

BIG DIVE

TECH. CUSTOM EDITION

A project by **TOP-IX**
designed for **Intesa Sanpaolo**

#DataScience



machine learning basics

ciro cattuto

ISI Foundation

www.cirocattuto.info, @ciro

BIG DIVE / INTESA
Torino, 10 November 2016

BIG DIVE | HACKING DEVELOPMENT,
VISUALIZATION & SCIENCE



ISI Foundation

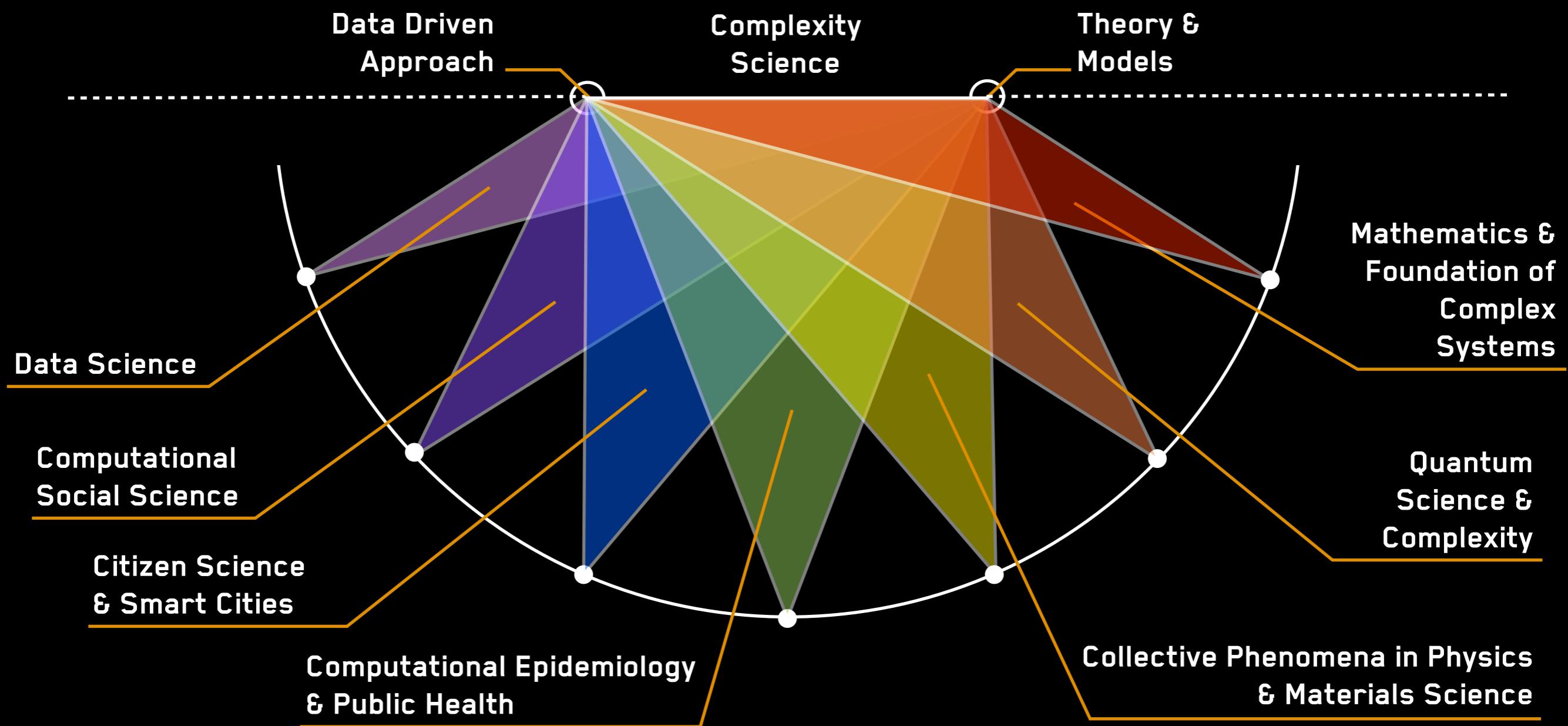
www.isi.it

- ▶ basic and applied research
- ▶ 30+ years of history
- ▶ 40+ researchers
- ▶ Turin, Italy & New York, USA
- ▶ international network

- ▶ supported by:
 - bank foundations
 - research grants
 - industrial research projects

- ▶ focus on
 - data & network science
 - complex systems science
 - mathematical modeling

research at ISI Foundation





Francesco Bonchi, PhD
computer science
ex- Yahoo! Labs



Laetitia Gauvin, PhD
theoretical physics
ex- École Normale Sup.



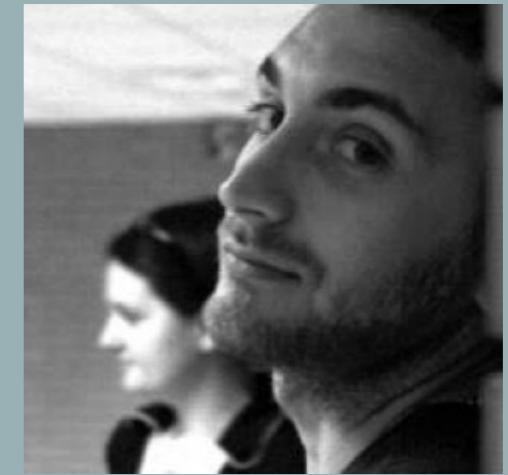
Michele Tizzoni, PhD
statistical physics
ex- PoliTO



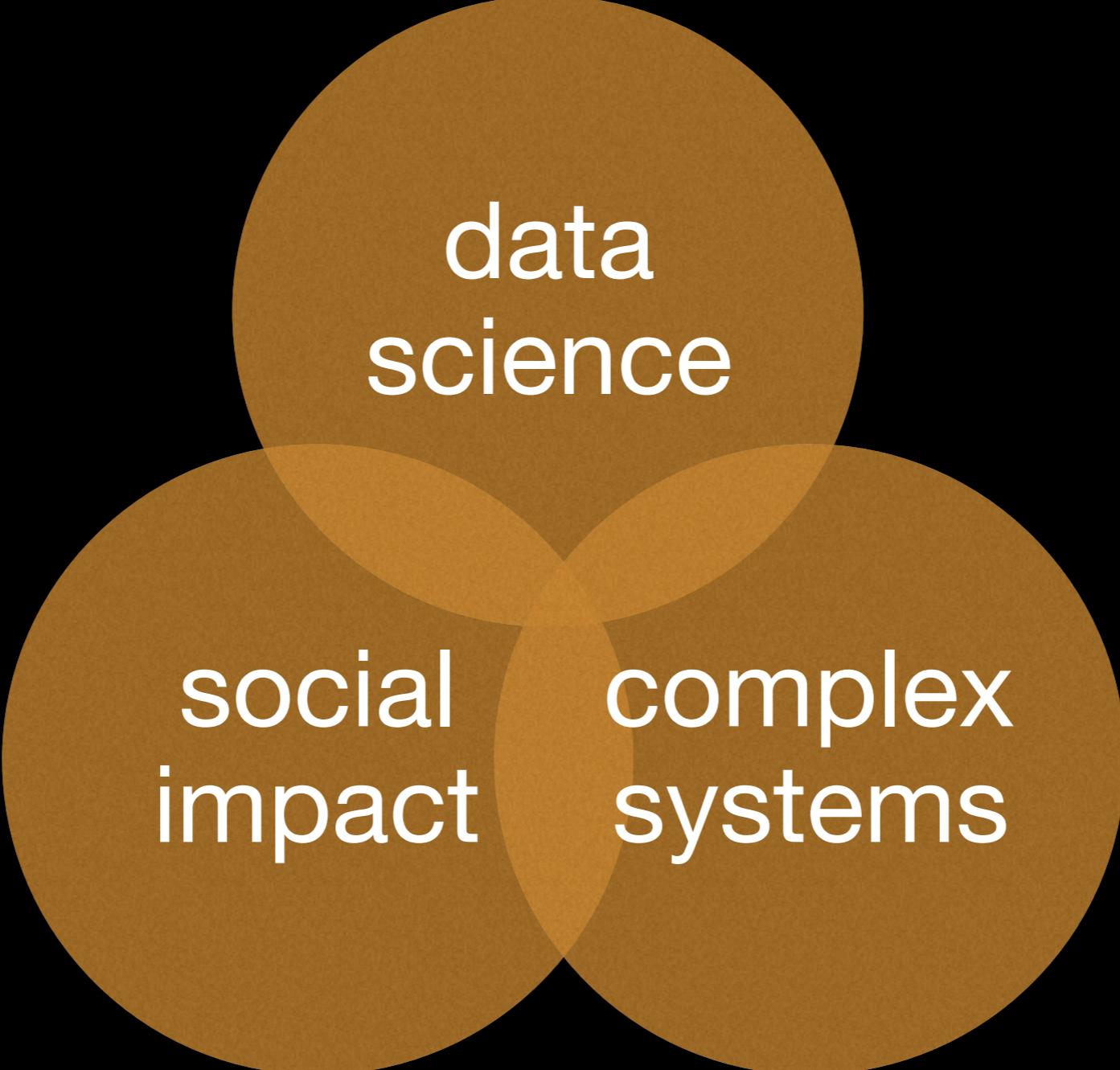
André Panisson, PhD
computer science
ex- Brazilian Govt.



Daniela Paolotti, PhD
complex systems
ex- Novartis V&D



Giovanni Petri, PhD
statistical physics
ex- Imperial College



A Venn diagram consisting of three overlapping circles, each containing a white text label. The top circle contains the text 'data science'. The bottom-left circle contains the text 'social impact'. The bottom-right circle contains the text 'complex systems'. The circles overlap in the center, representing the intersection of all three concepts.

