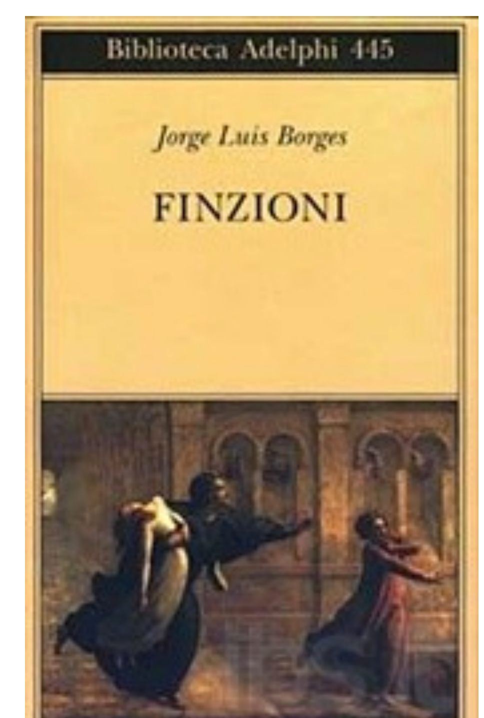
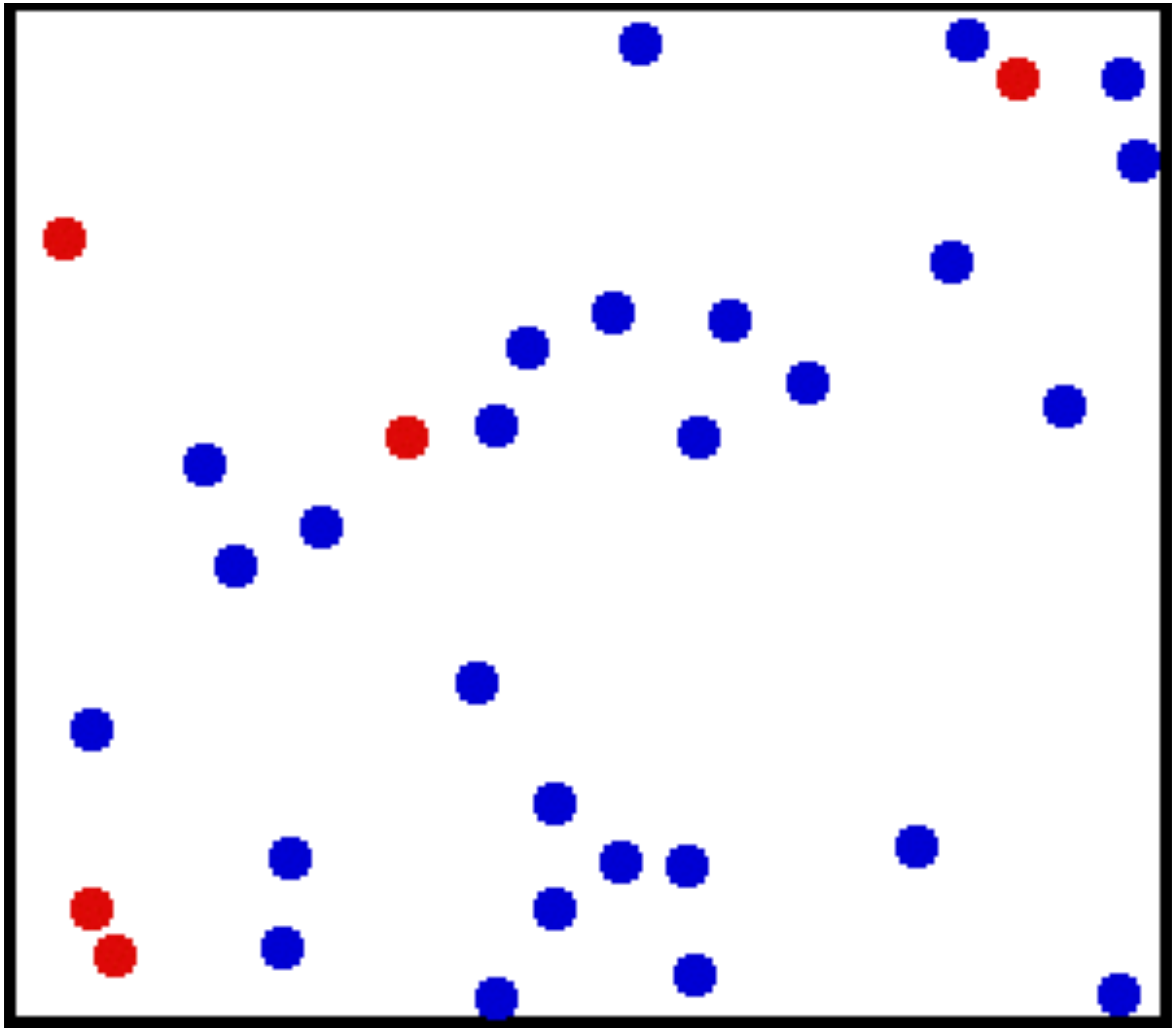
Big DATA
vs.
Small DATA

“Noi, in un’occhiata, percepiamo: tre bicchieri su una tavola. Funes: tutti i tralci, i grappoli e gli acini d’una pergola. Sapeva le forme delle nubi australi dell’alba del 30 aprile 1882, e poteva confrontarle, nel ricordo, con la copertina marmorizzata d’un libro che aveva visto una sola volta, o con le spume che sollevò un remo, nel Rio Negro, la vigilia della battaglia di Quebracho. Questi ricordi non erano semplici: ogni immagine visiva era legata a sensazioni muscolari, termiche ecc. Poteva ricostruire tutti i sogni dei suoi sonni, tutte le immagini dei suoi dormiveglia”.

“Funes o della memoria”

Silvia Hopenhayn per il quotidiano argentino La Nación

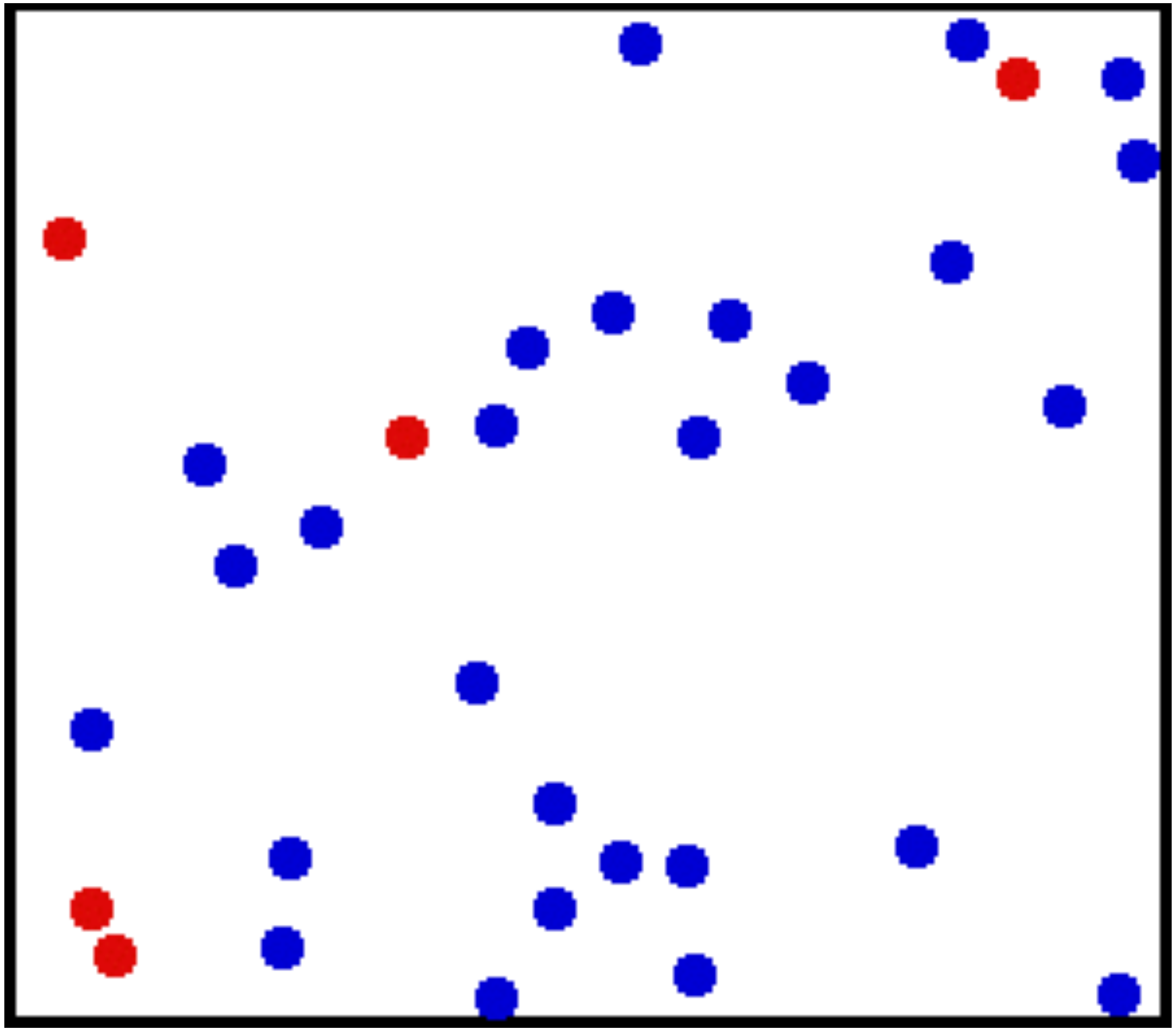




Memoria vs. sintesi (data compression)

Microscopico vs. macroscopico

$\sim 10^{23}$ $O(1)$

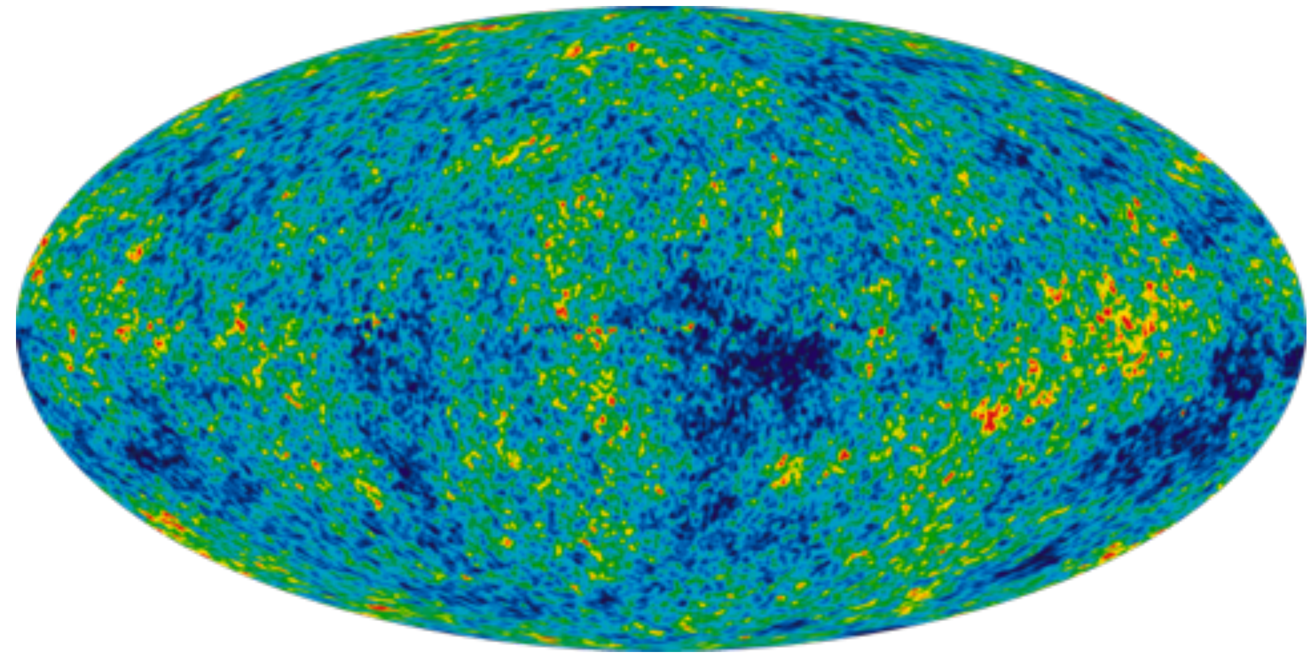


Memoria vs. sintesi (data compression)

