CS 6474/CS4803 Social Computing: Sociological Foundations I

Munmun De Choudhury

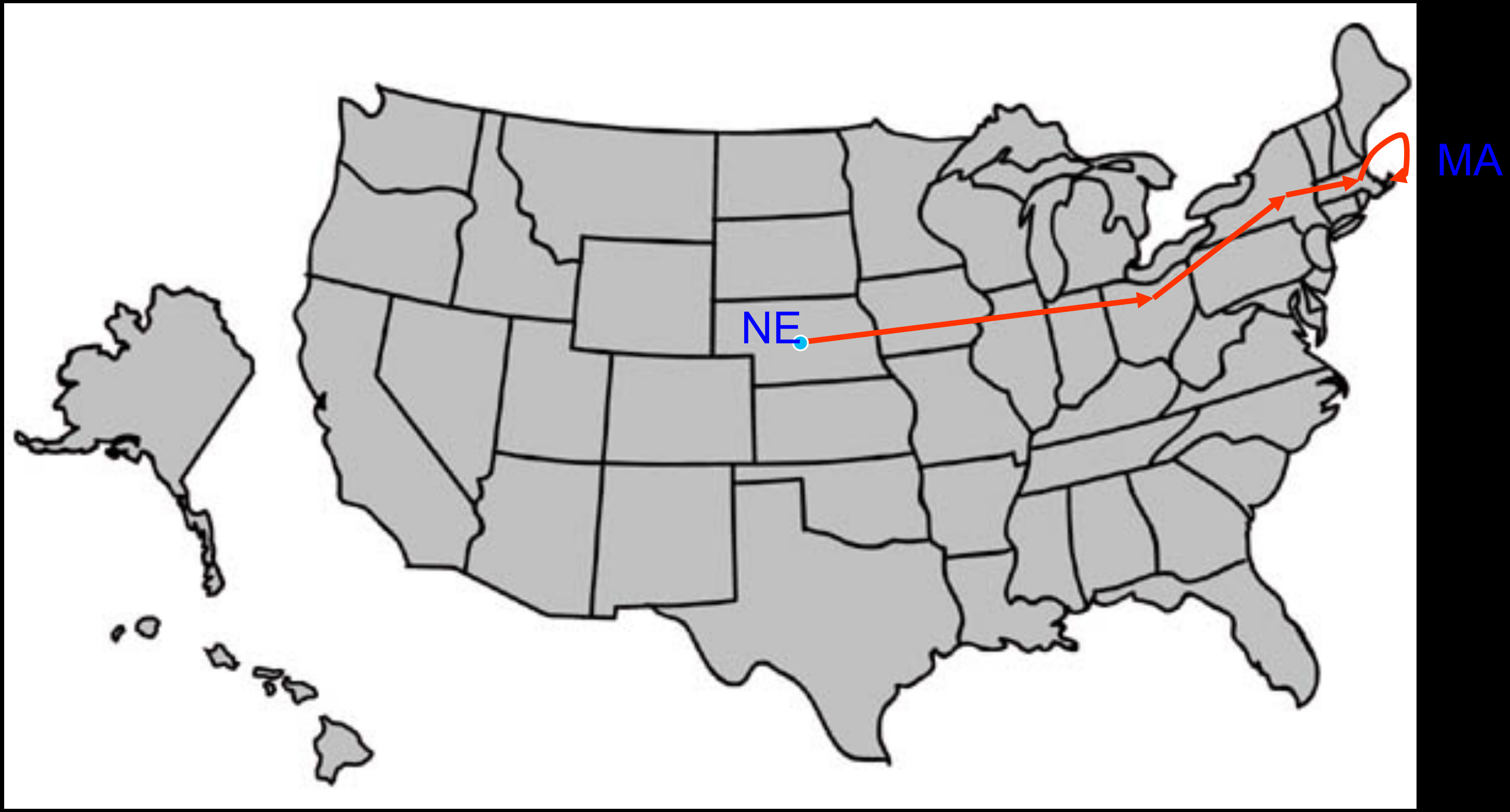
munmund@gatech.edu

Week 2 | January 21, 2021

- Frigyes Karinthy in 1929 published a volume of short stories called “Everything is Different”
- He was the first proponent of the six degrees of separation concept, in his 1929 short story, Chains (Láncszemek)
- In his book the characters created a game out of the notion that “the world is shrinking”:

> A fascinating game grew out of this discussion. One of us suggested performing the following experiment to prove that the population of the Earth is closer together now than they have ever been before. We should select any person from the 1.5 billion inhabitants of the Earth – anyone, anywhere at all. He bet us that, using no more than *five* individuals, one of whom is a personal acquaintance, he could contact the selected individual using nothing except the network of personal acquaintances

An Experimental Study of the Small World Problem



Summary

- First sociological study of the "six degrees of separation"
- Empirically determine the maximum number of intermediaries it would require to reach anybody in the US
- Experiment conducted through forwarding of a set of snail mail letters, all targeted to a target in Massachusetts
- N=296 for two groups in Nebraska and Boston
- Main strategies involved in selecting the next point of forwarding: geographic and business

Summary

- How many of the starters would be able to establish contact with the target?
- Well, that depends: the overall mean 5.2 links
  - Through hometown: 6.1 links
  - Through business: 4.6 links
  - Boston group faster than Nebraska groups
  - Nebraska stakeholders not faster than Nebraska random

Summary

- Results:
  - 64 chains reached target
- At least two facts about this study are somewhat remarkable:
  - First, that short paths appear to be abundant in the network
  - Second, that people are capable of discovering them in a "decentralized" fashion, i.e., they are somehow good at "guessing" which links will be closer to the target
- What really stood out:
  - Funneling - Presence of a set of "hubs"/sociometric stars, through which most letters went through near the final target

Collective dynamics of 'small-world' networks

Duncan J. Watts* & Steven H. Strogatz

Department of Theoretical and Applied Mechanics, Kimball Hall, Cornell University, Ithaca, New York 14853, USA

Networks of coupled dynamical systems have been used to model biological oscillators[1–4], Josephson junction arrays[5,6], excitable media[7], neural networks[8–10], spatial games[11], genetic control networks[12] and many other self-organizing systems. Ordinarily, the connection topology is assumed to be either completely regular or completely random. But many biological, technological and social networks lie somewhere between these two extremes. Here we explore simple models of networks that can be tuned through this middle ground: regular networks 'rewired' to introduce increasing amounts of disorder. We find that these systems can be highly clustered, like regular lattices, yet have small characteristic path lengths, like random graphs. We call them 'small-world' networks, by analogy with the small-world phenomenon[13,14] (popularly known as six degrees of separation[15]). The neural network of the worm *Caenorhabditis elegans*, the power grid of the western United States, and the collaboration graph of film actors are shown to be small-world networks. Models of dynamical systems with small-world coupling display enhanced signal-propagation speed, computational power, and synchronizability. In particular, infectious diseases spread more easily in small-world networks than in regular lattices.