data
science

social
impact

complex
systems

complexity: not like this

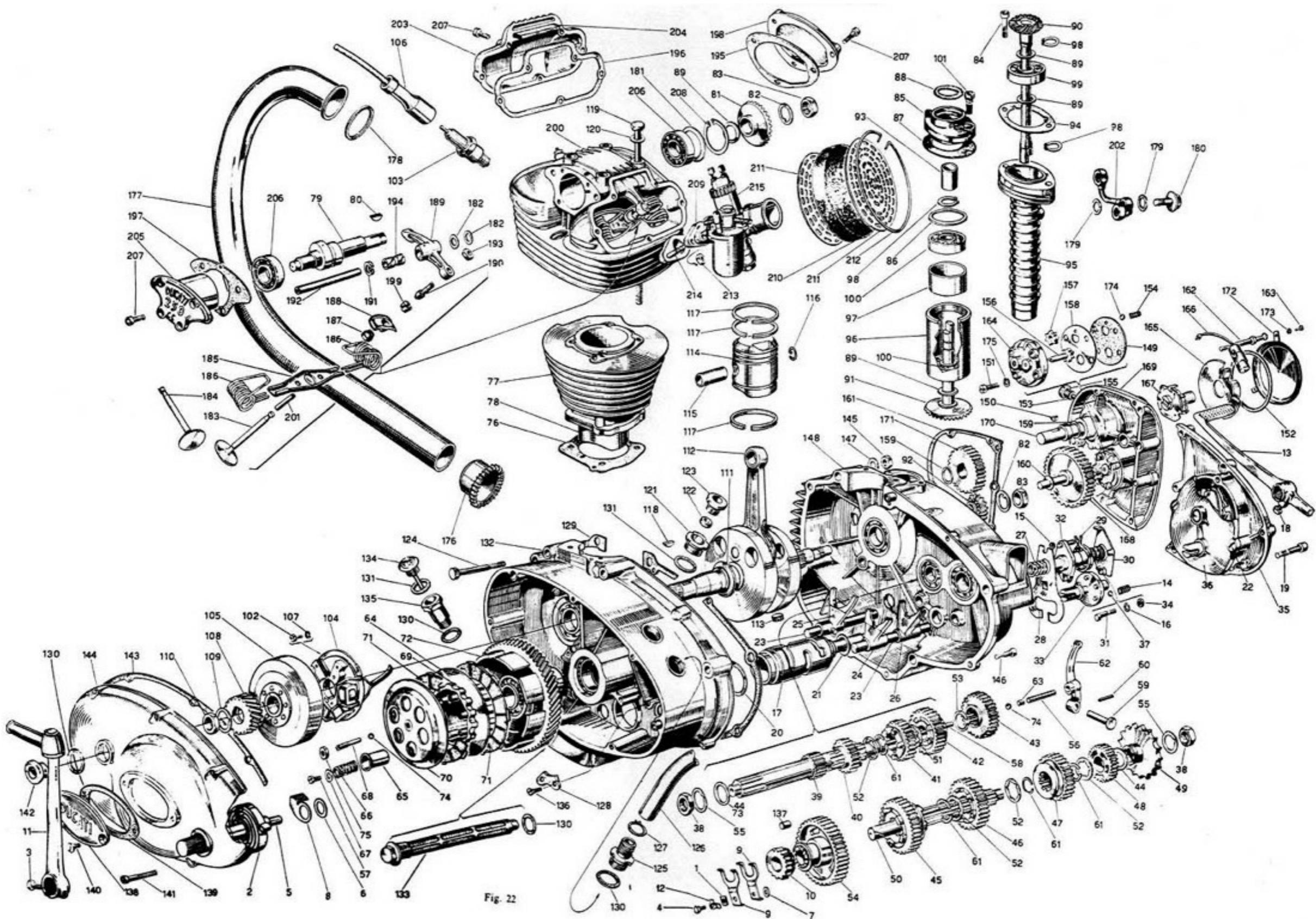


Fig. 22

shift[®] Obesity System Influence Diagram

Full Map

Click-drag to scroll
Double-click to zoom in/out

Clusters

Core Loop

Individual Psychology

Social Psychology

Individual Activity

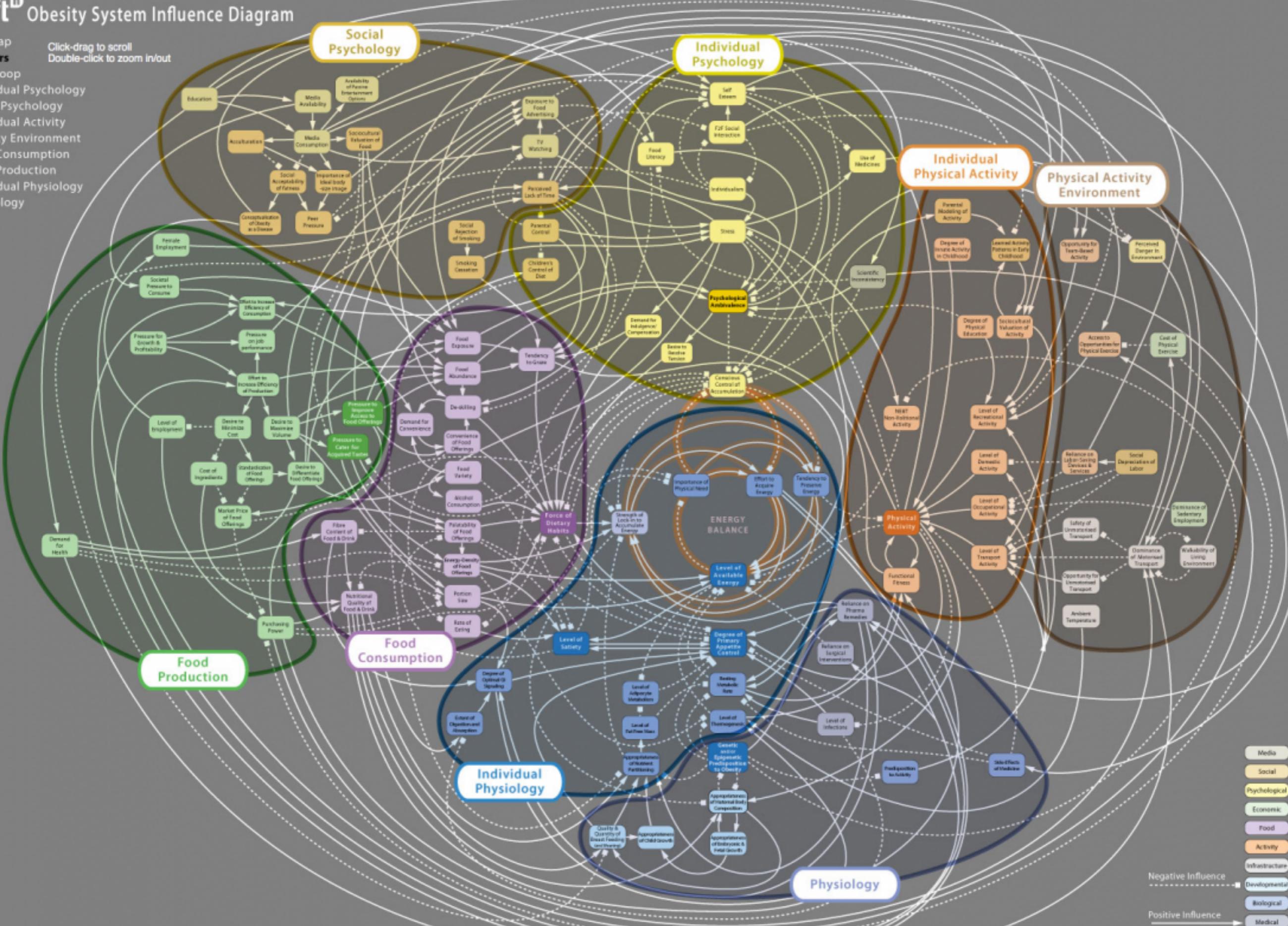
Activity Environment

Food Consumption

Food Production

Individual Physiology

Physiology



data to model to decision

decisions and policies

human-machine compositionally

computational
social science,
network science,
...
...

data mining,
machine learning,
natural language
processing

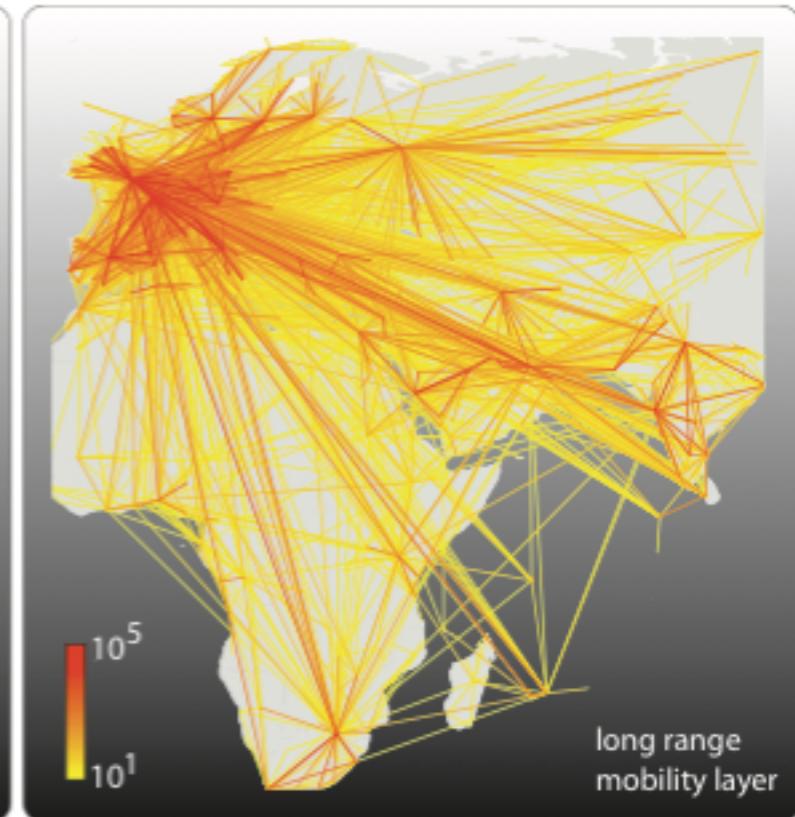
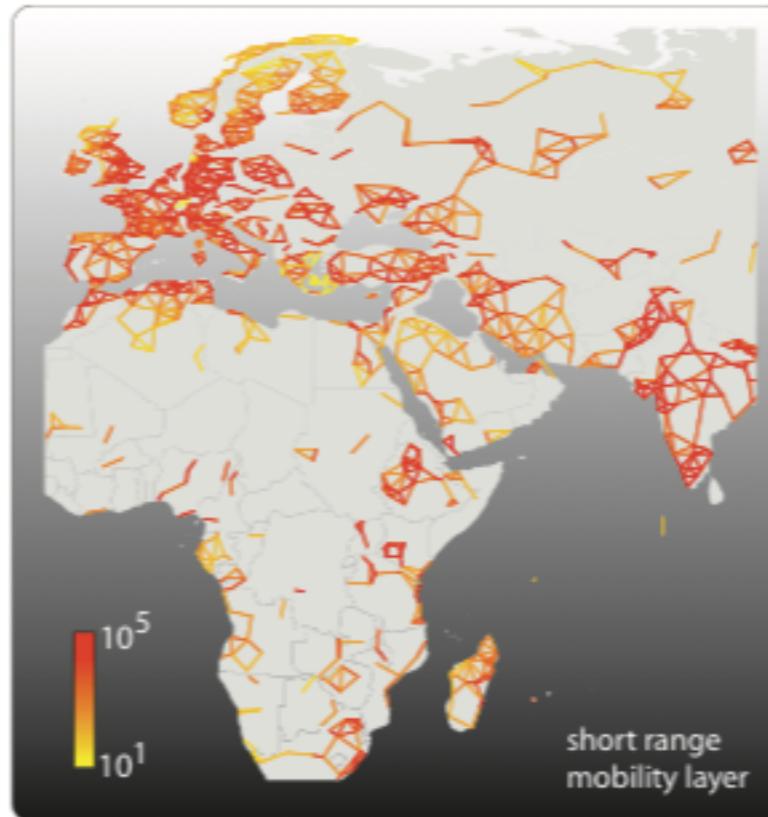
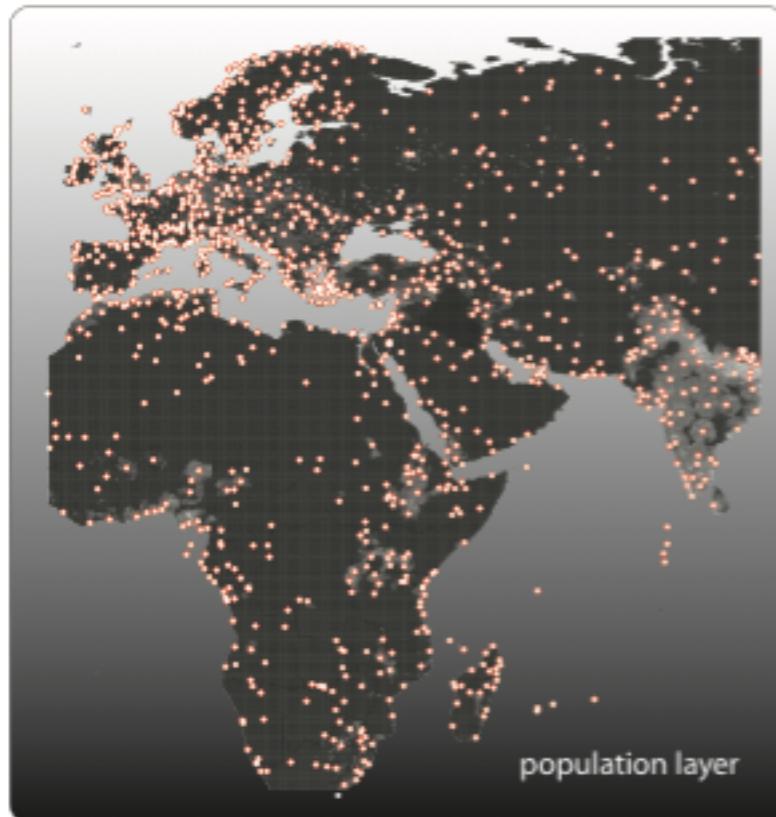
(big) data

