AN INTRODUCTION TO NETWORK SCIENCE

Nicola Perra
n.perra@greenwich.ac.uk
@net_science

REDUCTIONISM: DOMINANT APPROACH IN SCIENCE

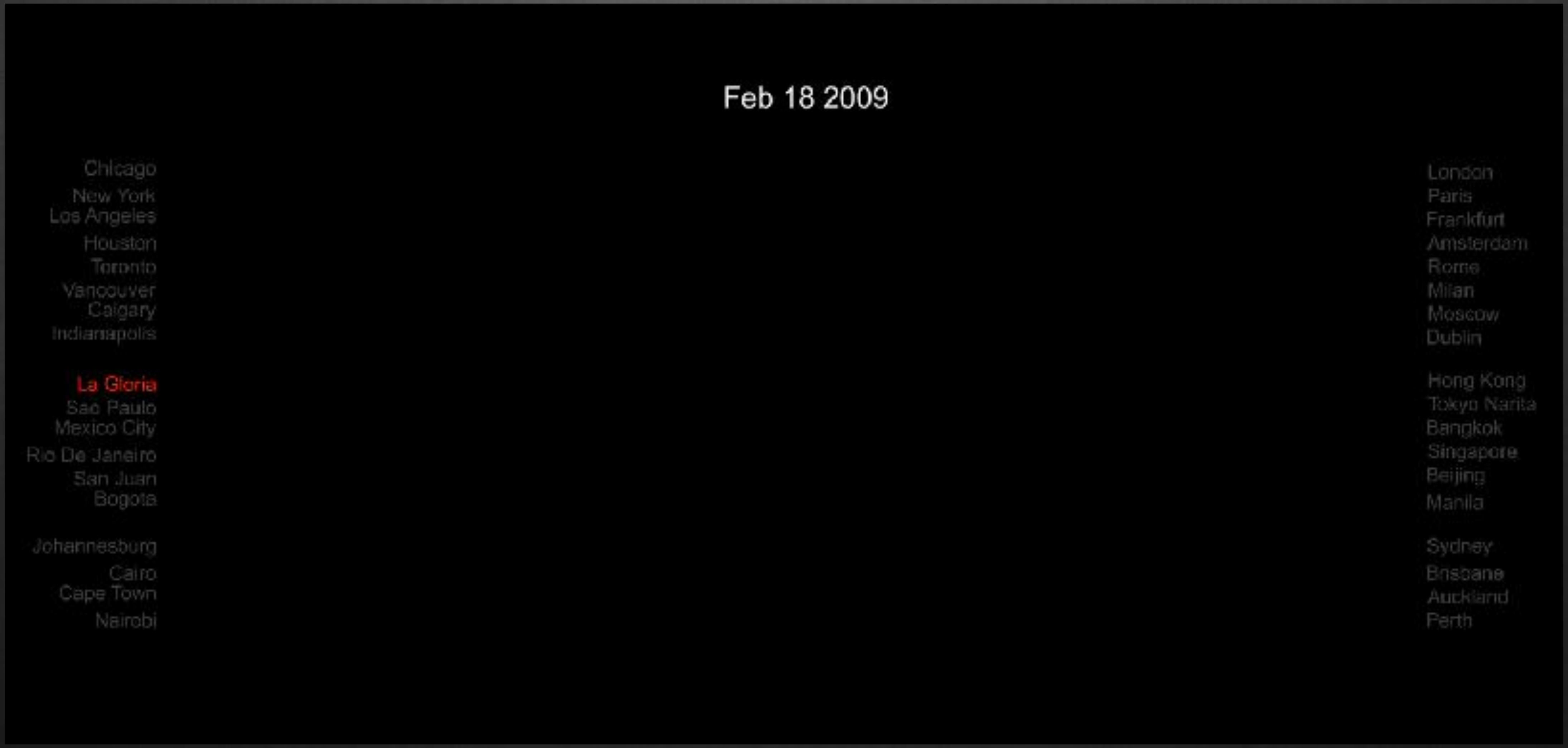
Systems are the nothing but the sum of their parts

NOT ALWAYS A GOOD APPROACH

By studying the interactions of single individuals can we understand the structure of a company?

NOT ALWAYS A GOOD APPROACH

By studying the interactions of single individuals can we understand the spreading of infectious diseases?



NOT ALWAYS A GOOD APPROACH

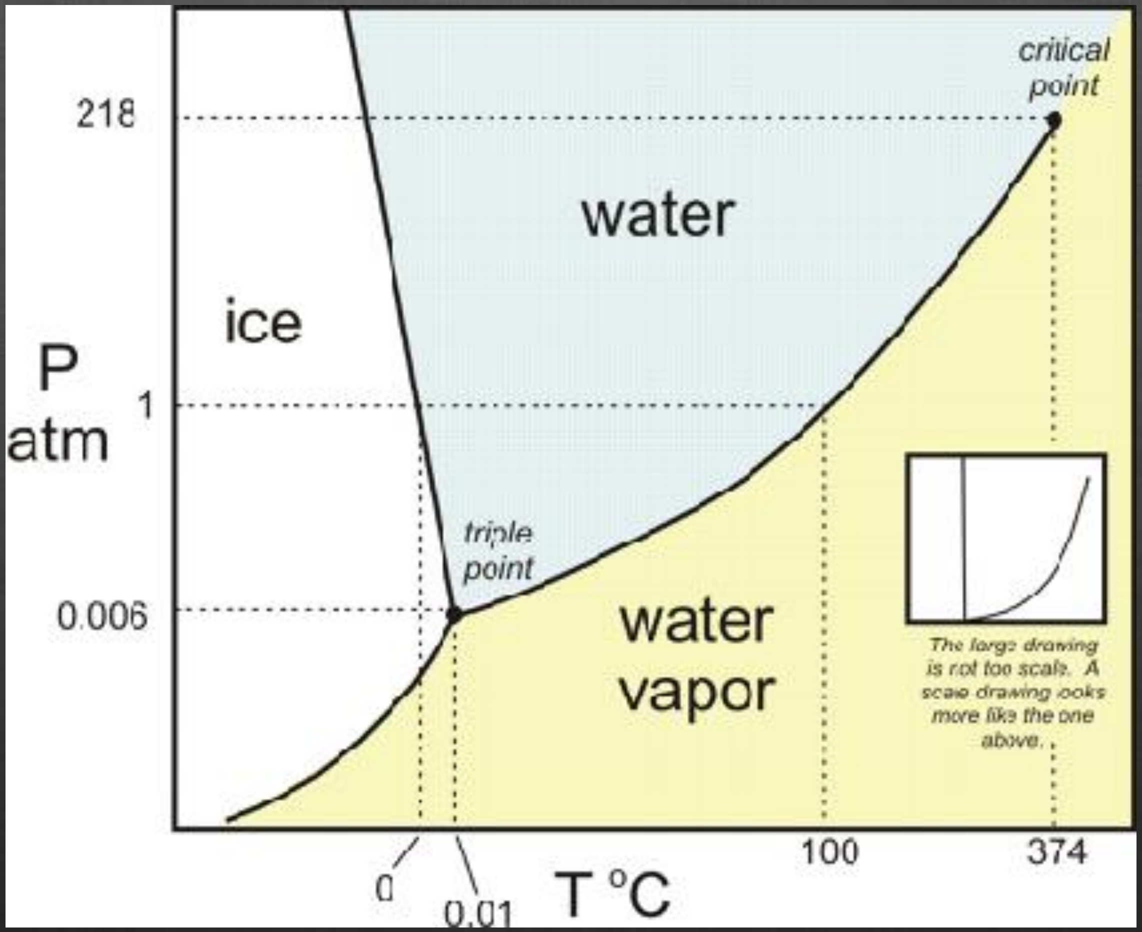
By studying the tweets of single Twitter users can we understand the emergence of social protests?

NOT ALWAYS A GOOD APPROACH

By studying the properties of single webpages can we build an efficient search engine?

NOT ALWAYS A GOOD APPROACH

By studying the properties of a single molecule of water can we understand the transition from ice to liquid water?



MORE IS DIFFERENT!

[...The main fallacy [of] the reductionist hypothesis [is that it] does not by any means imply a "constructionist" one: The ability to reduce everything to simple fundamental laws does not imply the ability to start from those laws and reconstruct the universe. In fact, the more the elementary particle physicists tell us about the nature of the fundamental laws, the less relevance they seem to have to the very real problems of the rest of science, much less to those of society...]

Anderson, P.W., "More is Different" in Science ,177, 4047. (1972)

COMPLEXITY

Holistic perspective

- Study systems as a whole
- Focus shifts on emergent phenomena

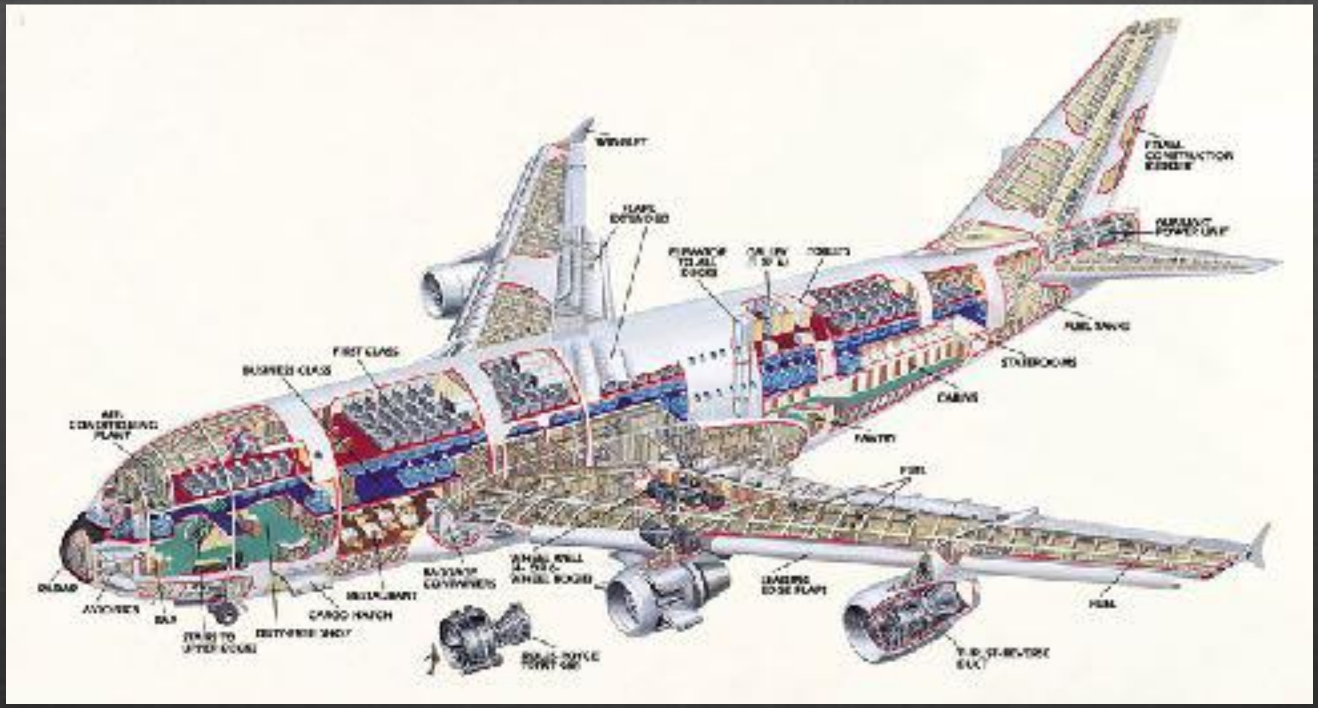
COMPLEX SYSTEMS

Properties:

- Complex systems are the spontaneous outcome of the interactions among the system constitutive units
- They are self-organizing systems. There is not blueprint, or global supervision
- Their behavior cannot be described from the properties of each constitutive units

COMPLEX SYSTEMS

Complex DOES NOT mean complicated!



COMPLEX SYSTEMS REPRESENTATION

Many complex systems can be described as a graph

- Nodes/vertices describe their constitutive units
- Links/edges describe the interaction between them

If, after this abstraction the complex features are still present

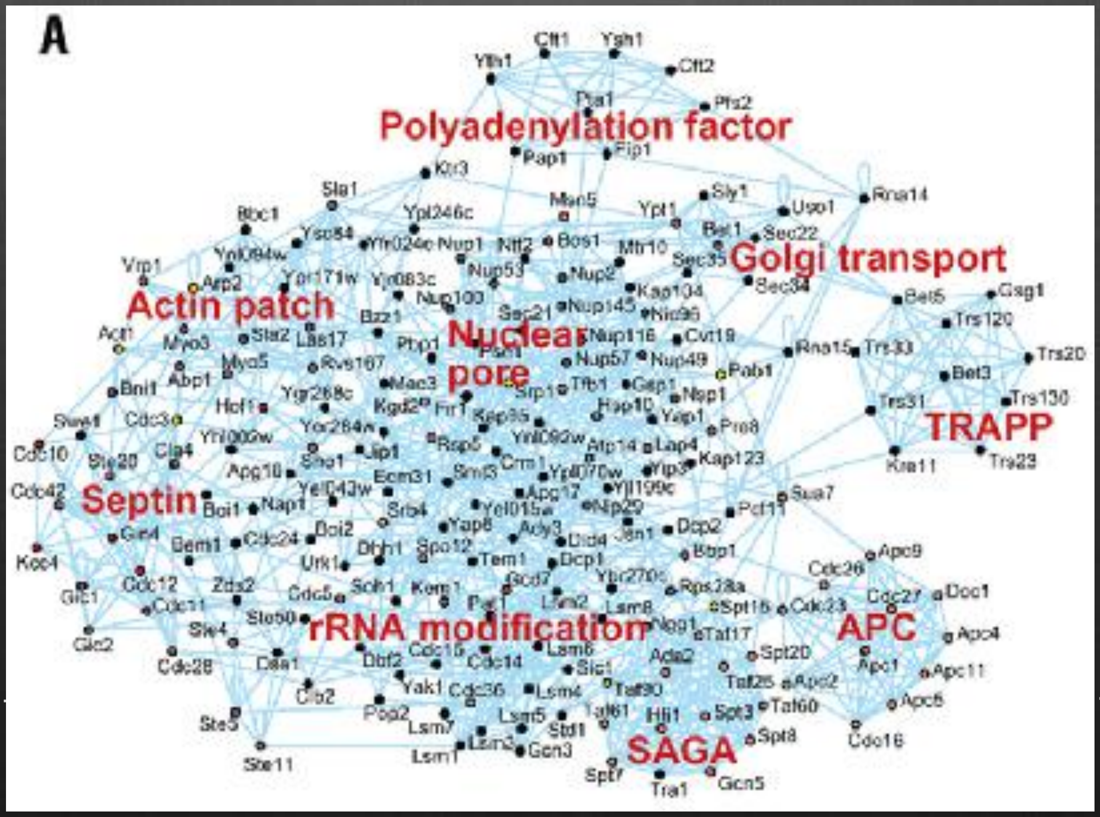
- Complex Networks!

WHY DO WE CARE?

Complex Networks are ubiquitous!

Biological networks

- Biochemical networks: molecular-level interactions and mechanisms of control in the cell
- Example 1) metabolic networks. Nodes are chemicals. Links describe the reactions
- Example 2) protein-protein interaction networks. Nodes are proteins. Links their interactions

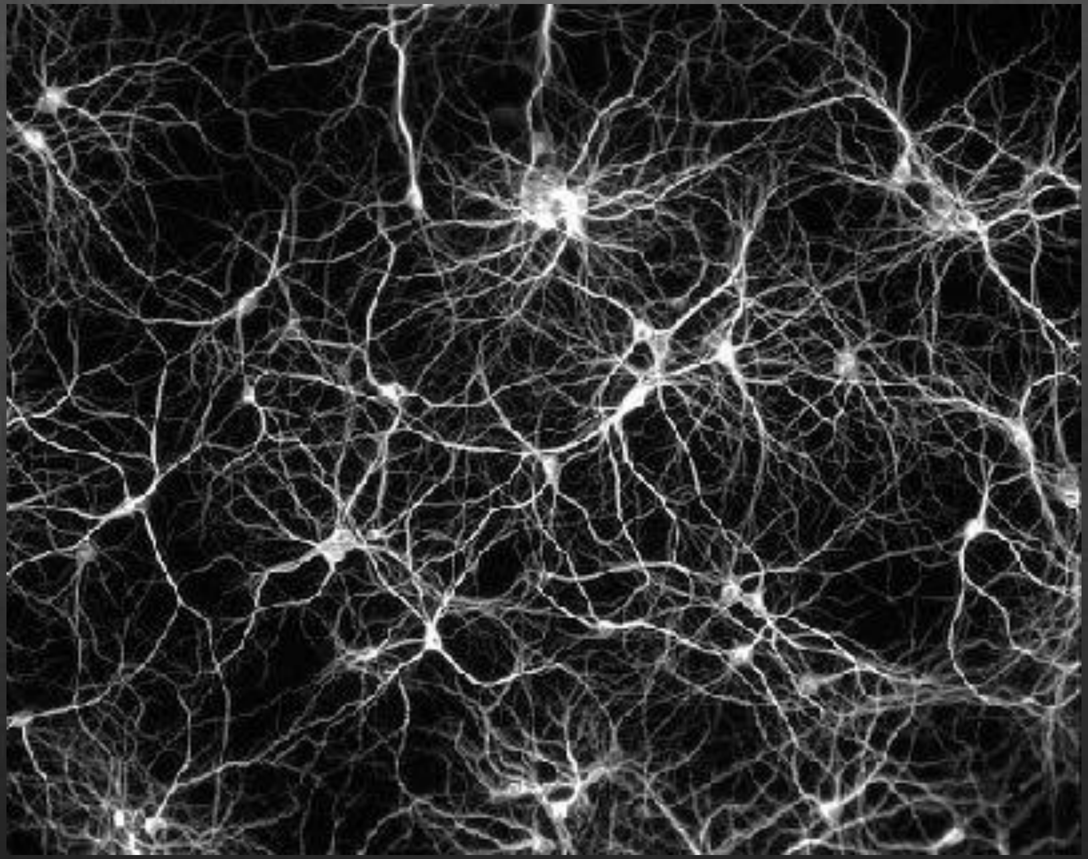


Nature Biotechnology 20, 991 - 997 (2002)

WHY DO WE CARE?