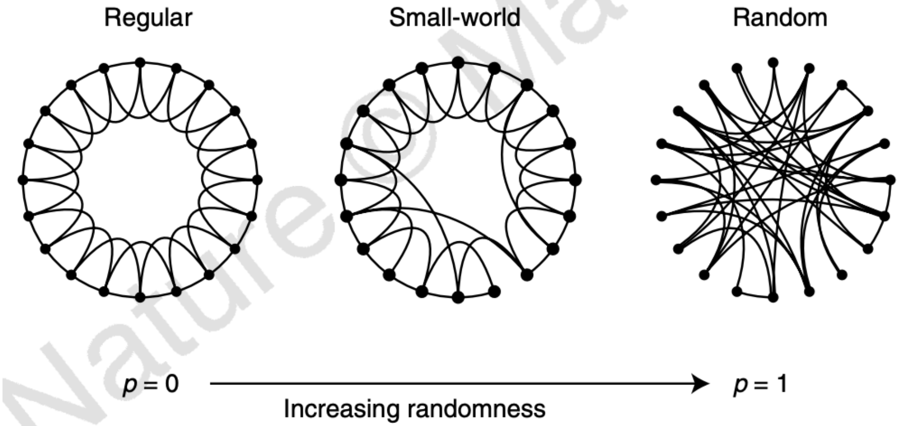


Figure 1 Random rewiring procedure for interpolating between a regular ring lattice and a random network, without altering the number of vertices or edges in the graph. We start with a ring of n vertices, each connected to its k nearest neighbours by undirected edges. (For clarity, $n = 20$ and $k = 4$ in the schematic examples shown here, but much larger n and k are used in the rest of this Letter.) We choose a vertex and the edge that connects it to its nearest neighbour in a clockwise sense. With probability p, we reconnect this edge to a vertex chosen uniformly at random over the entire ring, with duplicate edges forbidden; otherwise we leave the edge in place. We repeat this process by moving clockwise around the ring, considering each vertex in turn until one lap is completed. Next, we consider the edges that connect vertices to their second-nearest neighbours clockwise. As before, we randomly rewire each of these edges with probability p, and continue this process, circulating around the ring and proceeding outward to more distant neighbours after each lap, until each edge in the original lattice has been considered once. (As there are $nk/2$ edges in the entire graph, the rewiring process stops after $k/2$ laps.) Three realizations of this process are shown, for different values of p. For $p = 0$, the original ring is unchanged; as p increases, the graph becomes increasingly disordered until for $p = 1$, all edges are rewired randomly. One of our main results is that for intermediate values of p, the graph is a small-world network: highly clustered like a regular graph, yet with small characteristic path length, like a random graph. (See Fig. 2.)

Such small-world phenomena turn out to be abundant in a variety of network settings

The Erdös Number

Who was Erdös? http://www.oakland.edu/enp/

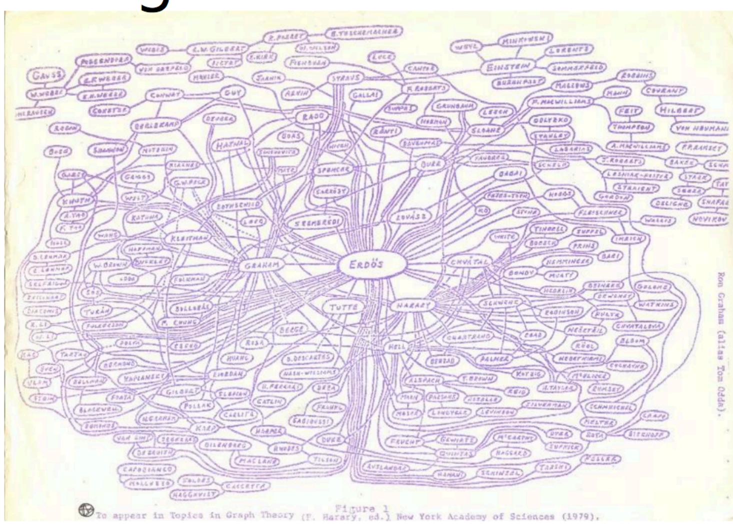
A famous Hungarian Mathematician, 1913-1996

Erdös posed and solved problems in number theory and other areas and founded the field of discrete mathematics.

- 511 co-authors (Erdös number 1)
- ~ 1500 Publications

Such small-world phenomena turn out to be abundant in a variety of network settings

e.g. Erdos numbers:

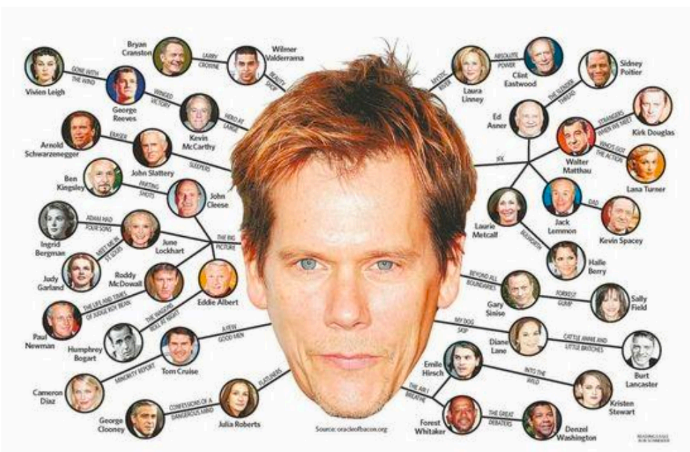


Erdös # 0	-	1	person
Erdös # 1	-	504	people
Erdös # 2	-	6593	people
Erdös # 3	-	33605	people
Erdös # 4	-	83642	people
Erdös # 5	-	87760	people
Erdös # 6	-	40014	people
Erdös # 7	-	11591	people
Erdös # 8	-	3146	people
Erdös # 9	-	819	people
Erdös #10	-	244	people
Erdös #11	-	68	people
Erdös #12	-	23	people
Erdös #13	-	5	people

http://www.oakland.edu/enp/trivia/

Such small-world phenomena turn out to be abundant in a variety of network settings

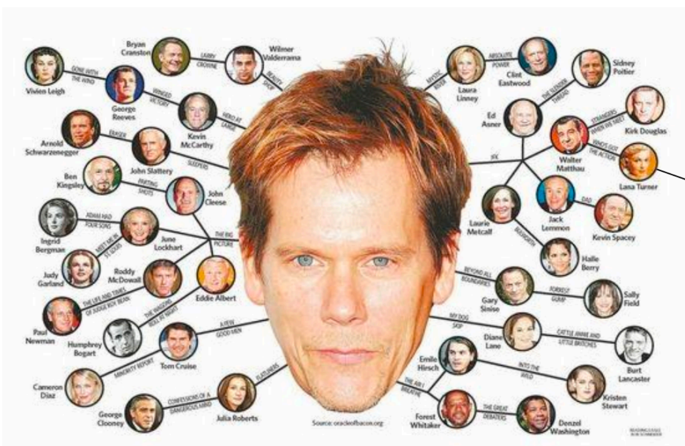
e.g. Bacon numbers:

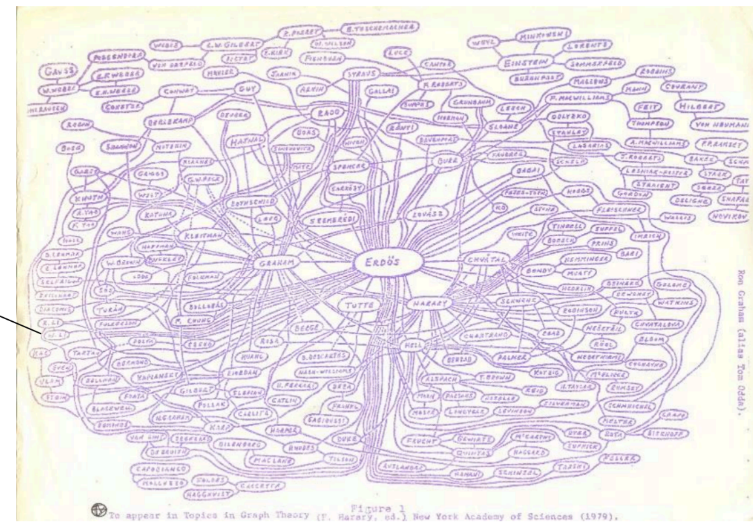


linkedscience.org & readingeagle.com

Such small-world phenomena turn out to be abundant in a variety of network settings

Bacon/Erdos numbers:





Kevin Bacon→Sarah Michelle Gellar→**Natalie Portman**→Abigail Baird→Michael Gazzaniga→J. Victor→Joseph Gillis→Paul Erdos

What were some of the biggest assumptions/constraints in the study that may have affected the outcomes?

What were some of the biggest limitations of the study? What could have been alternatives to address them?

What were your biggest surprises from the study?

A. There was a constant drop off rate as the letters traveled forward. What could be potential reasons behind this phenomenon?

B. Why do you think there were so few completed chains?

In Milgram’s chain letter experiment, men were 10 times more likely to forward the letters than women. Why do you think it was the case?

Milgram did not after all investigate whether tie strength might play a role. How do you think tie strength would impact the so called “small world phenomenon”?

