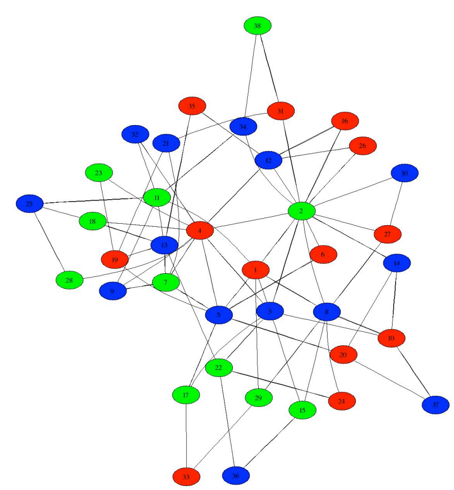
Navigation in Networks

Networked Life
NETS 112
Fall 2017
Prof. Michael Kearns

The Navigation Problem

- You are an individual (vertex) in a very large social network
- You want to find a (short) chain of friendships to another individual
- You don't have huge computers and a global/bird's-eye view
- All you (hopefully) know is who your neighbors/friends are
  - …and perhaps information about them (age, interests, religion, address, job,…)
- You can ask your friends to make introductions, which lead to more
- How would you do it?
- Also known as search in networks and the "small world problem"
- Small diameter is necessary but not sufficient!
  - …navigation is an algorithmic problem
- Related to the problem of routing data packets in the Internet



Small Worlds and the Law of the Few

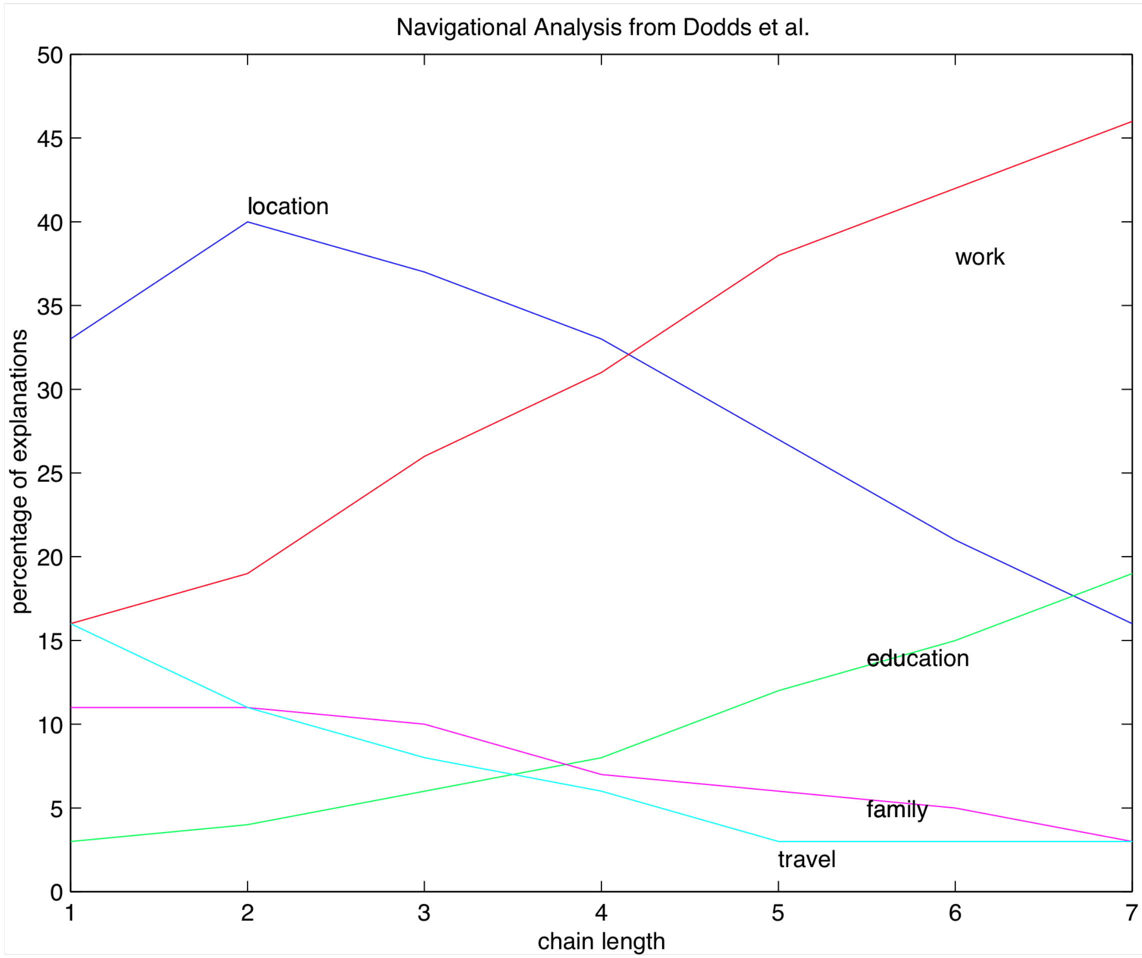
- Travers & Milgram 1969: classic early social network study
  - destination: a Boston stockbroker; lived in Sharon, MA
  - sources: Nebraska stockowners; Nebraska and Boston “randoms”
  - forward letter to a first-name acquaintance “closer” to target
  - target information provided:
    - name, address, occupation, firm, college, wife’s name and hometown
    - navigational value?
- Basic findings:
  - 64 of 296 chains reached the target
  - average length of *completed* chains: 5.2
    - interaction of chain length and navigational difficulties
  - main approach routes: home (6.1) and work (4.6)
  - Boston sources (4.4) faster than Nebraska (5.5)
  - no advantage for Nebraska stockowners