Biological networks

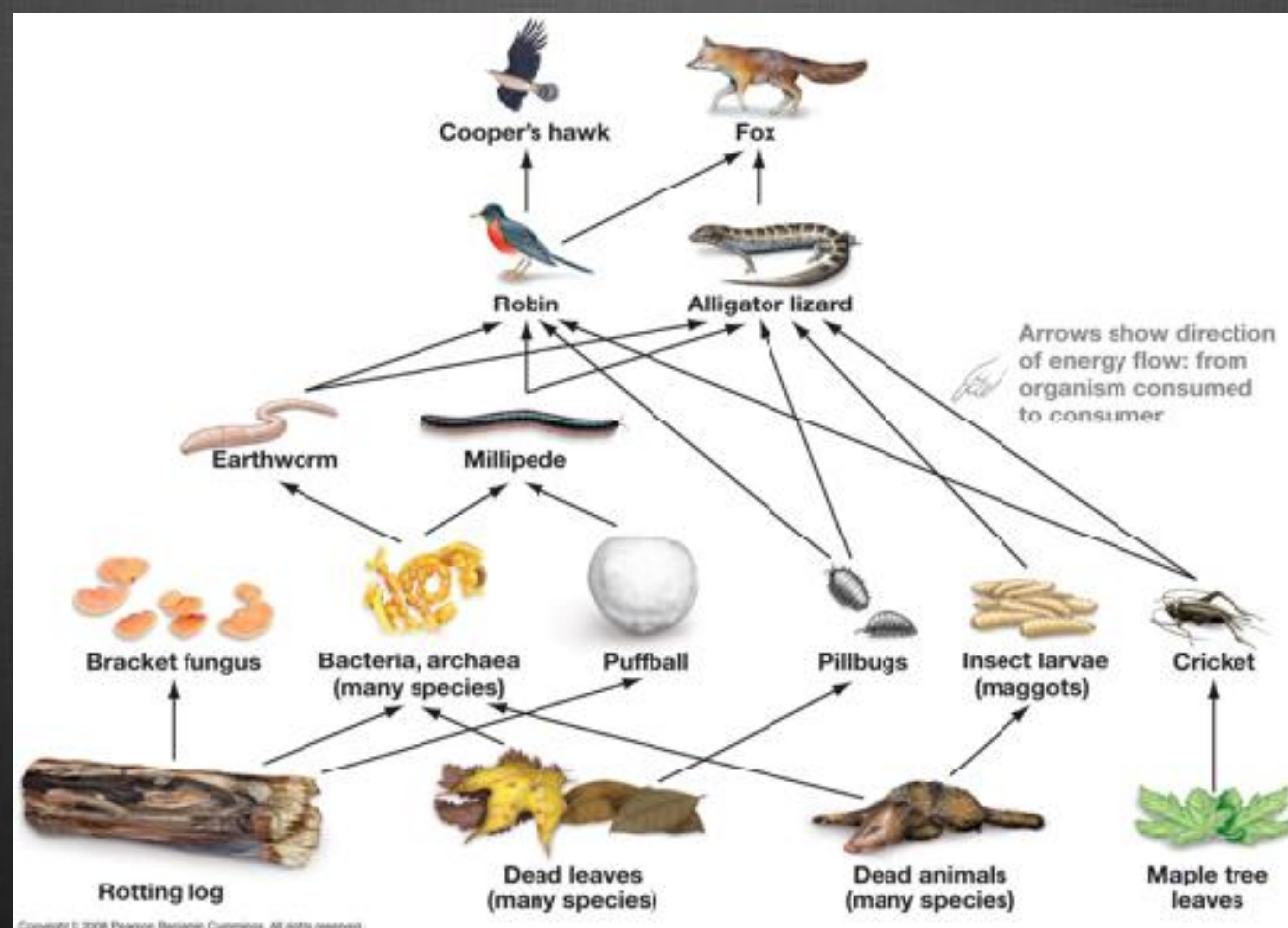
- Example 3) gene regulatory networks. Node are genes. A direct link between i and j implies that the first gene regulates the expression of the second
- Example 4) neural networks. Nodes are neurons. Links describe the synapses



WHY DO WE CARE?

Biological networks

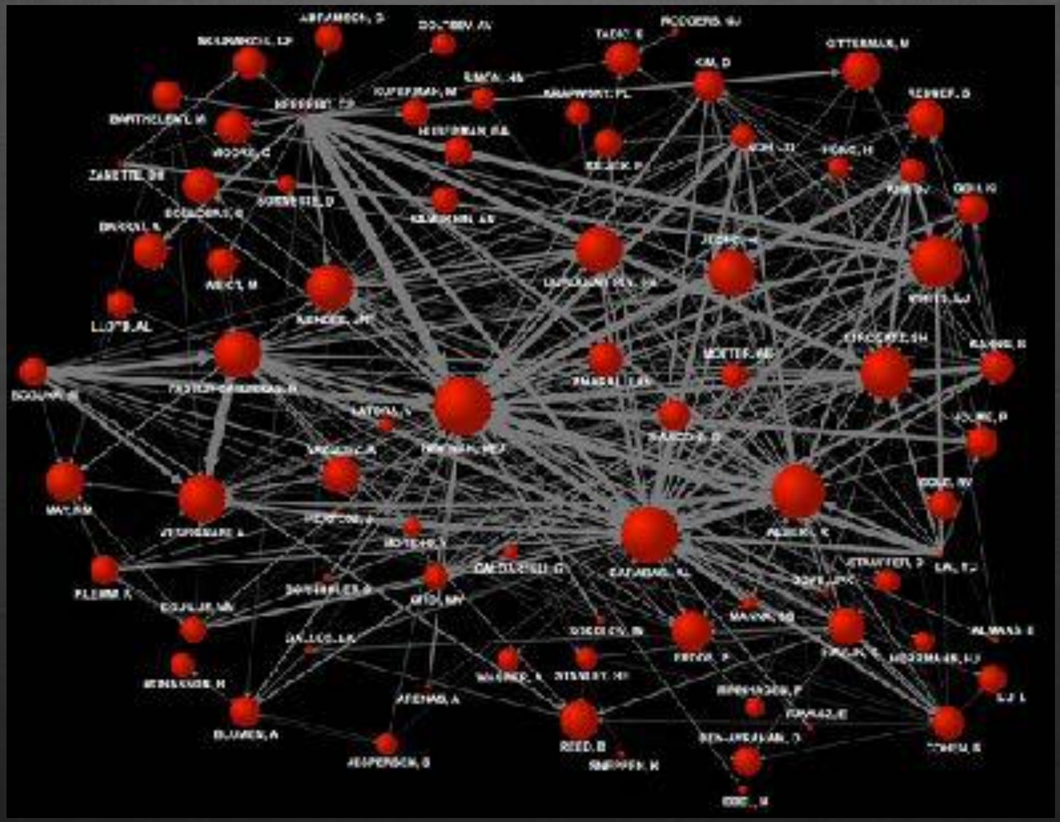
- Ecological networks. Nodes are species. Links their interactions
- Example 1) Food webs. Nodes are species. Links describe predator-prey interactions



WHY DO WE CARE?

Networks of information

- Data items, connected in some way
- World Wide Web. Nodes webpages. Links, connections between them
- Citation networks. Nodes papers (patents/legal documents). Links citations between them



WHY DO WE CARE?

Technological Networks

- Phone networks
- Internet
- Power grids
- Transportation networks

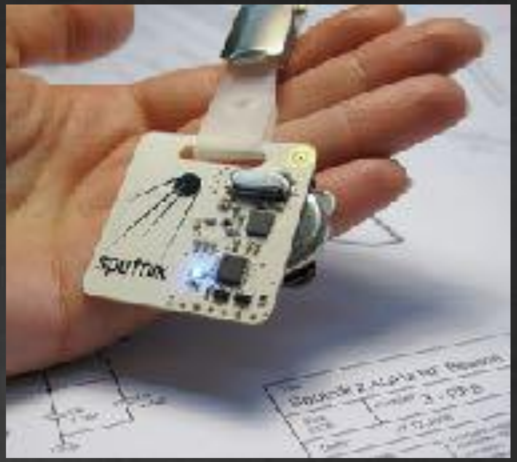
« »

WHY DO WE CARE?

Social Networks

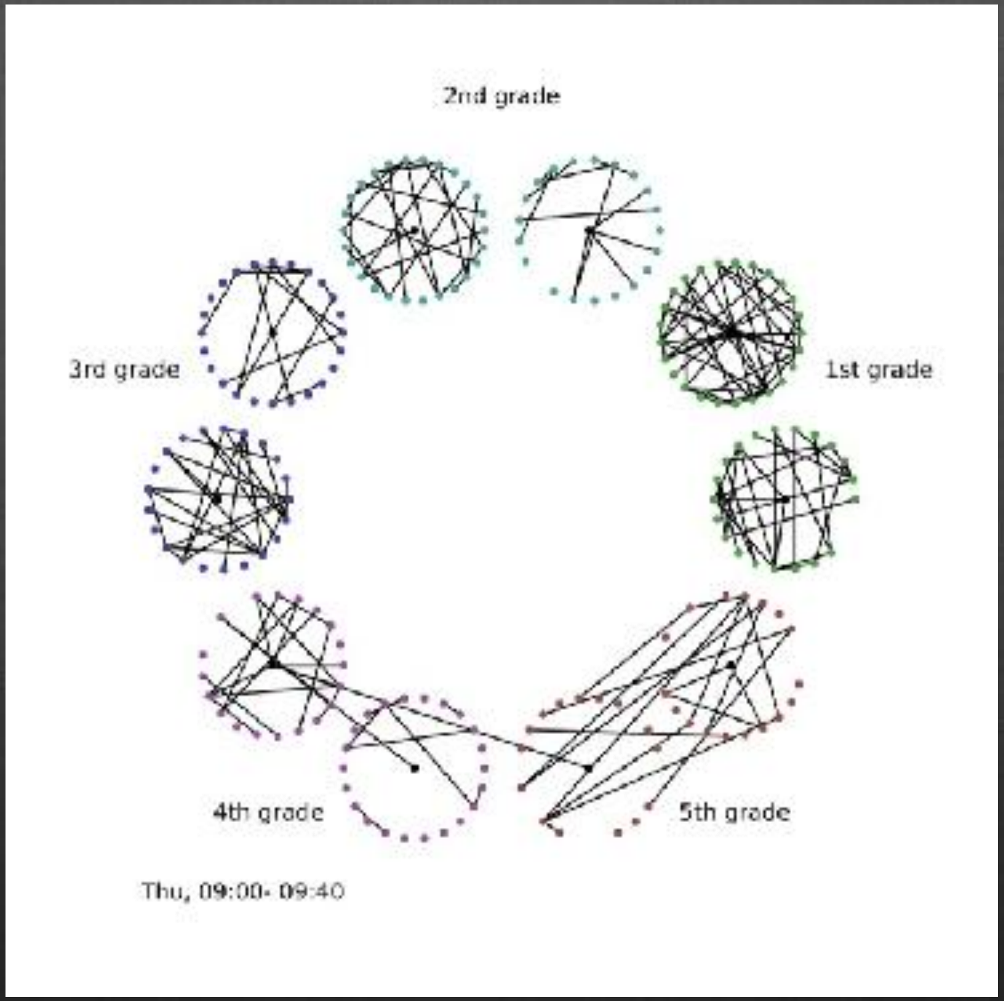
- Interviews and questionnaires
- Data from archival or third parties records

WHY DO WE CARE?



Social Networks

- Co-authorship networks
- Face-to-face networks



http://www.sociopatterns.org/

NETWORKS REPRESENTATION AND THEIR STATISTICAL FEATURES

NETWORKS AS GRAPHS

Basic Ingredients

- basic unites: nodes/vertices N
- their interactions: links, edges, connections E

$$G(N, E)$$

NETWORKS AS GRAPHS

Mathematical representation

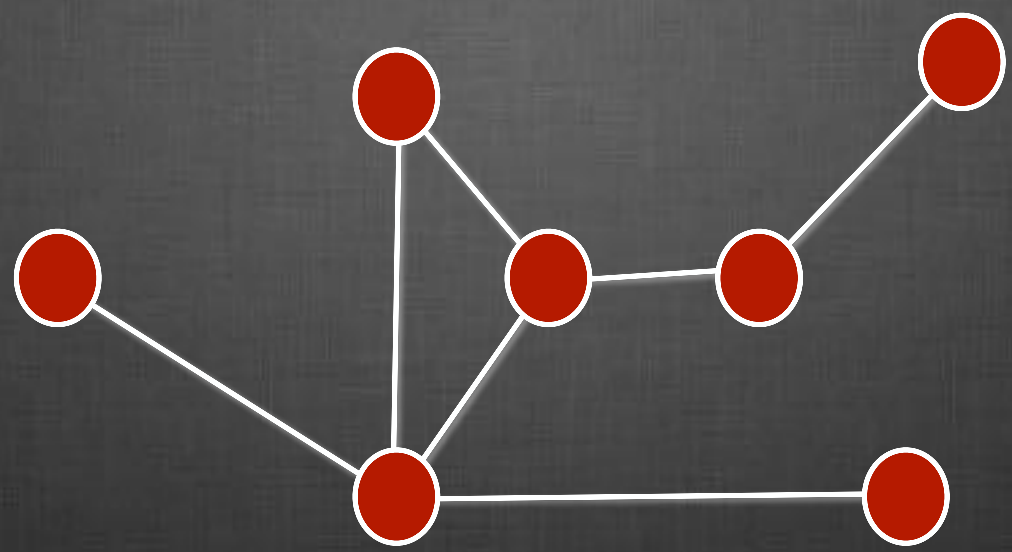
- adjacency matrix

$$A_{ij} = \begin{cases} 1 & \text{if there is a connection between i and j} \\ 0 & \text{otherwise} \end{cases}$$