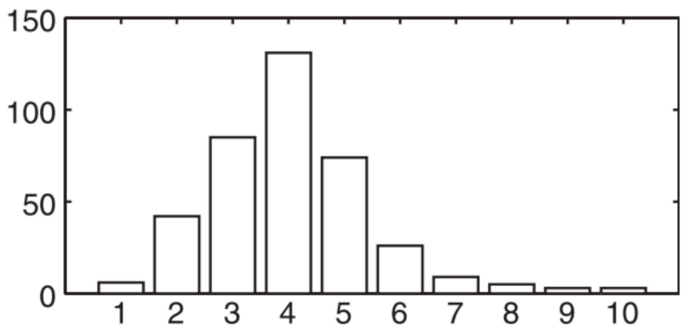
Class Exercise I: Validity + Parallels With OSNs

Assume that you have been asked to replicate/reproduce Milgram's small world study empirically (i.e., using data-driven observations). Answer the following questions:

1) Let's say you first want to do it observationally, that is, you have all of email data, or instant messaging data, or Facebook data. How would you go about assessing whether on an average, people are separated from each other by approximately six hops? What will you measure in this data? Why is it a good measurement? How does it contrast with what Milgram did?
2) Alternatively, let's say you have the ability to experimentally maneuver email exchanges in a large population. What would your experiment look like? How will it be different (or similar) from Milgram's? Do you think it will be six hops, validating Milgram's study? Or will it be less or more? Justify your answer.

Dodds, Muhamed, & Watts repeated Milgram's experiments using e-mail

- 18 "targets" in 13 countries
- 60,000+ participants across 24,163 chains
- Only 384 (!) reached their targets



Histogram of (completed) chain lengths – average is just **4.01!**

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

Reasons for choosing the next recipient at each point in the chain

from http://www.cis.upenn.edu/~mkearns/teaching/NetworkedLife/columbia.pdf

In Milgram's chain letter experiment, letter forwarding may imply a different notion of a friend compared to what we imply today in online settings. Can these differences affect the number of hops (i.e., people are separated by about 6 acquaintances)?

The study showed the funneling effect - presence of a set of “hubs”/sociometric stars, through which most letters went through near the final target. How does this relate to OSNs of today?

Four Degrees of Separation

Lars Backstrom* Paolo Boldi† Marco Rosa† Johan Ugander* Sebastiano Vigna†

January 6, 2012

Abstract

Frigyes Karinthy, in his 1929 short story "Láncszemek" ("Chains") suggested that any two persons are distanced by at most six friendship links.[1] Stanley Milgram in his famous experiment [20, 23] challenged people to route postcards to a fixed recipient by passing them only through direct acquaintances. The average number of intermediaries on the path of the postcards lay between 4.4 and 5.7, depending on the sample of people chosen.

We report the results of the first world-scale social-network graph-distance computations, using the entire Facebook network of active users (≈ 721 million users, ≈ 69 billion friendship links). The average distance we observe is 4.74, corresponding to 3.74 intermediaries or "degrees of separation", showing that the world is even smaller than we expected, and prompting the title of this paper. More generally, we study the distance distribution of Facebook and of some interesting geographic subgraphs, looking also at their evolution over time.

The networks we are able to explore are almost two orders of magnitude larger than those analysed in the previous literature. We report detailed statistical metadata showing that our measurements (which rely on probabilistic algorithms) are very accurate.

1 Introduction

At the 20th World–Wide Web Conference, in Hyderabad, India, one of the authors (Sebastiano) presented a new tool for studying the distance distribution of very large graphs: HyperANF [3]. Building on previous graph compression [4] work and on the idea of diffusive computation pioneered in [21], the new tool made it possible to accurately study the distance distribution of graphs orders of magnitude larger than it was previously possible.

One of the goals in studying the distance distribution is the identification of interesting statistical parameters that can be used to tell proper social networks from other complex networks, such as web graphs. More generally, the distance distribution is one interesting *global* feature that makes it possible to reject probabilistic models even when they match local features such as the in-degree distribution.