epidemic forecast



CHARTING THE NEXT PANDEMIC

THE MODEL: GLEAM



Population layer

Commuting

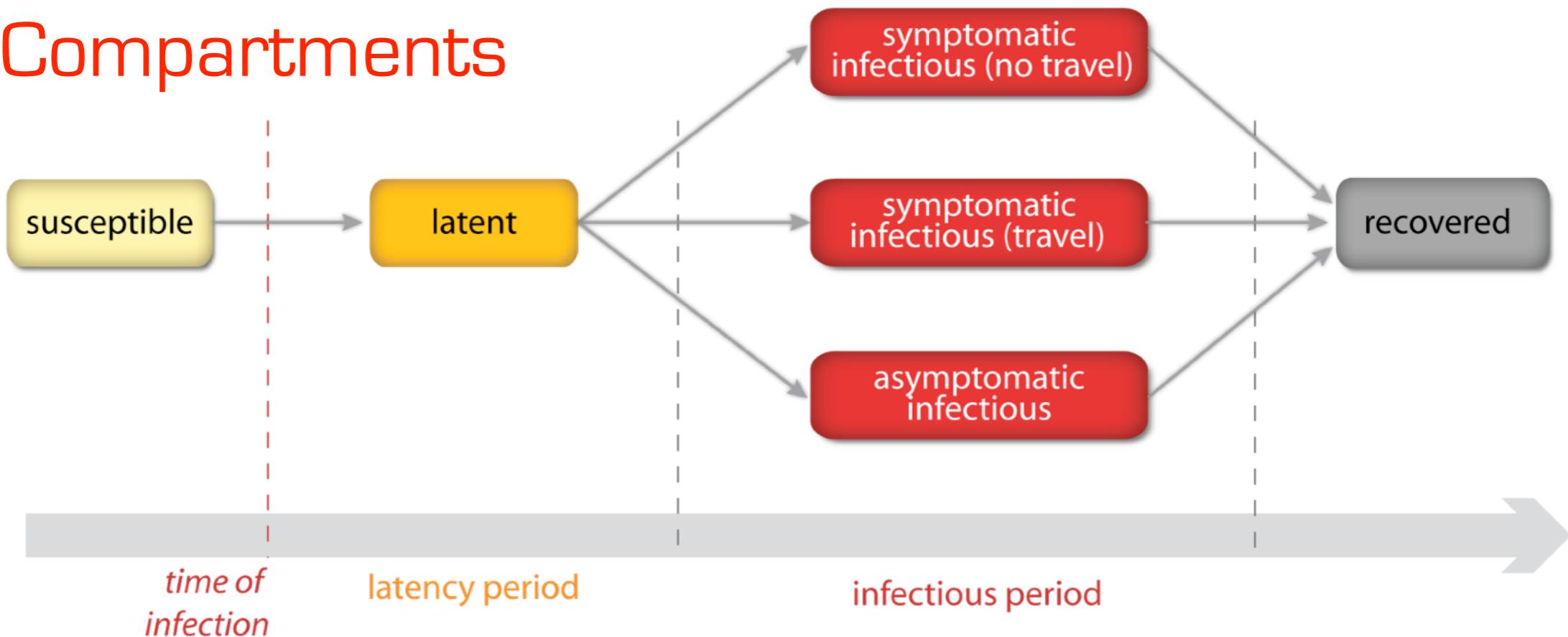
Human mobility layers

Air travel

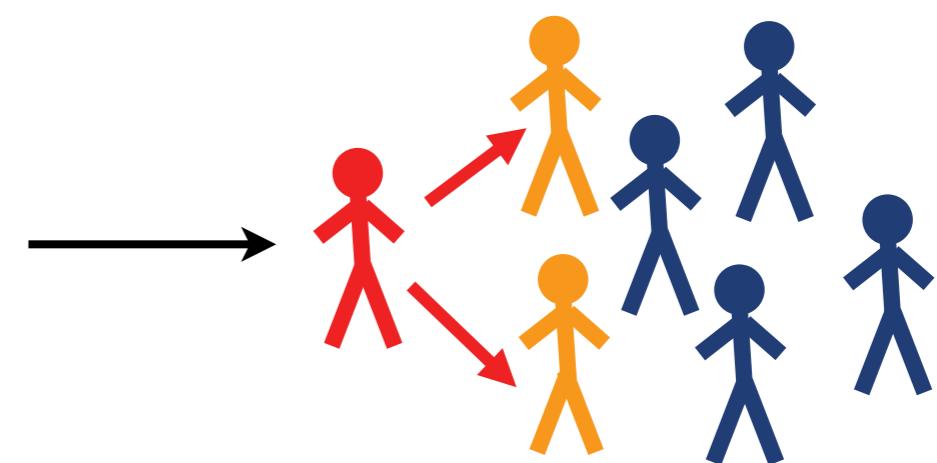
Geographic scale

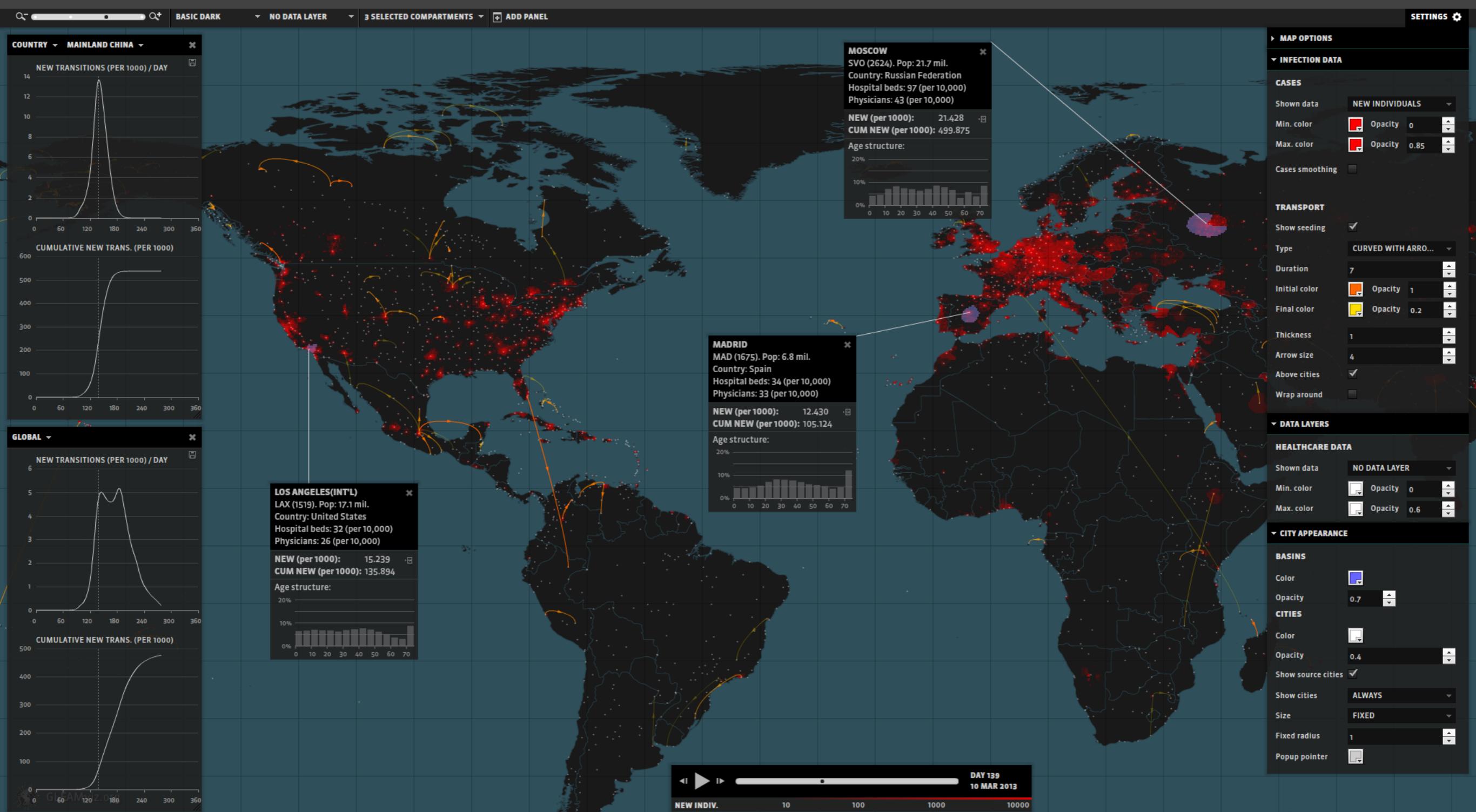
THE EPIDEMIC LAYER

Compartments

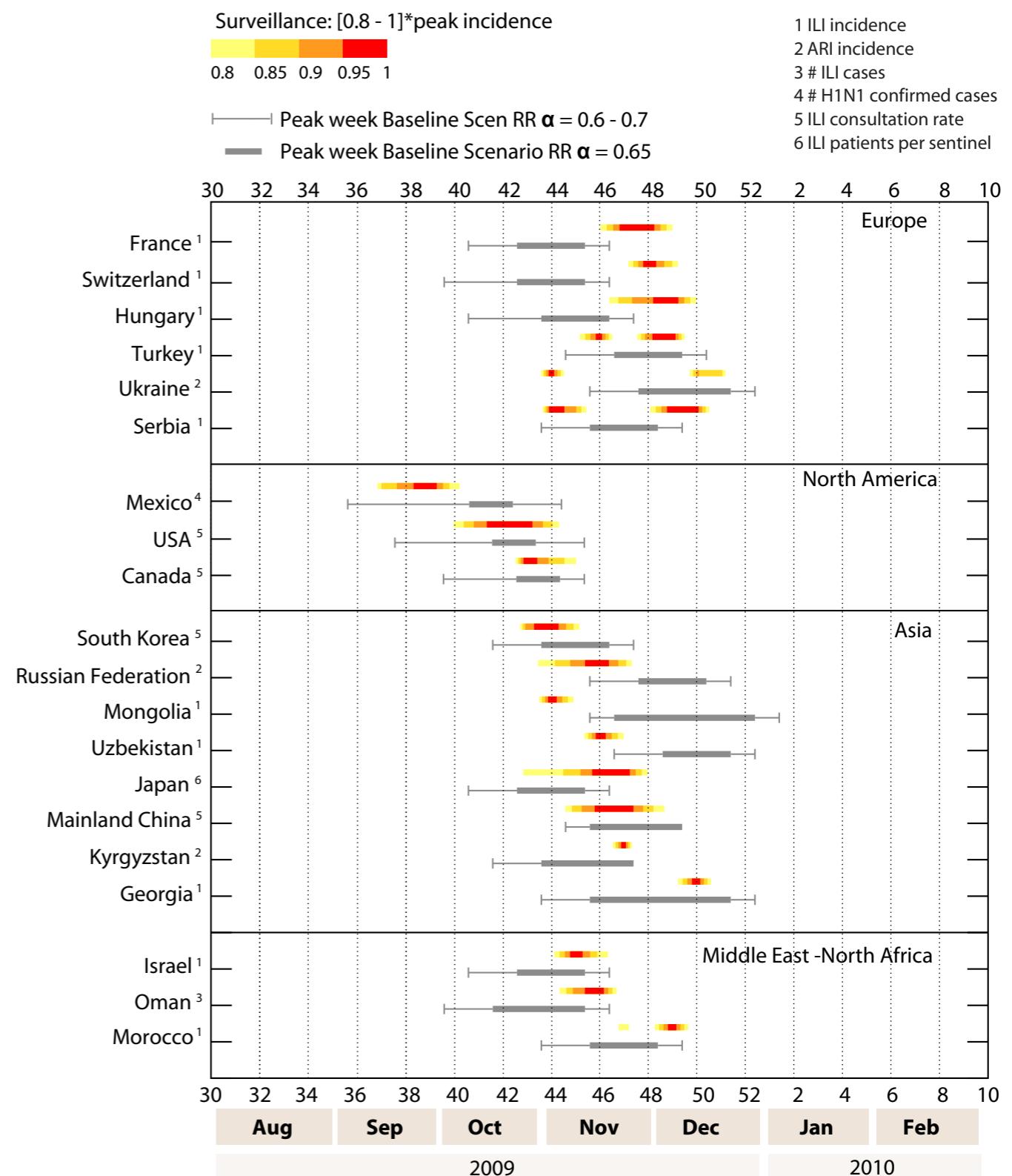
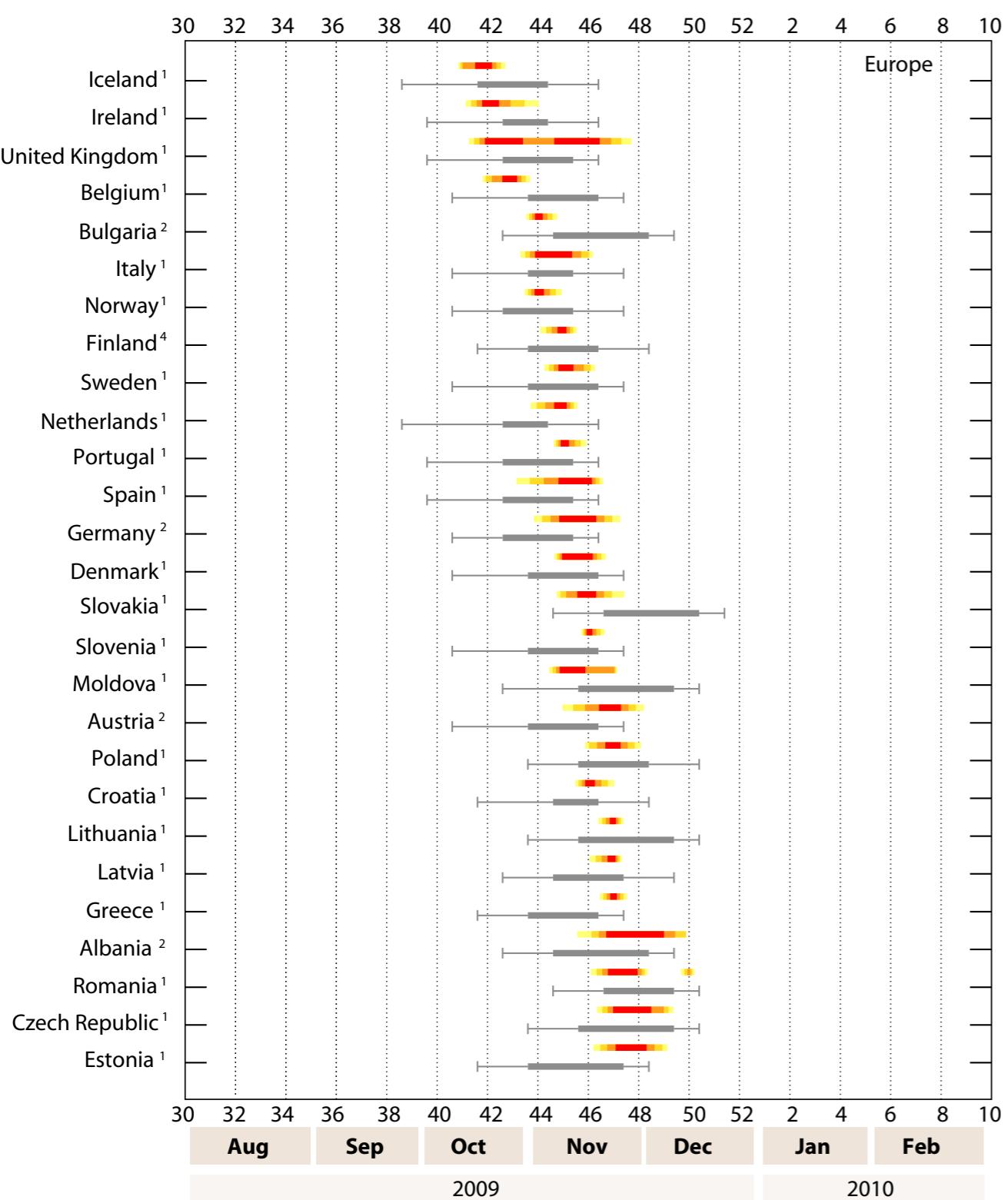


R_0
(average number of people infected by
an infectious individual during the
infectious period in a susceptible
population)





EPIDEMIC TIMING: VALIDATION



**Bill Gates**

@BillGates

**Following**

These pictures aren't just pretty. They can help us predict how an epidemic will spread: [b-gat.es/1GEG8bl](http://bit.gat.es/1GEG8bl)



RETWEETS

1,631

LIKES

2,008

10:31 PM - 7 Apr 2015



1.6K

2K

...