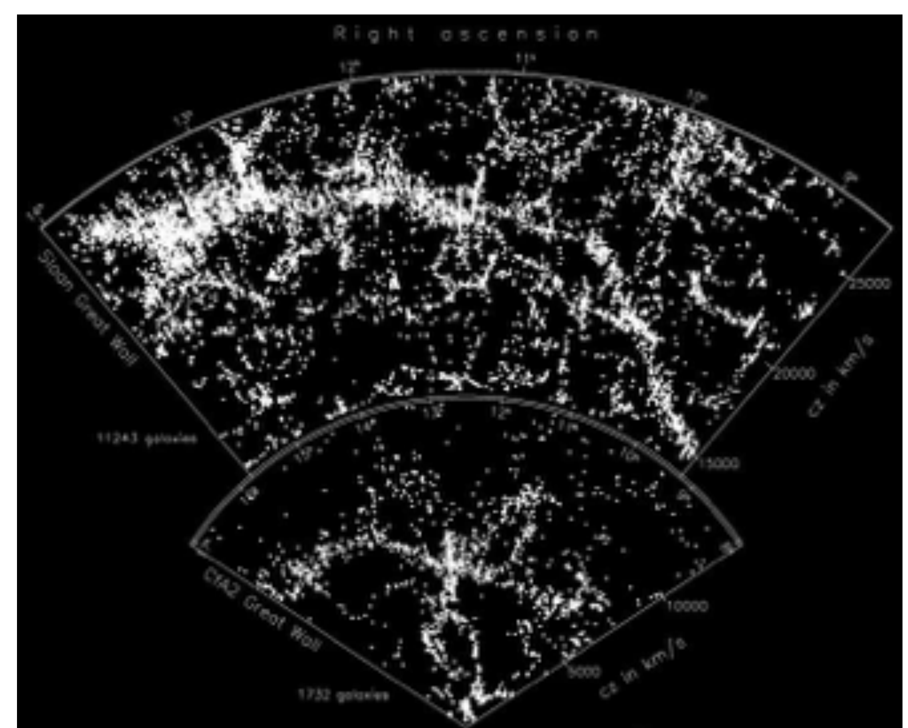
Microscopico vs. macroscopico

$\sim 10^{23}$ $O(1)$

Traditional DATA sources

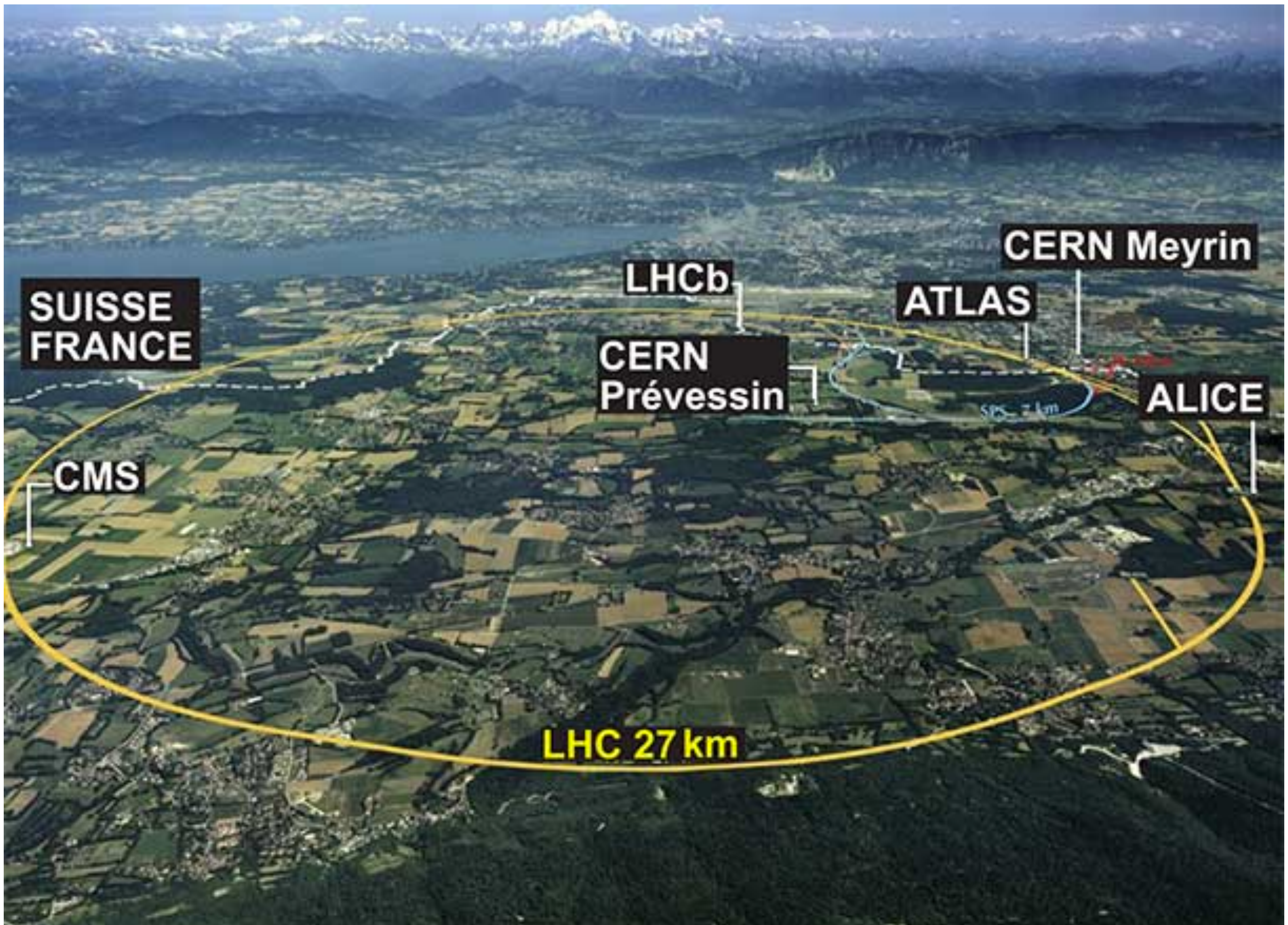


Observations

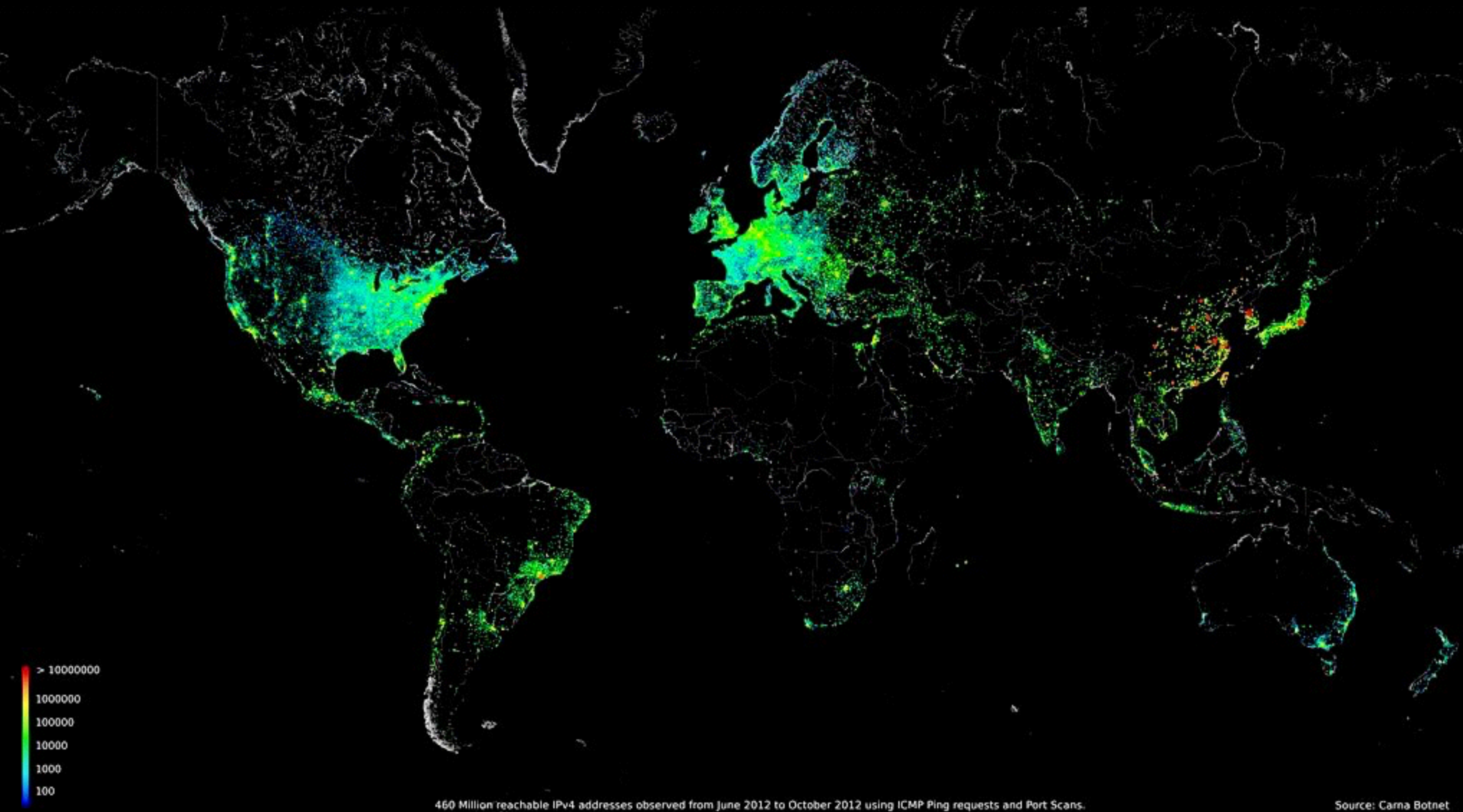


Traditional DATA sources

Experiments

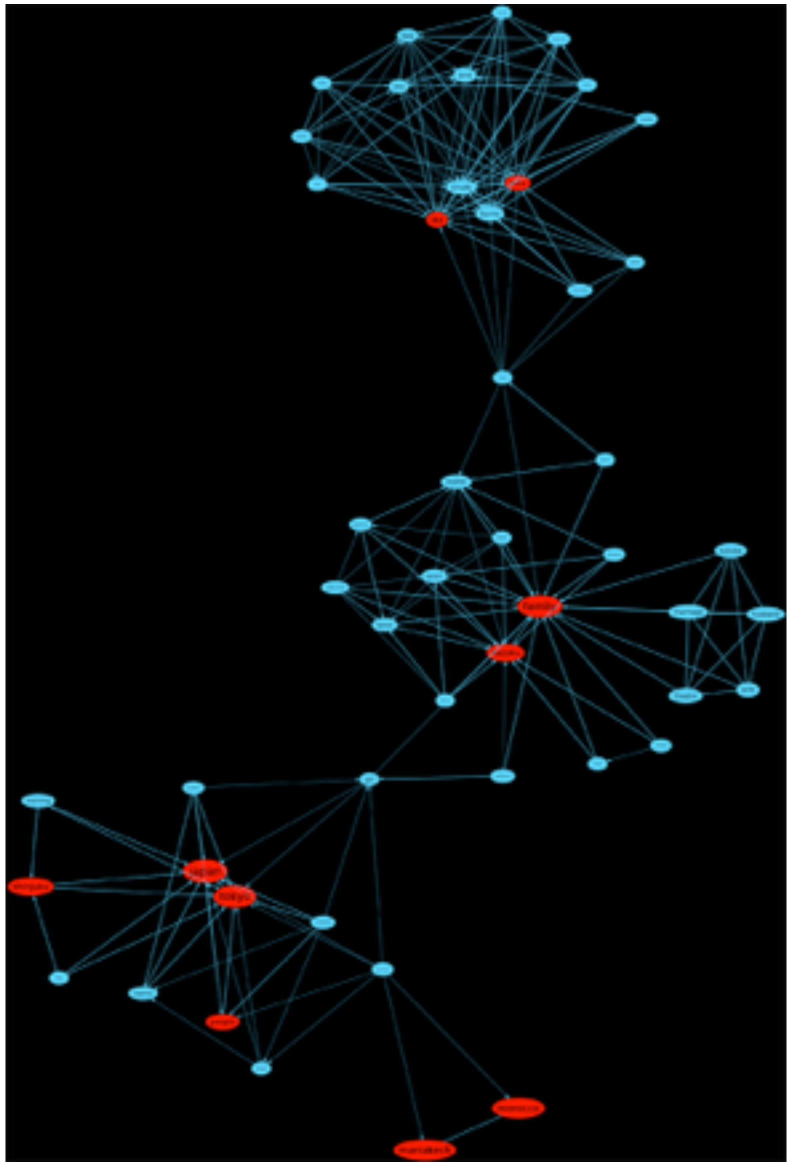


The Internet



The Largest Artificial Entity in History

Techno-social systems



user level

cognitive,
behavioural

community
level

social,
interactive

infrastructure
level

ICT, networks,
physical-digital

Executive Summary

Annual global IP traffic will pass the zettabyte ([ZB]; 1000 exabytes [EB]) threshold by the end of 2016, and will reach 2.3 ZB per year by 2020. By the end of 2016, global IP traffic will reach 1.1 ZB per year, or 88.7 EB per month, and by 2020 global IP traffic will reach 2.3 ZB per year, or 194 EB per month.

Global IP traffic will increase nearly threefold over the next 5 years. Overall, IP traffic will grow at a compound annual growth rate (CAGR) of 22 percent from 2015 to 2020. Monthly IP traffic will reach 25 GB per capita by 2020, up from 10 GB per capita in 2015.

Busy-hour Internet traffic is growing more rapidly than average Internet traffic. Busy-hour (or the busiest 60-minute period in a day) Internet traffic increased 51 percent in 2015, compared with 29-percent growth in average traffic. Busy-hour Internet traffic will increase by a factor of 4.6 between 2015 and 2020, and average Internet traffic will increase by a factor of 2.0.

Smartphone traffic will exceed PC traffic by 2020. In 2015, PCs accounted for 53 percent of total IP traffic, but by 2020 PCs will account for only 29 percent of traffic. Smartphones will account for 30 percent of total IP traffic in 2020, up from 8 percent in 2015. PC-originated traffic will grow at a CAGR of 8 percent, and TVs, tablets, smartphones, and machine-to-machine (M2M) modules will have traffic growth rates of 17 percent, 39 percent, 58 percent, and 44 percent, respectively.

Traffic from wireless and mobile devices will account for two-thirds of total IP traffic by 2020. By 2020, wired devices will account for 34 percent of IP traffic, and Wi-Fi and mobile devices will account for 66 percent of IP traffic. In 2015, wired devices accounted for the majority of IP traffic, at 52 percent.

Content delivery networks (CDNs) will carry nearly two-thirds of Internet traffic by 2020. Sixty-four percent of all Internet traffic will cross CDNs by 2020 globally, up from 45 percent in 2015.

The number of devices connected to IP networks will be more than three times the global population by 2020. There will be 3.4 networked devices per capita by 2020, up from 2.2 networked devices per capita in 2015. There will be 26.3 billion networked devices in 2020, up from 16.3 billion in 2015.

Broadband speeds will nearly double by 2020. By 2020, global fixed broadband speeds will reach 47.7 Mbps, up from 24.7 Mbps in 2015.

By the year 2020

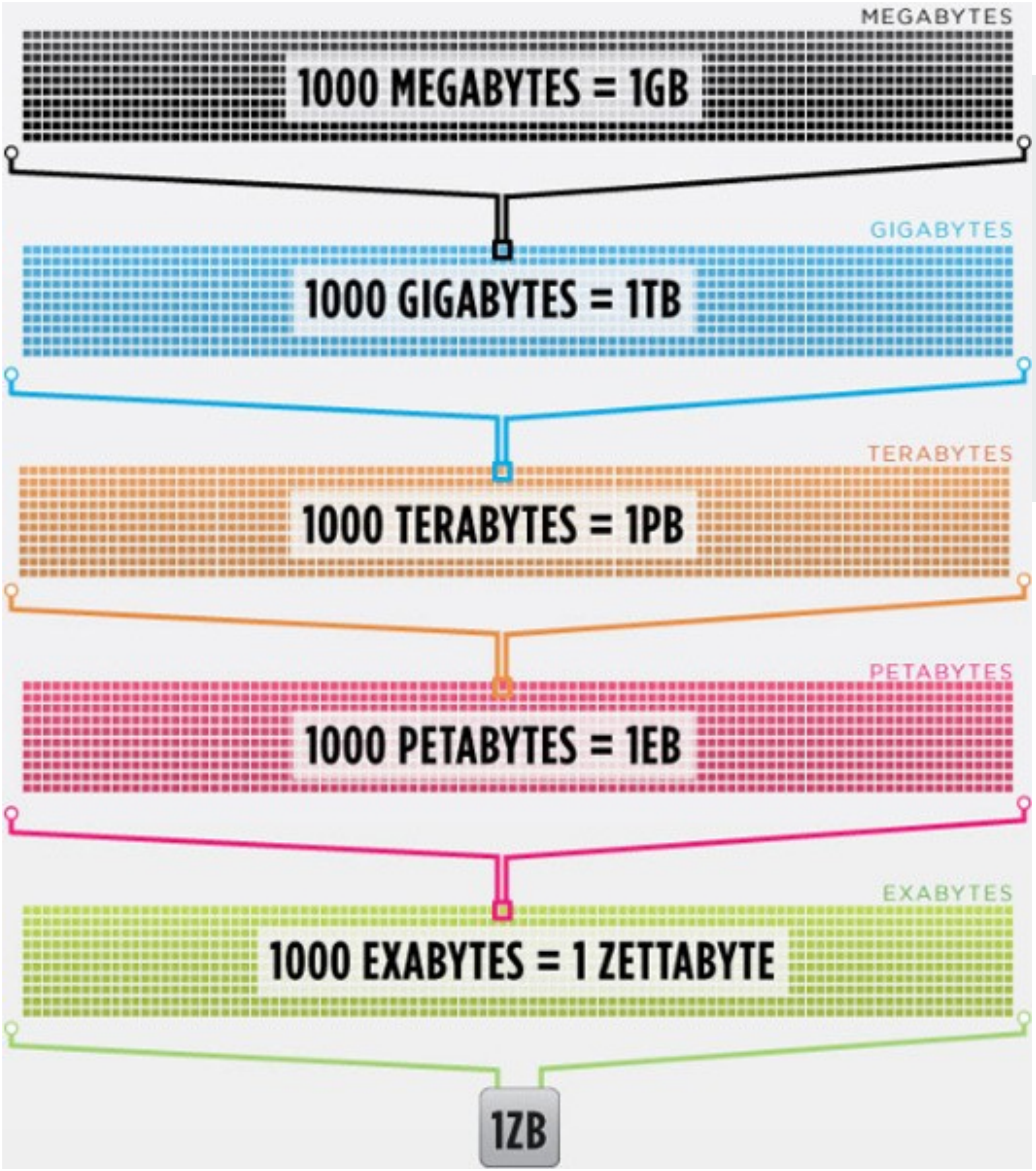
2.3
zettabytes will be the amount of annual global IP traffic.

47.7
Mbps of global fixed broadband speeds up from 24.7 Mbps in 2015.

30%
of total IP traffic will be from smartphones (exceeding PC traffic).

2/3
Traffic from wireless devices will account for two-thirds of total IP traffic by 2020.

3.4
devices per capita will be connected to IP networks (3X the global population).

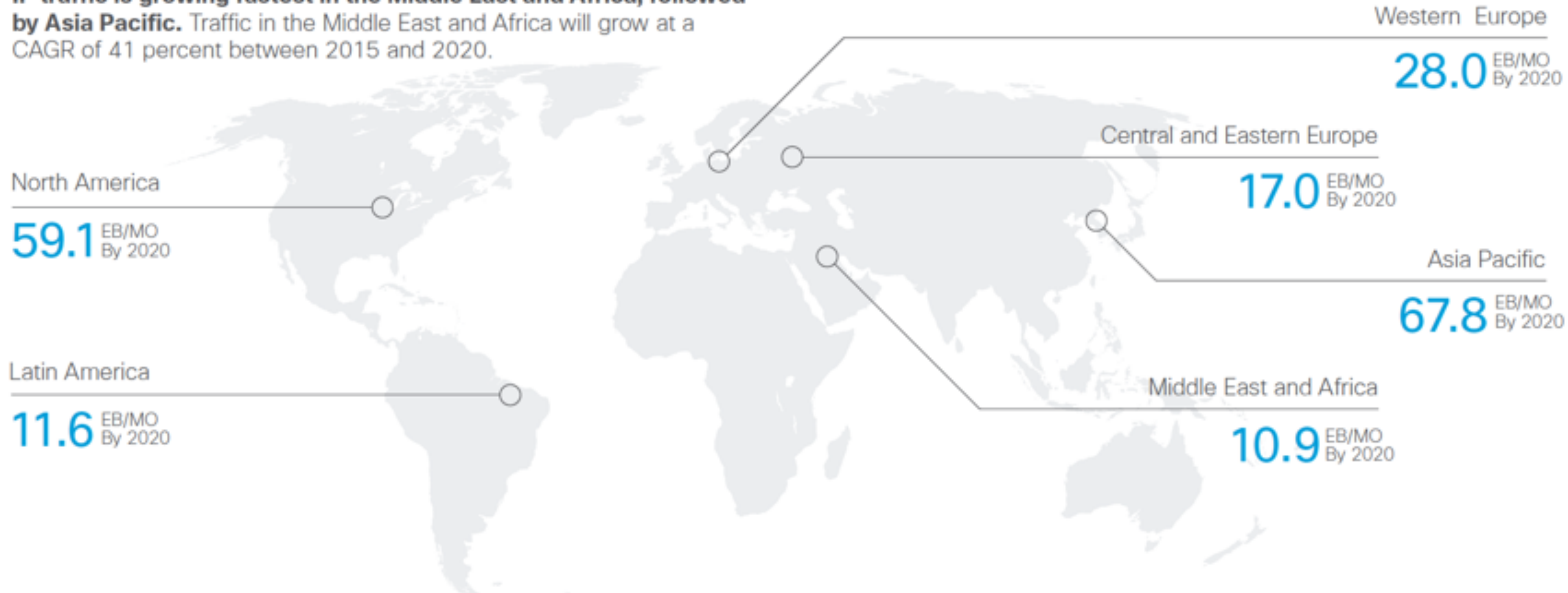


What is a Zettabyte?

1,000,000,000,000 gigabytes
1,000,000,000,000 terabytes
1,000,000,000,000 petabytes
1,000,000,000,000 exabytes
1,000,000,000,000 zettabyte

Regional Highlights

IP traffic is growing fastest in the Middle East and Africa, followed by Asia Pacific. Traffic in the Middle East and Africa will grow at a CAGR of 41 percent between 2015 and 2020.