UNDIRECTED NETWORKS

Symmetrical connections -> symmetrical adjacency matrix

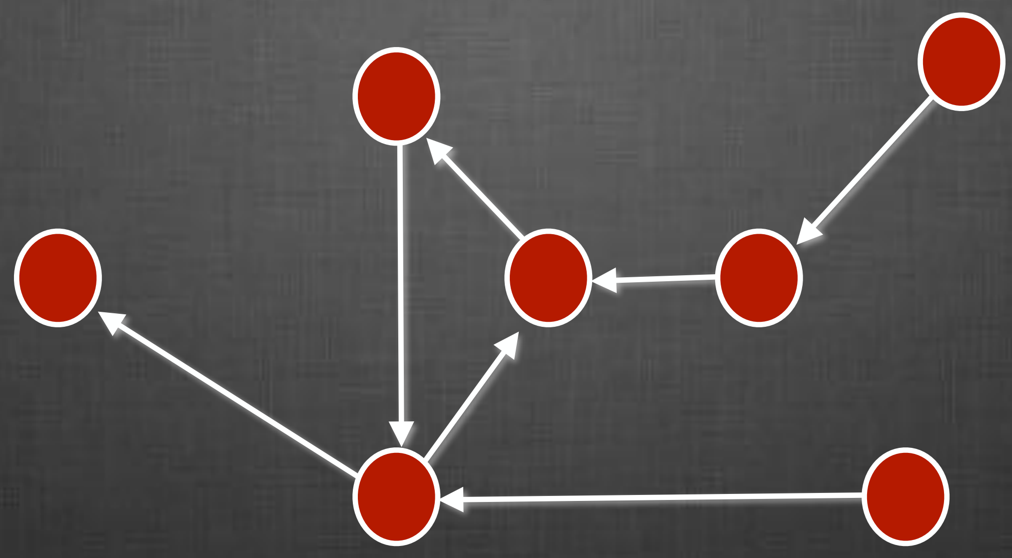
$$A = A^T$$



DIRECTED NETWORKS

Links (arcs) have direction

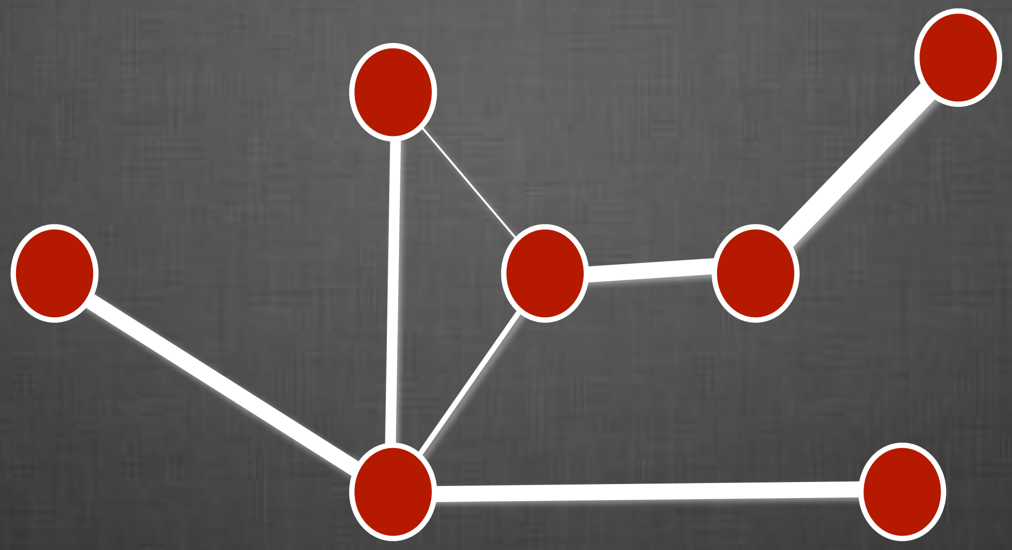
$$A \neq A^T$$



WEIGHTED NETWORKS

Links are not simply binary

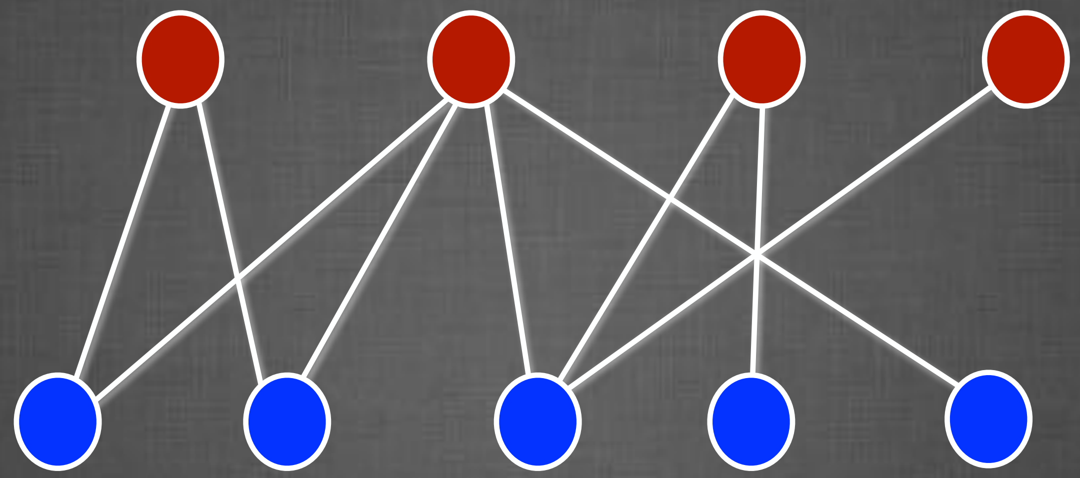
$$A_{ij} = \begin{cases} w_{ij} & \text{if i and j interacted w times} \\ 0 & \text{otherwise} \end{cases}$$



Typically weights are positive, but it is not necessary (signed networks)

BIPARTITE NETWORKS

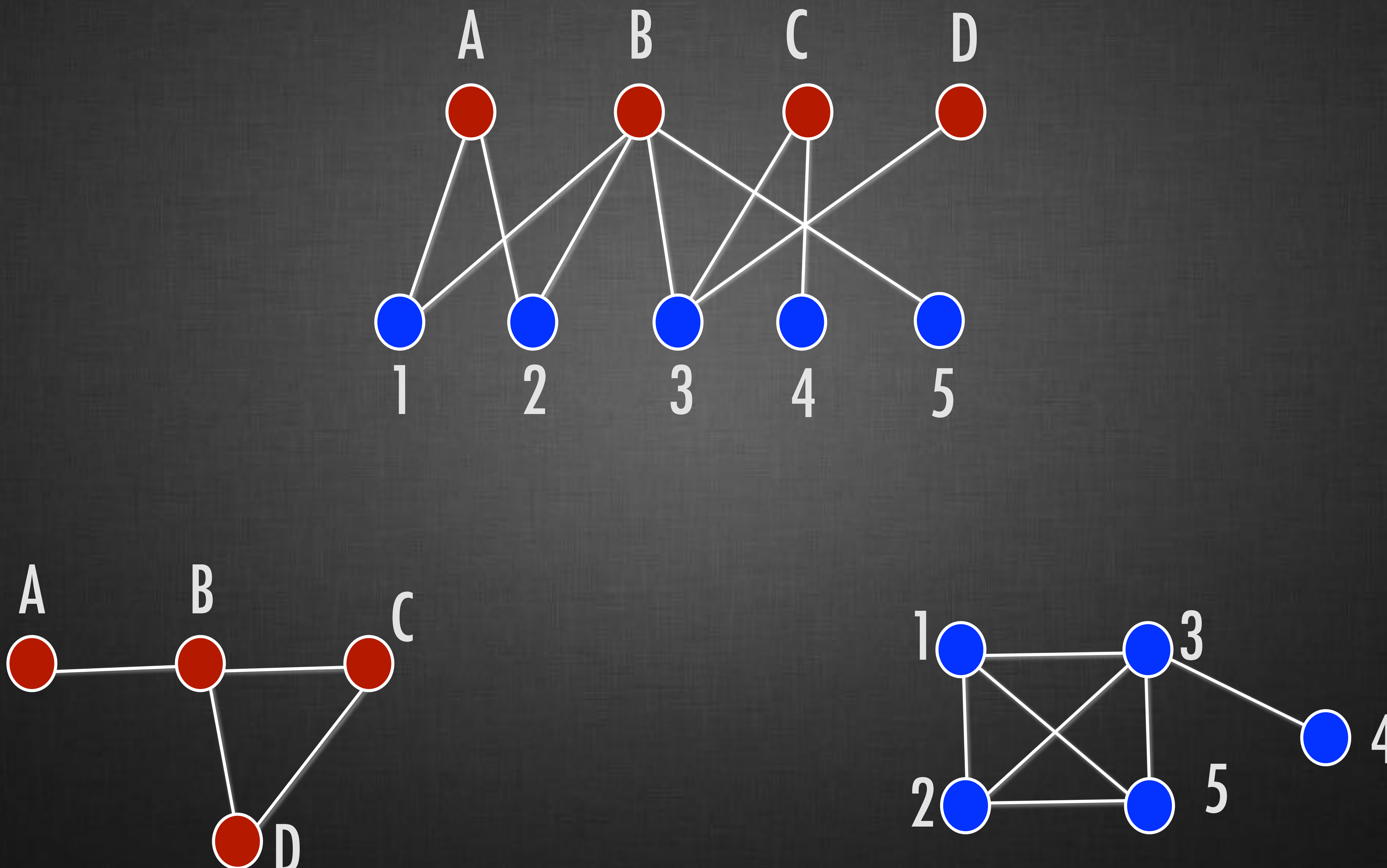
Two type of vertices



Incidence matrix [m,n]

$$B_{ij} = \begin{cases} 1 & \text{if j belongs to i} \\ 0 & \text{otherwise} \end{cases}$$

PROJECTIONS OF BIPARTITE NETWORKS



BASIC MEASURES

Degree

- number of connections of each node

$$k_i = \sum_j A_{ij}$$

Degree in directed networks

- in-degree
- out-degree

$$k_i^{IN} = \sum_j A_{ij}^T$$

$$k_i^{OUT} = \sum_j A_{ij}$$

Strength

- total number of interactions of each node

$$s_i = \sum_j A_{ij}$$

BASIC MEASURES

Degree

- what is the sum of all the degree?

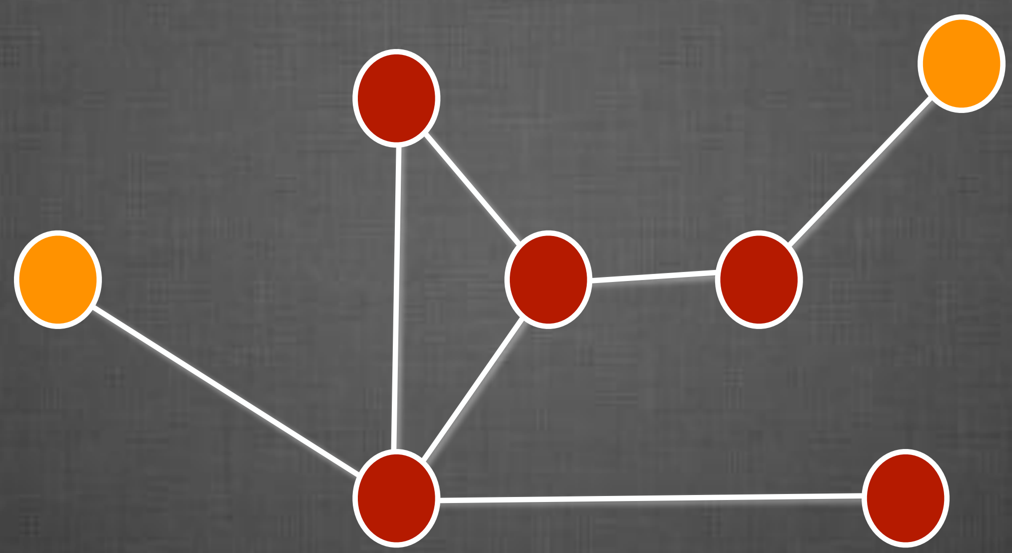
$$\sum_i k_i = 2E$$

$$\langle k \rangle = \frac{1}{N} \sum_i k_i = \frac{2E}{N}$$

BASIC MEASURES

Path

- sequence of nodes between i and j



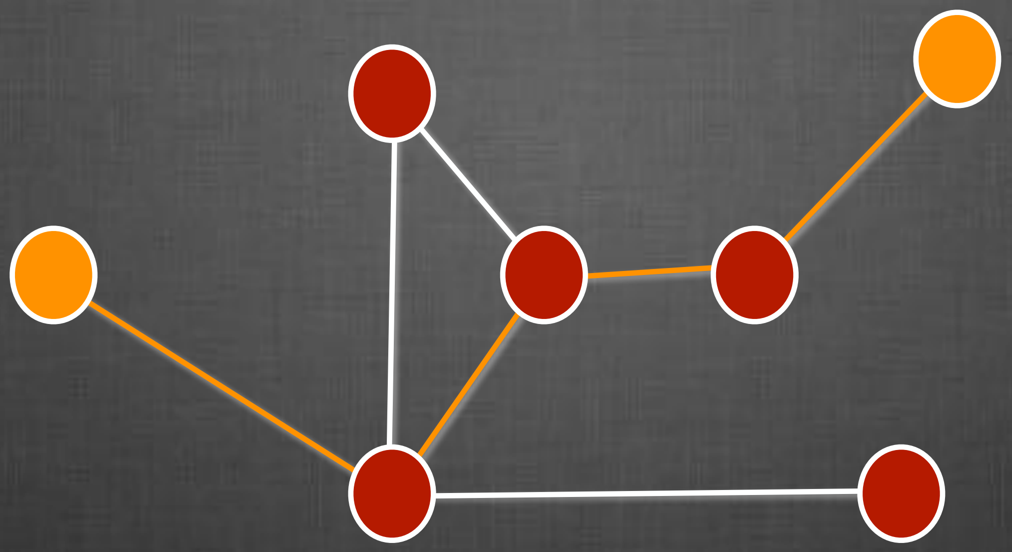
Path length

- number of hops between i and j

BASIC MEASURES

Geodesic Path

- the path with the shortest path length

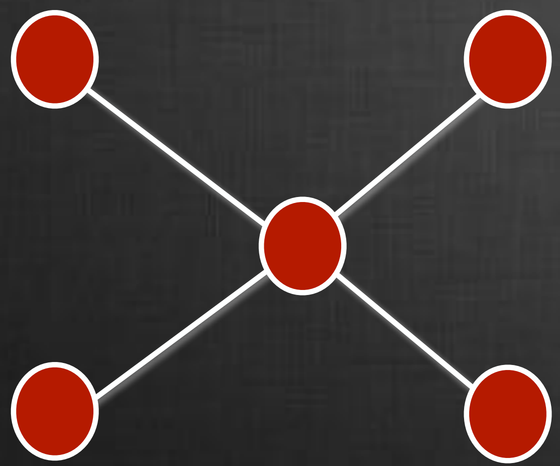


BASIC MEASURES

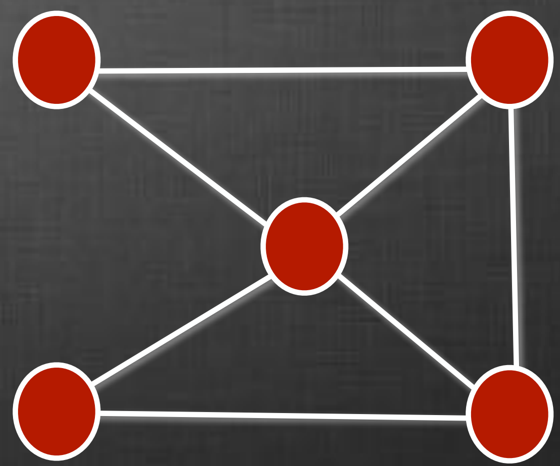
Local clustering

- for any i it is the fraction of the neighbours that are connected

$$c_i = \frac{e_i}{\frac{k_i(k_i-1)}{2}}$$



$c_i = 0$



$c_i = 0.5$

STATISTICAL DESCRIPTION OF NETWORKS MEASURES

In large systems statistical descriptions are necessary

- distributions

$$x \to P(x) \equiv \frac{N_x}{N}$$

$$\langle x \rangle = \sum_x x P(x)$$

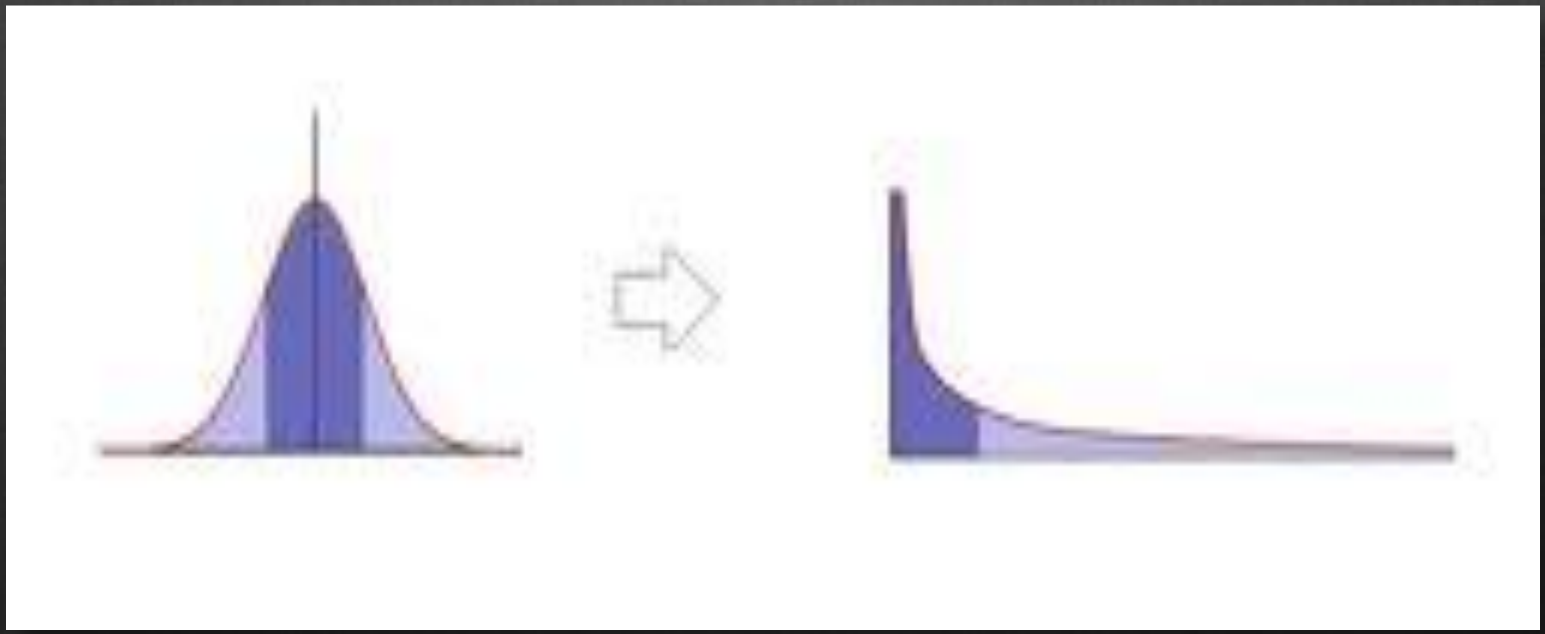
$$\langle x^n \rangle = \sum_x x^n P(x)$$

$$\sigma^2 = \sum_x (x-\mu)^2 P(x) = \langle x^2 \rangle - \mu^2 \equiv \langle x^2 \rangle - \langle x \rangle^2$$

DEGREE DISTRIBUTION IN REAL NETWORKS

Far from normal distributions

- the average is not a good descriptor of the distribution (absence of a characteristic scale)
- large variance -> large heterogeneity
- mathematically described by heavy-tailed (sometimes power-law) distributions