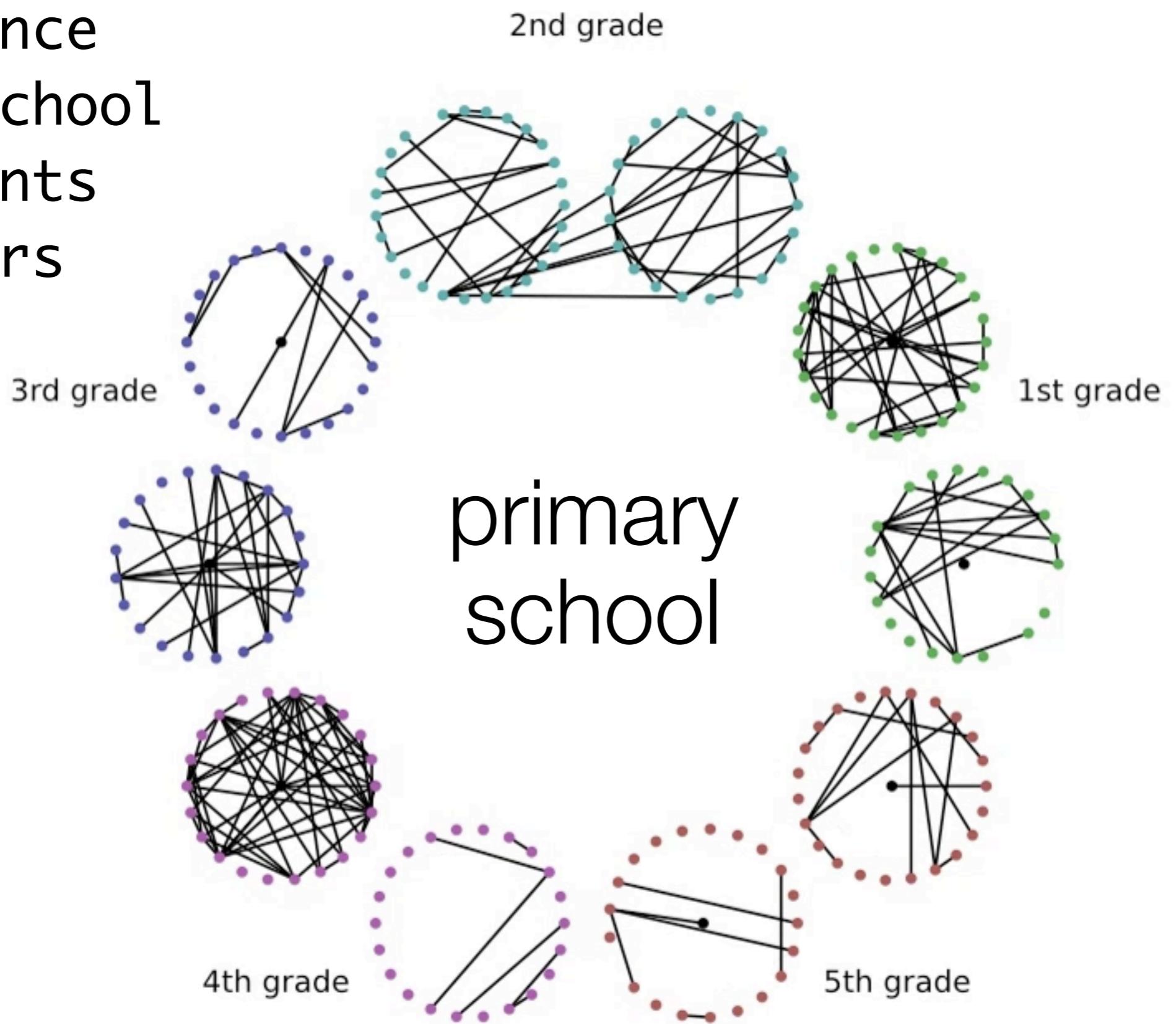
high-resolution social networks

- ▶ human mobility modeling
- ▶ social science & health
- ▶ organizational science
- ▶ **infectious disease dynamics**