The Connectors to the Target

- T & M found that many of the completed chains passed through a very small number of penultimate individuals
  - Mr. G, Sharon merchant: 16/64 chains
  - Mr. D and Mr. P: 10 and 5 chains
- Connectors are individuals with extremely high degree
  - why should connectors exist?
  - how common are they?
  - how do they get that way? (see Gladwell for anecdotes)
- Connectors can be viewed as the “hubs” of social traffic
- Note: no reason *target* must be a connector for small worlds
- Two ways of getting small worlds (low diameter):
  - truly random connection pattern → dense network
  - a small number of well-placed connectors in a sparse network

Small Worlds: A Modern Experiment

- The Columbia Small Worlds Project:
  - considerably larger subject pool, uses email
  - subject of Dodds et al. assigned paper
- Basic methodology:
  - 18 targets from 13 countries
  - on-line registration of initial participants, all tracking electronic
  - 99K registered, 24K initiated chains, 384 reached targets
- Some findings:
  - < 5% of messages through any penultimate individual
  - large "friend degree" rarely (< 10%) cited
  - Dodds et al: → no evidence of connectors!
    - (but could be that connectors are not cited for this reason…)
  - interesting analysis of reasons for forwarding
  - interesting analysis of navigation method vs. chain length



Navigational Analysis from Dodds et al.
50
45
40
35
30
25
20
15
10
5
0
percentage of explanations
1
2
3
4
5
6
7
chain length
location
work
education
family
travel

The Strength of Weak Ties

- Not all links are of equal importance
- Granovetter 1974: study of job searches
  - 56% found current job via a personal connection
  - of these, 16.7% saw their contact "often"
  - the rest saw their contact "occasionally" or "rarely"
- Your "closest" contacts might not be the most useful
  - similar backgrounds and experience
  - they may not know much more than you do
  - connectors derive power from a large fraction of weak ties
- Further evidence in Dodds et al. paper
- T&M, Granovetter, Gladwell: multiple "spaces" & "distances"
  - geographic, professional, social, recreational, political,…
  - we can reason about general principles without precise measurement

The Magic Number 150

- Social channel capacity
  - correlation between neocortex size and group size
  - Dunbar's equation: neocortex ratio → group size
- Clear implications for many kinds of social networks
- Again, a *topological* constraint on typical degree
- From primates to military units to Gore-Tex

