Make use of link structure (topology)

Decentralized search

- Efficiency/Speed depends on underlying topology
- Gossip algorithms (D. Kempe and J. Kleinberg): spreading shared information quickly through local exchanges (e.g., sums, local averages/consensus).
- Applications to sensor networks .... communications ... satellites