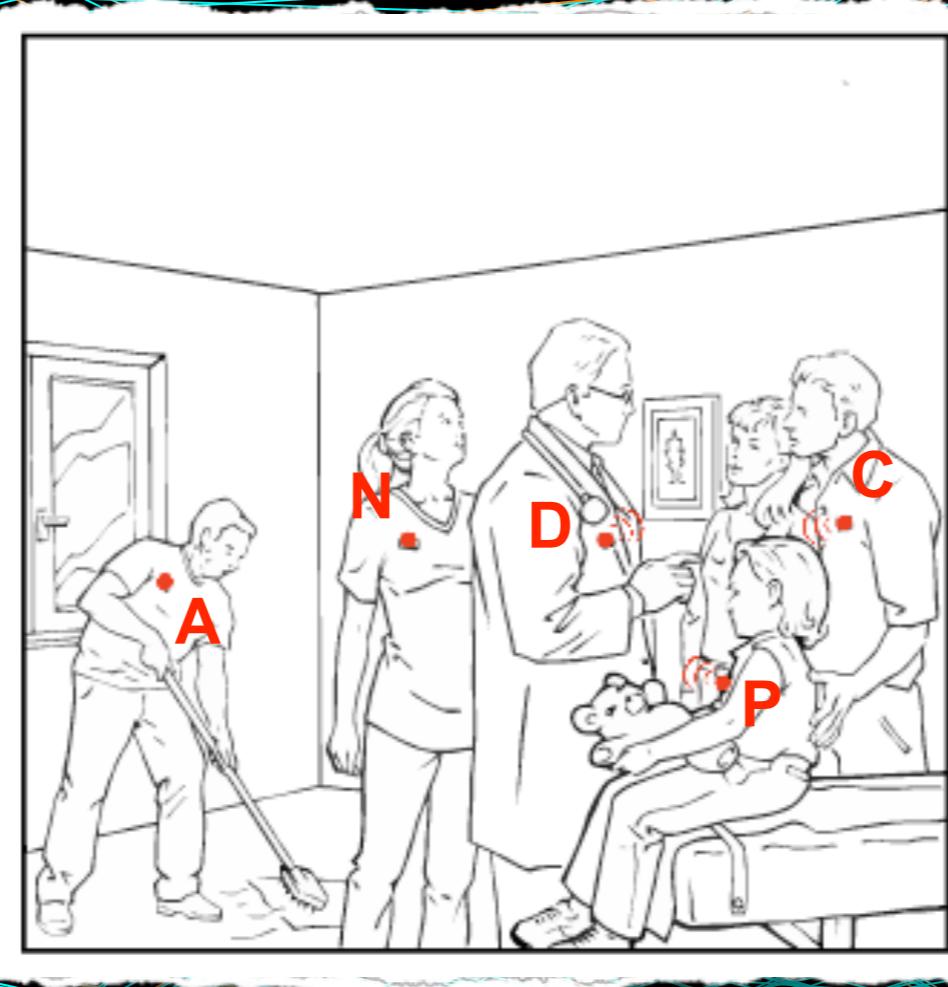
wearable sensors

Lyon, France
primary school
231 students
10 teachers



children

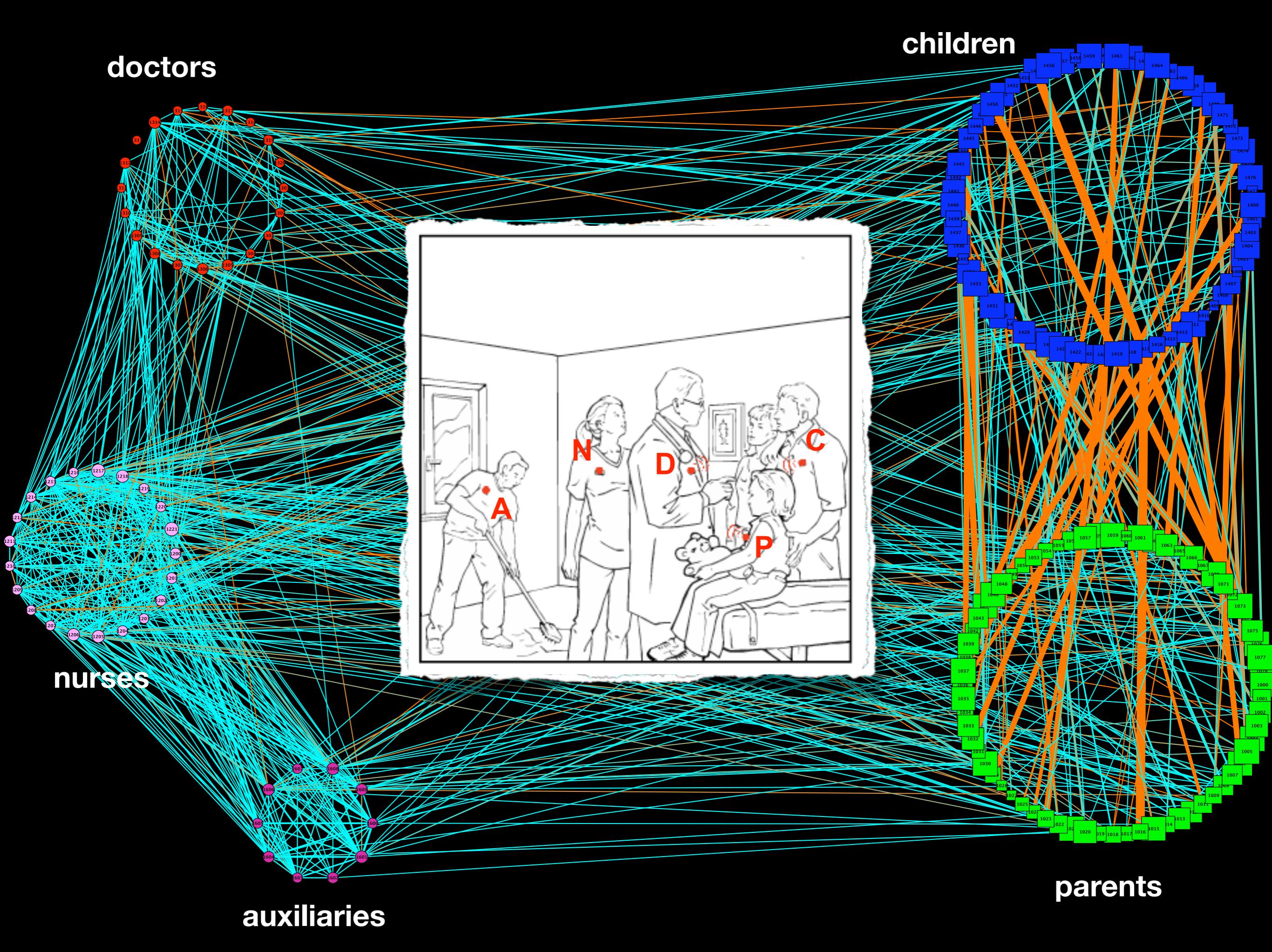
doctors



nurses

auxiliaries

parents



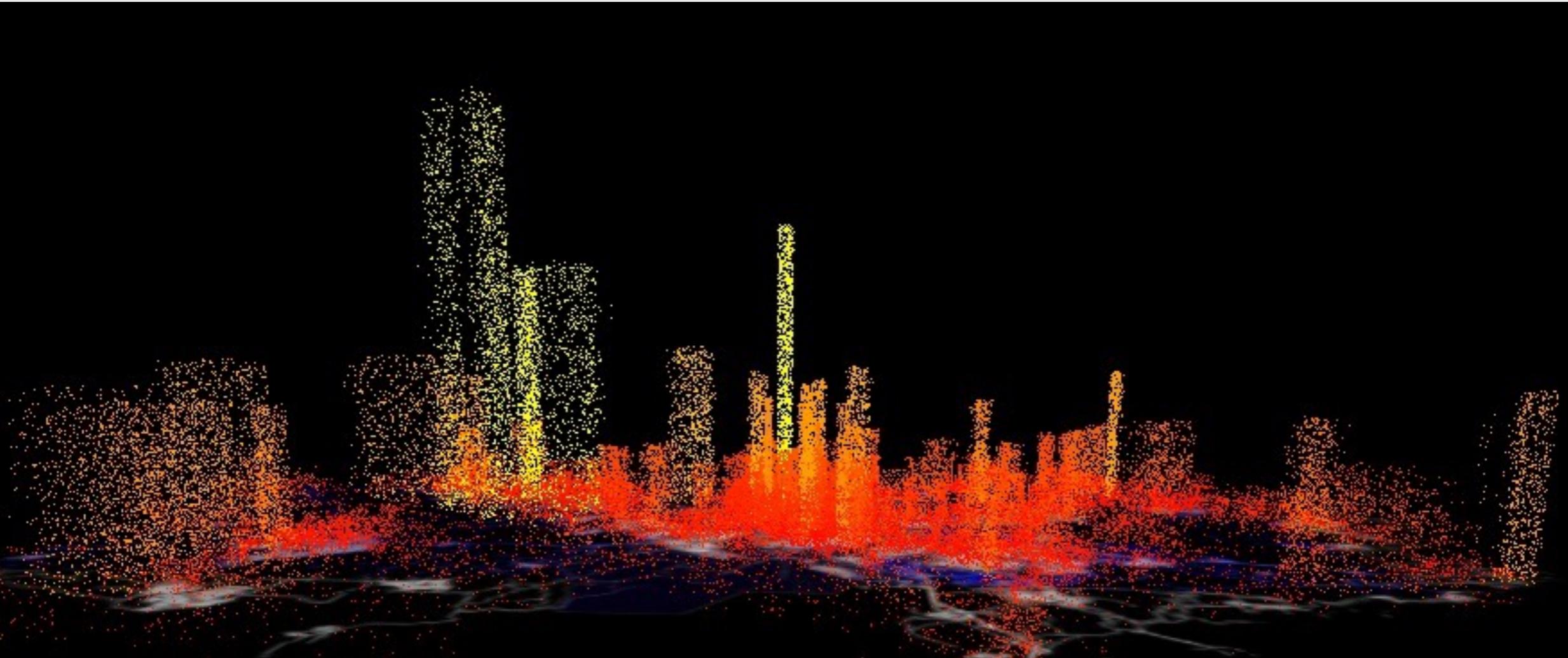


SocioPatterns.org

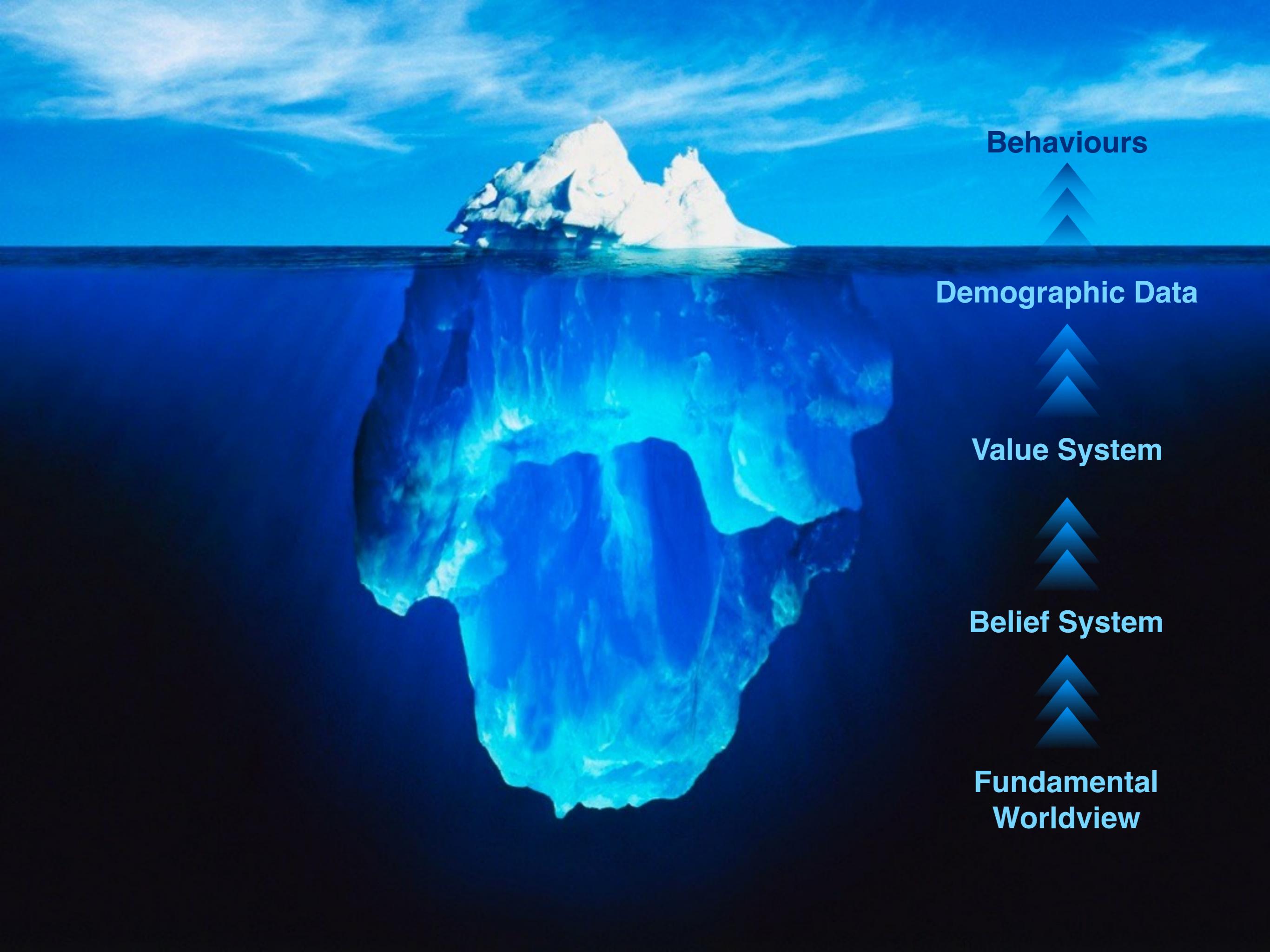


7 years, 30+ deployments, 12 countries

computational social science



aaronkoblin.com



Behaviours

Demographic Data

Value System

Belief System

**Fundamental
Worldview**

Profiling
population-scale analysis



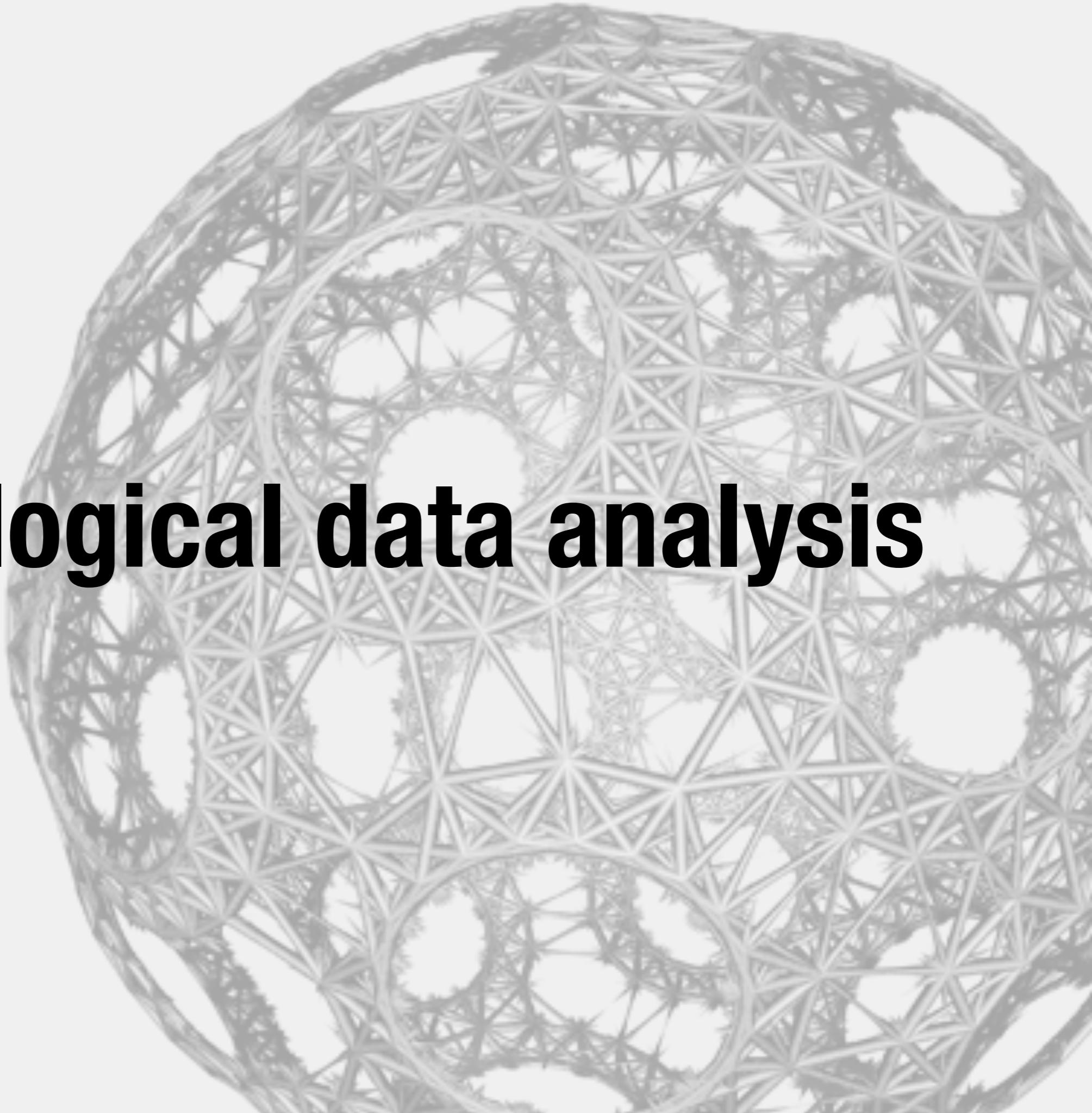
Segmentation
few prototypical profiles



Targeting & Validation
design of tailored interventions

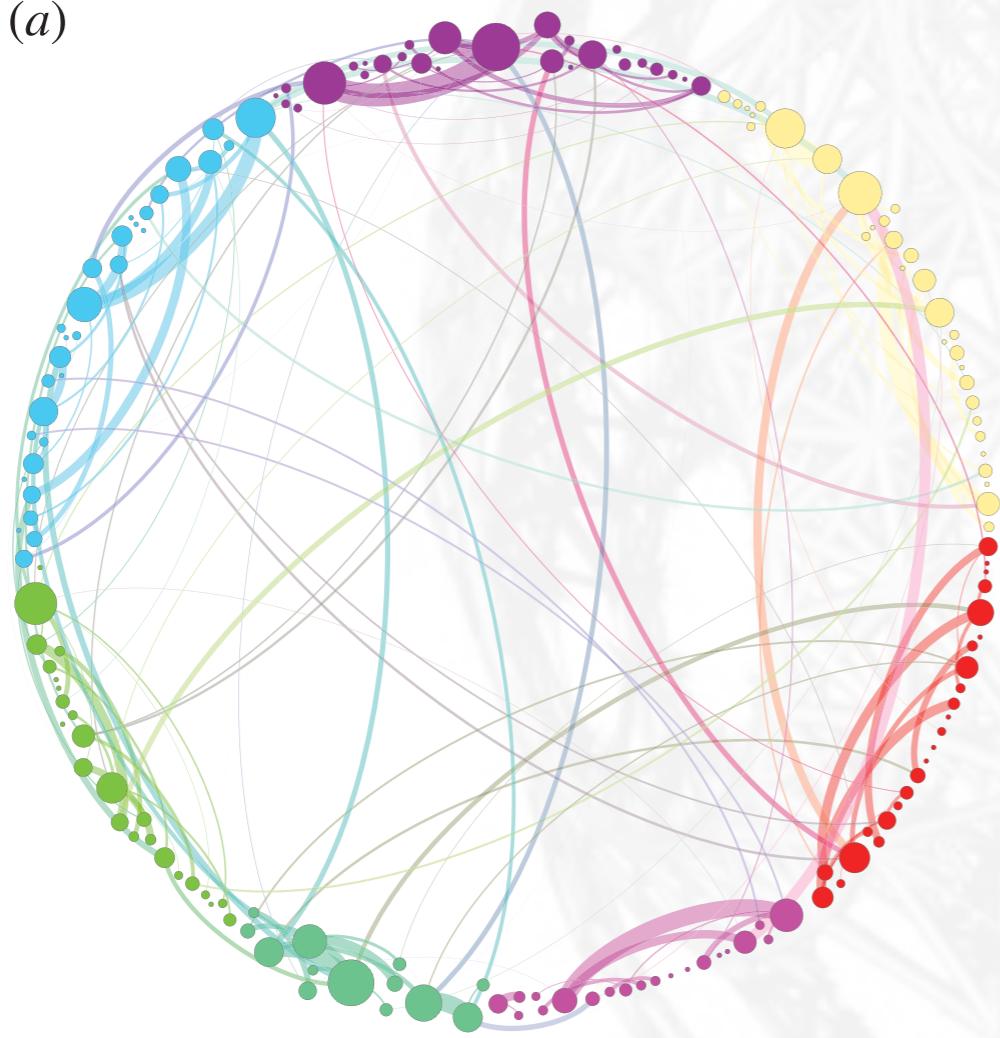


topological data analysis





(a)



(b)





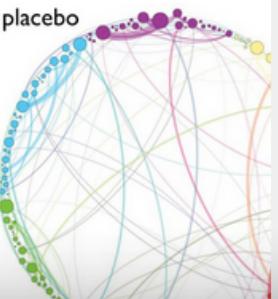
« Receipts May Expose Us to High Levels of BPA
Unlimited Free Lives! 900 Classic Arcade Games Resurrected Online »

This Is Your Brain on Magic Mushrooms

By Ben Thomas | October 29, 2014 4:00PM



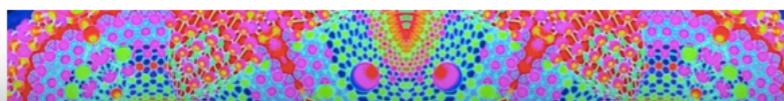
[Like](#) 21m [Follow](#) 166K followers [Follow](#) 421k



THE BRAIN

How Magic Mushrooms Change Your Brain

October 29, 2014 | by Lisa Winter



Science Graphic of the Week: How Magic Mushrooms Rearrange Your Brain

I RED

BUSINESS

DESIGN

ENTERTAINMENT

GEAR

SCIENCE

SUBSCRIBE



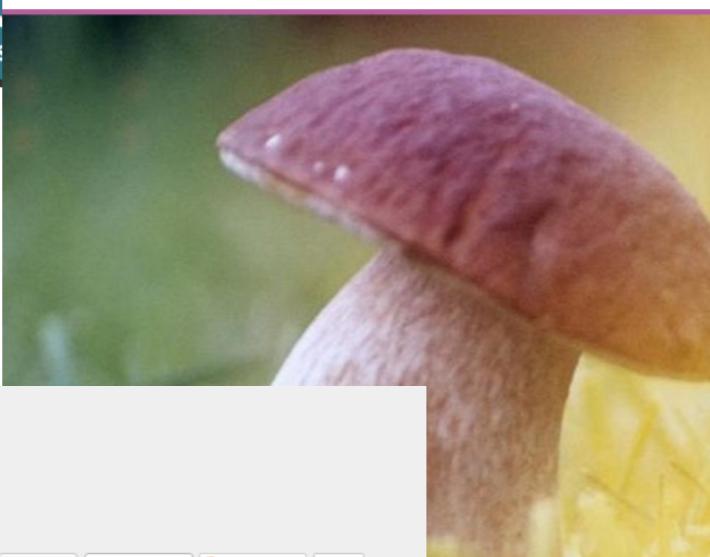
BRANDON KEIM SCIENCE 10.30.14 6:30 AM

SCIENCE GRAPHIC OF THE WEEK: HOW MAGIC MUSHROOMS REARRANGE YOUR BRAIN

SHARE
29065

TWEET
1393

PIN
100



How Tripping On Mushrooms Changes The Brain

ERIN BRODWIN



Oct. 29, 2014, 8:45 PM



1,858



FACEBOOK



LINKEDIN



TWITTER



linkedin

OV

UPI

TOP NEWS

ENTERTAINMENT

ODD NEWS

BUSINESS

SPORTS

SCIENCE

HEALTH

PHOTOS

HOME / SCIENCE NEWS

Psychedelic mushrooms enable a hyperconnected brain

"[Users] report it as one of the most profound experiences they've had in their lives," said Paul Expert.

By Brooks Hays | Oct. 30, 2014 at 2:08 PM



Share



Tweet



+1



Email



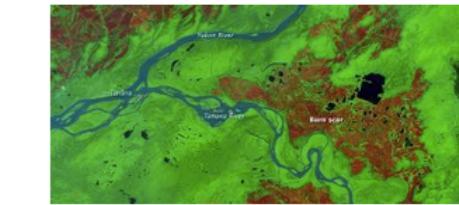
Print



PM



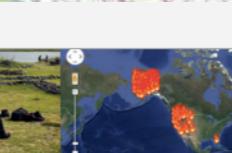
TOP STORIES



NASA satellite images Alaska's scorched earth



What Happens Now in the Search for MH370?



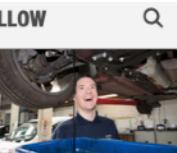
This Interactive Map Shows All the Places the US...



US Navy: We Want High-Powered Lasers and...



The F-35 Is (Finally) Cleared For Take-Off!