POWER LAWS

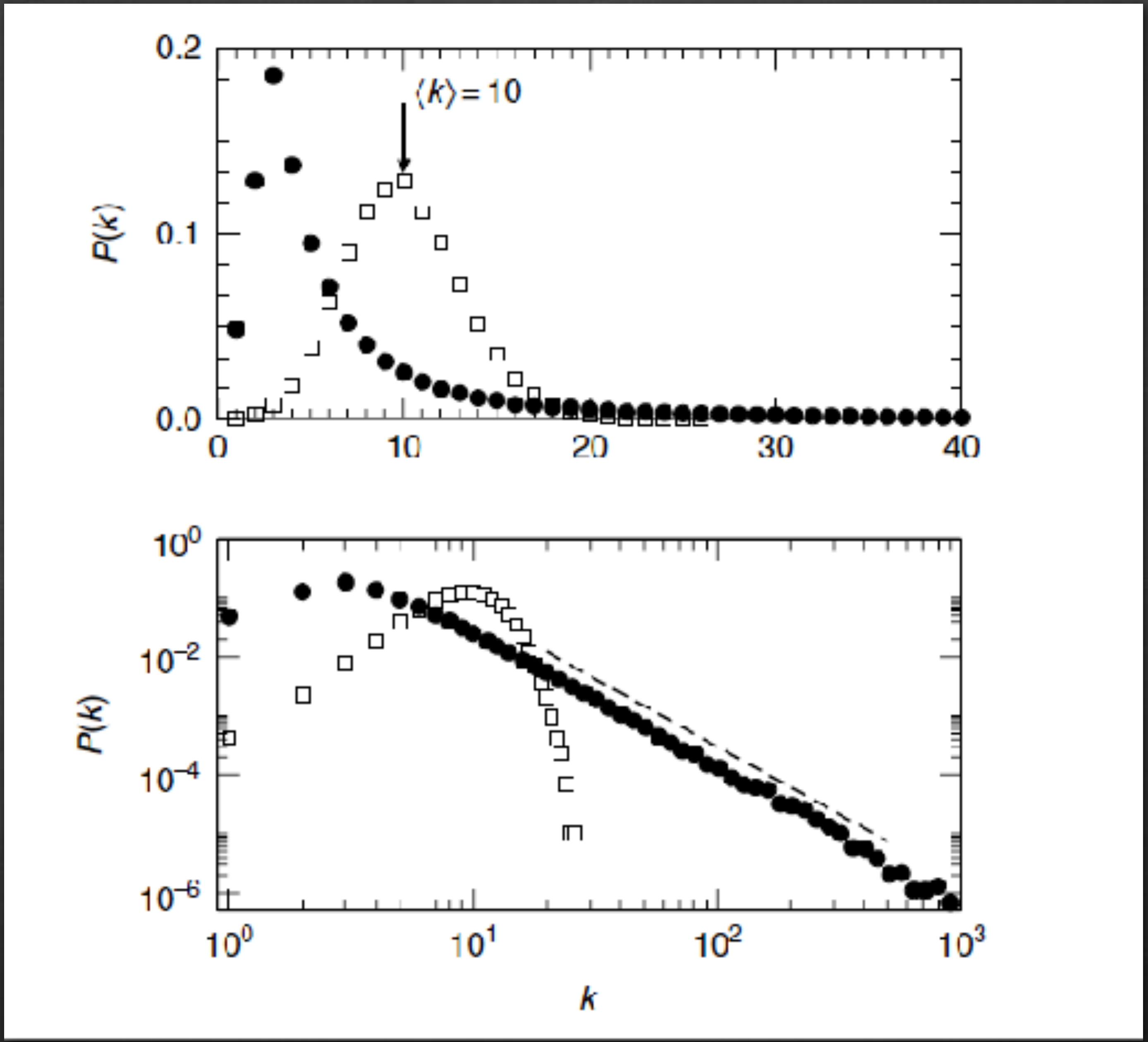
Power-laws

- scale invariance
- linear in log-log scale
- divergent moments depending on the exponent

$$f(x) = ax^{-\gamma} \rightarrow f(cx) = ac^{-\gamma}x^{-\gamma} \sim x^{-\gamma}$$

$$f(x) = ax^{-\gamma} \rightarrow \log(f(x)) = \log(a) - \gamma \log(x)$$

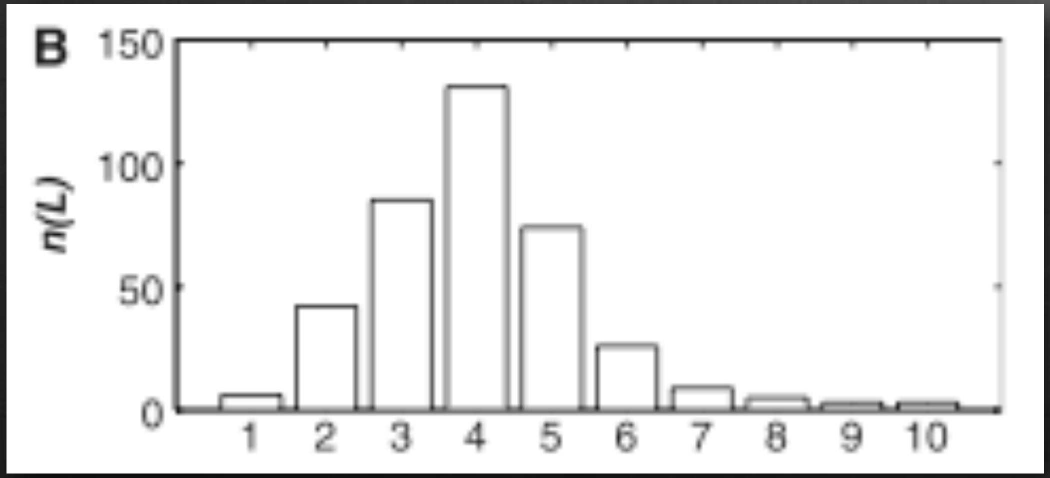
POWER LAWS



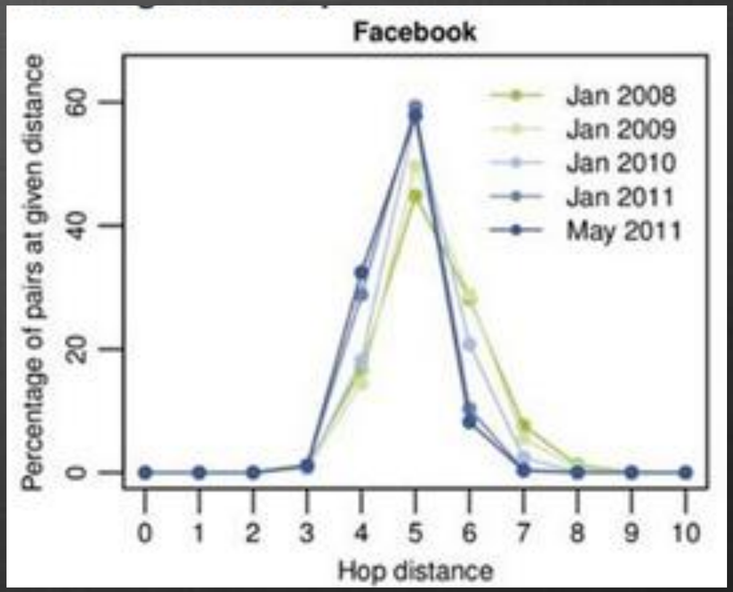
PATH LENGTH DISTRIBUTION IN REAL NETWORKS

Small-world phenomena

- even for very large graphs the average path length is very very small
- it scales logarithmically, or even slower, with networks' size
- the path length distribution is defined by a characteristic scale



Science, 301, 2003



https://www.facebook.com/notes/facebook-data-team/anatomy-of-facebook/10150388519243859

CLUSTERING IN REAL NETWORKS

Average local clustering

$$\langle C \rangle = \frac{1}{N} \sum_i C_i$$

Given a value, is it high or low?

- Null models
- typically high for social networks, typically low for technological networks
- still open and debated topic

REAL NETWORKS PROPERTIES

Generally speaking

- heavy-tailed degree distribution
- small-world phenomena
- large clustering (depends on the network type)

NETWORKS MODELS

Albert-Barabasi model (1999)

- based on preferential attachment (rich get richer), or Matthew effect (1968), Gibrat principle (1955), or cumulative advantage (1976)
- network growth

NETWORKS MODELS

The model

- network starts with m0 connected nodes
- at each time step a new node is added
- the node connects with m<m0 existing nodes selected proportionally to their degree

$$\Pi(k_i) = \frac{k_i}{\sum_l k_l}$$

NETWORKS MODELS

Albert-Barabasi model (1999)

- degree distribution

$$P(k) = 2m^2k^{-3}$$

NETWORKS MODELS

Albert-Barabasi model (1999)

- clustering

$$\langle C \rangle \sim \frac{(\ln N)^2}{N}$$

NETWORKS MODELS

Albert-Barabasi model (1999)

- path length

$$\langle l \rangle = \frac{\log N}{\log \log N}$$

NETWORKS MODELS

In summary

- the model creates scale-free networks
- small-world phenomena
- vanishing clustering

UNIVERSITY of GREENWICH

MODELING AND FORECASTING EPIDEMIC EVENTS

Nicola Perra
@net_science

DATA

Digital revolution

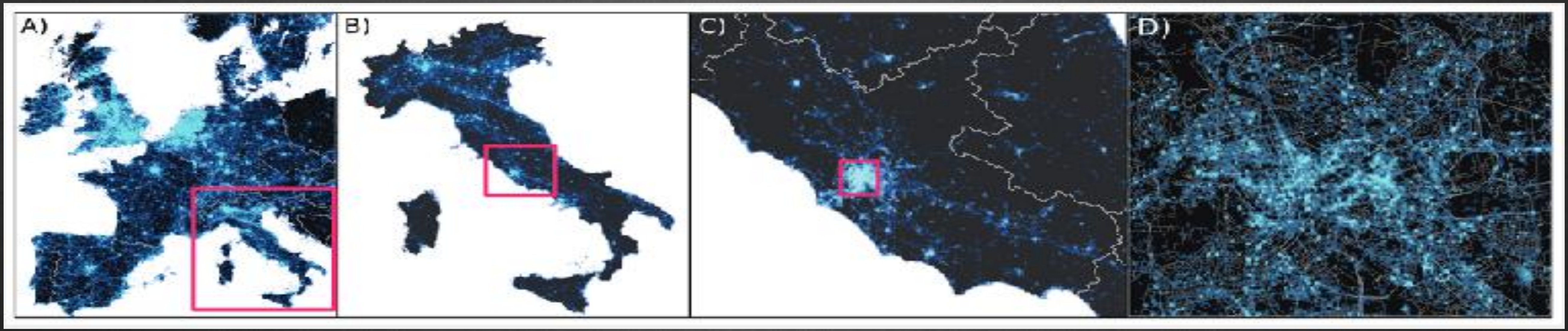
We are in a unique position in history

- unprecedented amount of data now available on human activities and interactions

From the "social atom" to "social molecules"

- dramatic shift in scale
- new phenomenology (More is different!)