Sources of data

Social Networks
(human-sourced information)

1100. Social Networks: Facebook, Twitter, Tumblr etc.
1200. Blogs and comments
1300. Personal documents
1400. Pictures: Instagram, Flickr, Picasa etc.
1500. Videos: Youtube etc.
1600. Internet searches
1700. Mobile data content: text messages
1800. User-generated maps
1900. E-Mail

Traditional Business systems
(process-mediated data):

- 21. Data produced by Public Agencies
  - 2110. Medical records
- 22. Data produced by businesses
  - 2210. Commercial transactions
  - 2220. Banking/stock records
  - 2230. E-commerce
  - 2240. Credit cards

Internet of Things
(machine-generated data)

- 31. Data from sensors
  - 311. Fixed sensors
    - 3111. Home automation
    - 3112. Weather/pollution sensors
    - 3113. Traffic sensors/webcam
    - 3114. Scientific sensors
    - 3115. Security/surveillance videos/images
  - 312. Mobile sensors (tracking)
    - 3121. Mobile phone location
    - 3122. Cars
    - 3123. Satellite images
- 32. Data from computer systems
  - 3210. Logs
  - 3220. Web logs

Task Team on Big Data, June 2013

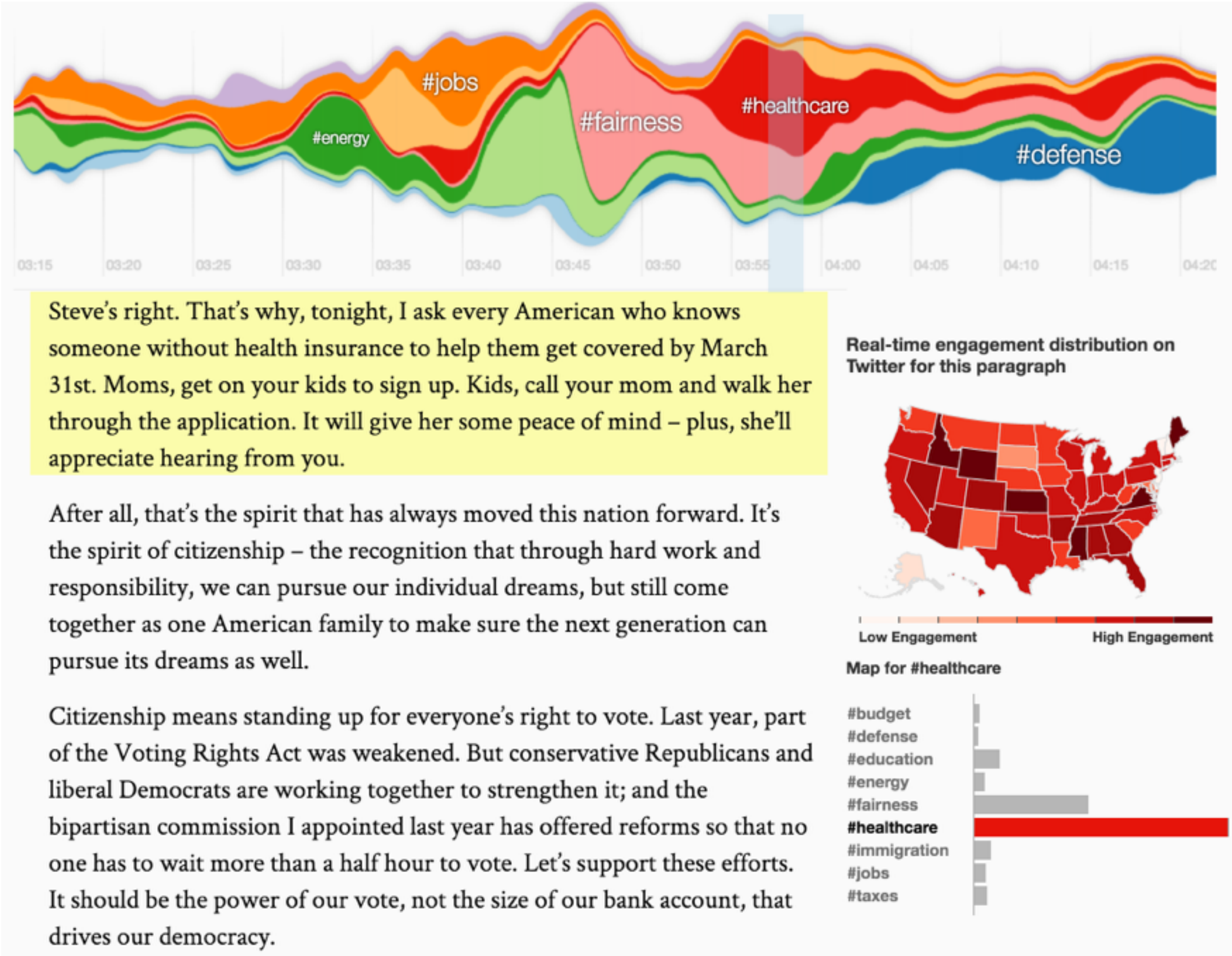
Data exhausts

data generated as a byproduct of people's
online actions and choices

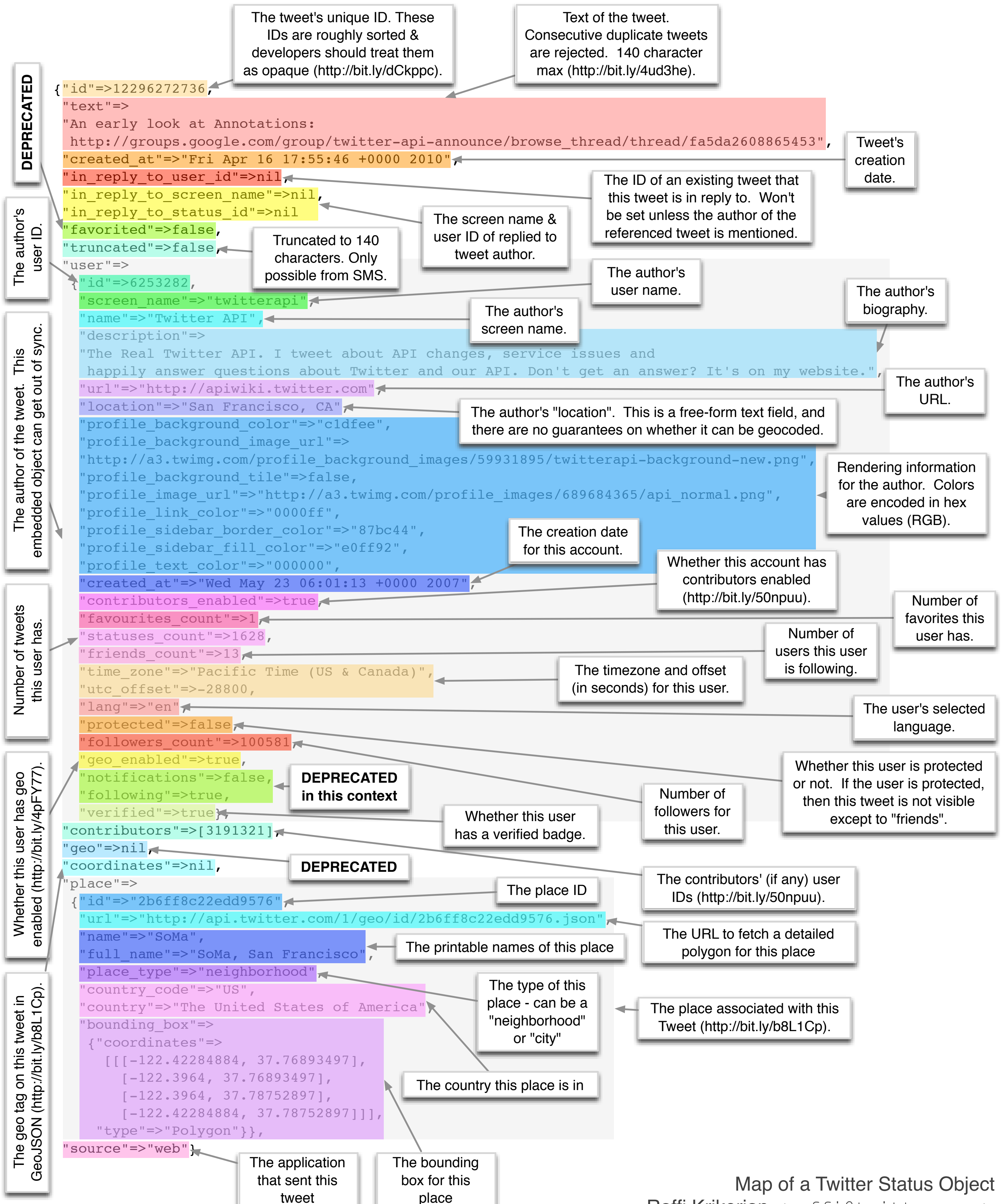
Frozen

Forced perspectives

Contexts ??


#jobs
#energy
#fairness
#healthcare
#defense
03:15
03:20
03:25
03:30
03:35
03:40
03:45
03:50
03:55
04:00
04:05
04:10
04:15
Steve's right. That's why, tonight, I ask every American who knows someone without health insurance to help them get covered by March 31st. Moms, get on your kids to sign up. Kids, call your mom and walk her through the application. It will give her some peace of mind – plus, she'll appreciate hearing from you.
After all, that's the spirit that has always moved this nation forward. It's the spirit of citizenship – the recognition that through hard work and responsibility, we can pursue our individual dreams, but still come together as one American family to make sure the next generation can pursue its dreams as well.
Citizenship means standing up for everyone's right to vote. Last year, part of the Voting Rights Act was weakened. But conservative Republicans and liberal Democrats are working together to strengthen it; and the bipartisan commission I appointed last year has offered reforms so that no one has to wait more than a half hour to vote. Let's support these efforts. It should be the power of our vote, not the size of our bank account, that drives our democracy.
Real-time engagement distribution on Twitter for this paragraph
Low Engagement
High Engagement
Map for #healthcare
#budget
#defense
#education
#energy
#fairness
#healthcare
#immigration
#jobs
#taxes

Metadata



Map of a Twitter Status Object
Raffi Krikorian <raffi@twitter.com>
18 April 2010


More than
4m
articles on the English language Wikipedia
About
250,000
new Wikipedia accounts are created each month
More than
131,000
people have performed an edit on Wikipedia within the last month
Wikipedia is in more than
280
different languages
Anyone
can edit Wikipedia
Sources:
Wikipedia, Alexa,
The Noun Project
W
Ω
И
維
ウィ
위

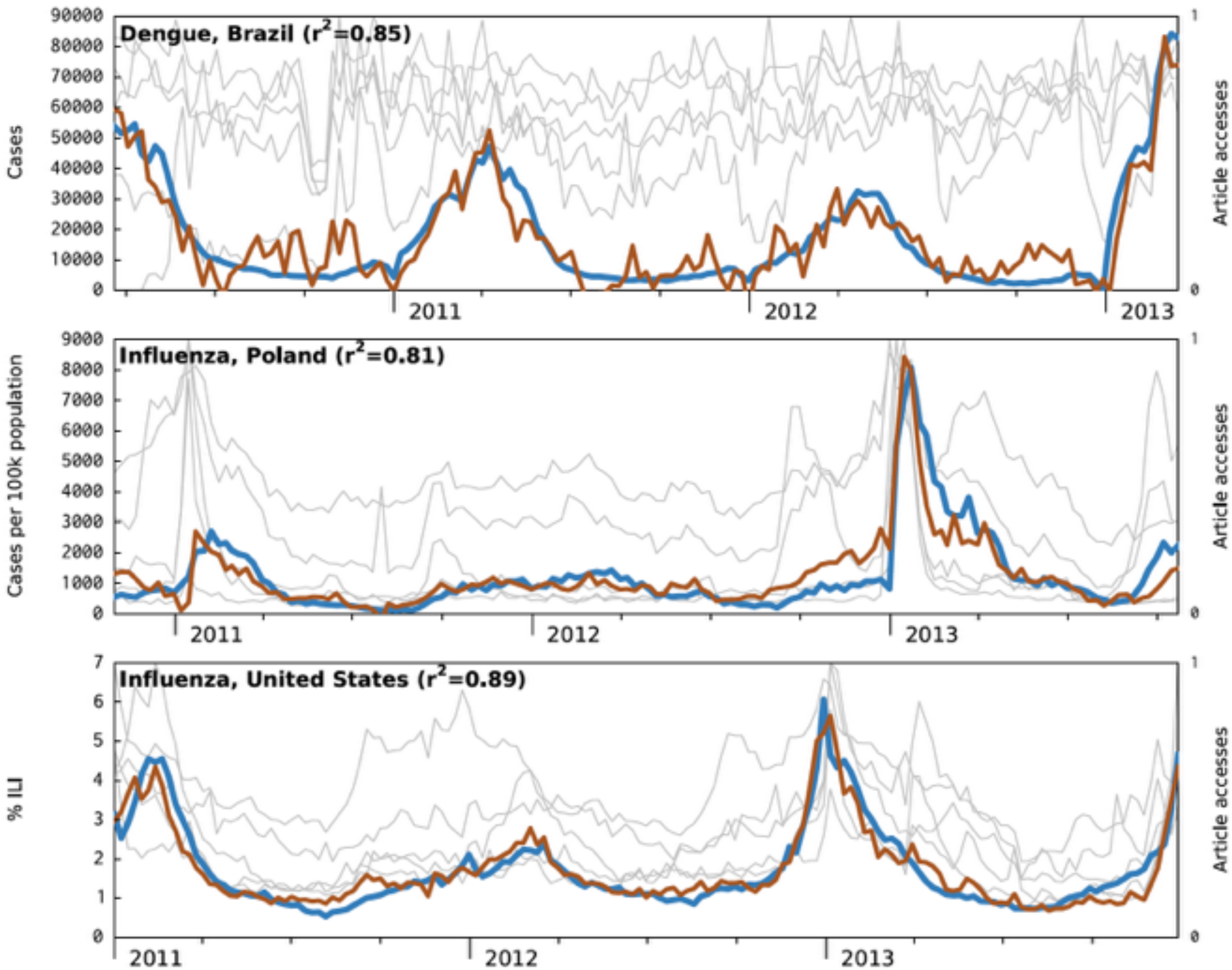
OPEN ACCESS Freely available online

PLOS | COMPUTATIONAL BIOLOGY

Global Disease Monitoring and Forecasting with Wikipedia

Nicholas Generous*, Geoffrey Fairchild, Alina Deshpande, Sara Y. Del Valle, Reid Priedhorsky

Defense Systems and Analysis Division, Los Alamos National Laboratory, Los Alamos, New Mexico, United States of America

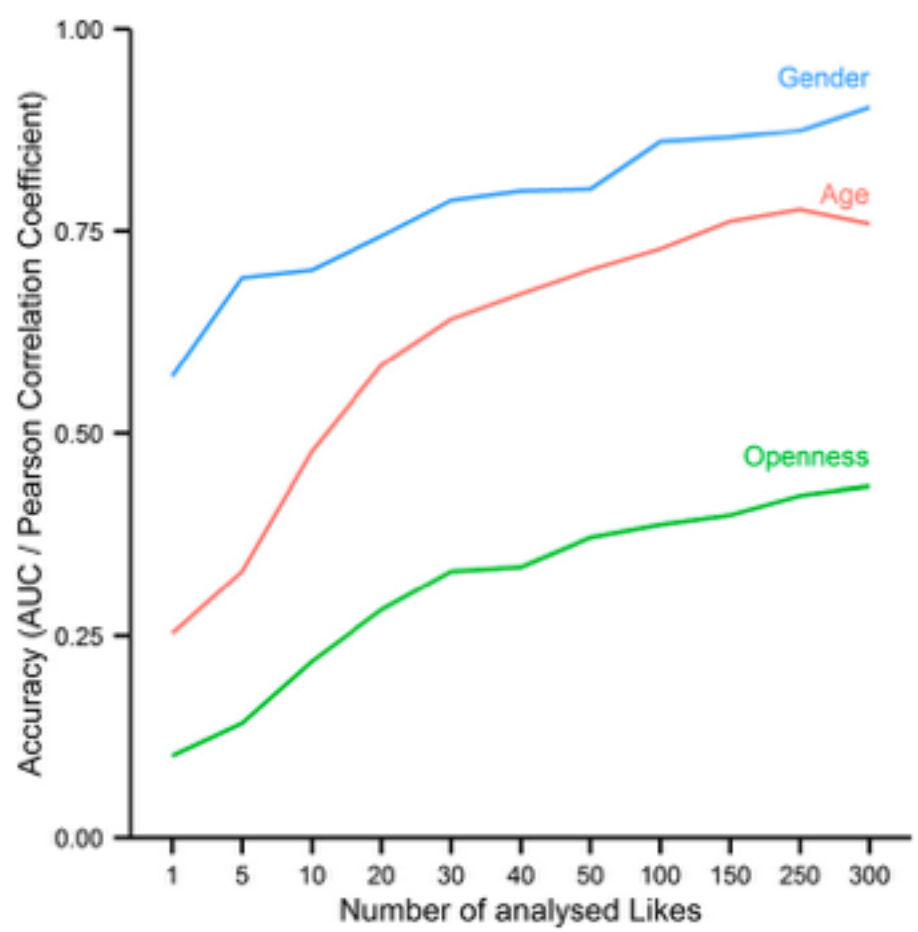


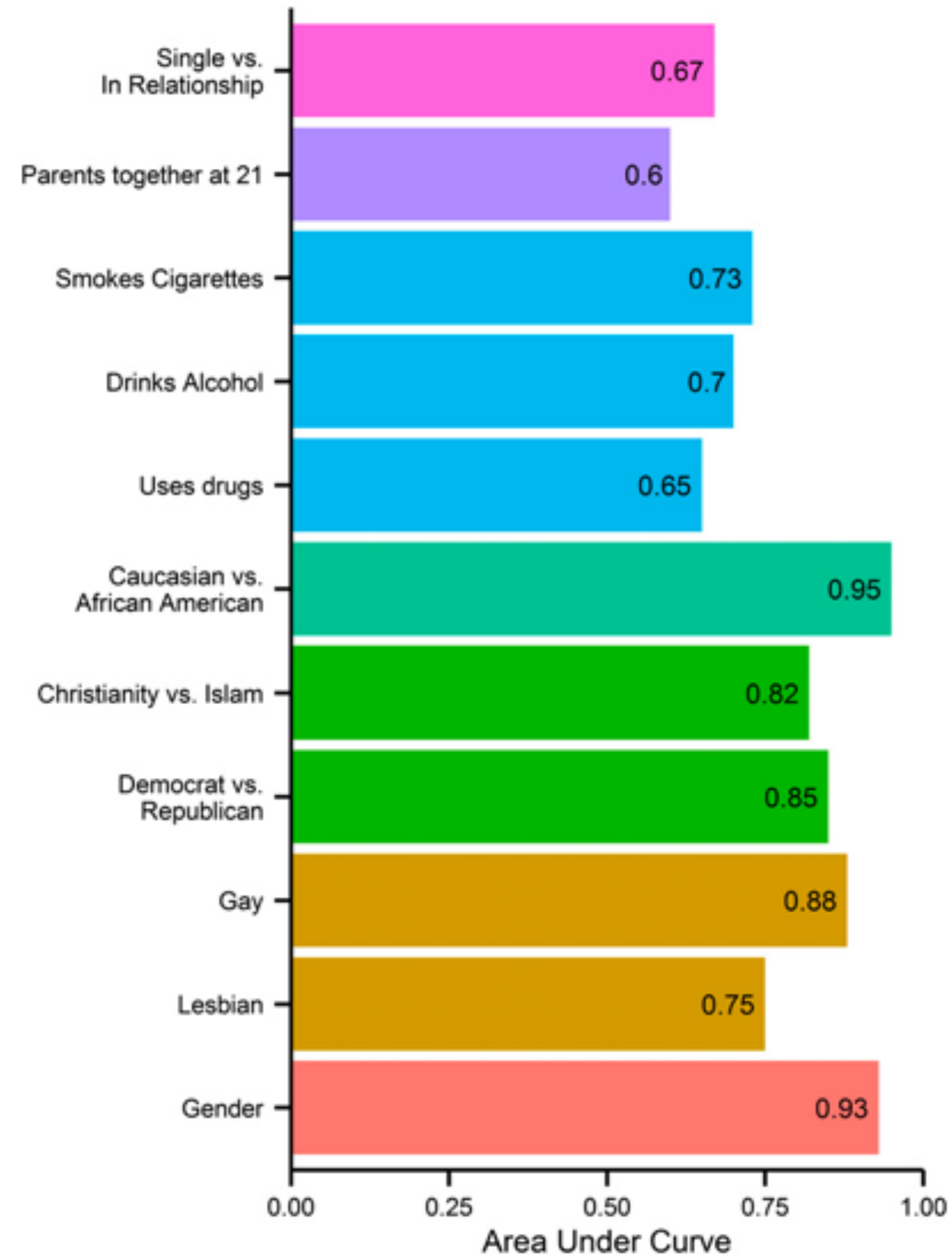
Implicit signals

Private traits and attributes are predictable from digital records of human behavior