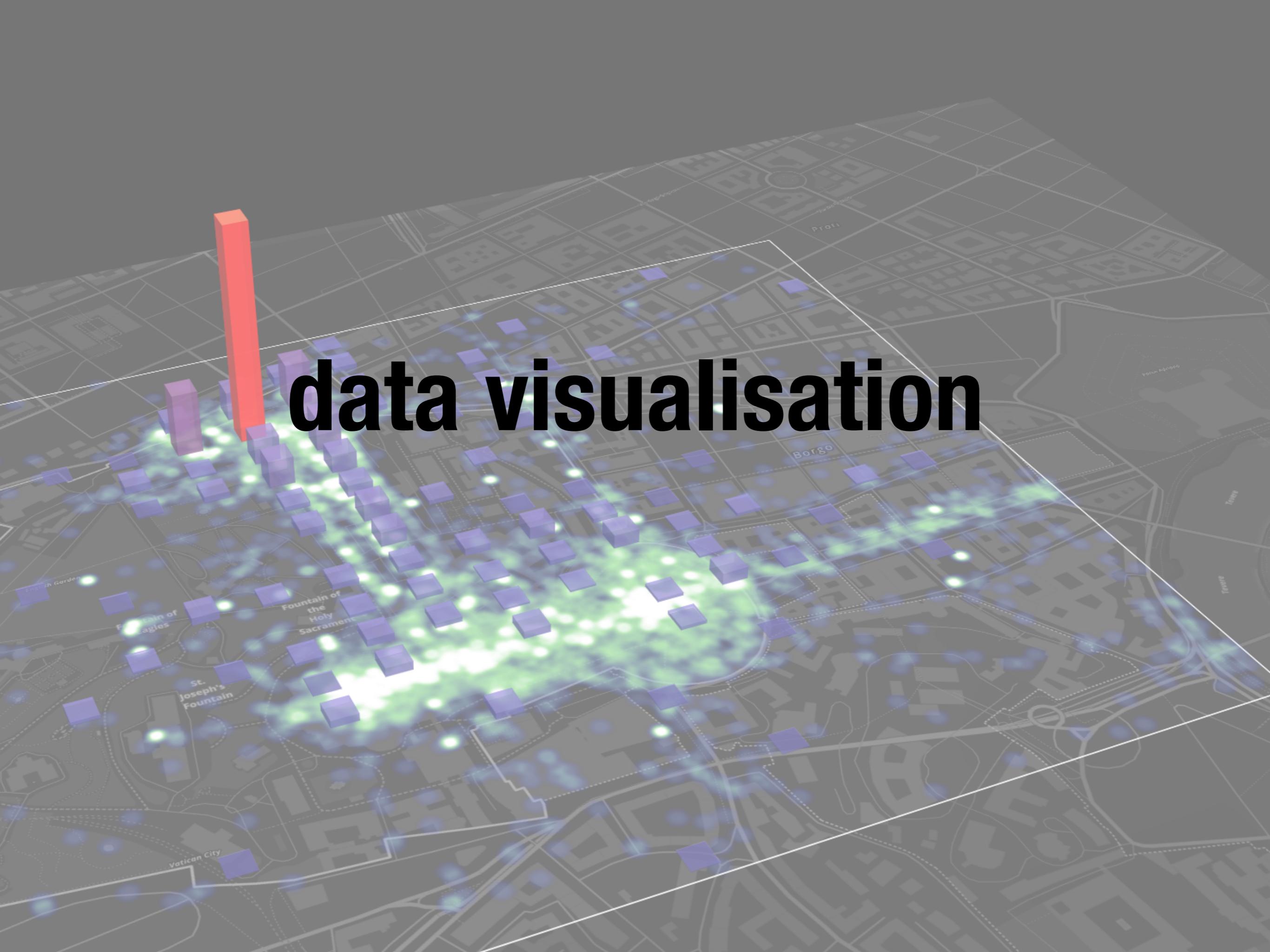
5 Simple Car Repairs You Should Know

SCIENCE HEALTH SCIENCE NEUROSCIENCE NEWS BRAIN

This Is Your Brain on Magic Mushrooms

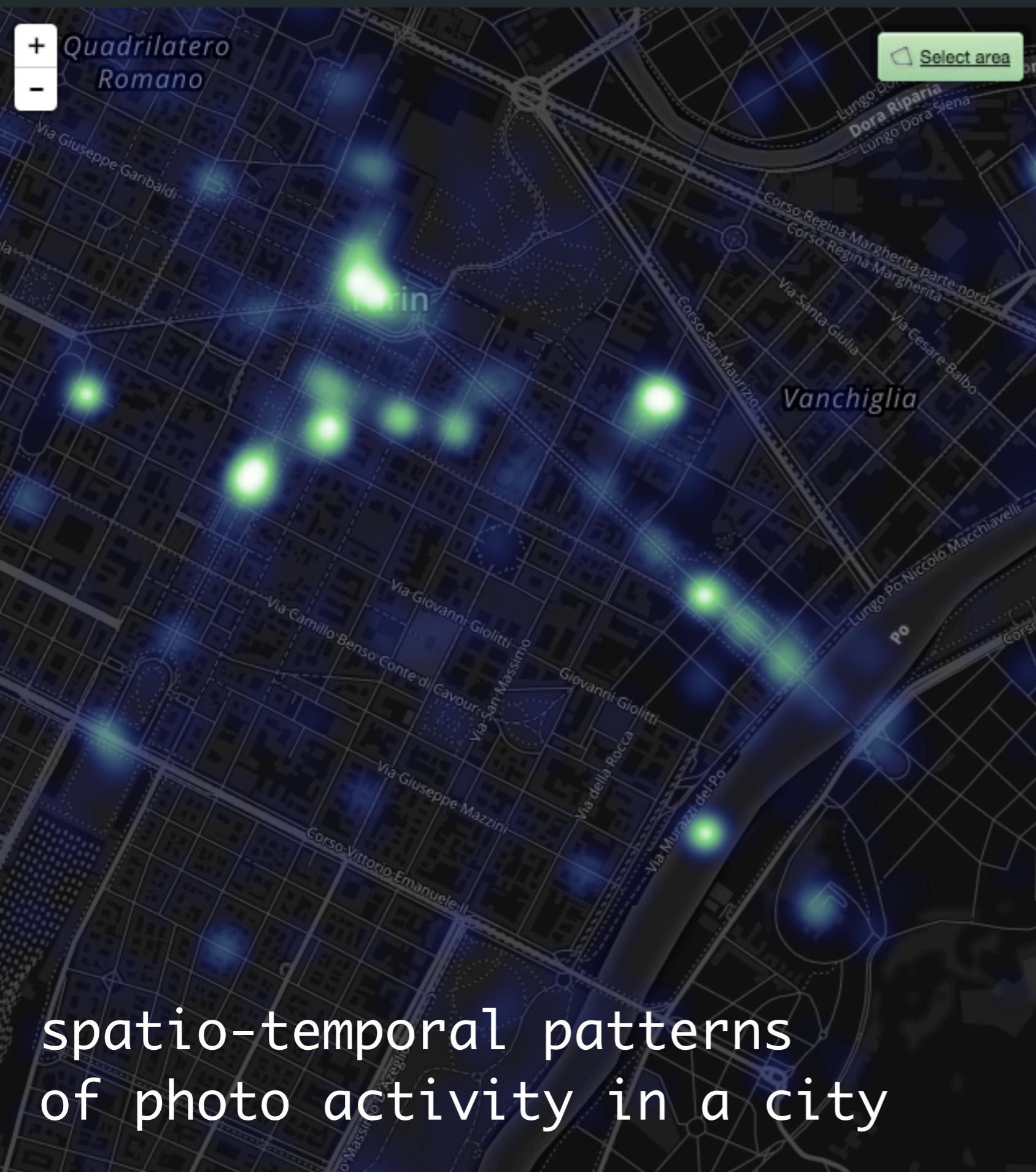
Brain scans of people injected with psilocybin, the active ingredient in magic mushrooms, explain what happens to connections across the brain when people take psychedelic drugs.





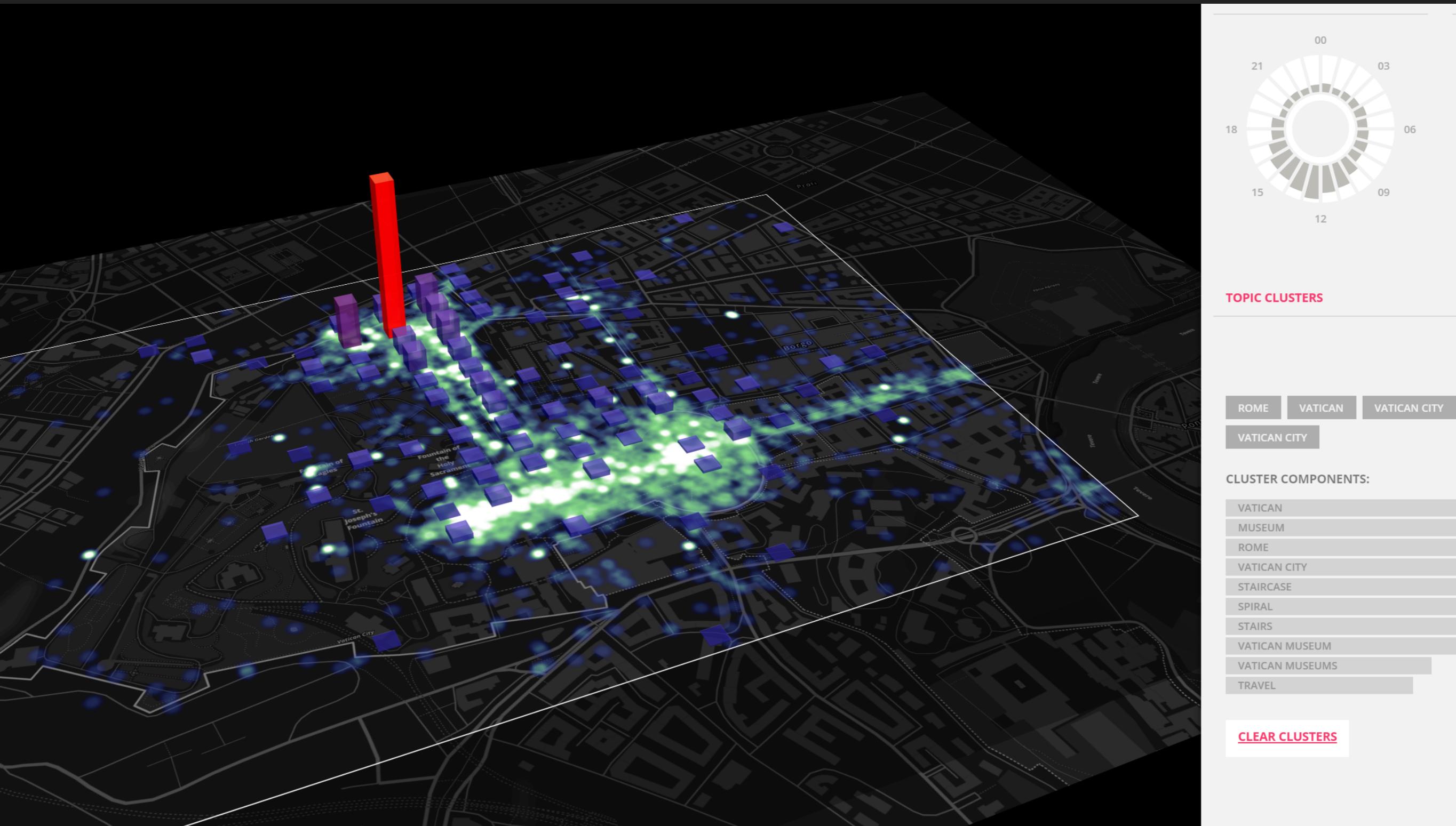
A 3D bar chart is overlaid on a grayscale map of Rome, Italy. The chart features several bars of varying heights and colors (red, purple, green) representing data values across different locations. The map shows major landmarks and street grids. The word "data visualisation" is prominently displayed in large black letters in the center of the chart.