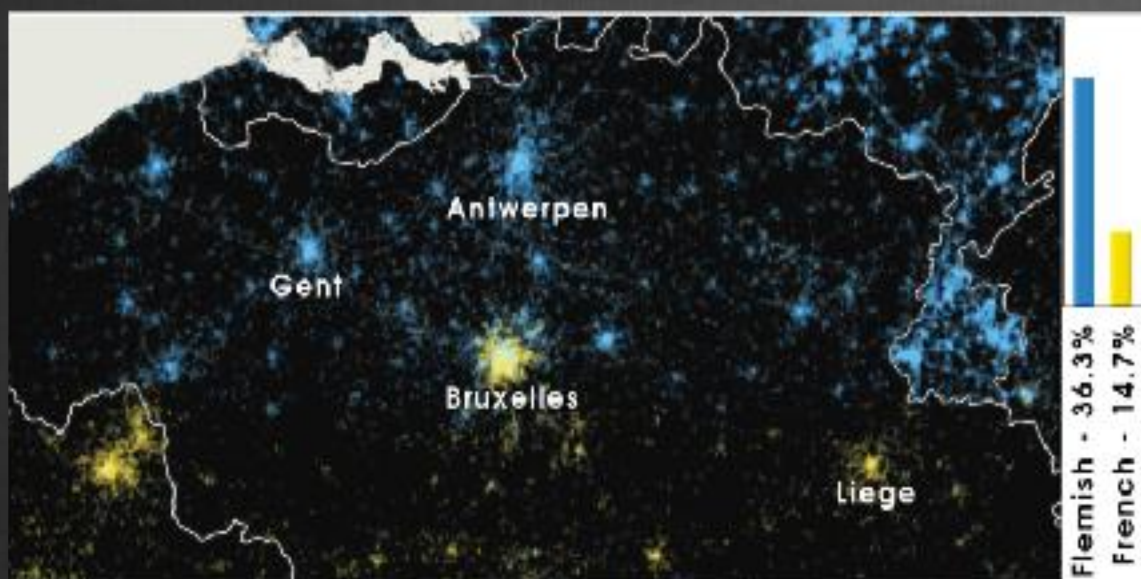
DATA

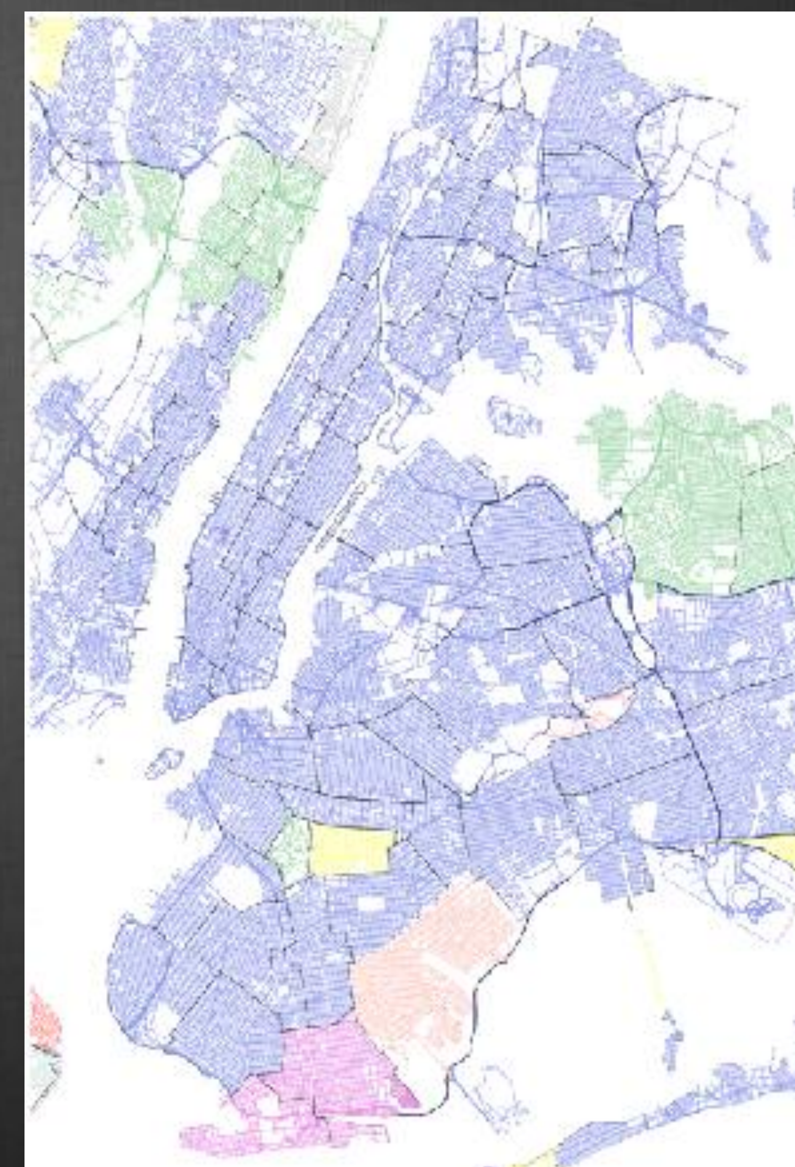


PLoS ONE, 8(4), 2013

PROBING SOCIO-DEMOGRAPHIC TREATS

Mapping language use at worldwide scale



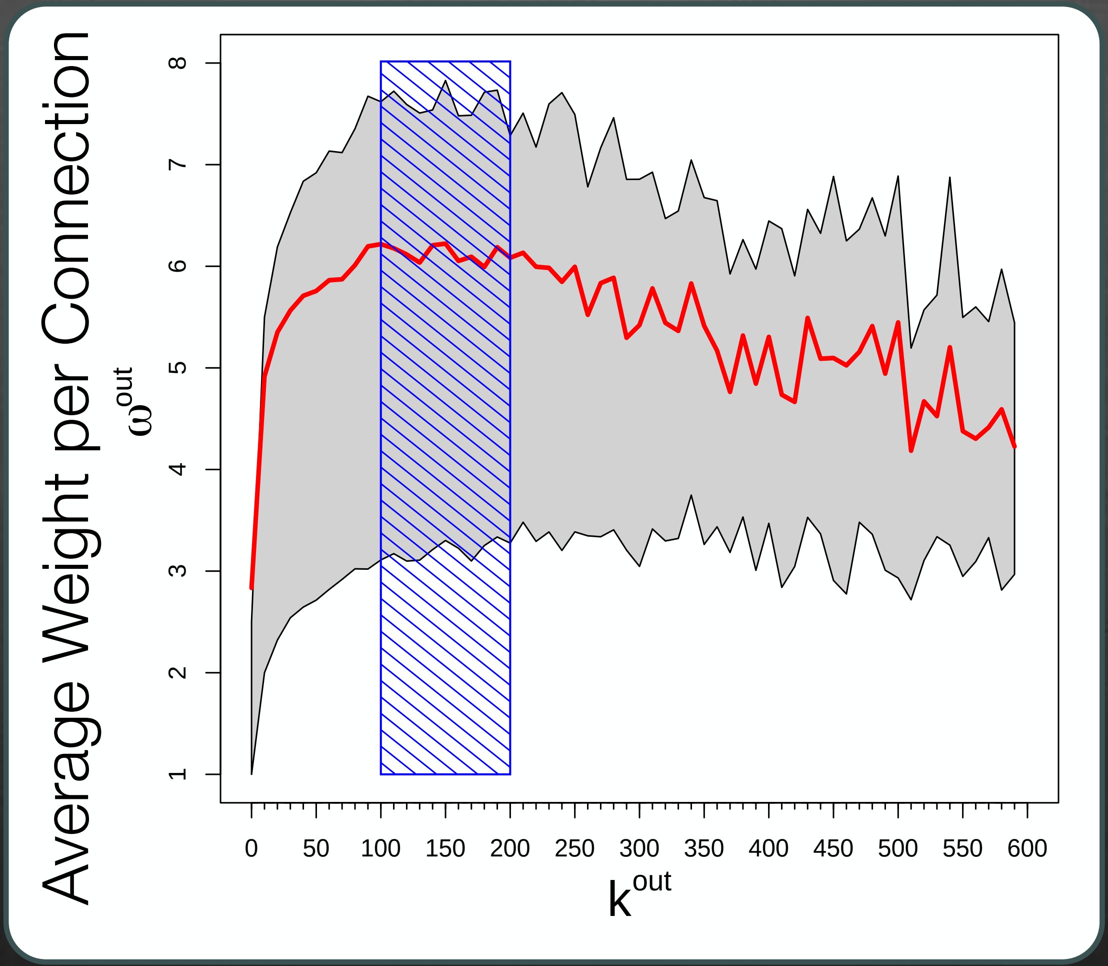


PLoS ONE, 8(4), 2013

PROBING COGNITIVE LIMITS

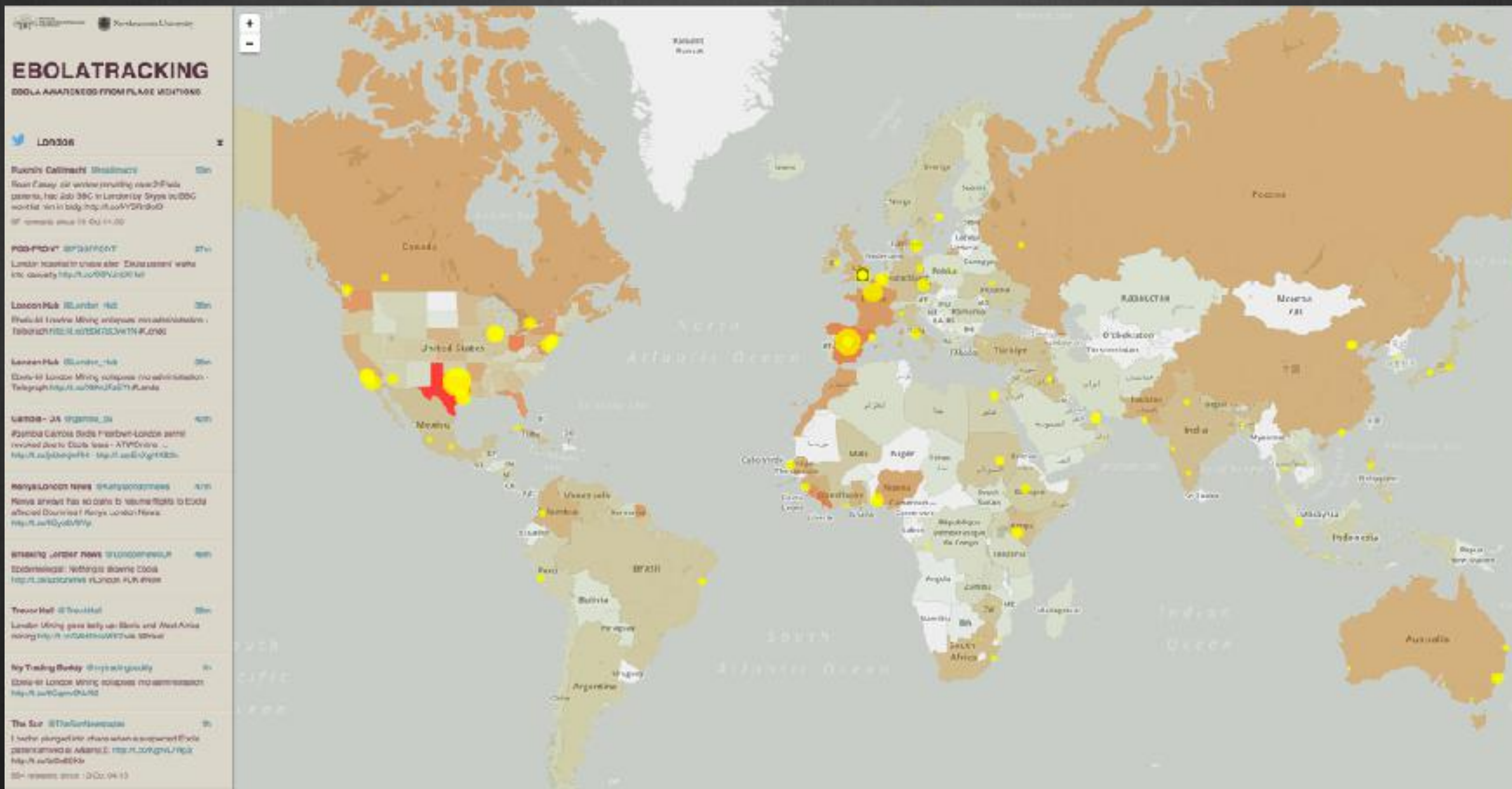
The social brain hypothesis

- typical social group size determined by neocortical size
- measured in various primates, extrapolated for humans: 100-200 (Dunbar's number)



PLoS ONE, 6(8), 2011

MAPPING THE GLOBAL DISCUSSION DURING EMERGENCIES



www.ebolatracking.org

UNIVERSITY of GREENWICH

PROBING HUMAN MOBILITY

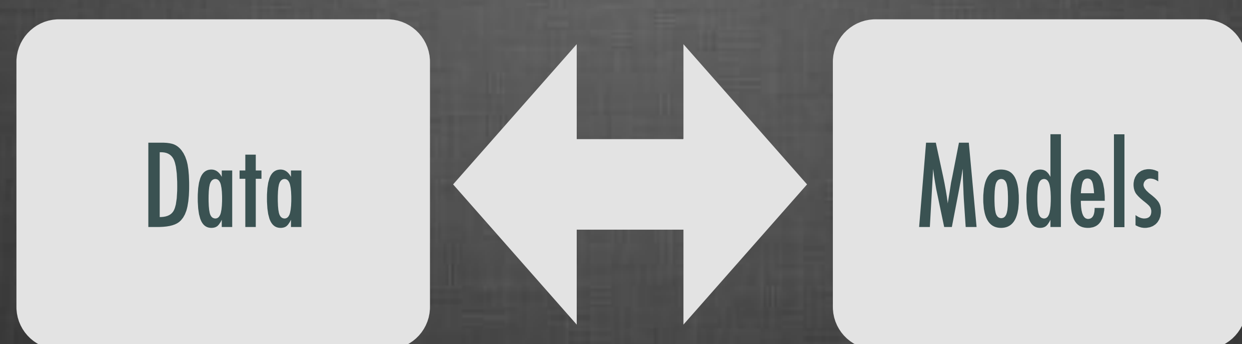
PROBING HEALTH STATUSES

Active and passive data collections

- (Active) participatory platforms
- (Passive) data harvesting

DATA ARE NOT ENOUGH!

WE NEED MODELS!

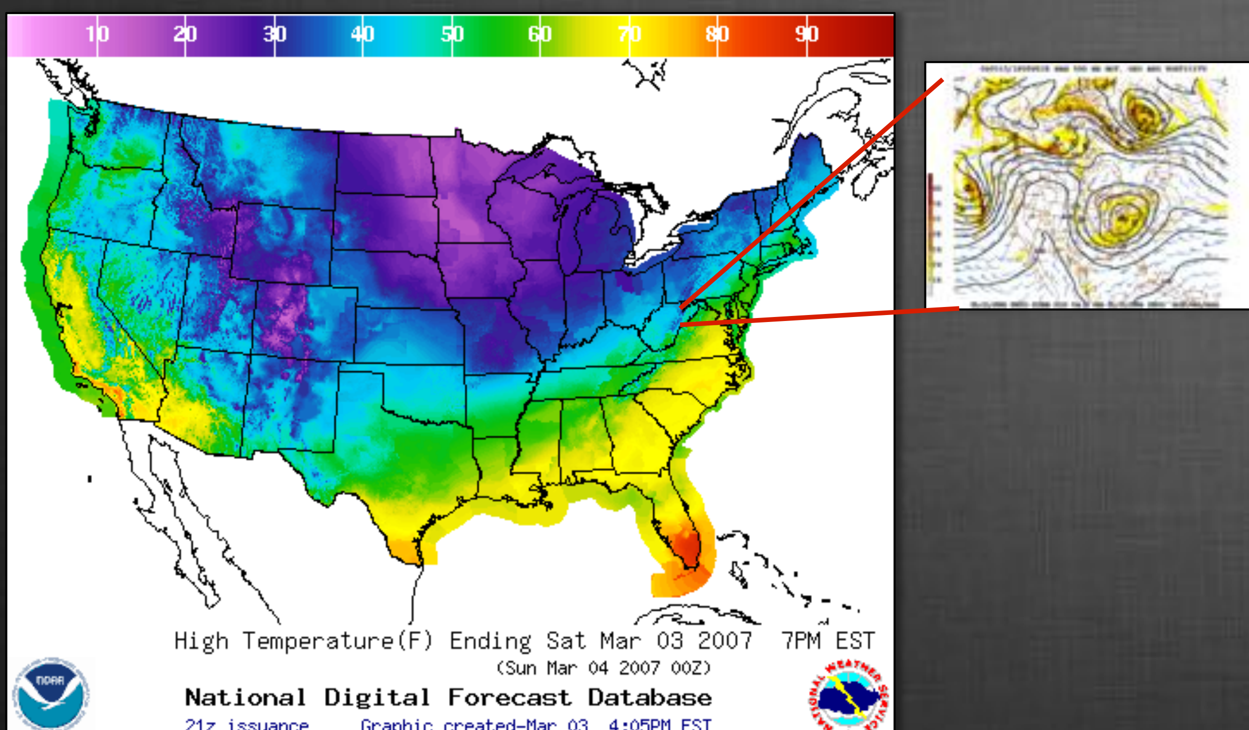


Holistic approach necessary --> Complex Systems/Networks

CAN WE FORECAST THE SPREADING OF INFECTIOUS DISEASES?

GOOD EXAMPLES

Weather Forecasts



WHY ARE WE ABLE TO FORECAST WEATHER?

Global collective effort

Large computational resources

Huge datasets

Deep knowledge of the Physical processes

FOR EPIDEMICS?