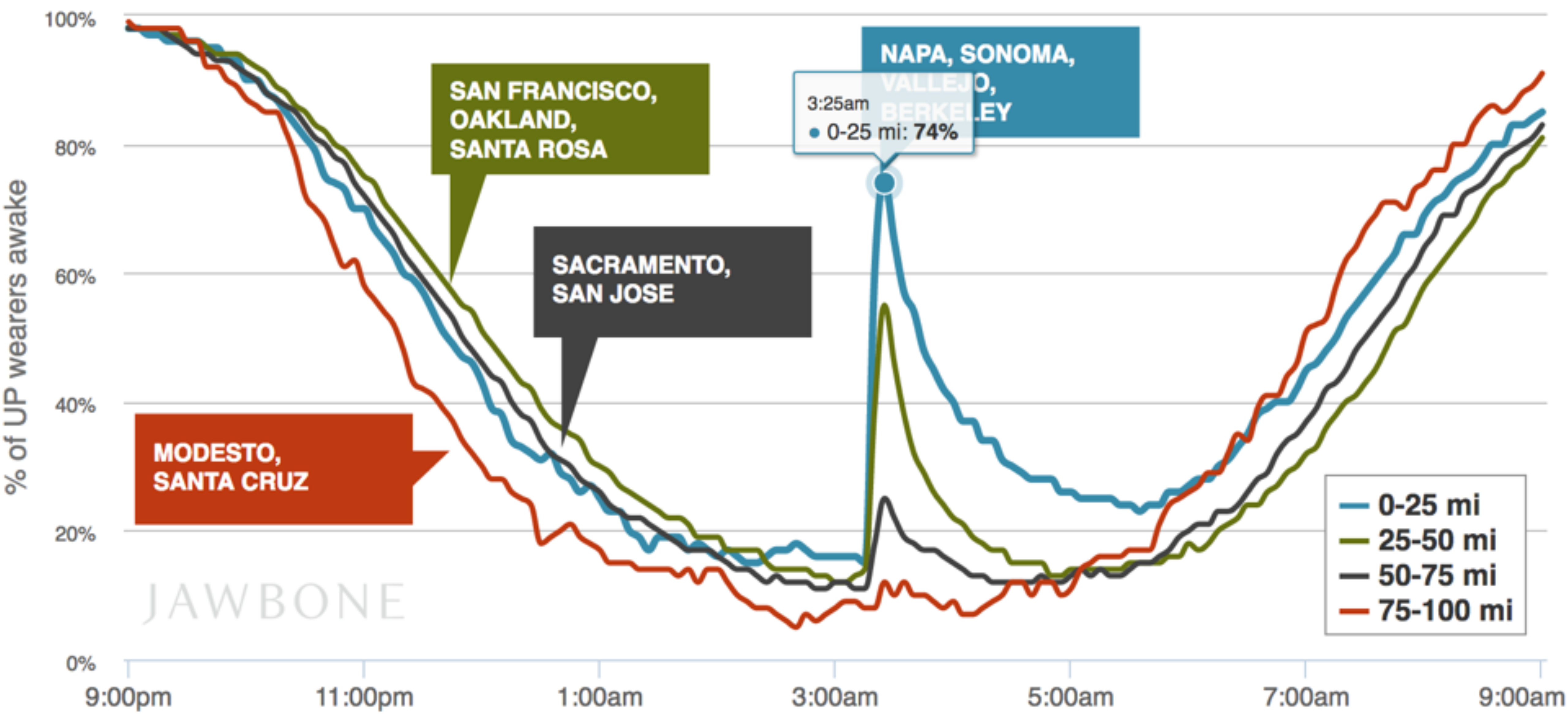
Michal Kosinski[a,1], David Stillwell[a], and Thore Graepel[b]

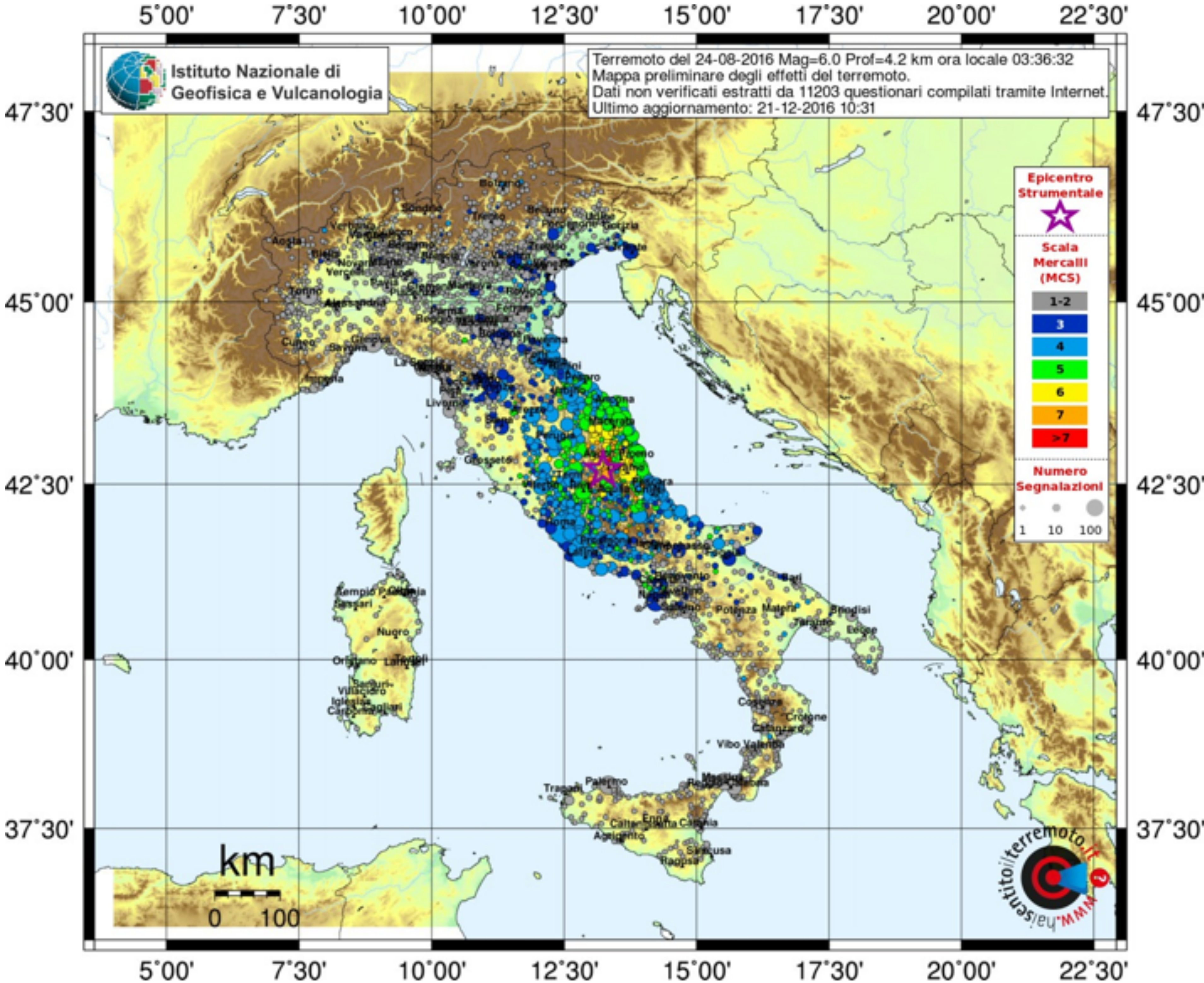
We show that easily accessible digital records of behaviour, Facebook Likes, can be used to automatically and accurately predict a range of highly sensitive personal attributes including: sexual orientation, ethnicity, religious and political views, personality traits, intelligence, happiness, use of addictive substances, parental separation, age, and gender.





HOW THE NAPA EARTHQUAKE AFFECTED BAY AREA SLEEPERS





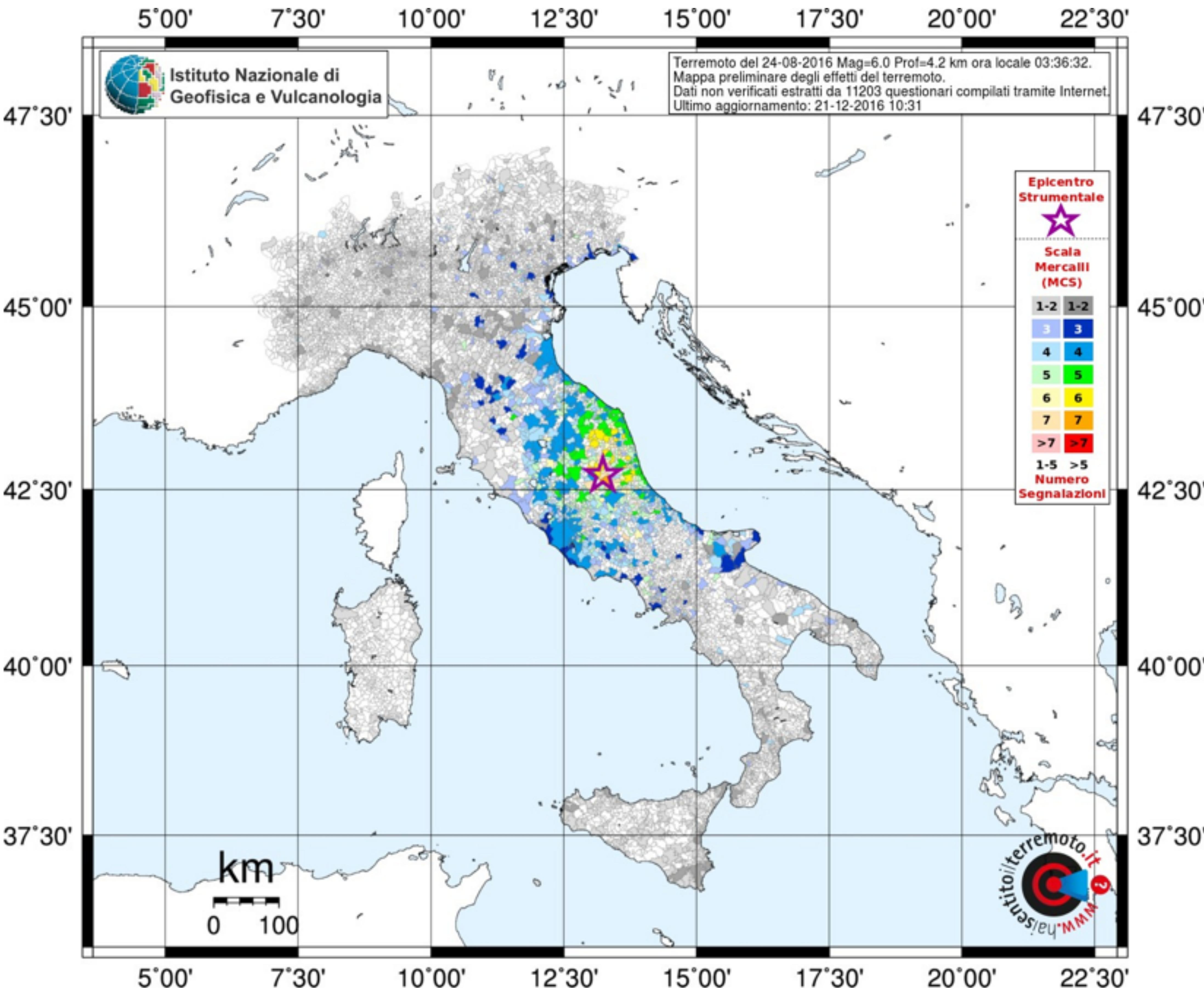
haisentitoilterremoto

Istituto Nazionale di Geofisica e Vulcanologia

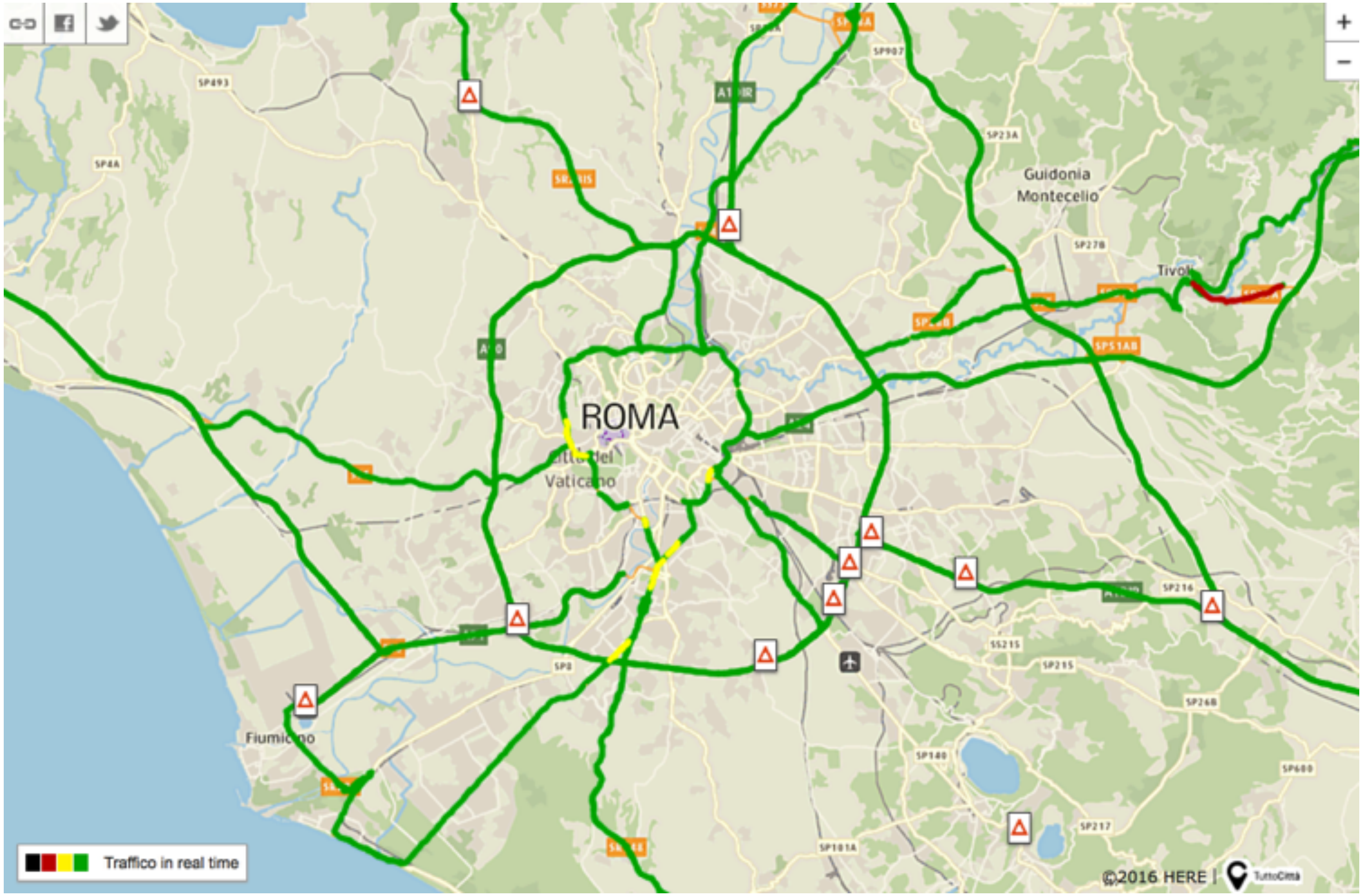
27234 corrispondenti fissi
10752 terremoti rappresentati
889961 questionari compilati

P. Tosi e V. De Rubeis (INGV)

www.haisentitoilterremoto.it

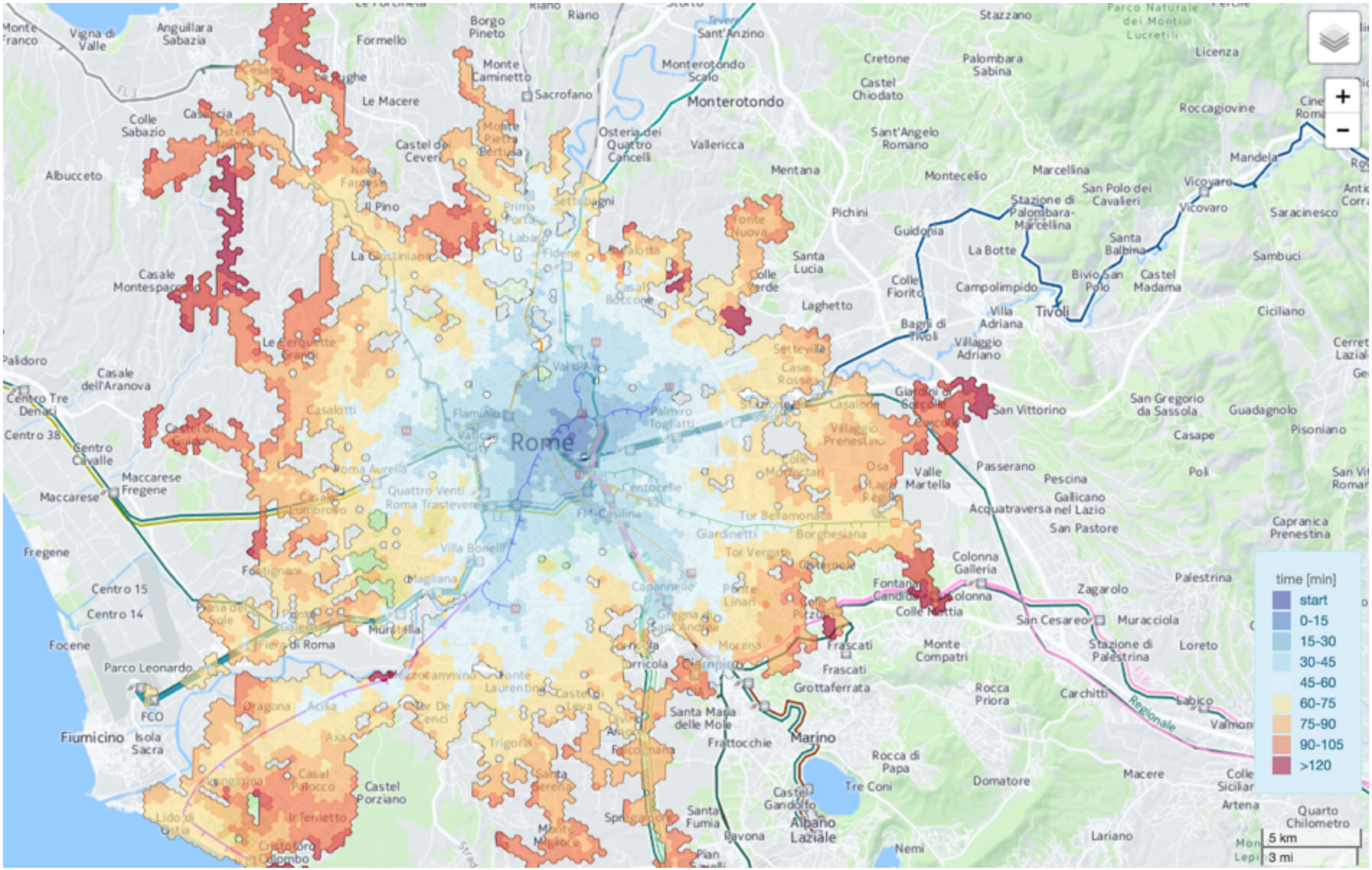

Istituto Nazionale di
Geofisica e Vulcanologia
Terremoto del 24-08-2016 Mag=6.0 Prof=4.2 km ora locale 03:36:32.
Mappa preliminare degli effetti del terremoto.
Dati non verificati estratti da 11203 questionari compilati tramite Internet.
Ultimo aggiornamento: 21-12-2016 10:31
Epicentro Strumentale
Scala Mercalli (MCS)
1-2 1-2
3 3
4 4
5 5
6 6
7 7
>7 >7
1-5 >5
Numero Segnalazioni
5°00' 7°30' 10°00' 12°30' 15°00' 17°30' 20°00' 22°30'
47°30' 45°00' 42°30' 40°00' 37°30'
km
0 100
www.haisentitoilterremoto.it