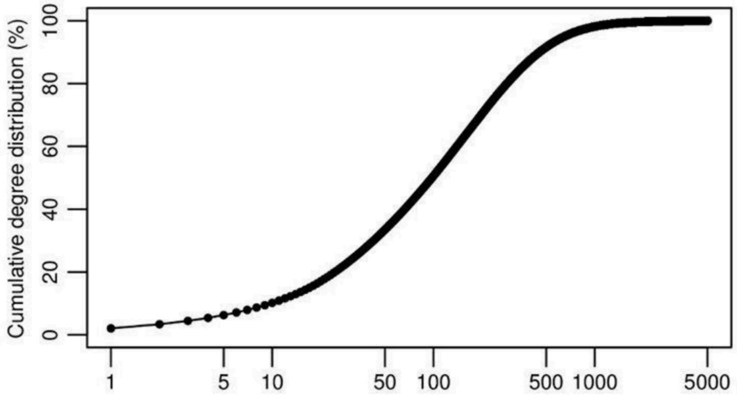
In particular, earlier work had shown that the *spid*[2], which measures the *dispersion* of the distance distribution, appeared to be smaller than 1 (underdispersion) for social networks, but larger than one (overdispersion) for web graphs [3]. Hence, during the talk, one of the main open questions was "What is the spid of Facebook?".

Lars Backstrom happened to listen to the talk, and suggested a collaboration studying the Facebook graph. This was of course an extremely intriguing possibility: beside testing the "spid hypothesis", computing the distance distribution of the Facebook graph would have been the largest Milgram-like [20] experiment ever performed, orders of magnitudes larger than previous attempts (during our experiments Facebook has ≈ 721 million active users and ≈ 69 billion friendship links).

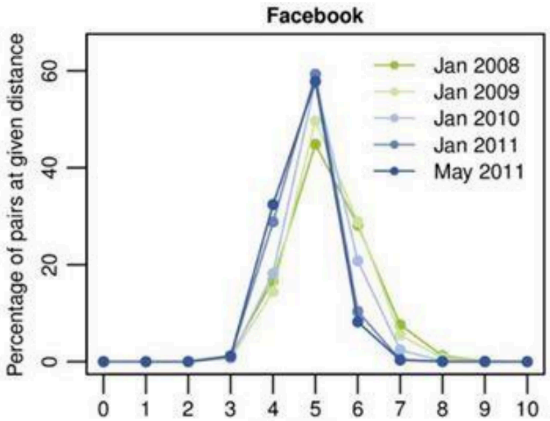
This paper reports our findings in studying the distance distribution of the largest electronic social network ever created. That world is smaller than we thought: the average distance of the current Facebook graph is 4.74. Moreover, the spid of the graph is just 0.09, corroborating the conjecture [3]

*Facebook.
†DSI, Università degli Studi di Milano, Italy. Paolo Boldi, Marco

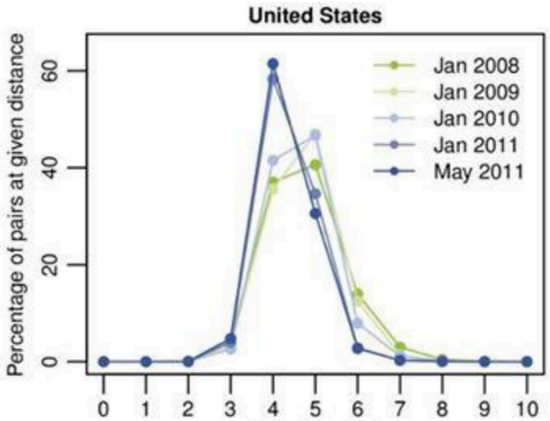
Actual shortest-path distances are similar to those in Dodds' experiment:



Cumulative degree distribution (# of friends) of Facebook users



Hop distance between Facebook users



Hop distance between users in the US

This suggests that people choose a reasonably good heuristic when choosing shortest paths in a decentralized fashion (assuming that FB is a good proxy for "real" social networks)

from "the anatomy of facebook": http://goo.gl/H0bkWY

So is the world really shrinking?