Neocortex size and group size in primates

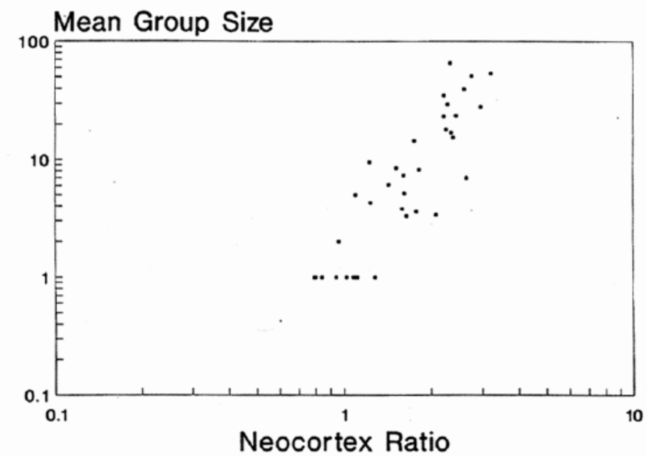


Figure 1. Group size plotted against neocortex ratio for nonhuman primates (redrawn from Dunbar 1992a).

Summary,and a Mathematical Digression

- So far:
  - large-scale social networks reliably have high-degree vertices
  - large-scale social networks have small diameter
  - furthermore, people can find or navigate the short paths from only local, distributed knowledge
  - these properties are true of other types of networks, too

- But there must be some limits to degrees
  - can't be "close friends" with too many people (150? 1000?)
- Large N, small diameter and limited degrees are in tension
  - not all combinations are possible
- Let N be population size, Delta be the maximum degree, and D be the diameter
- If Delta = 2 then must have D ~ N/4 (>> 6, >> log(N))

Summary,and a Mathematical Digression

- The relationship between D, Delta and N has been studied mathematically
- For fixed D and Delta, largest N can be is

$$N \leq \Delta^D$$

- For example: if N = 300M (U.S. population) and Delta = 150, get constraint on D:

$$300{,}000{,}000 \leq (150)^D$$

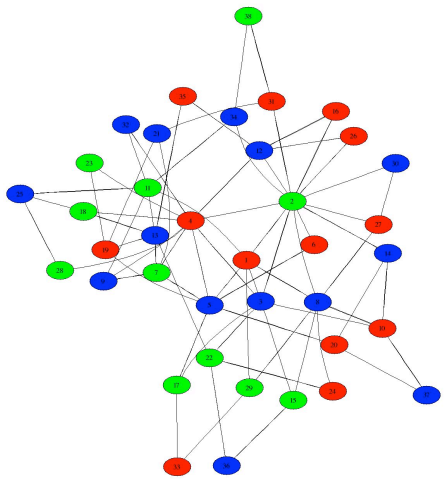
$$\log(300{,}000{,}000) \leq D\log(150)$$

$$D \geq 3.9$$

- So calculation consistent with reality (whew!)
- More generally: multiple structural properties may be *competing*

Two Aspects of Navigation

- In order for people (or machines) to find short paths in networks:
  - short paths must exist (structural; small diameter)
  - people must be able to find the short paths via only local forwarding (algorithmic)
- The algorithmic constraints are strong (Travers & Milgram)
  - only know your neighbors in the network
  - limited information about the target/destination (physical location, some background)
- Look at a model incorporating structural and algorithmic constraints