Mobility data and insurances



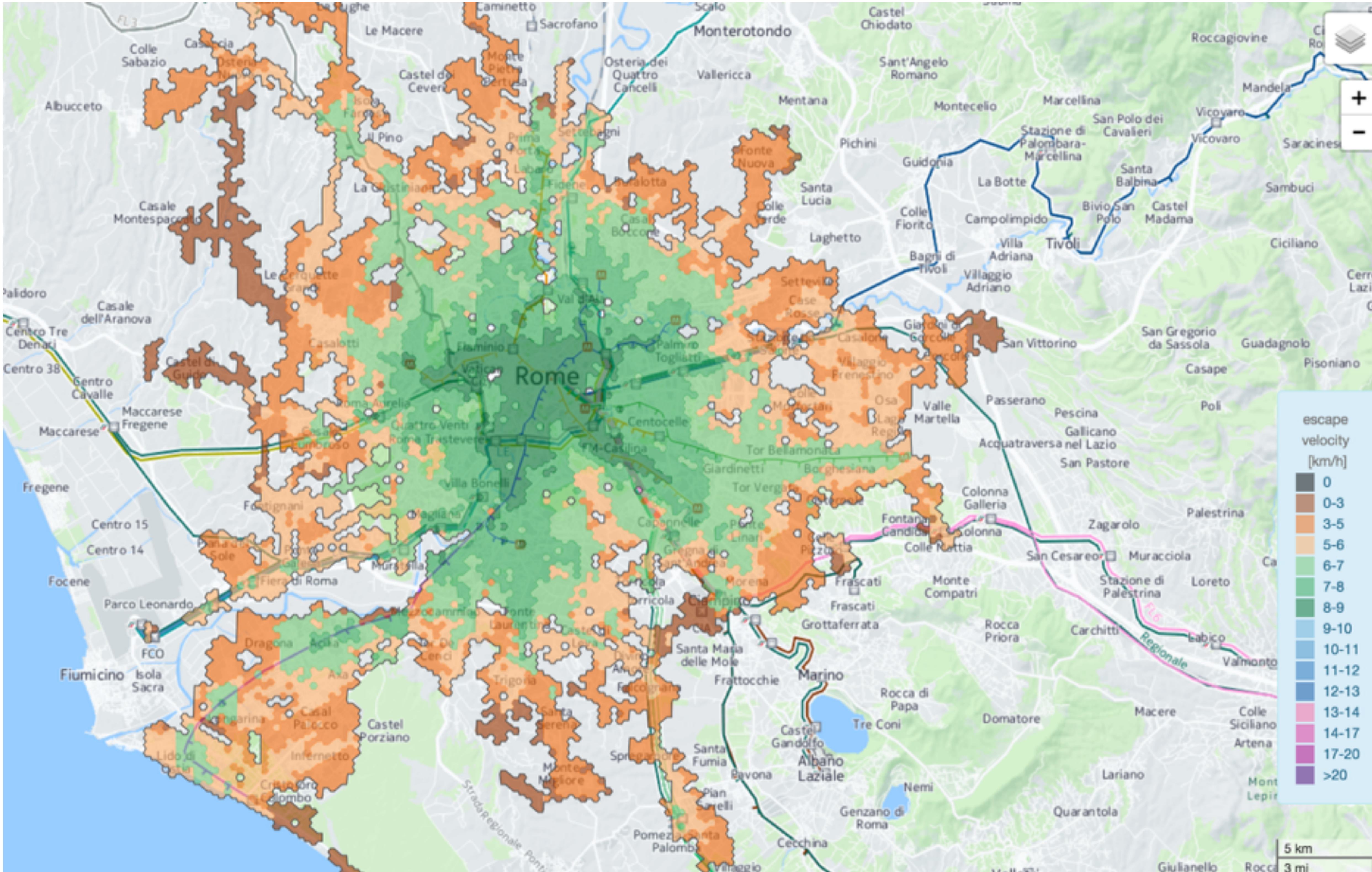
https://www.tuttocitta.it/traffico/roma

Citychrone - Isochrones

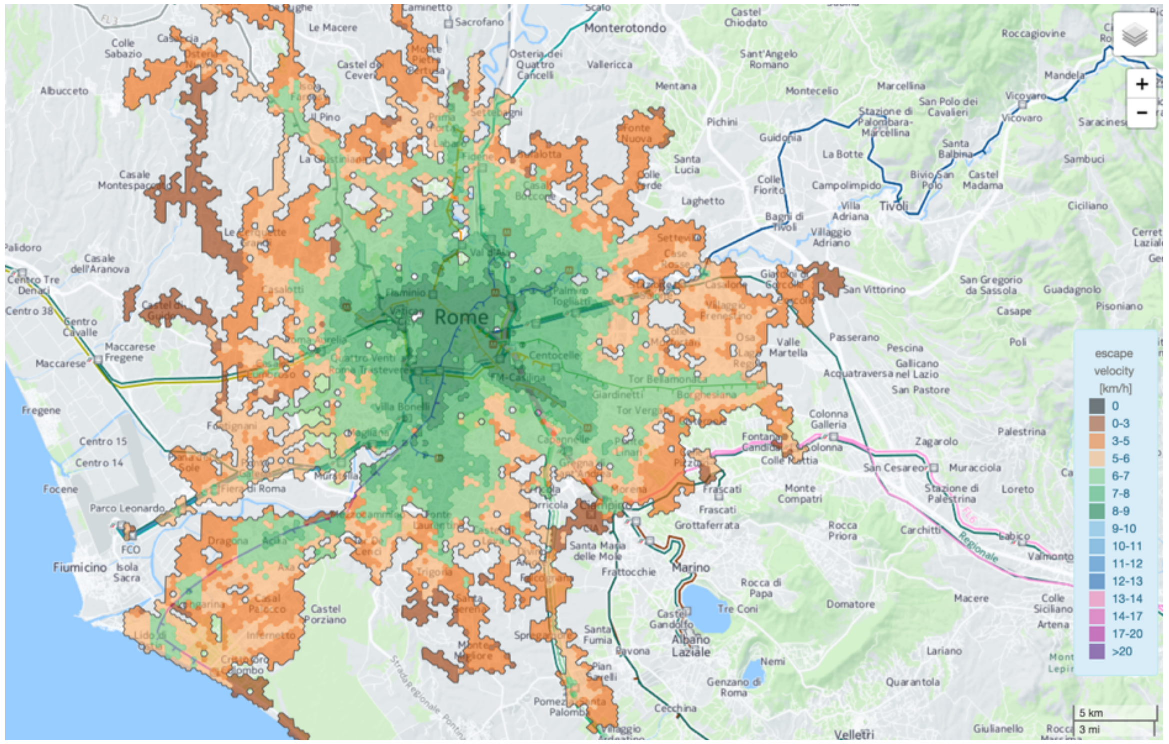


http://www.citychrone.org/

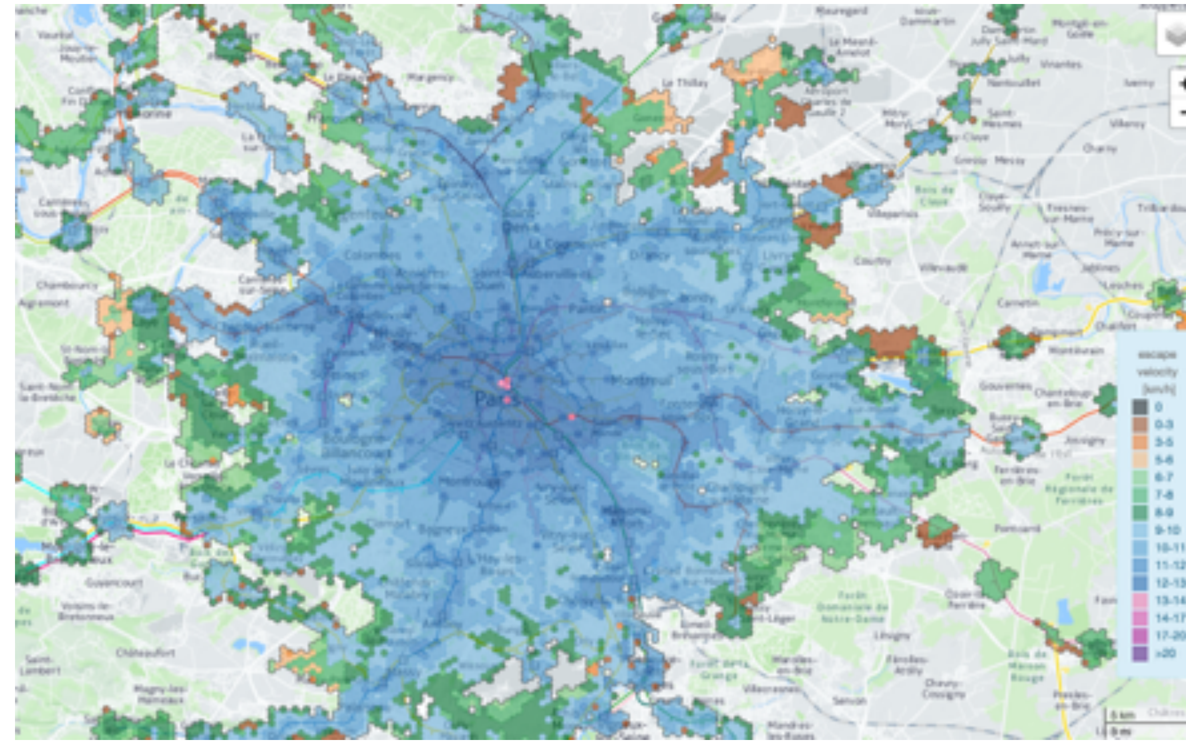
Citychrone - Escape velocities

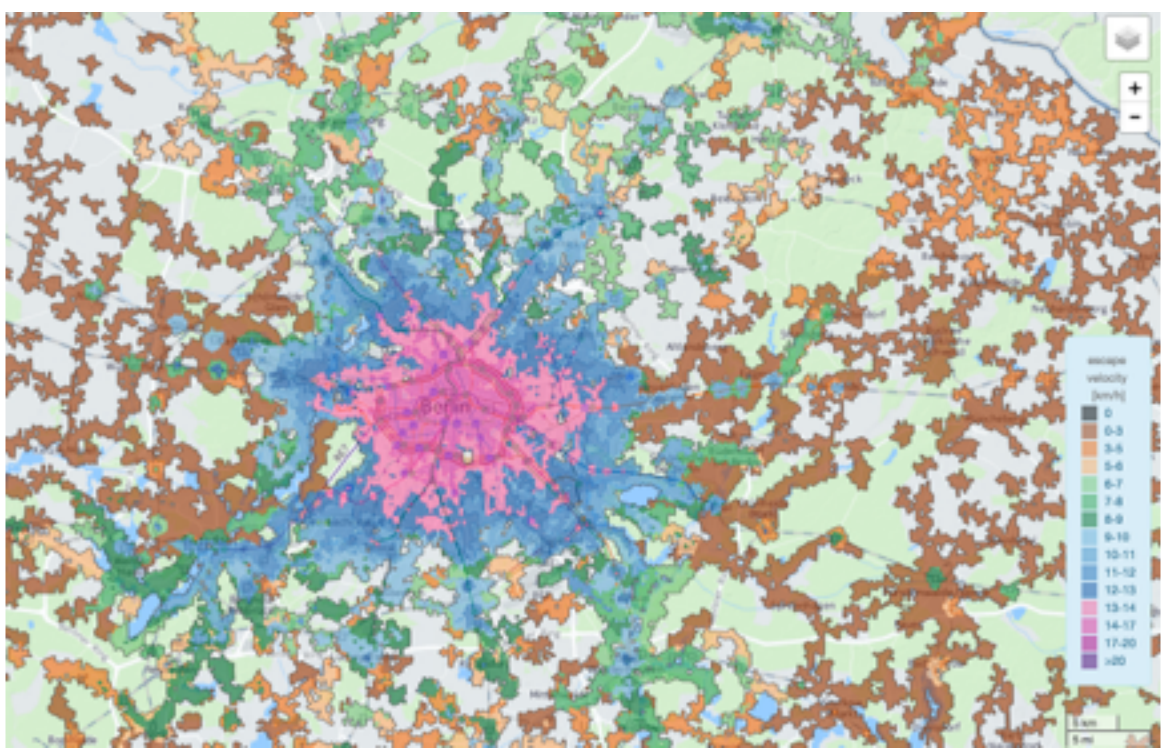


Rome

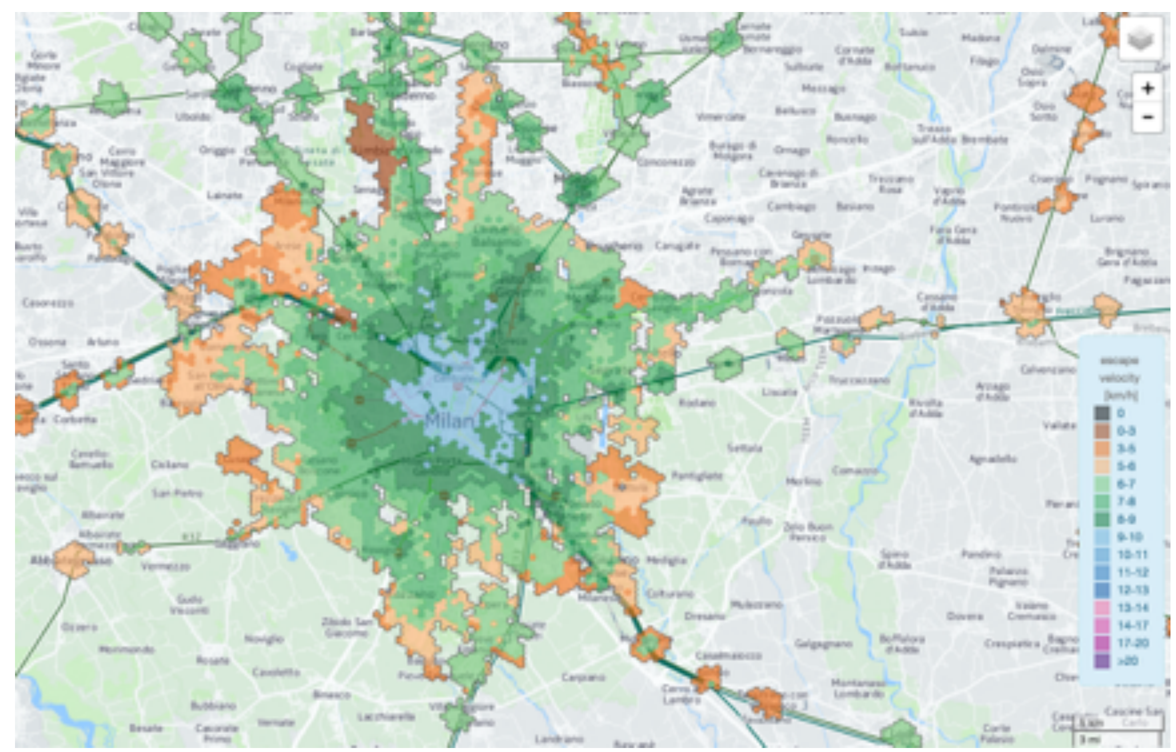


Paris



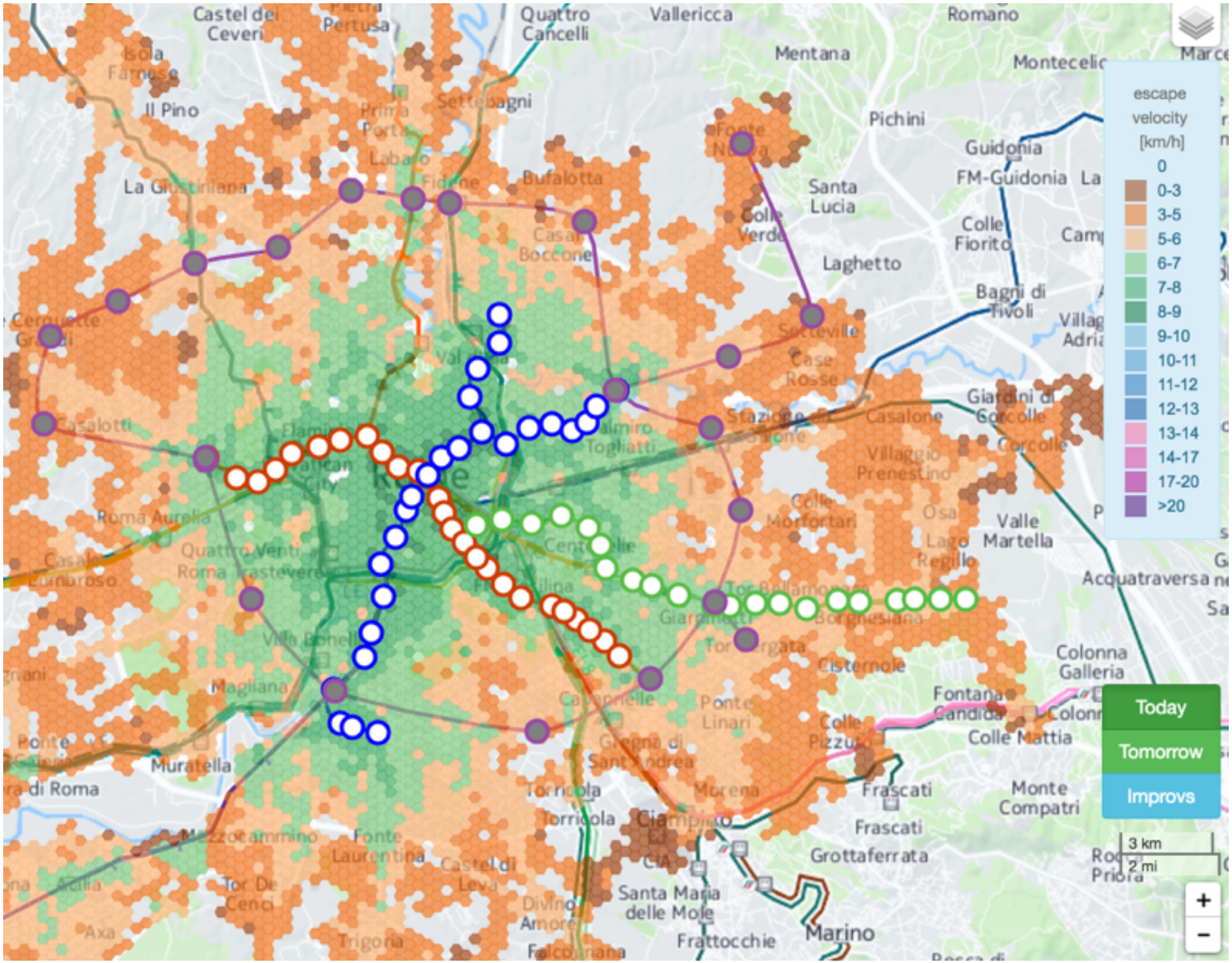


Berlin



Milan

Planning new lines



Big DATA
vs.
Small DATA

II

Making sense of all these data

Approaches

4.1 Decision tree learning
4.2 Association rule learning
4.3 Artificial neural networks
4.4 Deep learning
4.5 Inductive logic programming
4.6 Support vector machines
4.7 Clustering
4.8 Bayesian networks
4.9 Reinforcement learning
4.10 Representation learning
4.11 Similarity and metric learning
4.12 Sparse dictionary learning
4.13 Genetic algorithms
4.14 Rule-based machine learning
4.15 Learning classifier systems

Machine learning

BIG DATA

+

ENHANCED COMPUTATIONAL POWER

Credit Risk Analysis

Data:

Customer103: (time=t0)

Years of credit: 9
Loan balance: $2,400
Income: $52k
Own House: Yes
Other delinquent accts: 2
Max billing cycles late: 3
Profitable customer?: ?
...