data visualisation

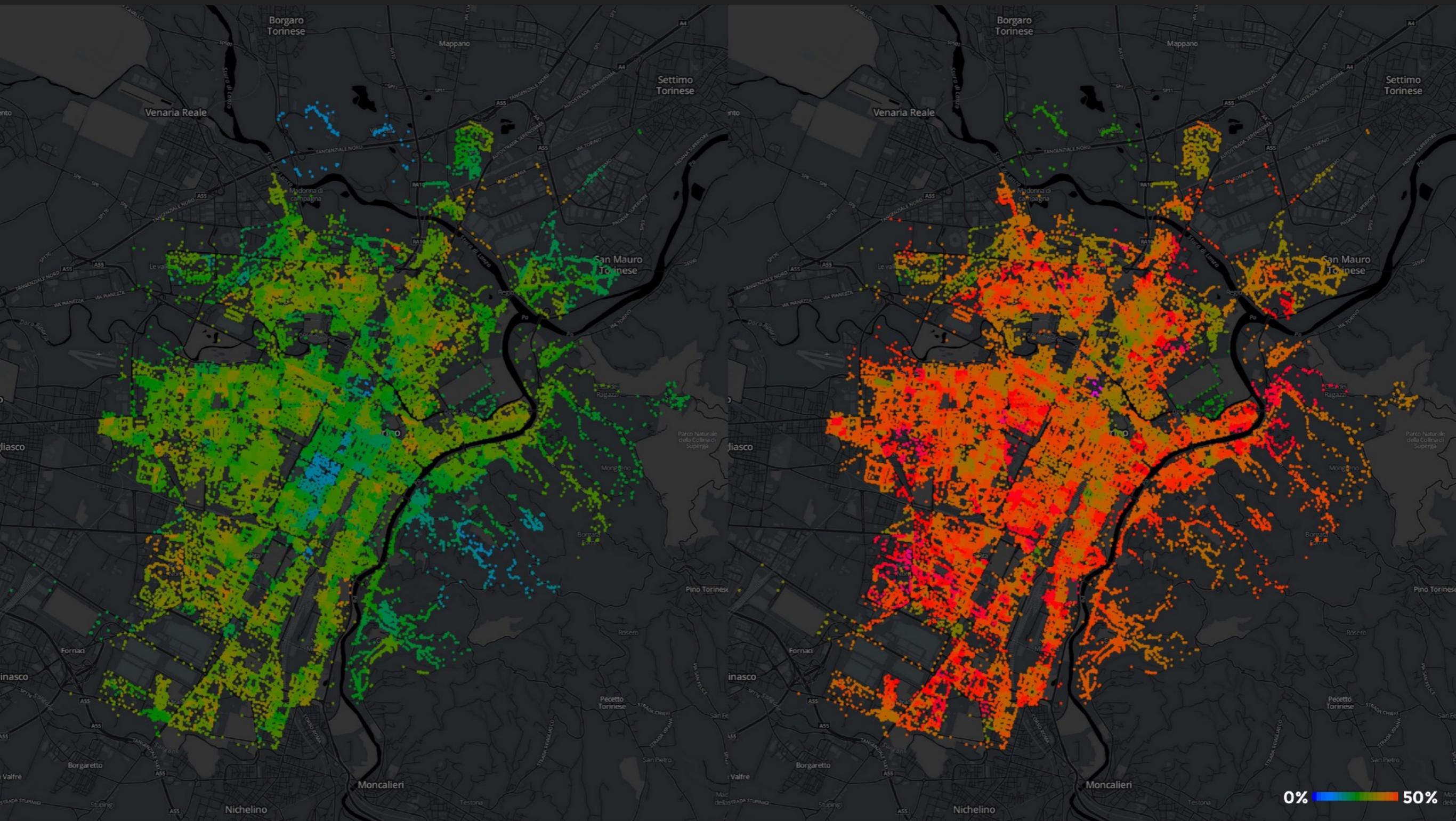


spatio-temporal patterns
of photo activity in a city

www.datainterfaces.org/projects/flickr/



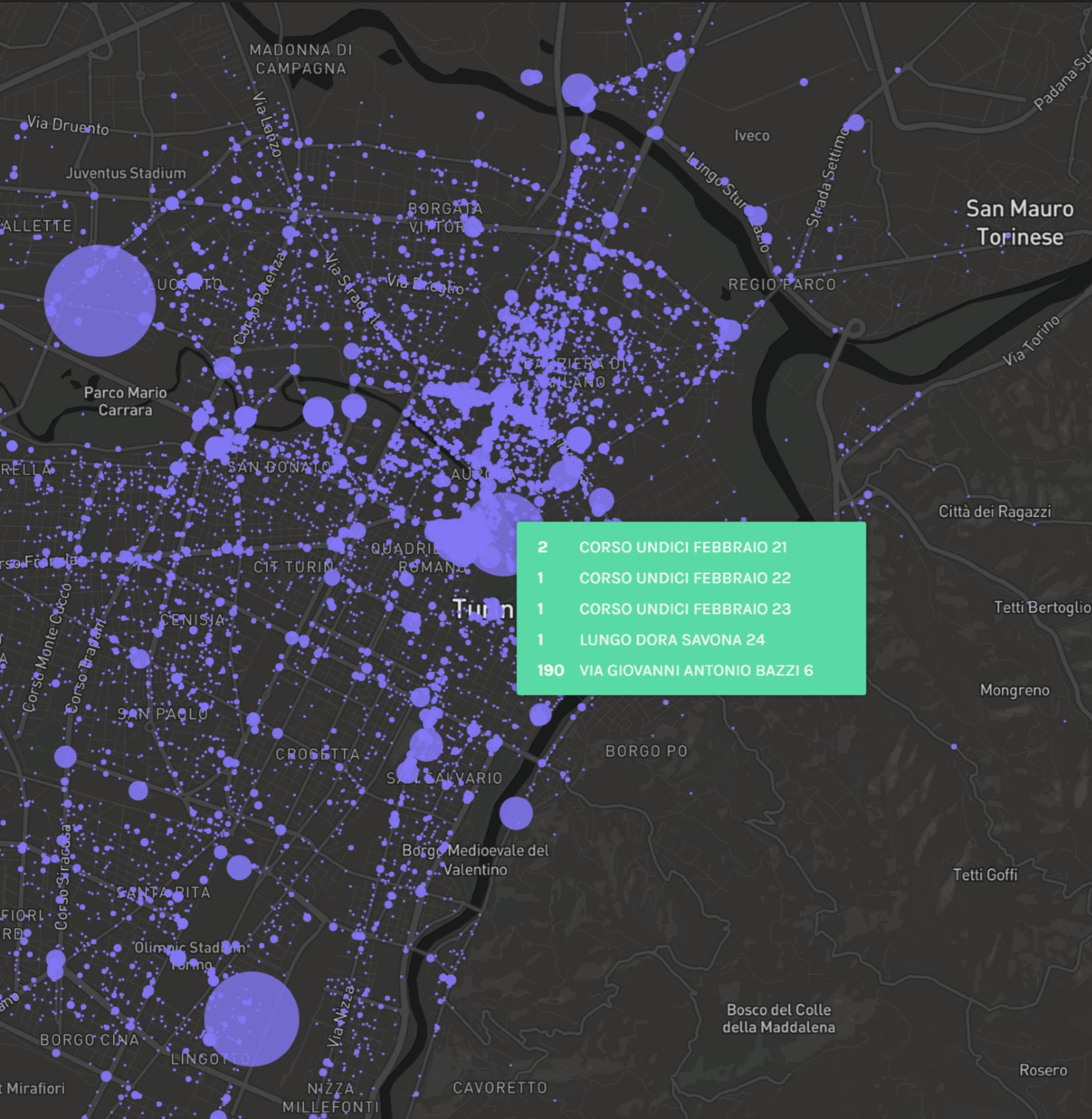
Open (electoral) Data



2013: PARTITO DEMOCRATICO

2014: PARTITO DEMOCRATICO





Infrazioni stradali della città di Torino.

Infrazioni è un'interfaccia per l'esplorazione dei dati relativi alle infrazioni stradali della città di Torino. I dati, disponibili su [Open Data](#) del comune di Torino, rappresentano più di quattromila infrazioni effettuati dal 2011 al 2015. Filtra i dati per luogo, data e ora, tipo di veicolo, e combina i vari parametri per esplorare i pattern nelle diverse infrazioni torinesi.

Infrazioni selezionate: 12.183

Totale infrazioni: 4.012.230

Infrazioni

Ora

Giorno

Sanzioni accessorie

SEQUESTRO AMM.VO DEL

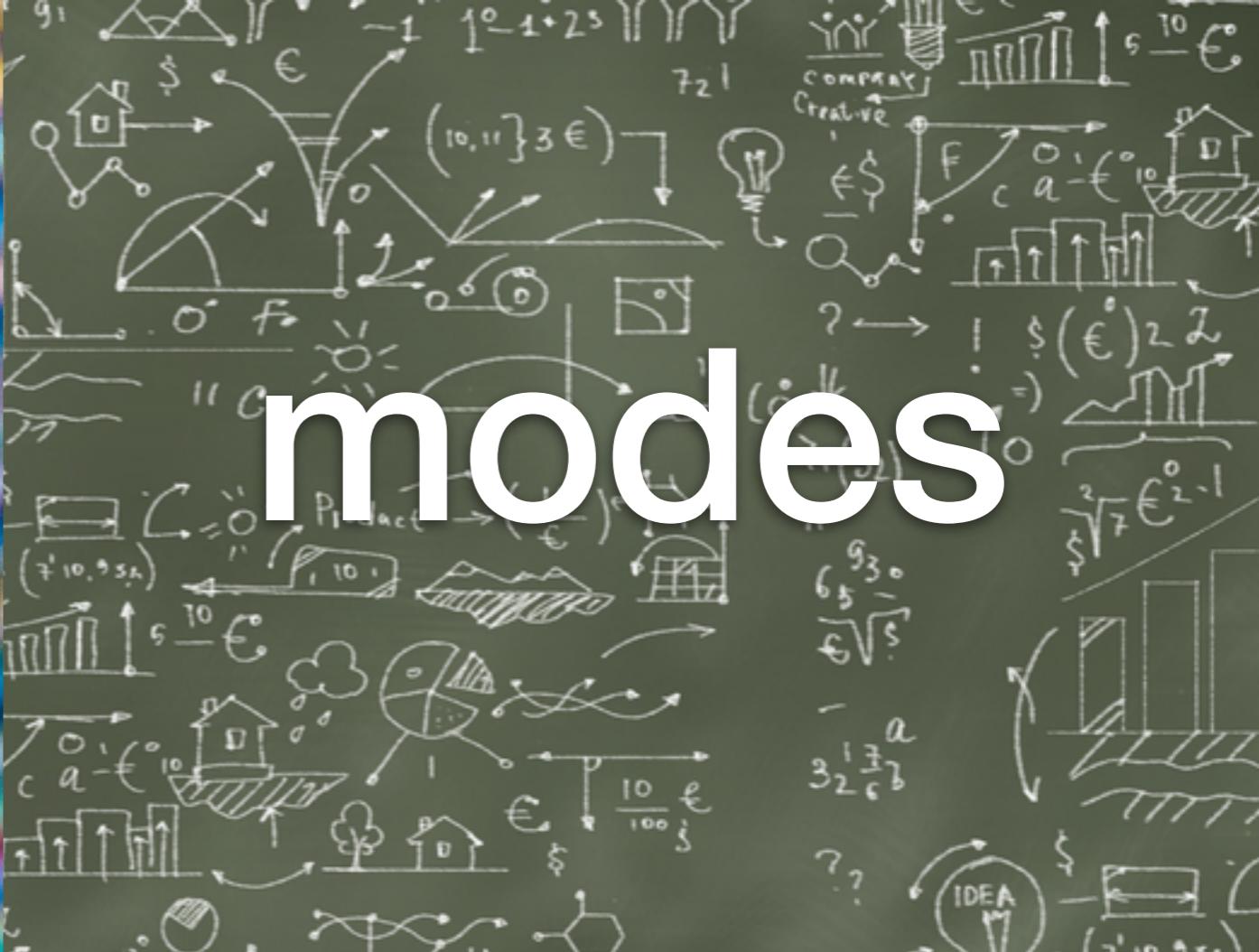
- CONFISCA AMMINISTRATIVA DEL VEICOLO O ALTRE COSE ? SEQUESTRO (ART. 321 C.P.P.) : 1
- CONFISCA AMMINISTRATIVA DEL VEICOLO O ALTRE COSE - SEQUESTRO (ART. 321 C.P.P.) : 1
- CONFISCA AMMINISTRATIVA DEL VEICOLO ? SEQUESTRO ? E RITIRO CARTA DI CIRCOLAZIONE (ART. 321 C.P.P.) : 1
- SEQUESTRO GIUDIZIARIO DEL VEICOLO (ART. 321 C.P.P.) : 24
- SEQUESTRO AMMINISTRATIVO FINALIZZATO ALLA CONFISCA DEL CASCO NON OMOL.
- SEQUESTRO GIUDIZIARIO DELLA TARGA: 7
- CONFISCA AMMINISTRATIVA DEL VEICOLO - SEQUESTRO - E RITIRO CARTA DI CIRCOLAZIONE (ART. 321 C.P.P.) : 1
- FERMO AMM.VO DEL VEICOLO E SEQUESTRO AMM.VO DEL CASCO NON OMOL.
- SEQUESTRO DEL VEICOLO E SOSPENSIONE CARTA DI CIRCOLAZIONE ? RITIRO ?

Veicolo





data



modes



?



decisions

A blurred background of colorful digital data or code, consisting of a grid of small, multi-colored pixels.

(big) data

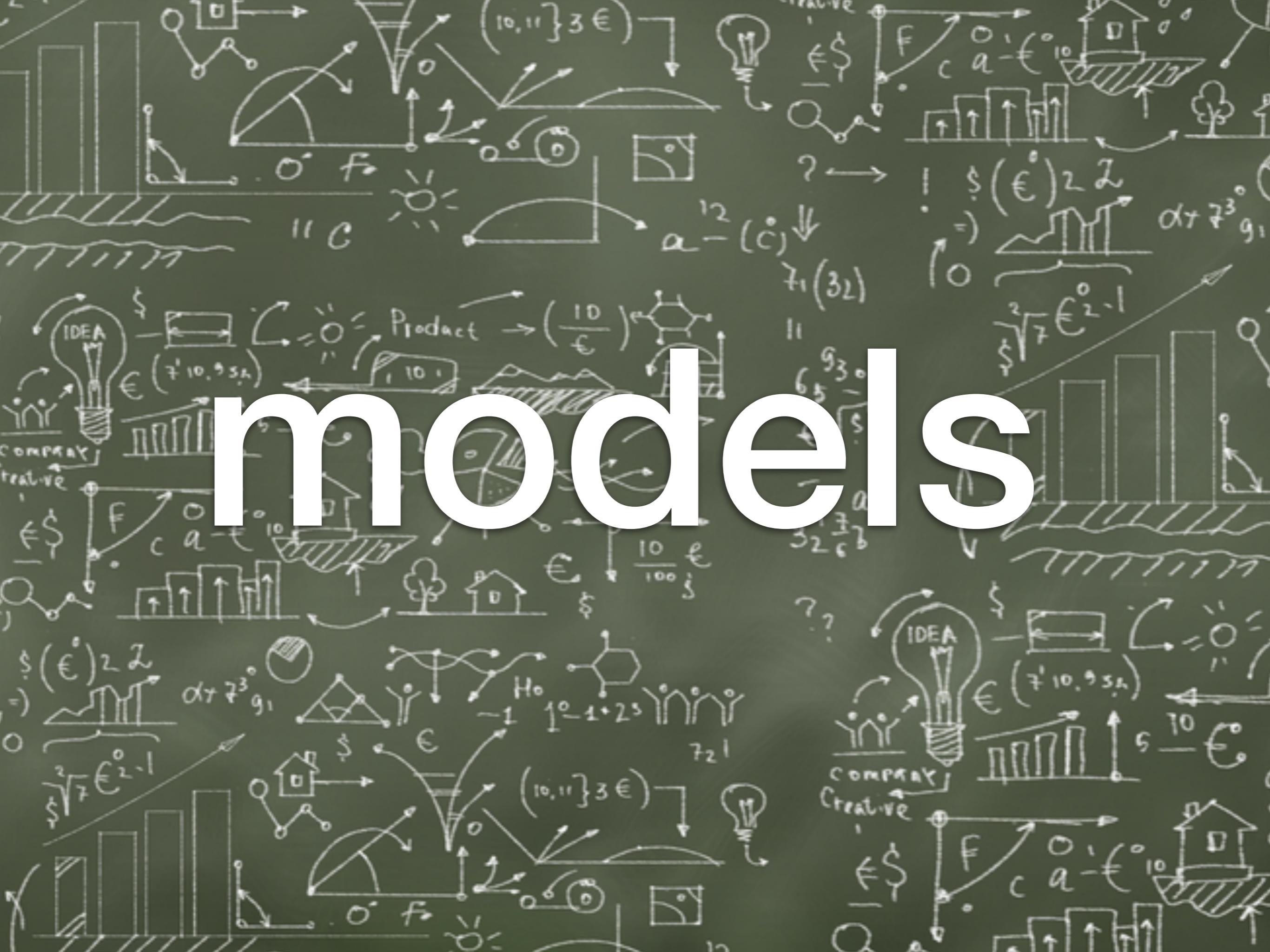
digital traces



*“This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear. **Out with every theory of human behavior**, from linguistics to sociology. Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. **With enough data, the numbers speak for themselves.**”*

Chris Anderson (2008)

models



what is a “model” ?