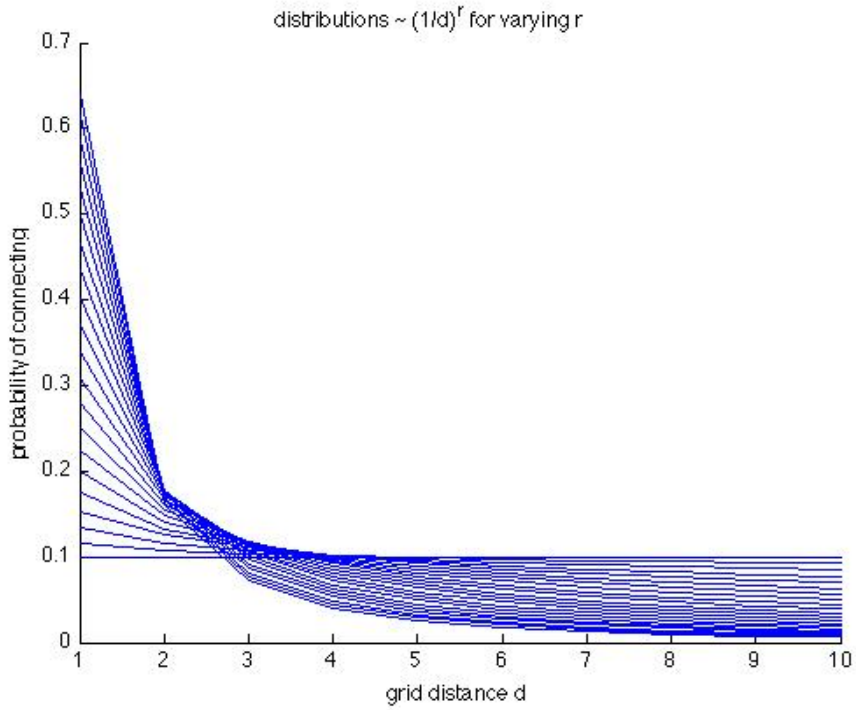


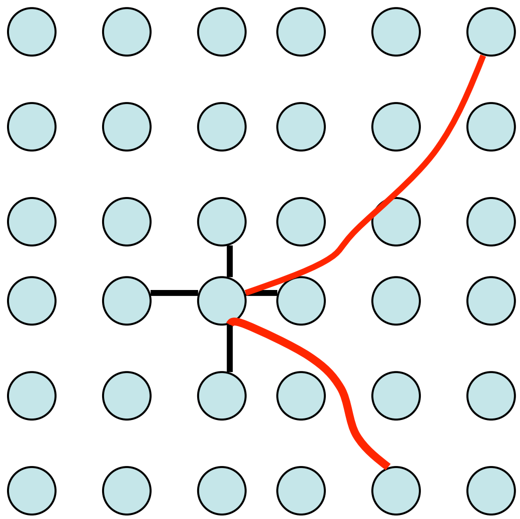
Kleinberg's Model

- Start with an k by k *grid* of vertices (so N = k^2)
  - each vertex connected to compass neighbors
  - add a few random "long-distance" connections to each vertex
  - probability p(d) of connecting to a vertex at grid distance d:

$$p(d) \propto (1/d)^r, r \geq 0$$

  - large r: heavy bias towards "more local" long-distance connections
  - small r: approach uniformly random



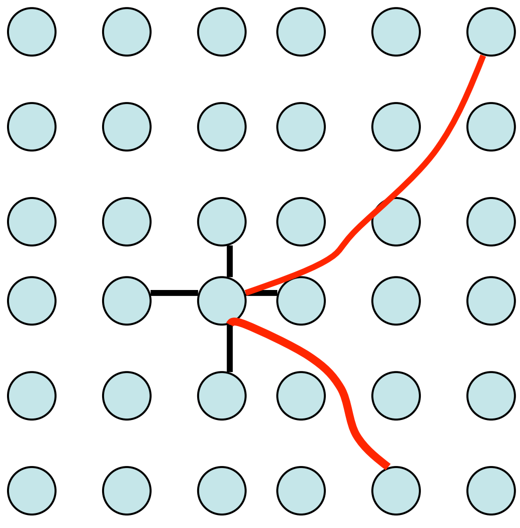


Kleinberg's Question

- Which values of r:

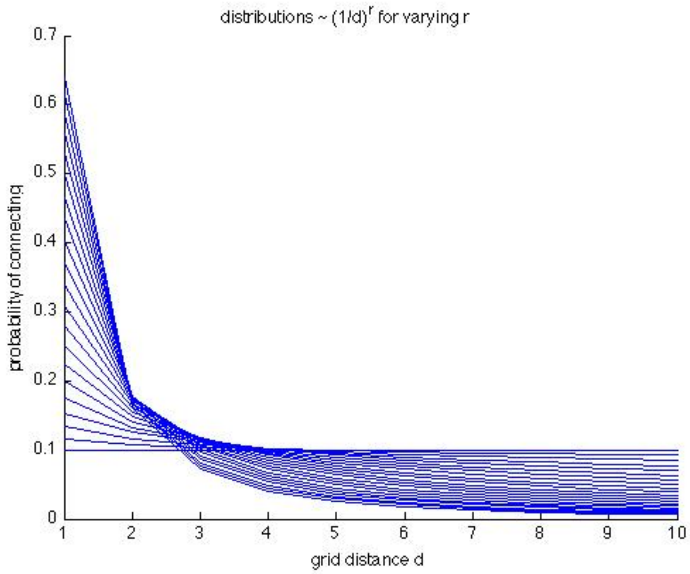
$$p(d) \propto (1/d)^r, r \geq 0$$

  permit efficient navigation?
- Efficient: number of hops << N, e.g. log(N)
- Algorithmic assumption:
  - vertices know the grid addresses of their neighbors
  - vertices know the grid address of the target (Sharon, MA)
  - vertices always forward the message to neighbor closest to the target in grid distance
  - no "backwards" steps, even if helpful
  - purely geographic information



Kleinberg's Result

- Intuition:
  - if r is too *large* (strong local bias), then "long-distance" connections never help much; short paths may not even *exist*
  - if r is too *small* (no local bias), we may quickly get close to the target; but then we'll have to use grid links to finish
  - effective search requires a delicate *mixture* of link distances
- The result (informally): as N becomes large:
  - r = 2 is the *only value* that permits rapid navigation (~log(N) steps)
  - a "knife's edge" result; very sensitive
- Note: *locality of information* crucial to this argument
  - At r <= 2, will have small diameter, but local algorithms can't find the short paths



www.wheresgeorge.com/b:aSEMWEQic&entry=15

WWW.WHERESGEORGE.COM®

Home | Top 10 | Public Forum | Tools/Fun | Join/Login | Store | FAQs/About | Contact Us

Logon

Where's George? Bill Tracking Report

Printer Friendly Version

One Dollar Bill, Serial# **K24------I** Series: **1999**

This bill has traveled at least 4,183 Miles in 3 Yrs, 11 Days, 17 Hrs, 26 Mins at an average of 3.8 Miles per day.
The bill is now 452 Miles from its starting location. This travel history below is in reverse-chronological order

Entry Time (Local Time of Zip)	Location, State/Province (Green=USA, Blue=Canada, Purple=International)	Travel Time (from previous entry)	Distance (Miles)*	Average Speed (Miles Per Day)	Social Networking
26-Mar-2005 08:34 PM	Rudyard, MI	212 Days, 14 Hrs, 37 Mins	8.6	0.04	
User's Note	this bill is getting pretty old looking				
26-Aug-2004 06:57 AM	Kincheloe, MI	112 Days, 7 Hrs, 10 Mins	1,539	14	
User's Note	Vote!				
05-May-2004 09:48 PM	Panguitch, UT	104 Days, 3 Hrs, 51 Mins	937	9.0	
User's Note	i FOUND THIS BILL AT THE FLYING M				
22-Jan-2004 05:57 PM	Irving, TX	25 Days, 23 Hrs, 57 Mins	30	1.1	
User's Note	Mr K's Food Mart, Irving Tx				
27-Dec-2003 06:00 PM	Rockwall, TX	3 Days, 3 Hrs, 57 Mins	152	48	
User's Note	its condition is good got it at jack in the box in rockwall tx.				
24-Dec-2003 02:03 PM	Shreveport, LA	8 Days, 1 Hr, 52 Mins	160	20	
16-Dec-2003 12:11 PM	Garland, TX	13 Days, 18 Hrs, 13 Mins	17	1.2	
02-Dec-2003 05:59 PM	Dallas, TX	22 Days, 20 Hrs, 8 Mins	15	0.67	
User's Note	Found on the floor at the Penthouse Key Club.				
09-Nov-2003 09:51 PM	Grapevine, TX	11 Days, 15 Hrs, 11 Mins	25	2.1	
User's Note	[Hit #328, 346, 359, 362, 419, 640, 667][Hit #740 dlt'd 20 July 04][Hit #803, #858]. Rec'd marked bill at the racetrack. My 72nd Wild.				
29-Oct-2003 06:41 AM	Fort Worth, TX	292 Days, 15 Hrs, 52 Mins	632	2.2	
User's Note	Bill is still in good shape. I got it as change at a McDonalds in Keller TX.				
09-Jan-2003 02:49 PM	Milton, FL	177 Days, 6 Hrs, 45 Mins	348	2.0	
User's Note	Not sure where I received it, in change from somewhere. The condition of the bill is good.				
16-Jul-2002 09:04 AM	Unionville, TN	13 Days, 20 Hrs, 16 Mins	7.3	0.53	
User's Note	This bill was at a country store at Halls Mill community in Unionville, Tennessee USA				
02-Jul-2002 12:48 PM	Chapel Hill, TN	48 Days, 51 Mins	84	1.7	
User's Note	Came into my possession at the Shell Food Mart in Chapel Hill, TN.				
15-May-2002 11:57 AM	Scottsville, KY	61 Days, 8 Hrs, 49 Mins	229	3.7	
User's Note	I work at Sonic and reiceved it as a tip. It's in preety good condition.				
15-Mar-2002 03:08 AM	Dayton, OH	Initial Entry	n/a	n/a	
User's Note	Thanks for entering this bill, and welcome to Where's George! [wb]				