Customer103: (time=t1)

Years of credit: 9
Loan balance: $3,250
Income: ?
Own House: Yes
Other delinquent accts: 2
Max billing cycles late: 4
Profitable customer?: ?
...

...

Customer103: (time=tn)

Years of credit: 9
Loan balance: $4,500
Income: ?
Own House: Yes
Other delinquent accts: 3
Max billing cycles late: 6
Profitable customer?: No
...

Rules learned from synthesized data:

```
If   Other-Delinquent-Accounts > 2, and
     Number-Delinquent-Billing-Cycles > 1
Then Profitable-Customer? = No
     [Deny Credit Card application]

If   Other-Delinquent-Accounts = 0, and
     (Income > $30k)  OR  (Years-of-Credit > 3)
Then Profitable-Customer? = Yes
     [Accept Credit Card application]
```

All that glitters ain't gold!

BIG DATA:

SEIZING OPPORTUNITIES, PRESERVING VALUES

Executive Office of the President

MAY 2014

THE WHITE HOUSE

WASHINGTON

“big data analytics have the potential to eclipse longstanding civil rights protections in how personal information is used in housing, credit, employment, health, education, and the marketplace.”

profiling

segmentation

Big Data Comes With the Biases of Its Creators

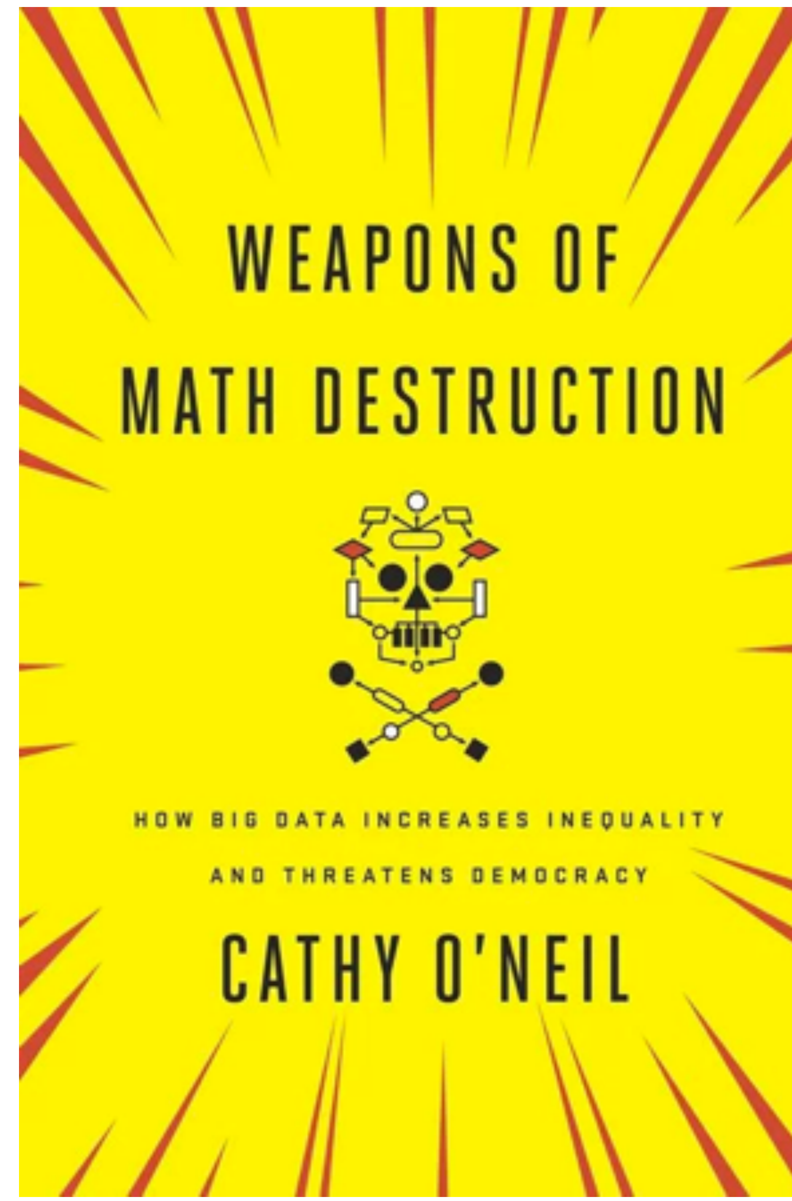
Rana Foroohar @RanaForoohar | Sept. 7, 2016

TIME

Algorithmic discrimination

Opacity

Filtering bubble



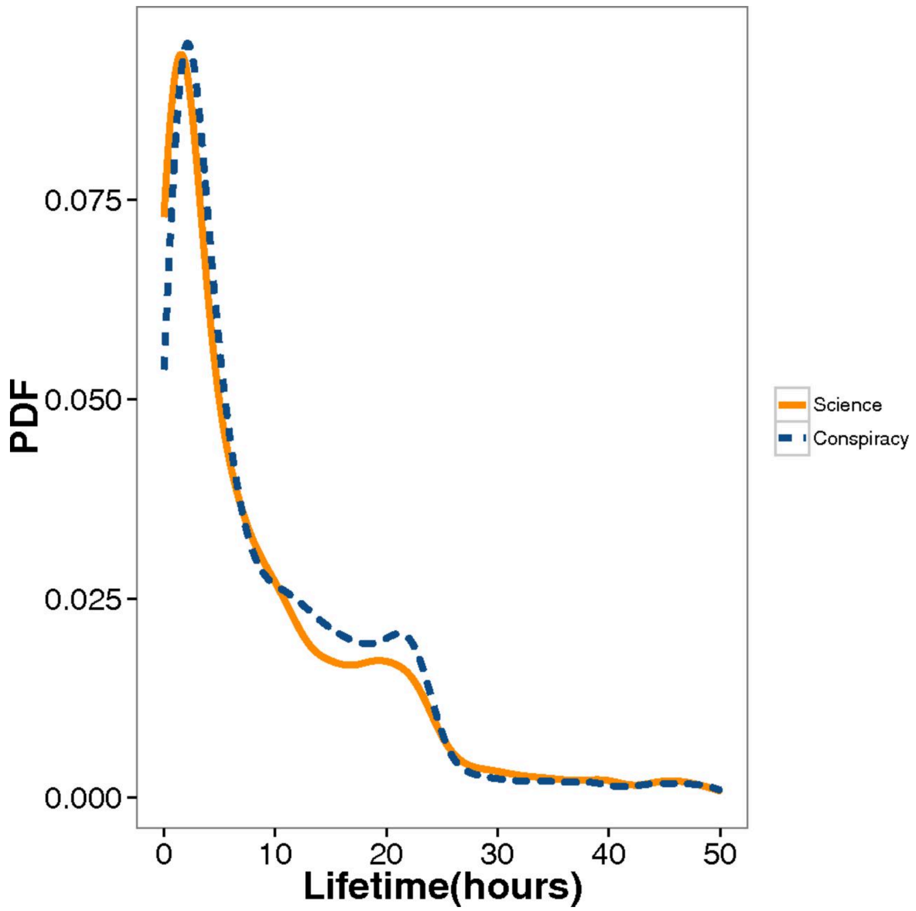
World Privacy Forum

The Scoring of America: How Secret Consumer Scores Threaten Your Privacy and Your Future

By Pam Dixon and Robert Gellman
April 2, 2014

Consumer scores

Consumer scores are built using predictive modelling. Predictive modelling uses copious amounts of information fed through analytical methods to predict the future, based on past information. Predictive consumer scores are important because they affect the lives, privacy, and wellbeing of individuals. Many people know about credit scores, but few know about the broader range of new consumer scores. Consumer scores are already abundant and are in active use. Consumer scores are not just an online phenomenon. Consumer scores are found in a wide array of “offline” arenas, including **businesses**, **health care providers**, **financial institutions**, **law enforcement**, **retail stores**, **federal and state government**, and many other locations. Some social consumer scores may have online applications, but mostly, consumer scores are not solely focused on just online activities. **And unlike credit scores, consumer scores remain largely secret and unregulated.**



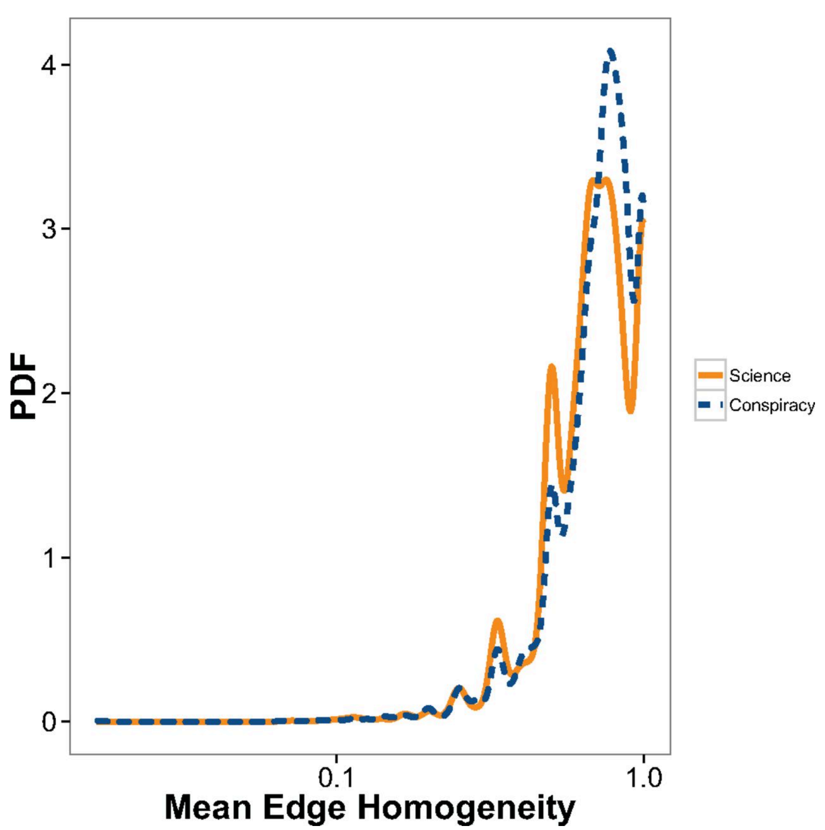
Michela Del Vicario et al. PNAS 2016;113:554-559

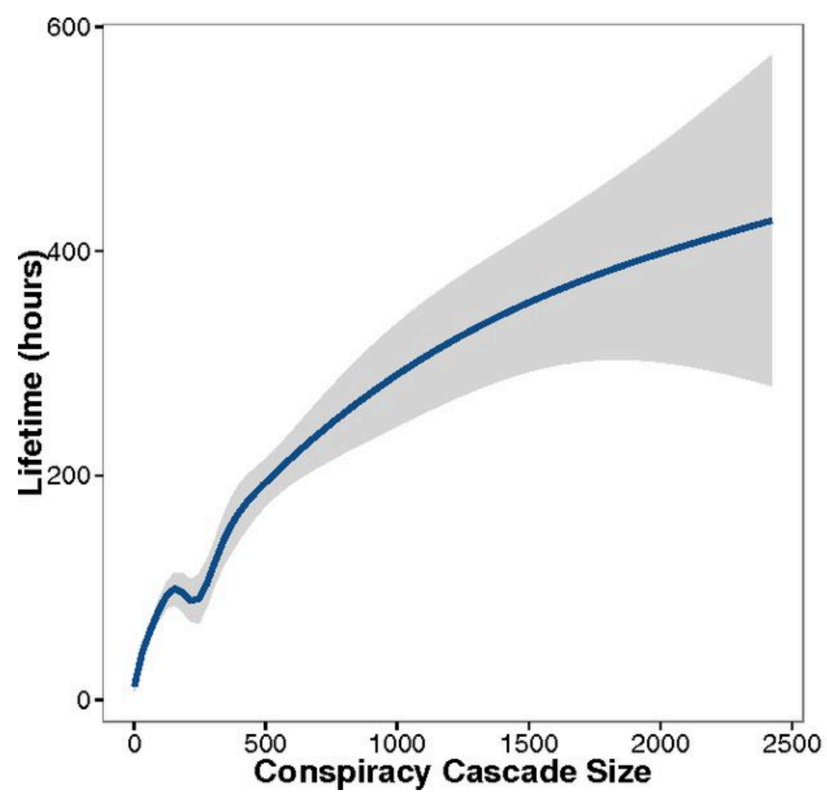
The spreading of misinformation online

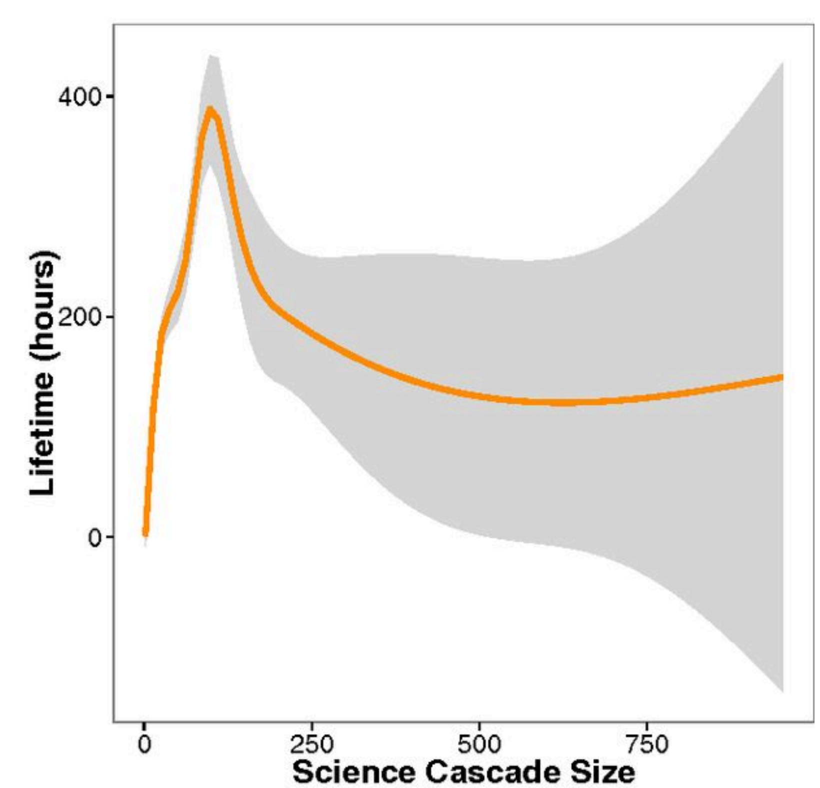
Michela Del Vicario[a], Alessandro Bessi[b], Fabiana Zollo[a], Fabio Petroni[c], Antonio Scala[a,d], Guido Caldarelli[a,d], H. Eugene Stanley[e], and Walter Quattrociocchi[a,1]

PNAS

"Contents tend to circulate only inside the echo chamber."








Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights

Executive Office of the President

May 2016

Challenge 1: Inputs to an Algorithm

Poorly selected data

Incomplete, incorrect, or outdated data,

Selection bias

Unintentional perpetuation and promotion of historical biases

Challenge 2: The Design of Algorithmic Systems and Machine Learning

Poorly designed matching systems

Personalization and recommendation services that narrow instead of expand user options

Decision-making systems that assume correlation necessarily implies causation

Data sets that lack information or disproportionately represent certain populations

PHILOSOPHICAL TRANSACTIONS
OF THE ROYAL SOCIETY A

MATHEMATICAL, PHYSICAL AND ENGINEERING SCIENCES

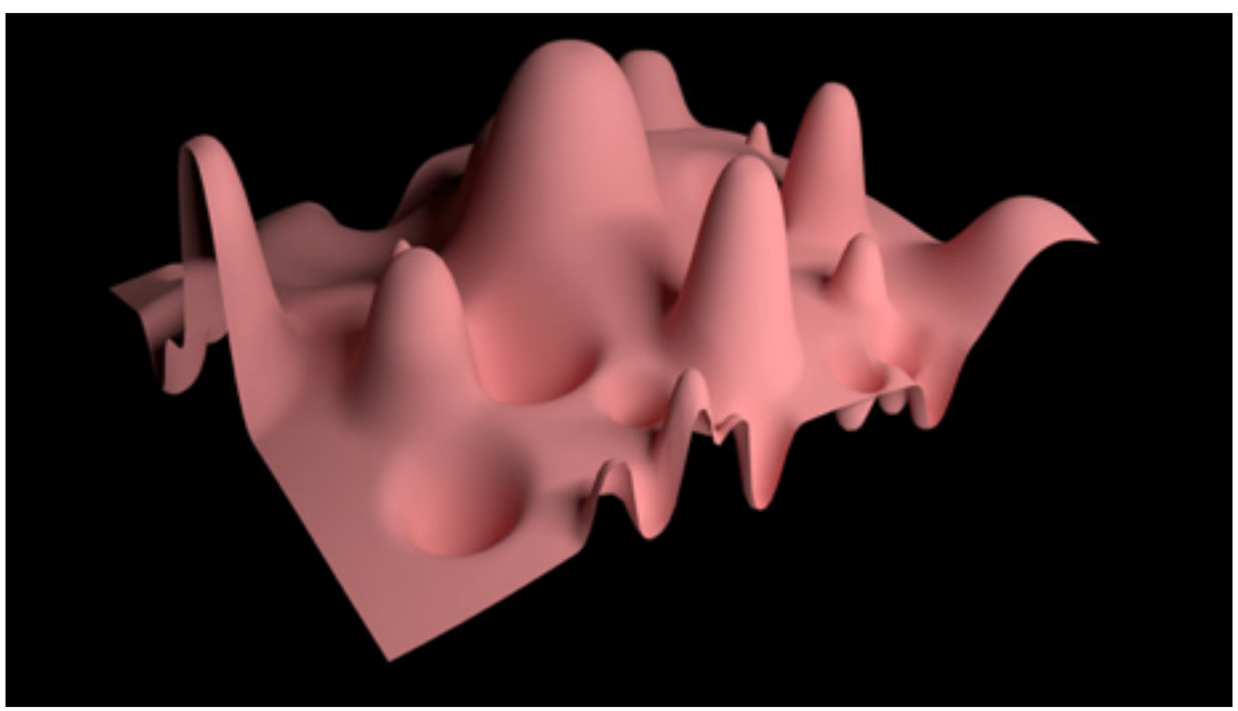
Home | Content | Information for | About us | Sign up | Propose an issue

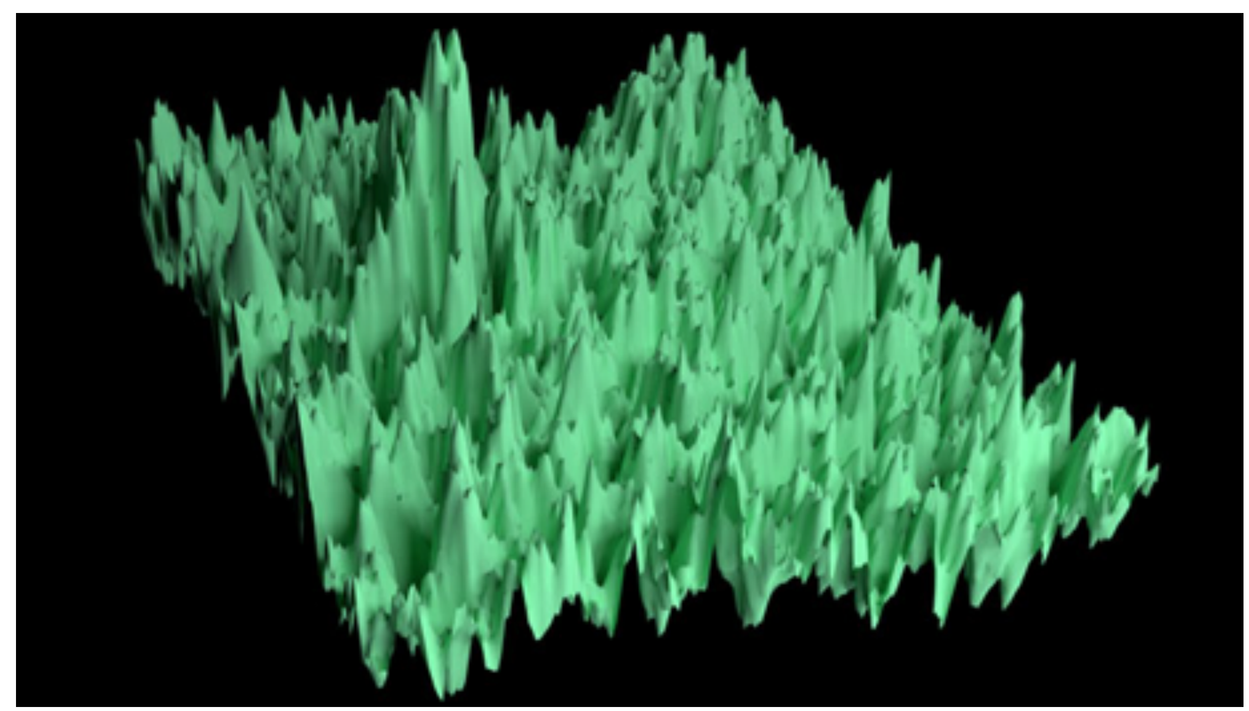
Big data need big theory too

Peter V. Coveney, Edward R. Dougherty, Roger R. Highfield

Published 3 October 2016. DOI: 10.1098/rsta.2016.0153

BIG DATA as Radical empiricism





New ICT-driven opportunities

Understand and predict complex phenomena

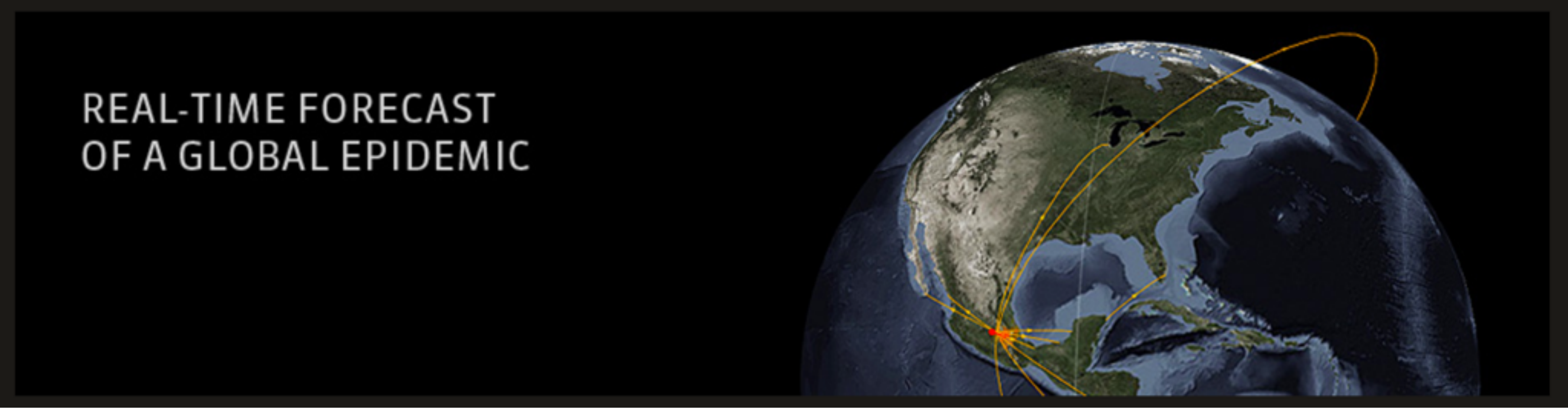
- mobility and urban dynamics
- information, culture, opinion dynamics
- epidemic spreading
- ...

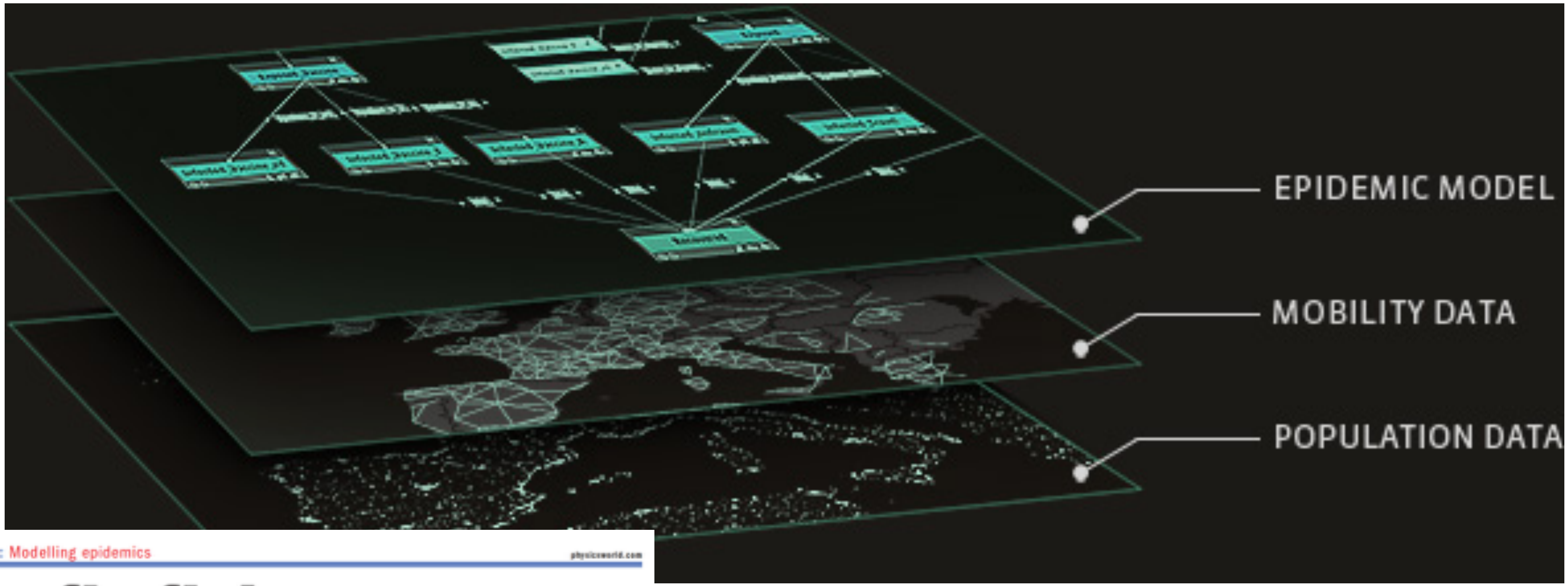
Games/experiments in social sciences

- opinions and norms formation
- consumers behaviours, marketing strategies
- cultural trends, globalisation
- dynamics of innovation
- language evolution
- ...

Learning, awareness and behavioural changes

- New learning paths
- Management of common resources and environment
- Feedback to policy makers
- Sustainable development
- ...





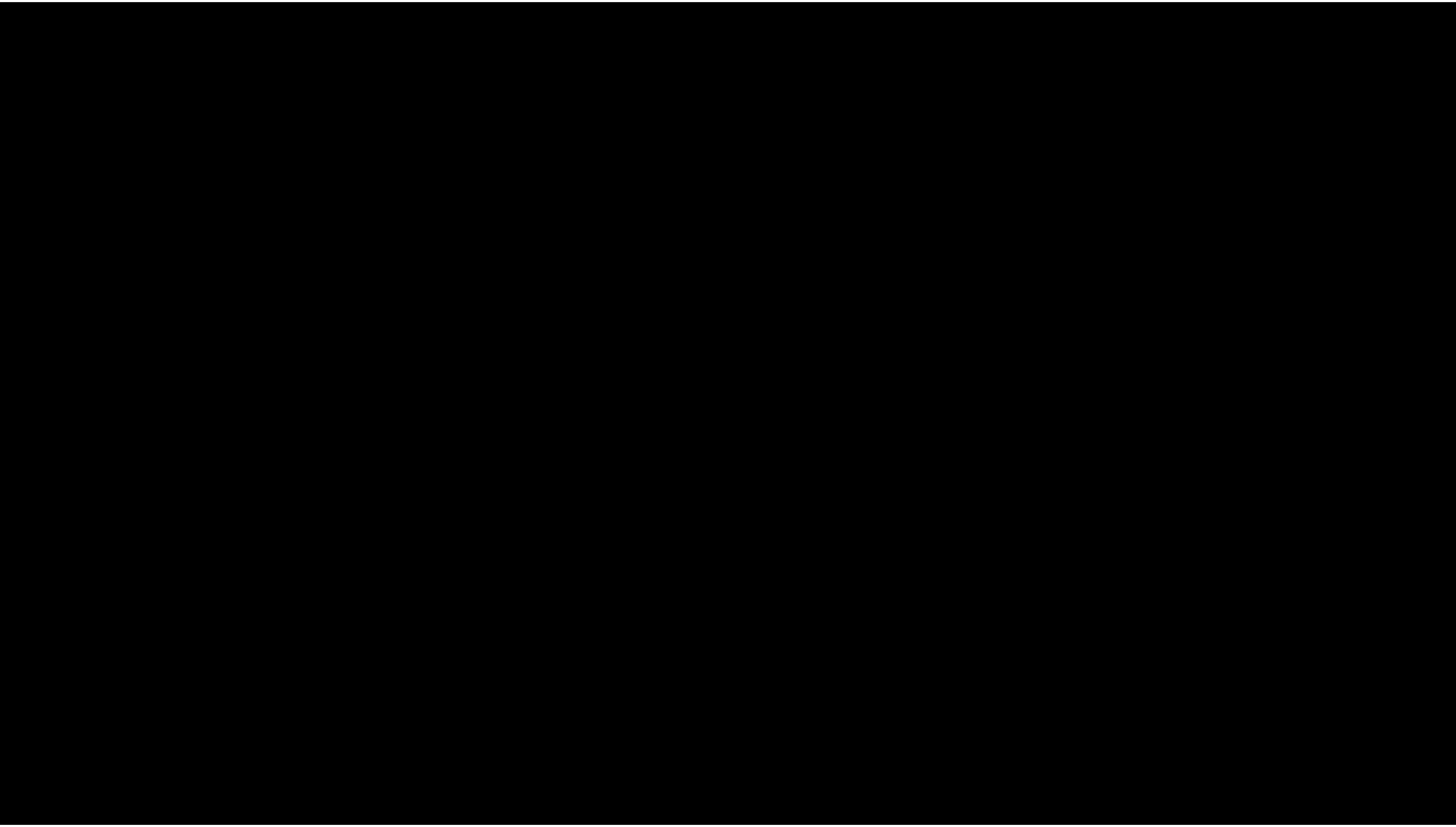
Complexity: Modelling epidemics

The flu fighters

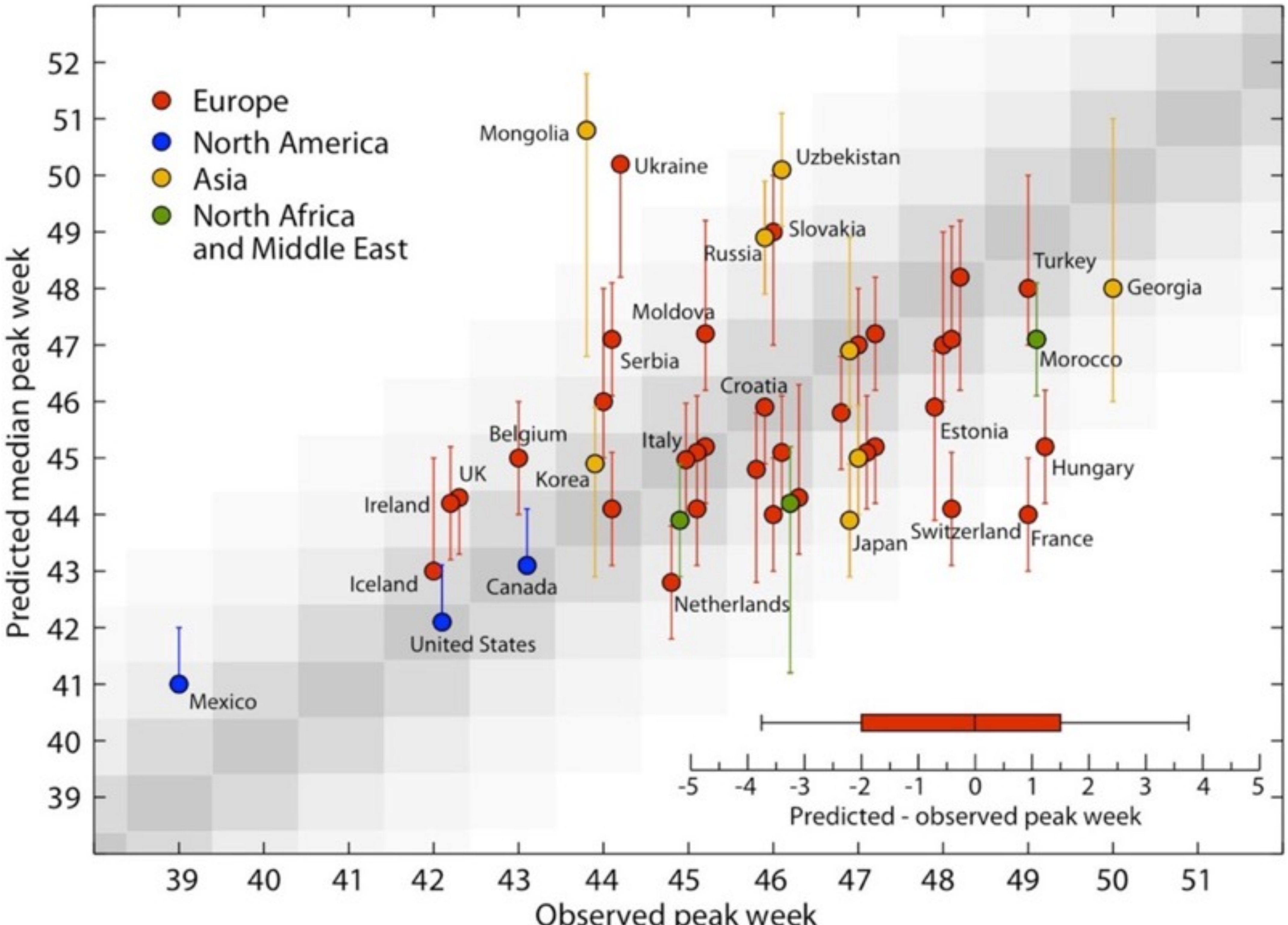
Quick and easy global air travel aids the spread of infectious diseases like the current H1N1 flu pandemic, and makes modelling them a complicated task. But as **Vittoria Colizza** and **Alessandro Vespignani** explain, physicists and computer scientists are at the forefront of such studies, using tools developed from fields such as statistical mechanics and complex networks

http://www.gleamviz.org/

www.gleamviz.org



validation



Balcan et al. BMC Med 2009
Tizzoni et al BMC Med 2012

Projected spread of Zika virus in the Americas

Qian Zhang, Kaiyuan Sun, Matteo Chinazzi, Ana Pastore-Piontti, Natalie E Dean, Diana P Rojas, Stefano Merler, Dina Mistry, Piero Poletti, Luca Rossi, Margaret Bray, M. Elizabeth Halloran, Ira M Longini, Alessandro Vespignani