- mathematical model
- statistical model
- generative model
- machine learning model
- descriptive model
- dynamical model
- agent-based model
- predictive model (of the future)
- predictive model (of unknown features)

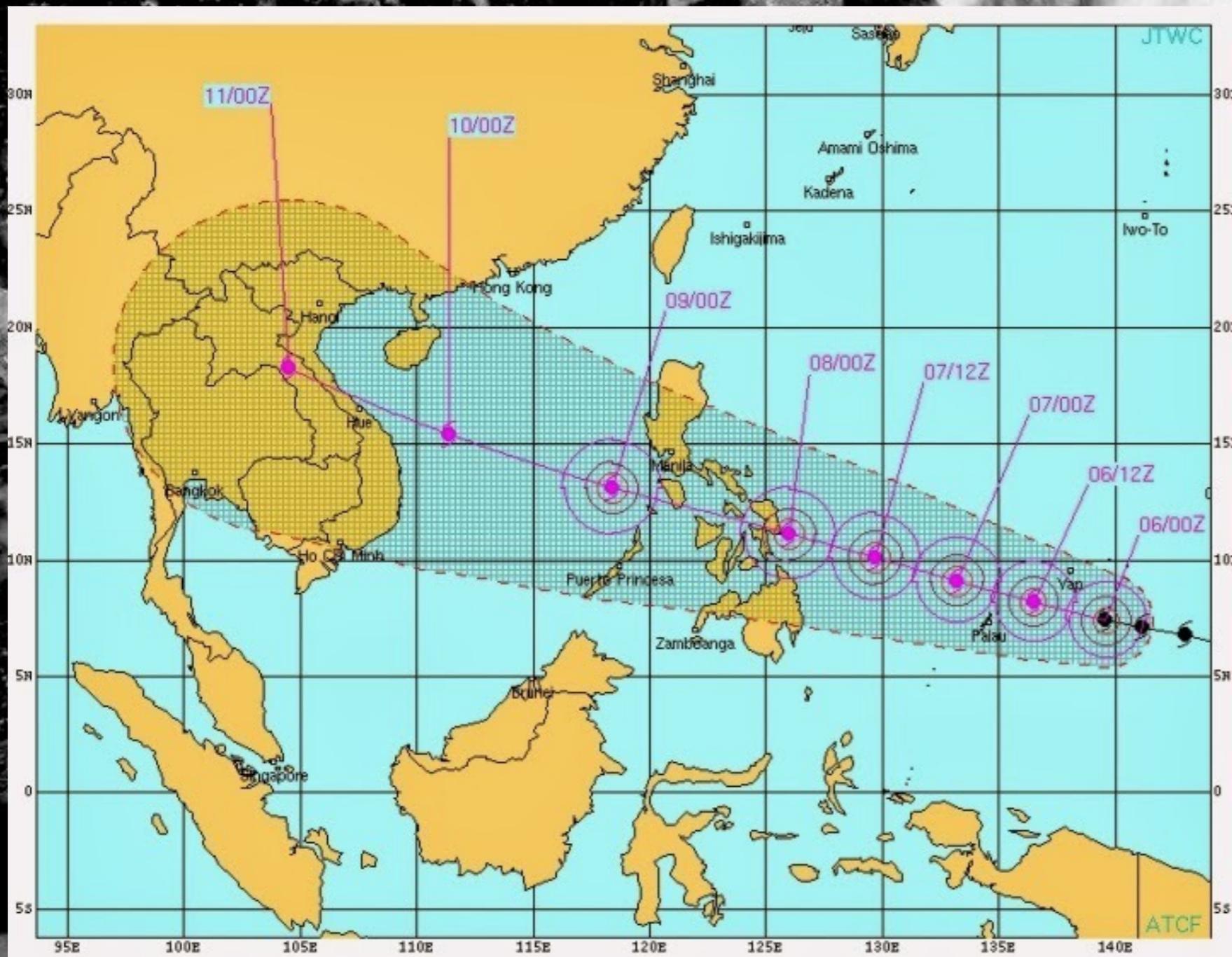
a mathematical model

$$\nabla \cdot \mathbf{E} = 4\pi\rho$$

$$\nabla \cdot \mathbf{B} = 0$$

$$\nabla \times \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}$$

$$\nabla \times \mathbf{B} = \frac{1}{c} \left(4\pi \mathbf{J} + \frac{\partial \mathbf{E}}{\partial t} \right)$$



The
**SOCIAL
ATOM**



WHY THE RICH GET RICHER,
CHEATERS GET CAUGHT,
AND YOUR NEIGHBOR USUALLY
LOOKS LIKE YOU



MARK BUCHANAN

MARK BUCHANAN



Mark Buchanan

The Social Atom

Bloomsbury, USA, 2007

machine learning

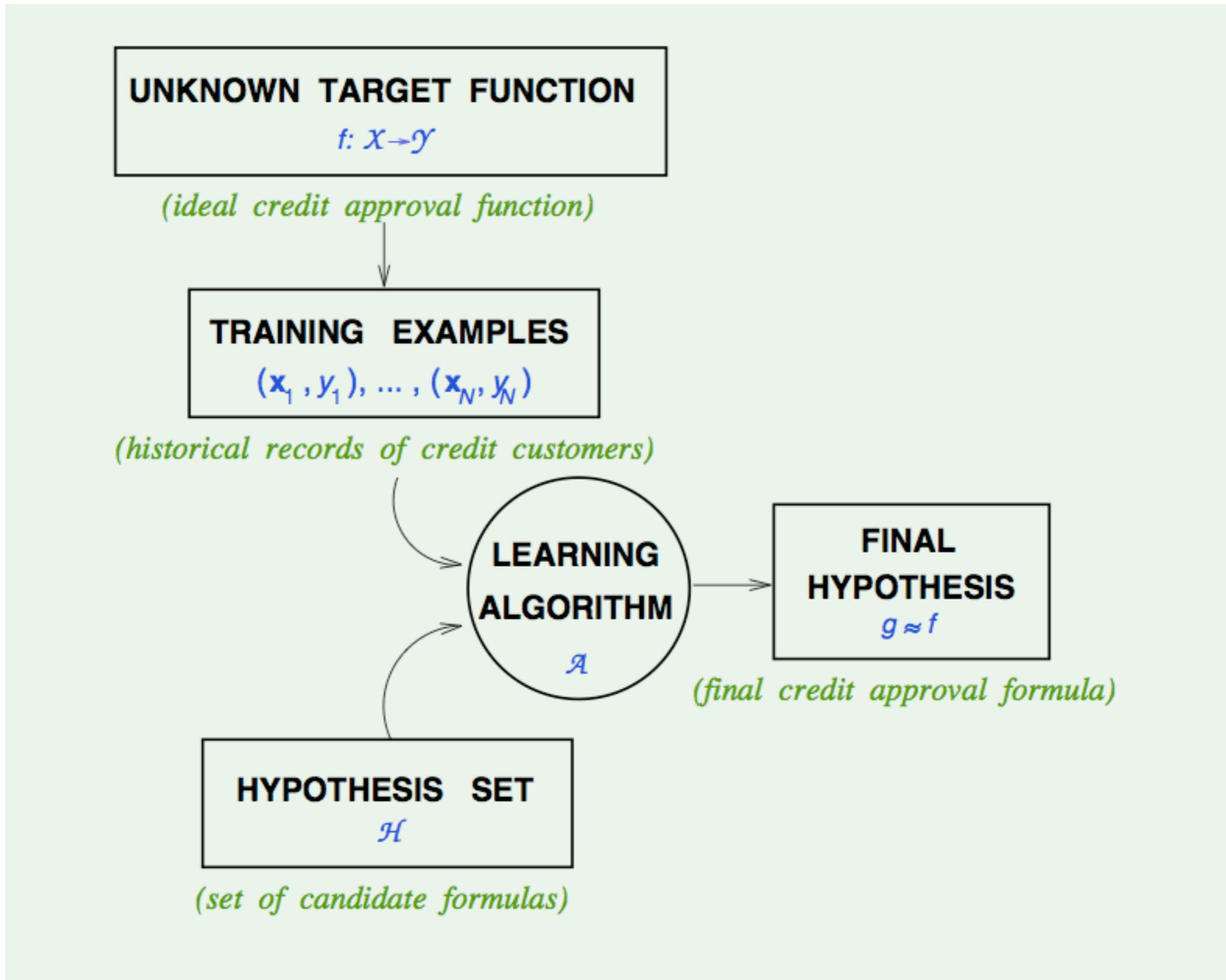
Formalization:

- Input: \mathbf{x} (*customer application*)
- Output: y (*good/bad customer?*)
- Target function: $f : \mathcal{X} \rightarrow \mathcal{Y}$ (*ideal credit approval formula*)
- Data: $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_N, y_N)$ (*historical records*)



- Hypothesis: $g : \mathcal{X} \rightarrow \mathcal{Y}$ (*formula to be used*)

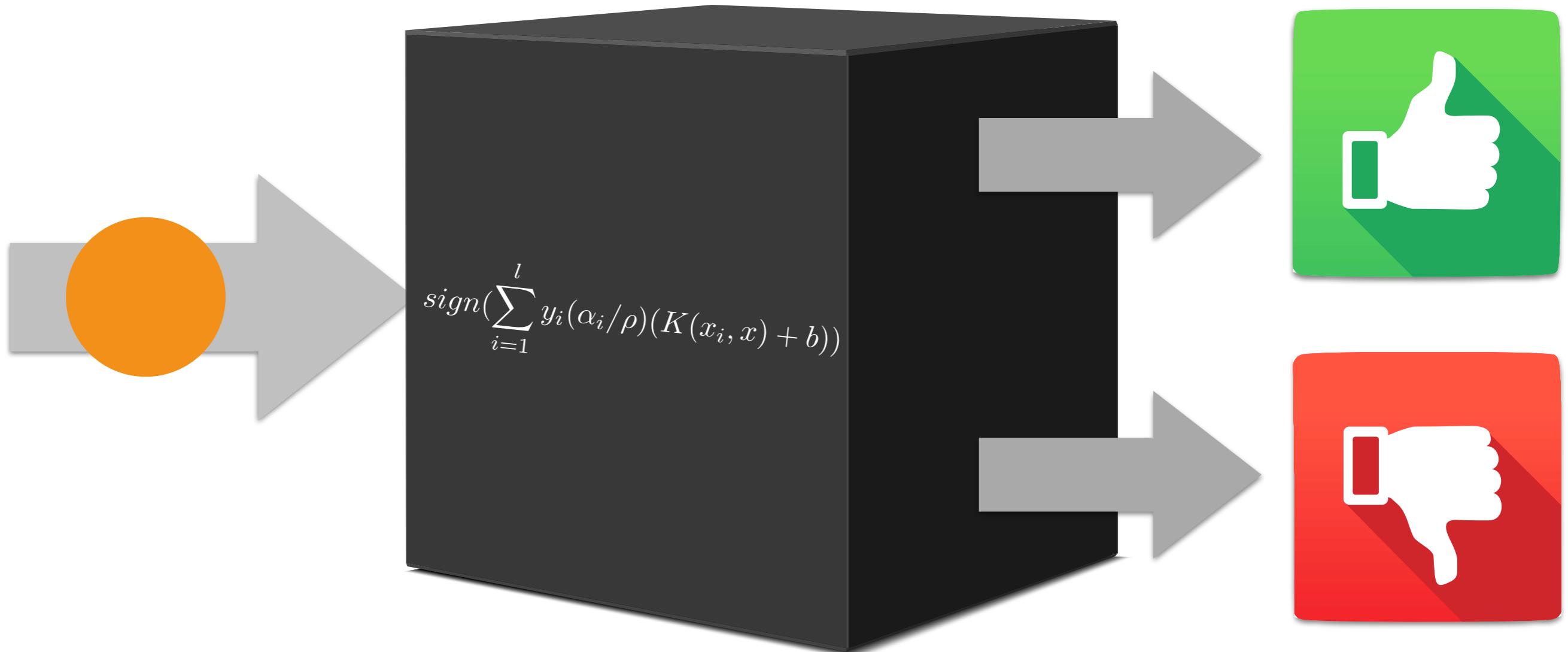
machine learning



algorithmic decision



a black box



Algorithmic bias: from discrimination discovery to fairness-aware data mining

A Tutorial at KDD'16



Welcome to the mini-website on the tutorial titled *Algorithmic bias: from discrimination discovery to fairness-aware data mining*, which will take place at **KDD'16** in San Francisco, California

Abstract

Algorithms and decision making based on Big Data have become pervasive in all aspects of our daily (offline and online) lives, as they have become essential tools in personal finance, health care, hiring, housing, education, and policies. Data and algorithms determine the media we consume, the stories we read, the people we meet, the places we visit, but also whether we get a job, or whether our loan request is approved. It is therefore of societal and ethical importance to ask whether these algorithms can be discriminative on grounds, such as gender, ethnicity, marital or health status. It turns out that the answer is positive: for instance, recent studies have shown that Google's online advertising system displayed ads for high-income jobs to men much more often than it did to women; and ads for arrest records were significantly more likely to show up on searches for distinctively black names or a historically black fraternity.

This *algorithmic bias* exists even when there is no discrimination intention in the developer of the algorithm. Sometimes it may be inherent to the data sources used (software making decisions based on data can reflect, or even amplify, the results of historical discrimination), but even when the sensitive attributes have been suppressed from the input, a well trained machine learning algorithm may still discriminate on the basis of such sensitive attributes because of correlations existing in the data. One approach is to develop data mining systems which are discrimination-conscious by-design. This is a novel and challenging research area for the data mining community.



Yaser S. Abu-Mostafa
Malik Magdon-Ismail
Hsuan-Tien Lin

LEARNING FROM DATA

A SHORT COURSE