Global collective effort

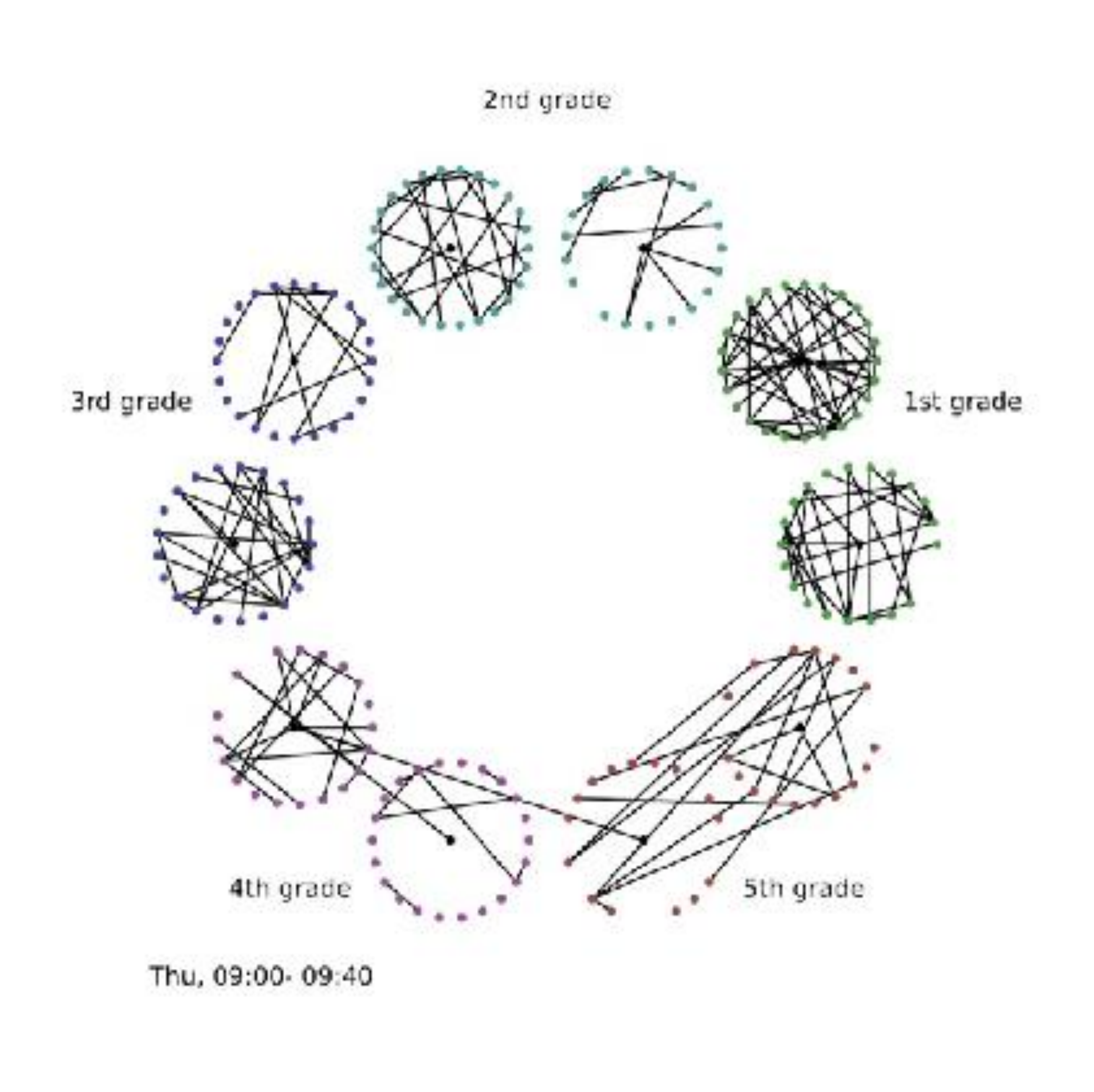
Large computational resources

Huge datasets

Deep knowledge of the Physical processes

NETWORK THINKING

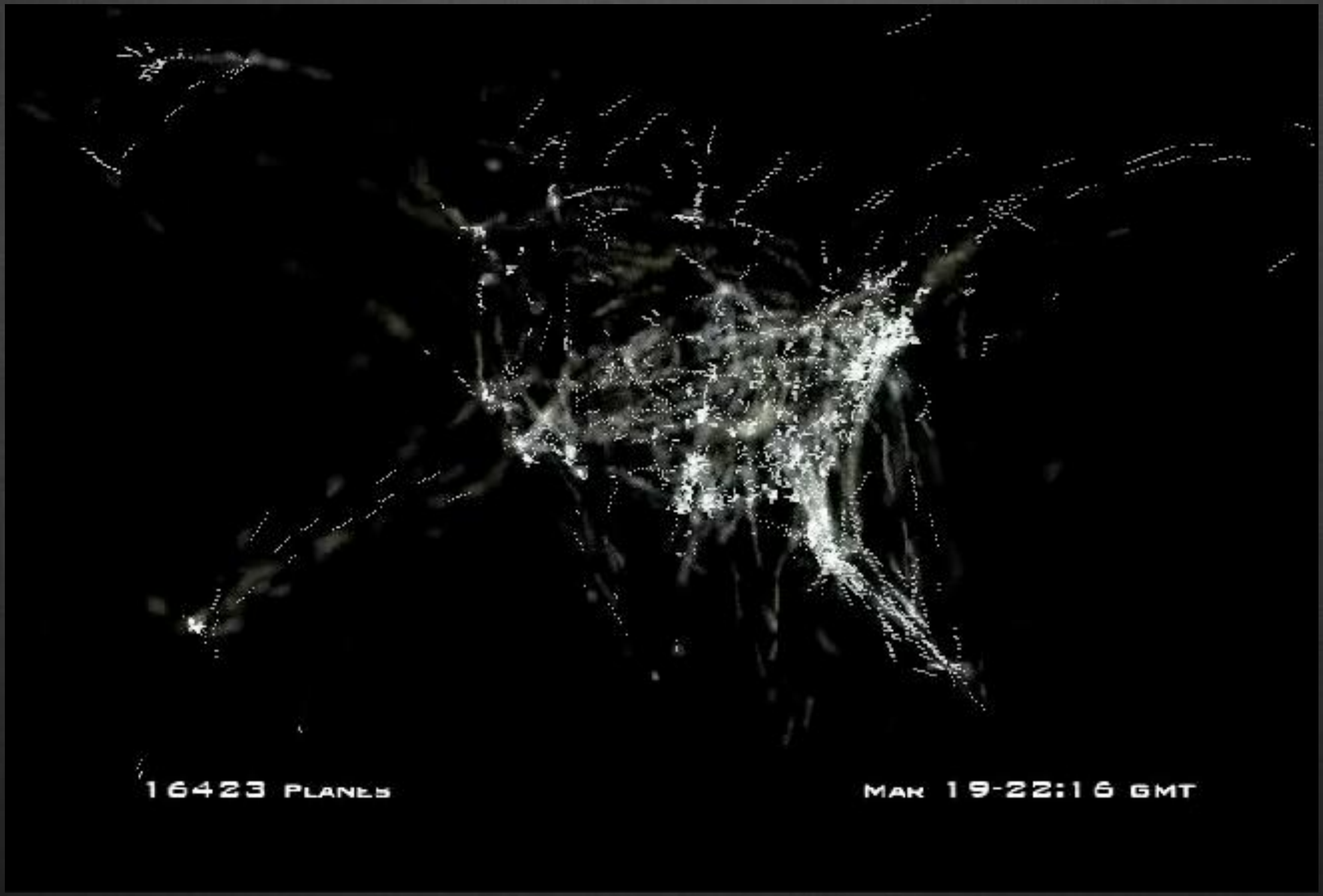
Human interactions are contact networks



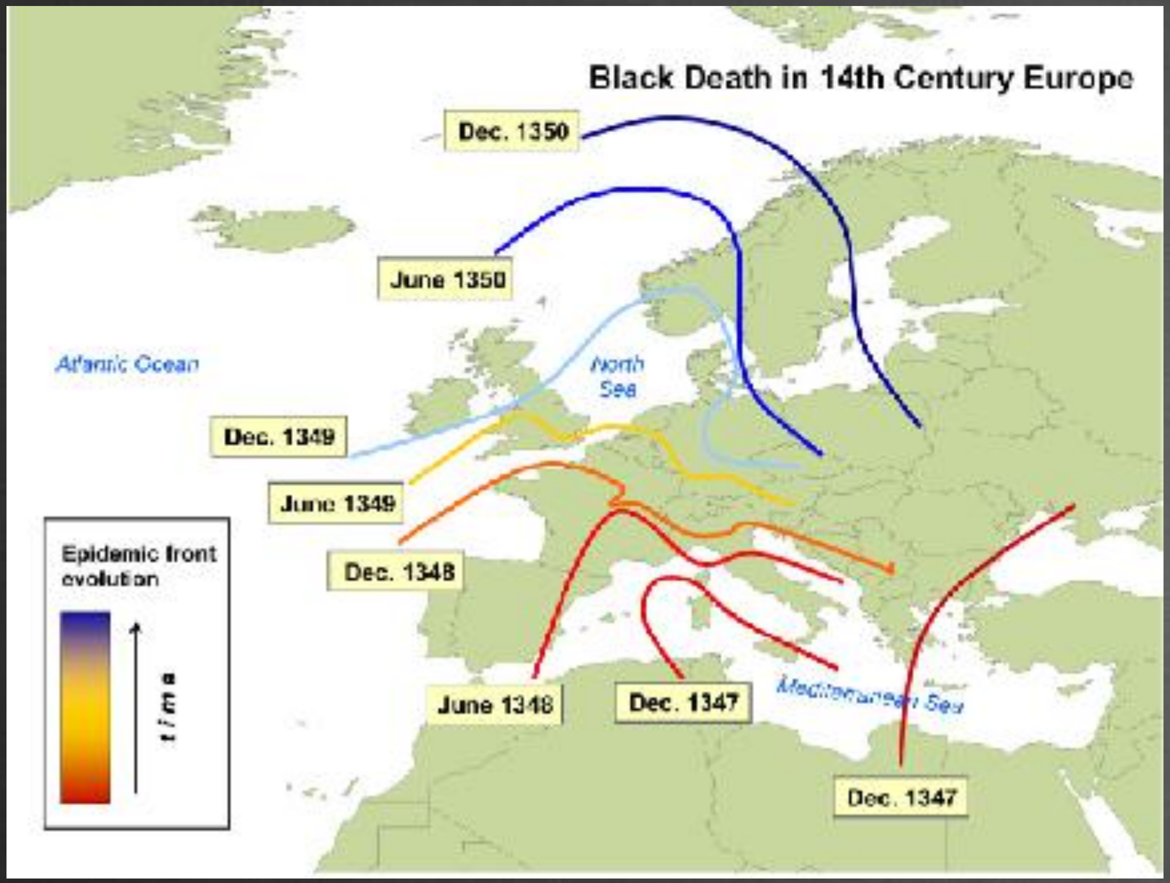
Within school contact patterns

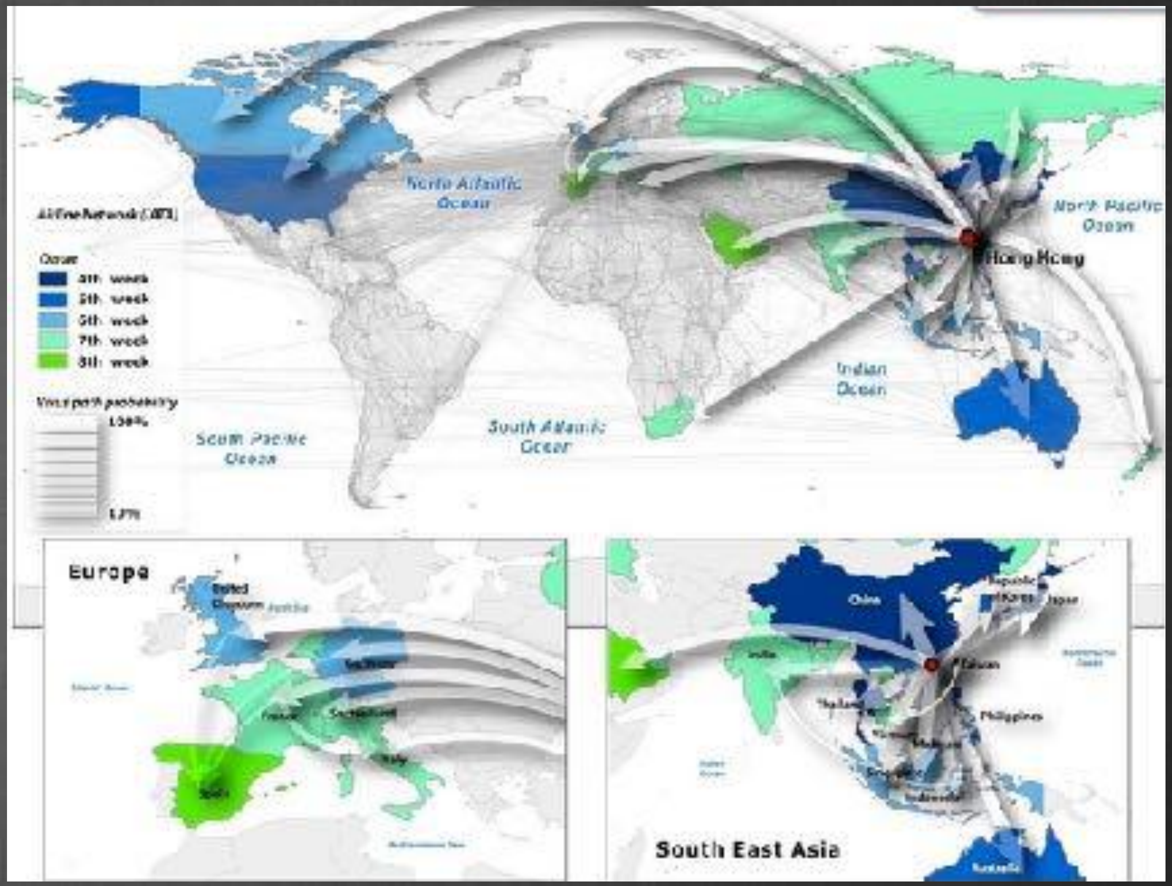
NETWORK THINKING

Mobility and epidemic spreding



NETWORK THINKING





Black death in1347: a continuous diffusion process

(Murray 1989)

SARS epidemics: a discrete network driven process

(Colizza et al. 2007; Brockmann&Helbing 2013)

NETWORKS ARE CENTRAL IN THE ANALYSIS OF CONTAGION PROCESSES

DISEASES SPREAD IN MULTI-LAYER NETWORKS

GLEAM

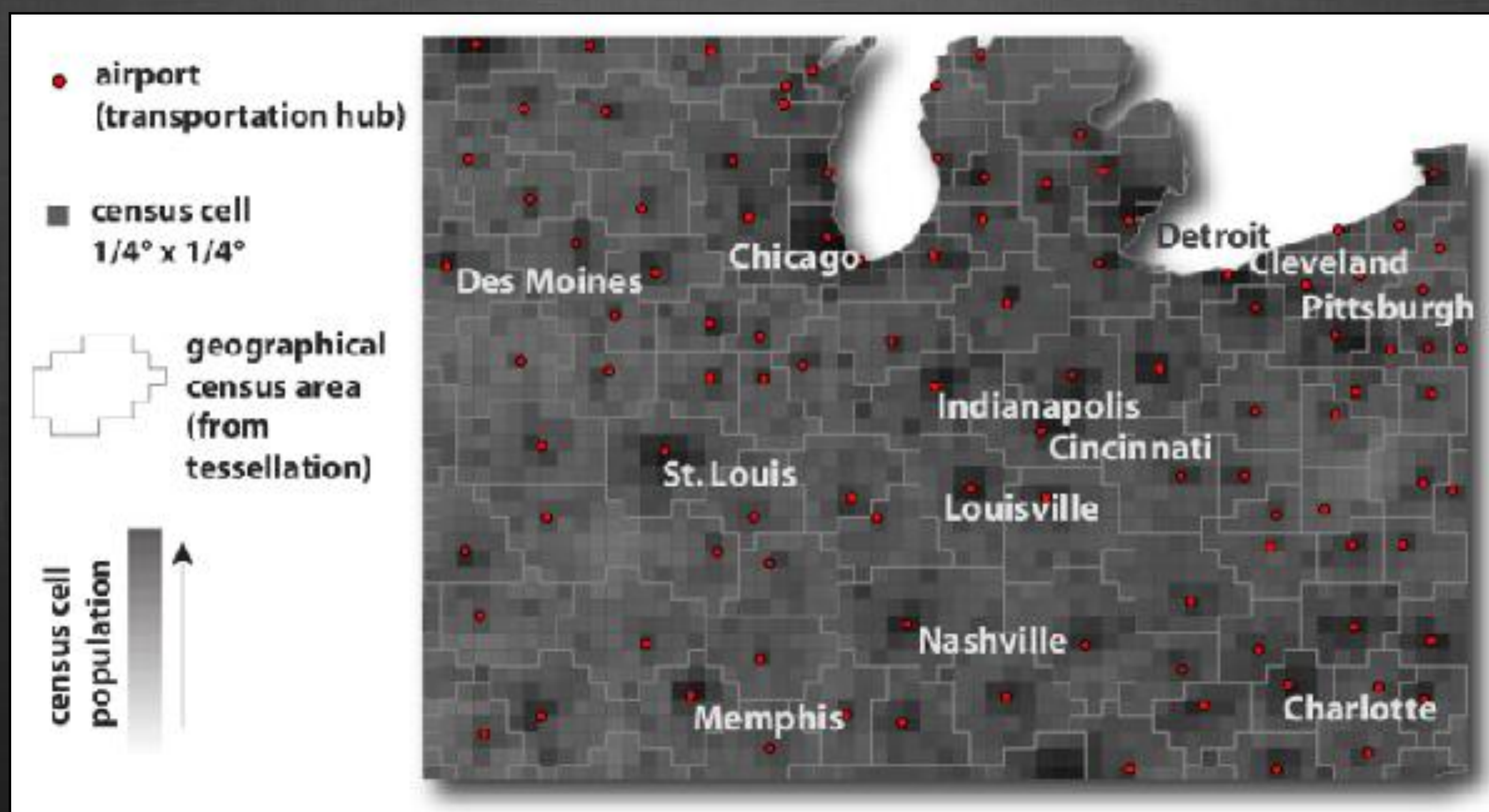
GLOBAL EPIDEMIC AND MOBILITY MODEL

WWW.GLEAMVIZ.ORG

POPULATION LAYER

Division of the earth in ~800K cells

Voronoi tessellation



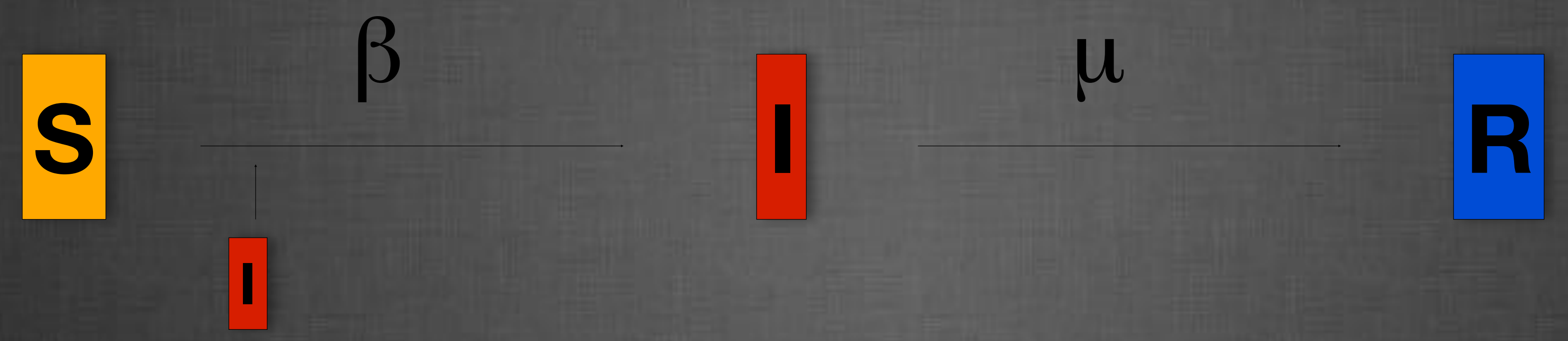
MOBILITY LAYER

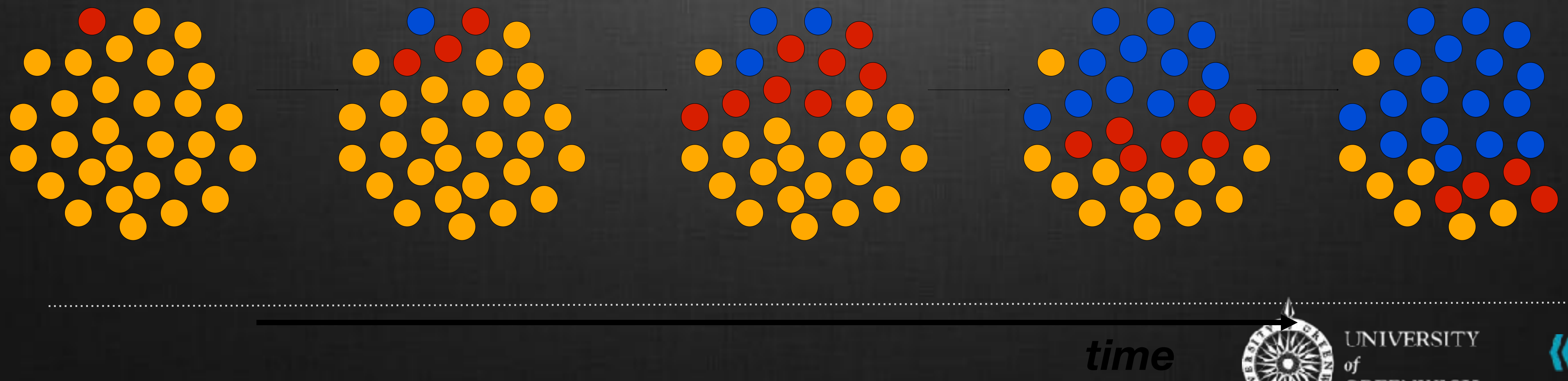
Long distance: 99% of the world wide air network

Short distance: real data+"gravity law"