Size: **Small Large Hide** Type: **Roadmap Terrain Satellite Hybrid**

Click here for Dynamic Google Map (Pop-up)

Real Time Hit Feed

Roswell, GA
1st hit
One:B3888---6J
I got this bill from a Chinese restaurant in Roswell GA. The bill is in great condition and is very crisp.

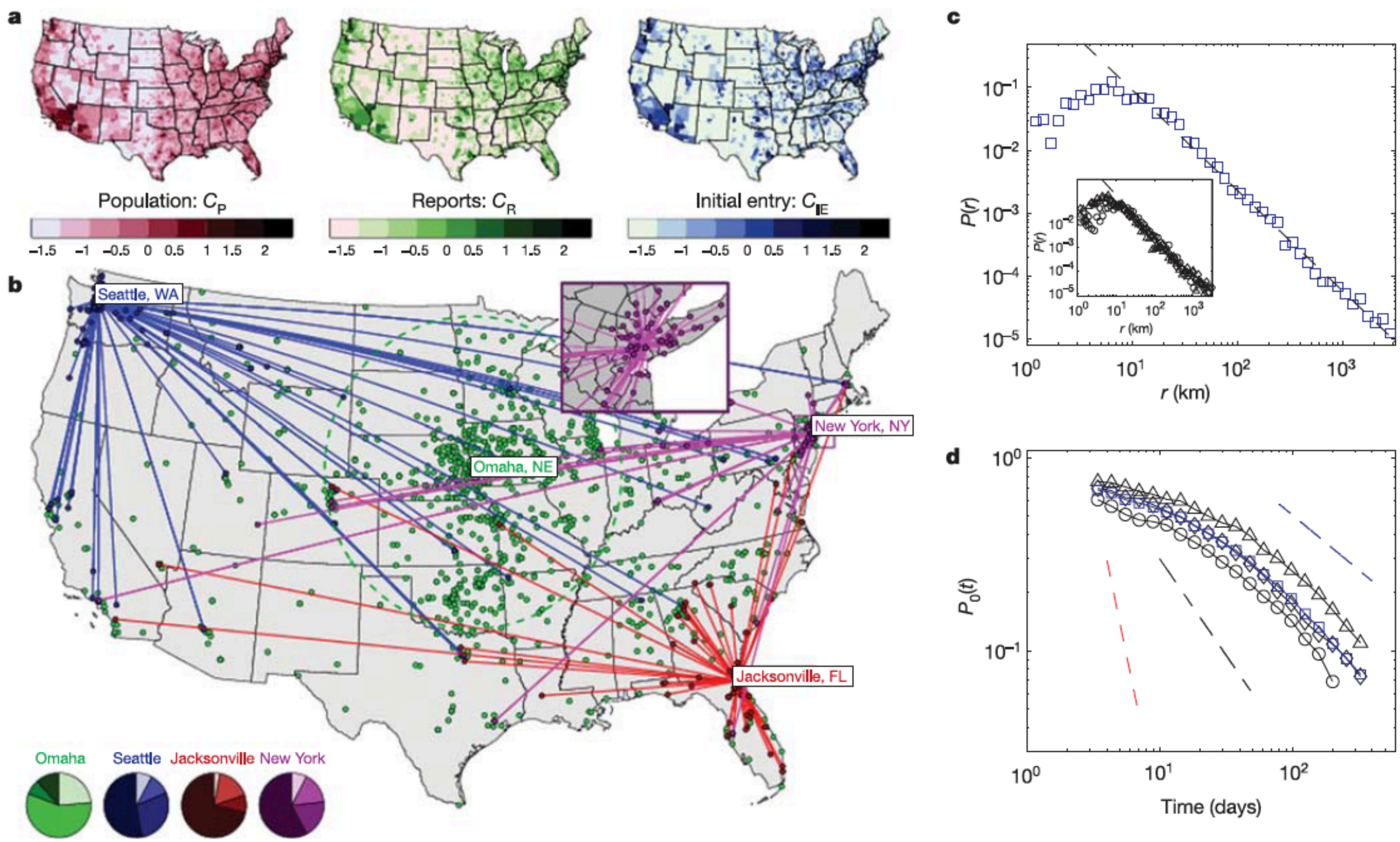
Columbia, MO
1st hit
One:B2201---6G
I received from it a retail store in Columbia - still in great condition.

