The Algorithmics of Information Diffusion

Alessandro Panconesi
Dipartimento di Informatica

DAY 1
Computer Science is about computers no more than astronomy is about telescopes

E.W. Dijkstra
THE STARS
PEOPLE
The INTERNET is an observatory on Crowds
No one would have believed in the first years of the XXI century that this world was being watched keenly and closely by awesome computing and financial powers; that as people busied themselves about their various concerns they were scrutinised and studied, perhaps almost as narrowly as a person with a microscope might scrutinise the transient creatures that swarm and multiply in a drop of water. ...

Adapted from “The War of the Worlds” by H.G.Wells
A Paradigm Shift
A Paradigm Shift
COURSE OUTLINE

• We will use a couple of “case studies” to..
• ..hopefully show interesting results
• ..illustrate the main thrust of this research area
• I will give:
  – An overview of results
  – But also delve deep into some results, to get the (mathematical) flavour of the kind of results one can hope to establish
LECTURE OUTLINE

• We will revisit a classic social psychology experiment in order to see..
• ...how BIG DATA opens new opportunities for making social science more rigorous
• ...and the virtuous interplay between mathematical models and empirical observations
COMPUTATIONAL SOCIAL SCIENCE?
A CLASSIC REVISITED
Omaha vs Boston
The chains progress from the starting position (Omaha) to the target area (Boston) with each remove. Diagram shows the number of miles from the target area, with the distance of each remove averaged over completed and uncompleted chains.
Outcome

30% of the letters reached the target.

Average (median) chain length was roughly six.
Outcome

1. **Average path length**, 5.2
2. **Bimodality is not accident**: target reached through
   - hometown, 6.1
   - business contacts, 4.6
3. **Role of geography**
   - Boston, 4.4
   - Nebraska, 5.5
4. **Role of occupation**
   - random, 5.7
   - stockholders, 5.4
Conjectures

Take any two people in the world, and they will be connected by a very short chain of acquaintances
Given any two people in the world, they will always be connected by a short chain of acquaintances.
Is Milgram’s conjecture true?
A matter of scale
A matter of scale

300

60,000
A matter of scale

300  60,000  230,000,000
A matter of scale

300  60,000  230,000,000  750,000,000
A matter of scale

300
60,000
230,000,000
750,000,000

Six orders of magnitude!!
Models
Salient properties

• Social networks exhibit special properties, eg they have many triangles and are “small worlds”

• ...can we come up with simple mathematical models (stochastic graphs) that reproduce them?
Watts & Strogatz
Watts & Strogatz
Watts & Strogatz
Watts & Strogatz
Watts & Strogatz
Chaos out of order
Coexistence
Coexistence

Ratio between clustering coefficients

Ratio between average path lengths
Coexistence

Graphs in this region have both high clustering coefficient and small diameter.
NEW INSIGHTS
The true mystery
The true mystery
The true mystery
Rebel King

There is a unique distribution of long links compatible with Milgram’s experiment.
Kleinberg’s Model
Long Links

- A long link $uv$ is inserted with probability (proportional to)

\[
\frac{1}{d(u, v)^\alpha}
\]

Local greedy works fast iff $\alpha = \text{dimension}$

Routing time $\sim c \log^2 N$ otherwise polynomial
Long Links

• A long link $uv$ is inserted with probability (proportional to)

$$\frac{1}{d(u, v)^\alpha}$$

Local greedy works fast iff $\alpha = \text{dimension}$

Routing time $\sim c \log^2 N$ otherwise polynomial

Later we will analyze Kleinberg’s result in some detail
Disappointments

The networks of Watts&Strogatz do have small diameter (every pair of nodes is connected by a short path) but Milgram’s experiment cannot succeed in them!
PREDICTIONS
Non-uniform density

Figure 1: US population density of geolocated Facebook users.
Non-uniform density

[Liben-Nowell et al]
- Dot for every inhabited location
- Each circle represents 50,000 nodes
- Centered at Ithaca, NY
For homogeneous densities we recover Kleinberg’s distribution.
Test with LIVE JOURNAL

$r = 1.30$
Test with LIVE JOURNAL

\[ P(r) \propto \frac{1}{r^{1.05}} \]

**West Coast**

**East Coast**

\[ P(r) \propto \frac{1}{r} \]
Figure 7: Probability of friendship as a function of distance. By computing the number of pairs of individuals at varying distances, along with the number of friends at those distances, we are able to compute the probability of two people at distance $d$ knowing each other. We see here that it is a reasonably good fit to a power-law with exponent near $-1$. 

Test with FACEBOOK
Recap

• We have seen how BIG DATA has the potential to change dramatically the social sciences.
• Social processes can be observed with a wealth of data and mathematical models can be developed to capture and predict (to some extent) their evolution.
Recap

• There is a virtuous interplay between empirical observations and mathematical modelling
• Some of the problems are inherently algorithmic
PARADOXES
The Grand Challenge
The Grand Challenge

What can we reconstruct the original diffusion process from the huge, and yet scanty, digital traces?
A nice example
Internet Activism
FINAL REMARKS

• Paradigm shift in the social sciences?
• Interesting algorithmic questions
• Interplay between data mining (observations) and modelling:
  – Observations inspire models
  – Models guide observations
  – Relatively good quantitative predictions are possible
THANKS