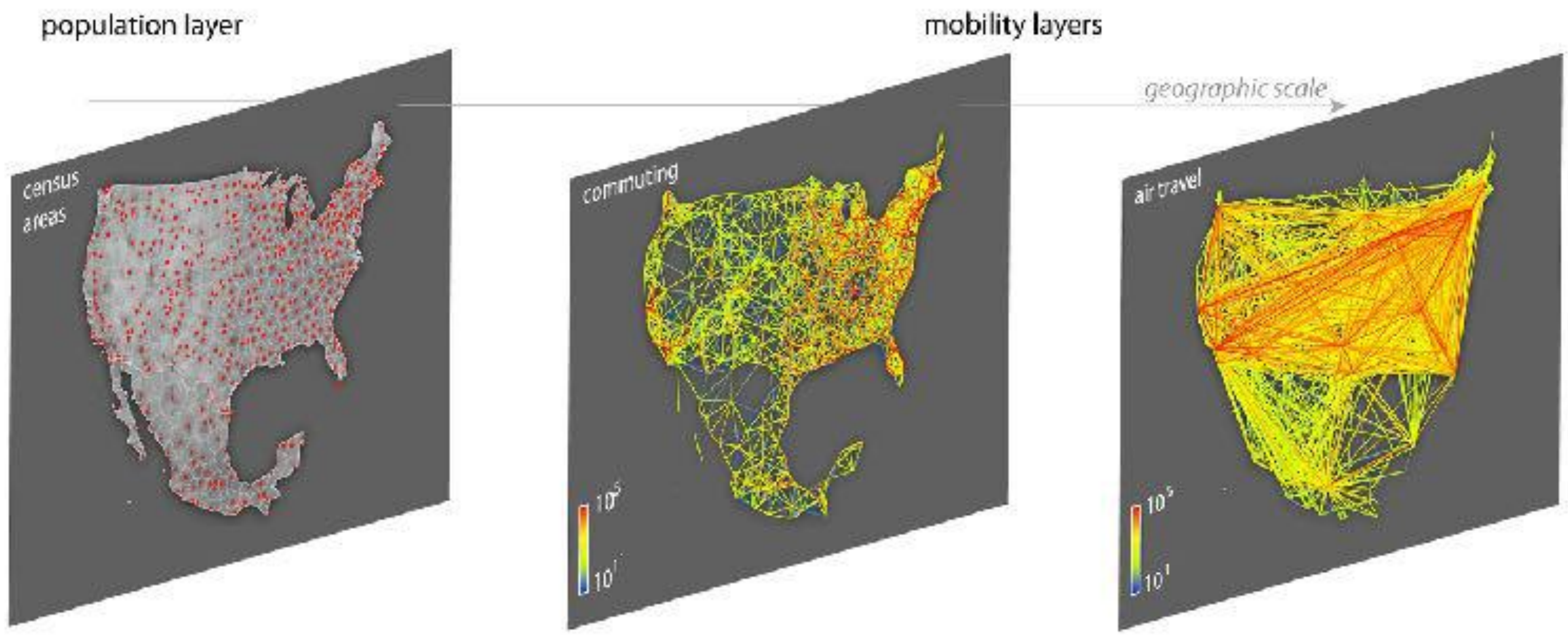
EPIDEMIC LAYER

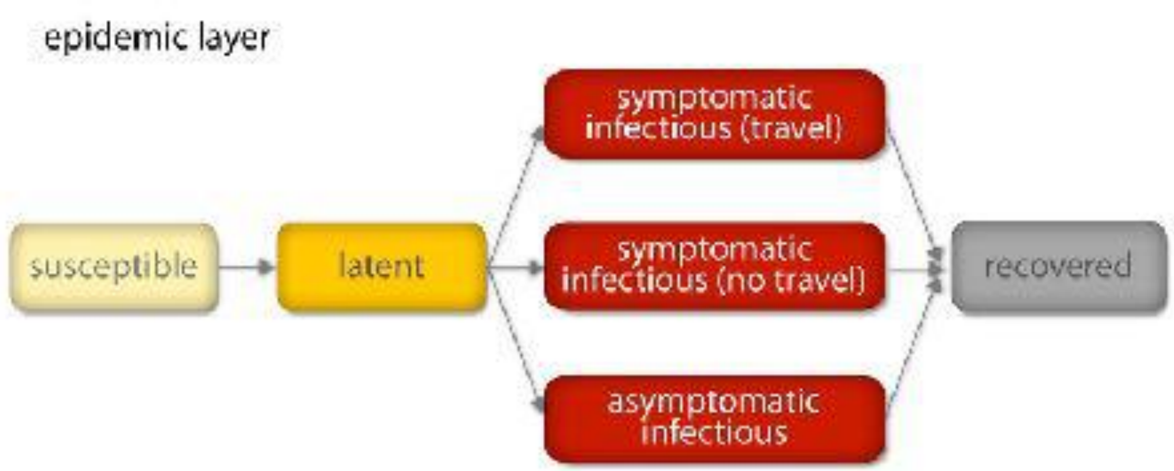
Any general model: according to the disease under study





DATA STRUCTURE

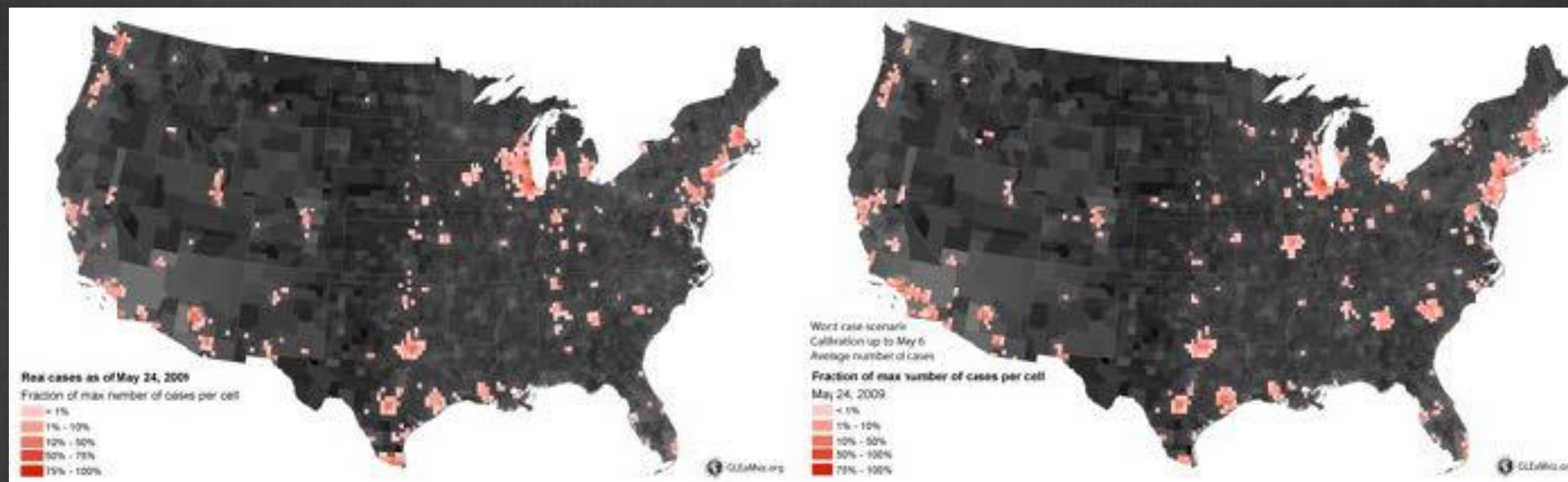


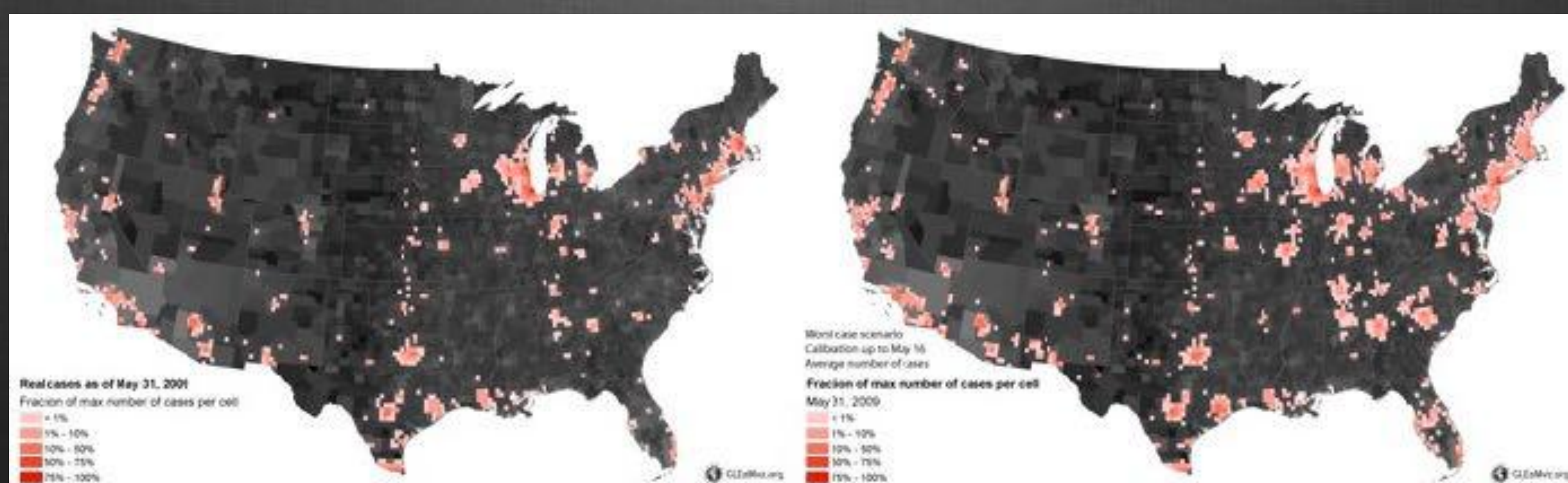


Parameter	Value	Description
β	from R_0	transmission probability
ε^{-1}	1.9 [1.1-2.5] d	average latency period
μ^{-1}	3 [3-5] d	average infectious period
p_t	50%	probability of traveling for infectious individuals
p_a	33%	probability of being asymptomatic
r_β	50%	relative infectiousness of asymptomatic infectious individuals