Elmhurst, IL
1st hit
One:B6274---4H
wrinkly

Announcements

Where's George? 4.0!

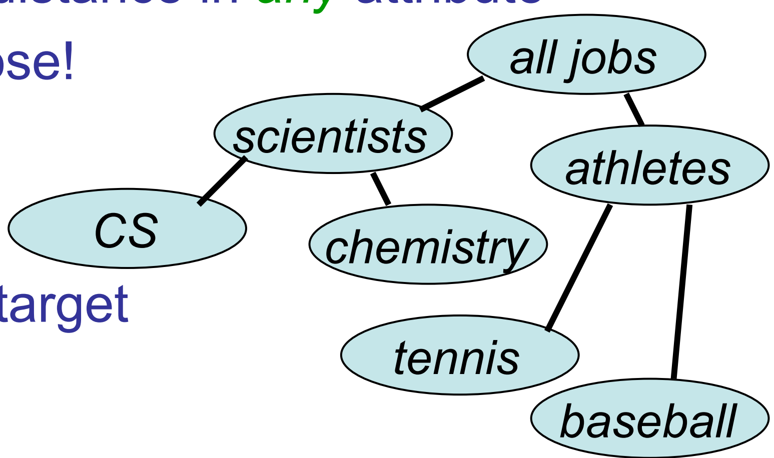
Welcome to the all new *Where's George?* website. To see all the new changes at *Where's George?* 4.0, please click to see --> this forum



From Brockmann, Hufnagel, Geisel (2006) *Best-fit value of r = 1.59*

Navigation via Identity

- Watts et al.:
  - we don’t navigate social networks by purely “geographic” information
  - we don’t use any *single* criterion; recall Dodds et al. on Columbia SW
  - different criteria used at different points in the chain
- Represent individuals by a *vector* of attributes
  - profession, religion, hobbies, education, background, etc…
  - attribute values have distances between them (tree-structured)
  - distance between individuals: minimum distance in *any* attribute
  - only need *one thing in common* to be close!
- Algorithm:
  - given attribute vector of target
  - forward message to neighbor closest to target
- Let’s look a bit at the paper
- Permits fast navigation under broad conditions
  - not as sensitive as Kleinberg’s model



Summary

- Efficient navigation has both structural and algorithmic requirements
- Kleinberg’s model and question captures both
- Result predicts delicate mixture of connectivity for success
- Not too far from reality? (Where’s George? data)
- Watts et al. provide more “sociological” model
- More complex, but less sensitive