GLEAM AT WORK

SHORT TERM PREDICTIONS





Quantification of current risks

LONG TERM PREDICTIONS

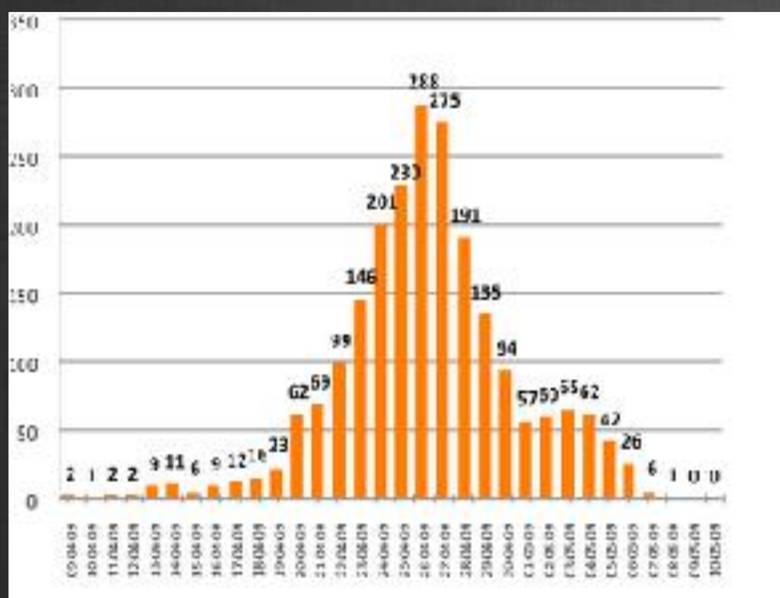
Crucial for vaccination campaigns

Characterisation of the unknown parameters

- Basic reproductive number, R0

LONG TERM PREDICTIONS

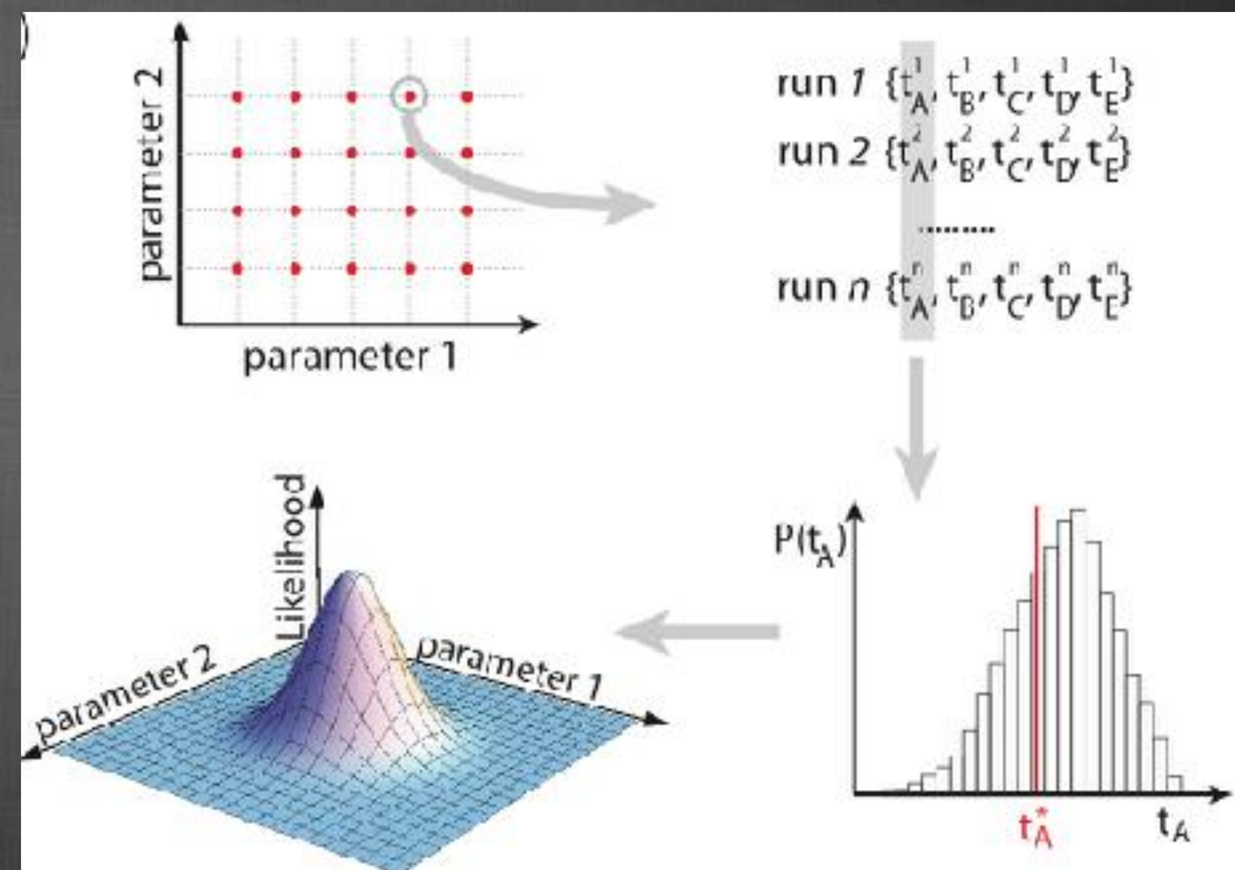
R0 estimation



Traditional approach
Fit the exponential phase

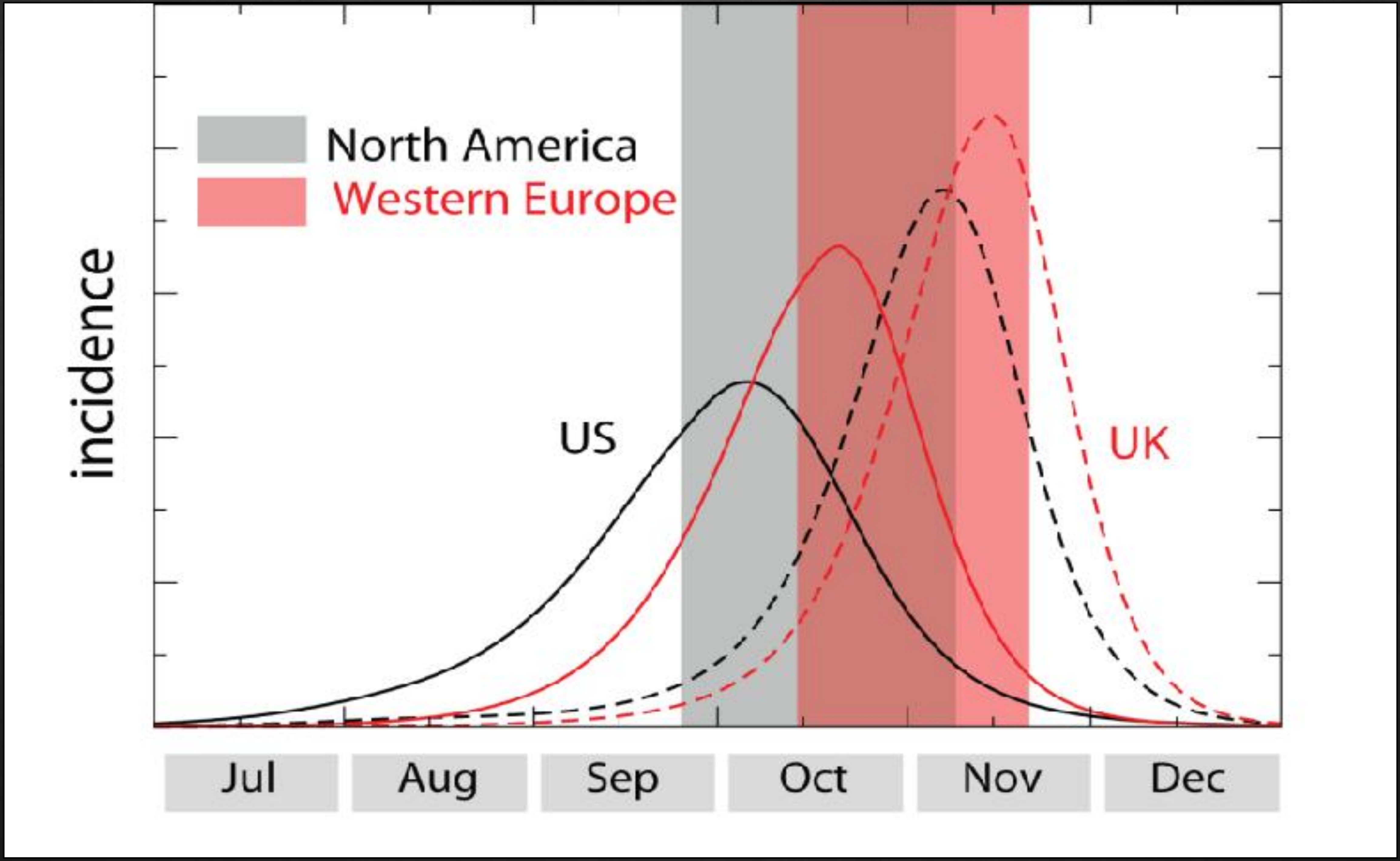
Our approach

Maximum Likehood on the arrival times



BMC, 7, 45, 2009

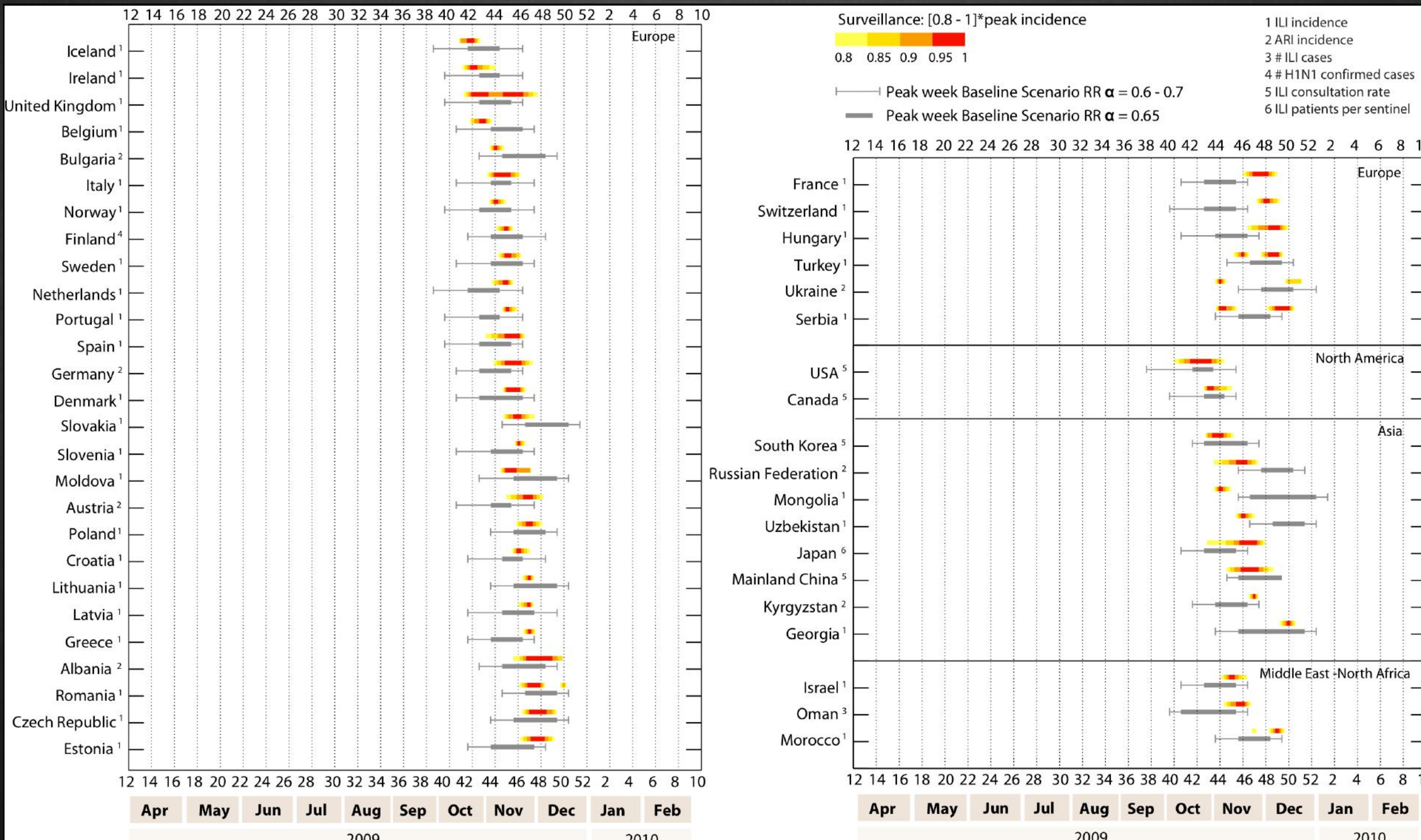
LONG TERM PREDICTIONS



BMC, 7, 45, 2009

UNIVERSITY of GREENWICH

MODEL'S ACCURACY



BMC, 10, 165, 2012

UNIVERSITY of GREENWICH

WHAT ABOUT THE SEASONAL FLU?

PREDICTING THE SEASONAL FLU

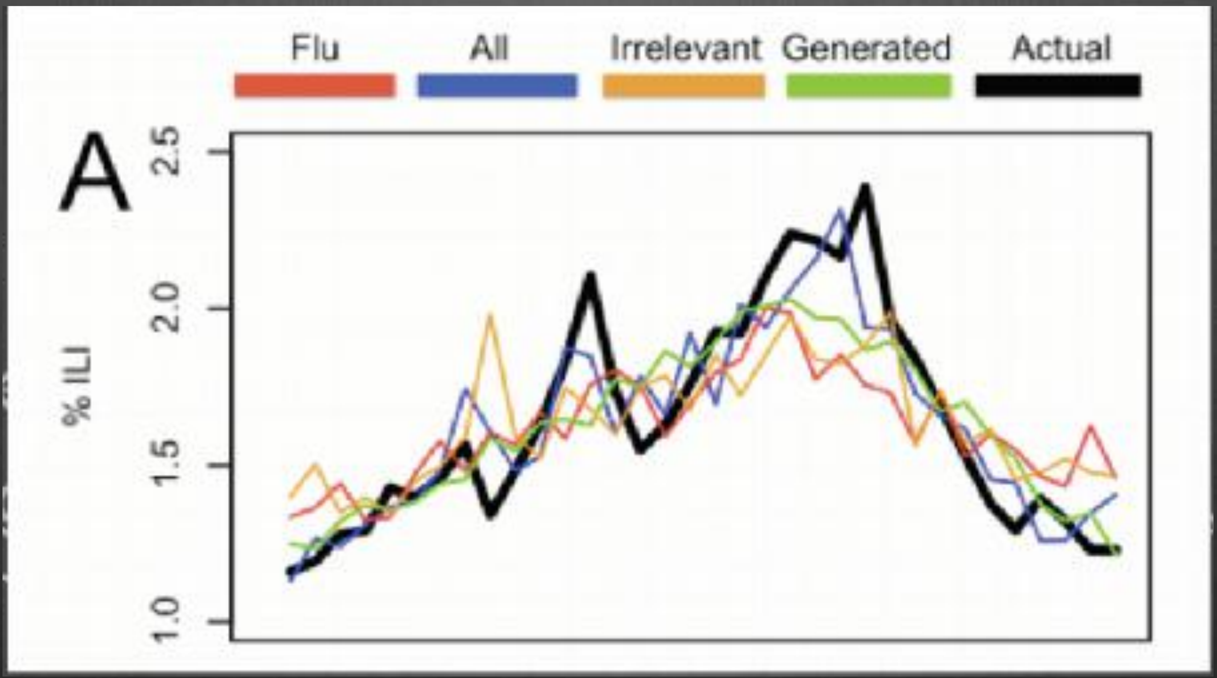
Major public health concern

- two modeling techniques: fits VS generative models

PREDICTING THE SEASONAL FLU

Classic time-series approach

- The goal is to find a correlation between a surveillance and another (more refined) data source such as Twitter or queries on google
- The parable of Google Flu Trends reveals the issues with this approach



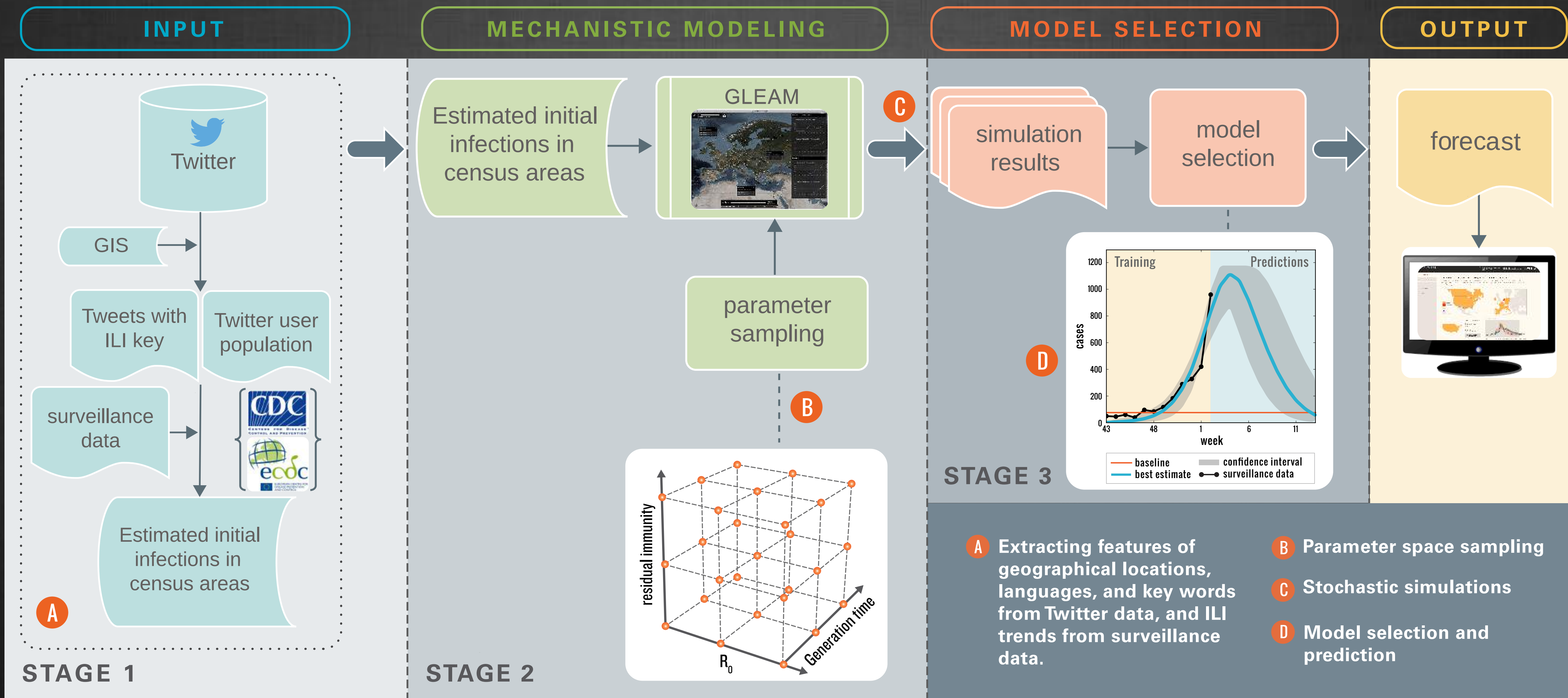
PREDICTING THE SEASONAL FLU

Generative models

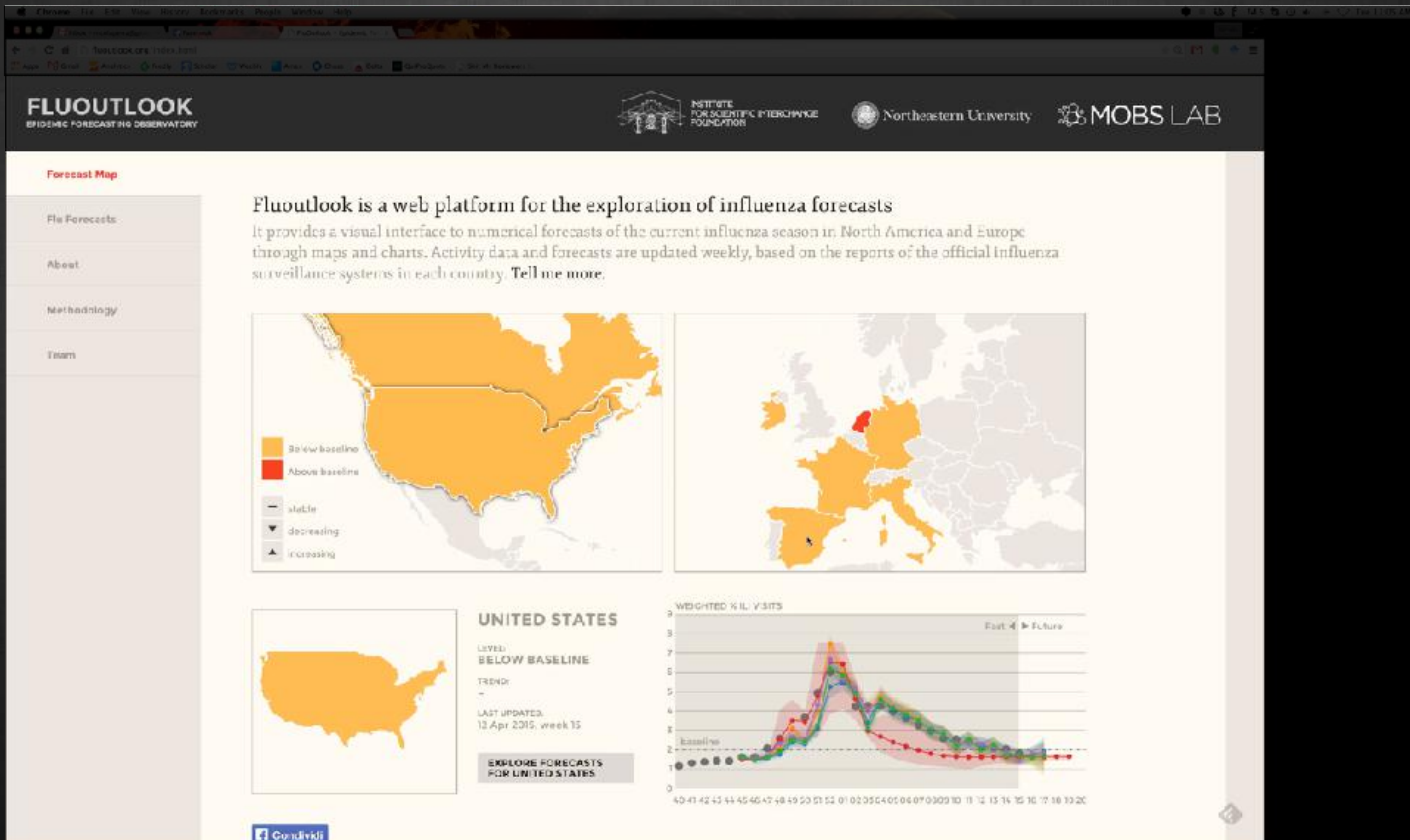
- Simulate the actual infection process
- They requires a lot of data as “initial conditions” that are typically not available during the outbreak

CAN WE MERGE THE TWO?

MODELING THE SEASONAL FLU



MODELING THE SEASONAL FLU



www.fluoutlook.org

THANKS TO

A. Vespignani
D. Mistry
K. Sun
Q. Zhang
C. Cattuto
M. Quaggiotto
M. Delfino
A. Panisson
D. Paolotti
M. Tizzoni
L. Rossi
S. Meloni
Y. Moreno
L. Weng
A. Flammini
F. Menczer
A. Baronchelli
M. Starnini
B. Goncalves
C. Castillo
E. Ubaldi
F. Ciulla
T.S. Lu

F. Bonchi
L.M. Aiello
J. Ratkiewicz
M. Martino
C. Dunne
B. Riberio
M.V. Tommasello
C. Tessone
F. Schweitzer
M. Karsai
V. Colizza
C. Poletto
D. Chao
H. M. Halloran
I. Longini
V. Loreto
G. Caldarelli
A. Chessa
R. Pastor-Satorras
J. Borge-Holthoefer
R. Burioni
S. Liu
D. Mocanu
R. Compton

Computational Social Sciences

Bruno Gonçalves · Nicola Perra *Editors*

Social Phenomena
From Data Analysis to Models

This book focuses on the new possibilities and approaches to social modeling currently being made possible by an unprecedented variety of datasets generated by our interactions with modern technologies. This area has witnessed a veritable explosion of activity over the last few years, yielding many interesting and useful results. Our aim is to provide an overview of the state of the art in this area of research, merging an extremely heterogeneous array of datasets and models. Social Phenomena: From Data to Models is divided into two parts. Part I deals with modeling social behavior under normal conditions: How we live, travel, collaborate and interact with each other in our daily lives. Part II deals with societal behavior under exceptional conditions: Protests, armed insurgencies, terrorist attacks, and reactions to infectious diseases. This book offers an overview of one of the most fertile emerging fields bringing together practitioners from scientific communities as diverse as social sciences, physics and computer science. We hope to not only provide an unifying framework to understand and characterize social phenomena, but also to help foster the dialogue between researchers working on similar problems from different fields and perspectives.

Physics
ISBN 978-3-319-14010-0
9 783319 140100
▸ springer.com

CSS
Gonçalves · Perra *Eds.*
Social Phenomena

Computational Social Sciences

Bruno Gonçalves
Nicola Perra *Editors*

Social Phenomena

From Data Analysis to Models

Springer