doi: https://doi.org/10.1101/066456

Cold Spring Harbor Laboratory

bioRxiv beta

THE PREPRINT SERVER FOR BIOLOGY

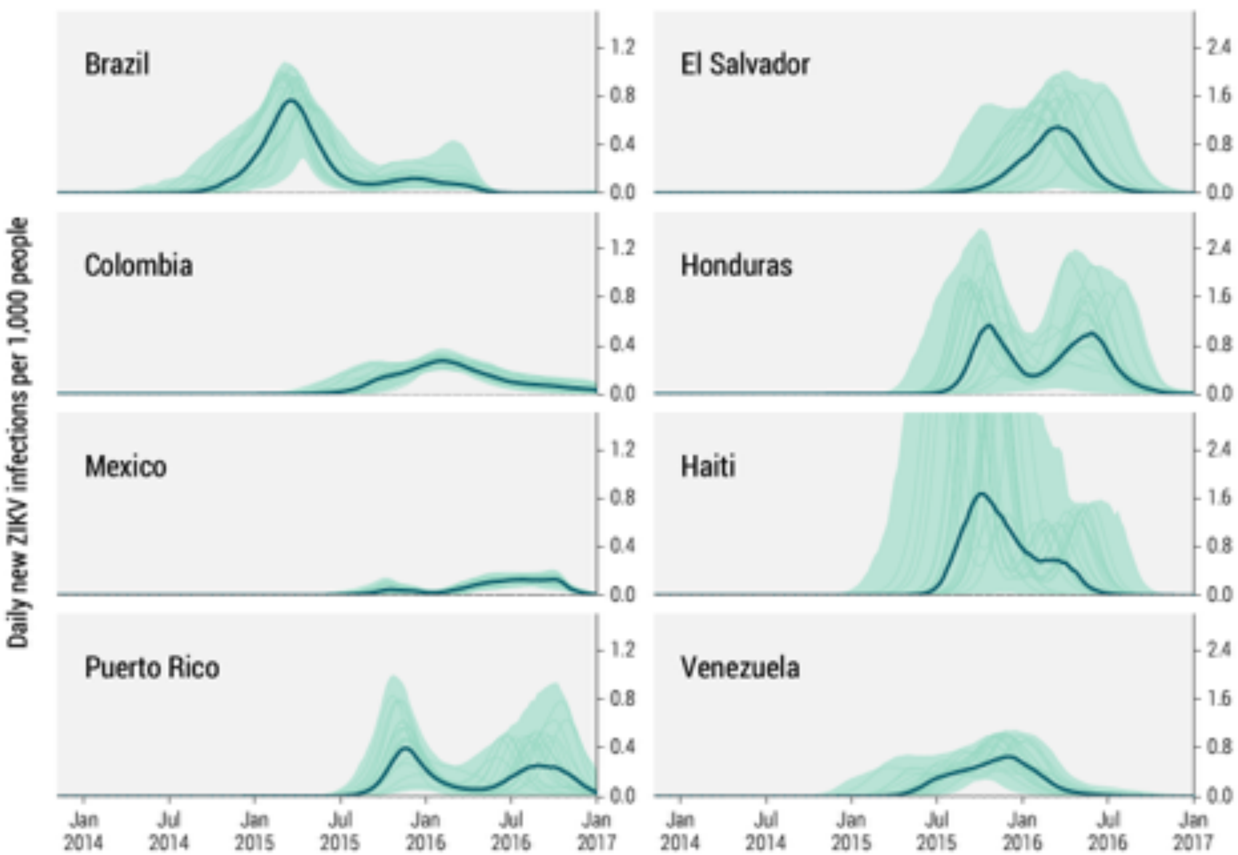


Fig. 2: Estimated daily number of new Zika infections (per 1000 people) in eight affected countries in the Americas between January 2014 and January 2017. The bold line and shaded area refer to the estimated median number of infections and 95%CI of the model projections, respectively. Rates include asymptomatic infections. The median incidence is calculated each week from the stochastic ensemble output of the model and may not be representative of specific epidemic realizations. Thin lines represent a sample of specific realizations.

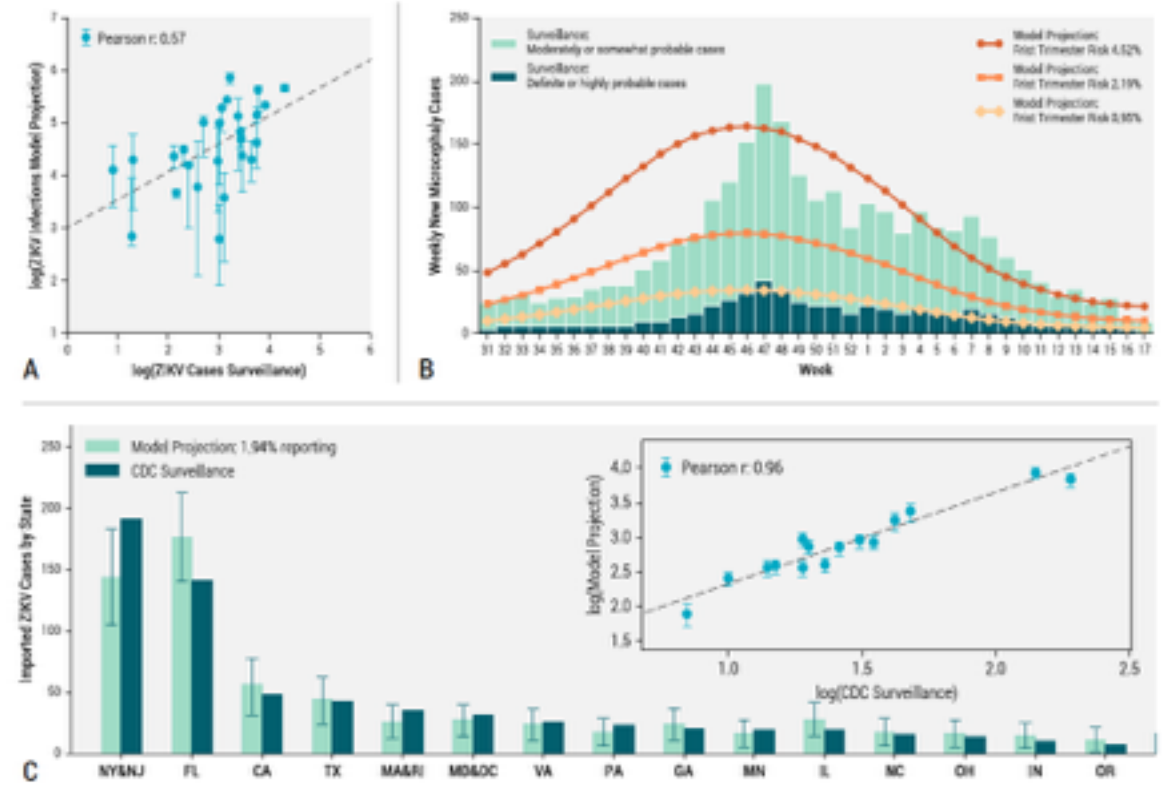
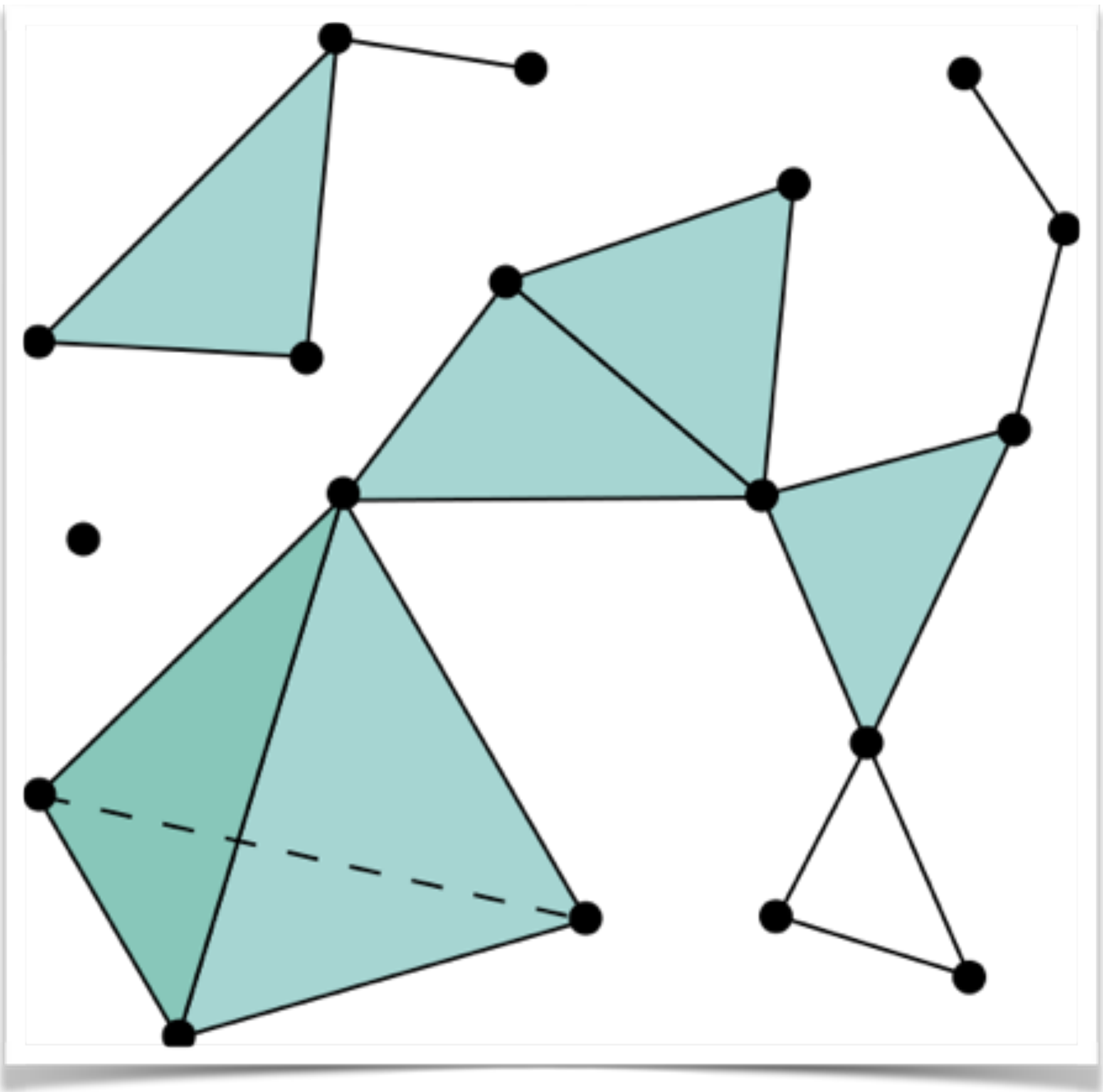


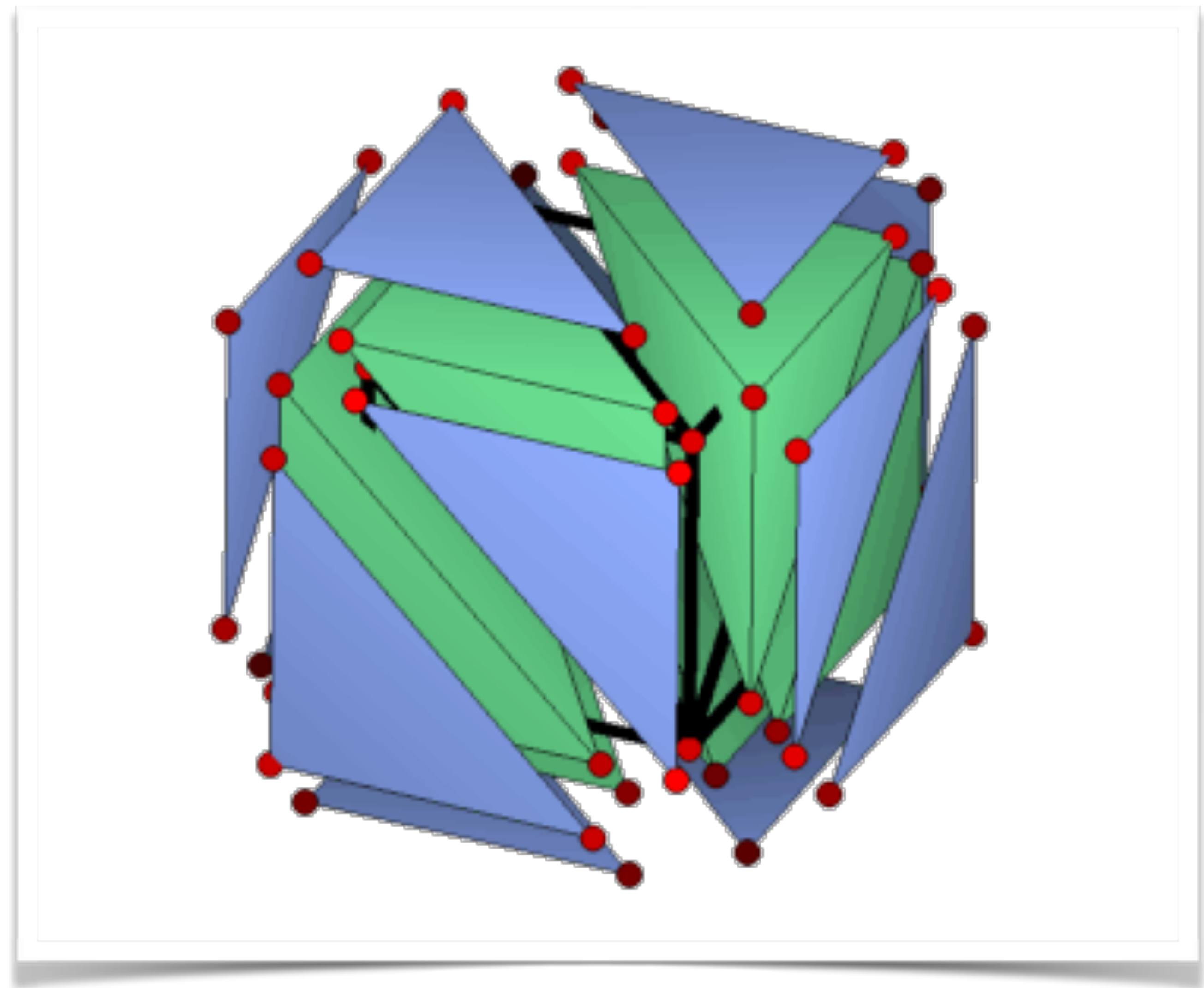
Fig. 5: A) Correlation between the number of ZIKV cases by states in Colombia as reported by surveillance data through June 5, 2016, compared with state-level model projections of infections (median with 95%CI). Pearson's r correlation coefficient is reported for the linear association on the log scale. B) Timeline of microcephaly cases in Brazil though April 30, 2016. Bar plot shows weekly definite (or highly probable cases) and moderately (or somewhat probable cases) from surveillance data[26]. Line plots indicate estimated weekly new microcephaly cases given three levels of first trimester risk: 4.52% (round) [25], 2.19% (square) [25], and 0.95% (diamond) [24]. C) Bar plot of ZIKV infections imported into the continental USA by state(s) as reported by CDC surveillance through June 15, 2016, and compared to model projections (median with 95% CI) for the same period assuming 1.94% reporting/detection. The insert shows the correlation between CDC surveillance data and model projections (median with 95%CI).

Topological approach to big data

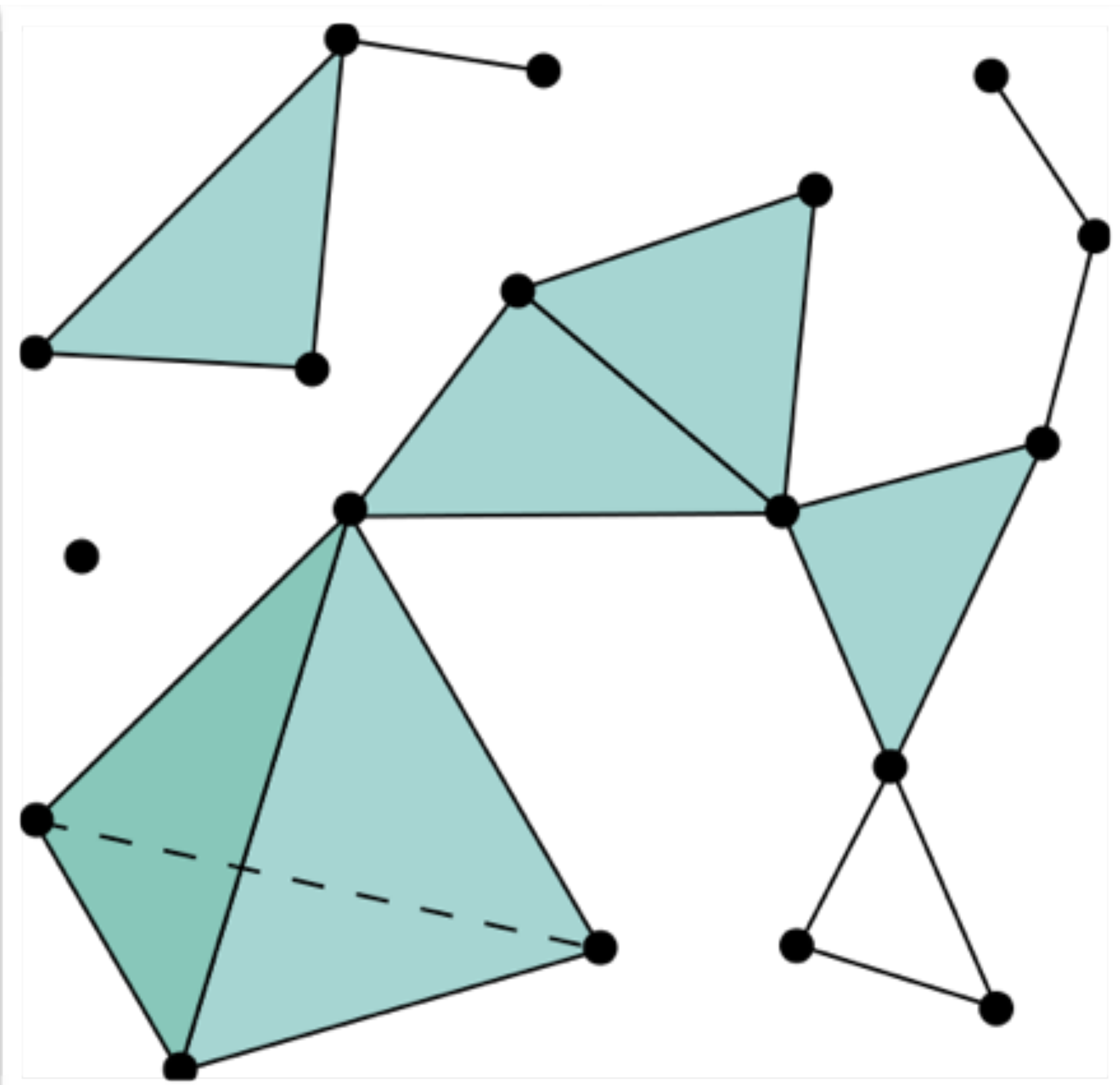
Topology
It's the branch of math that can't tell the difference between a coffee cup and a donut

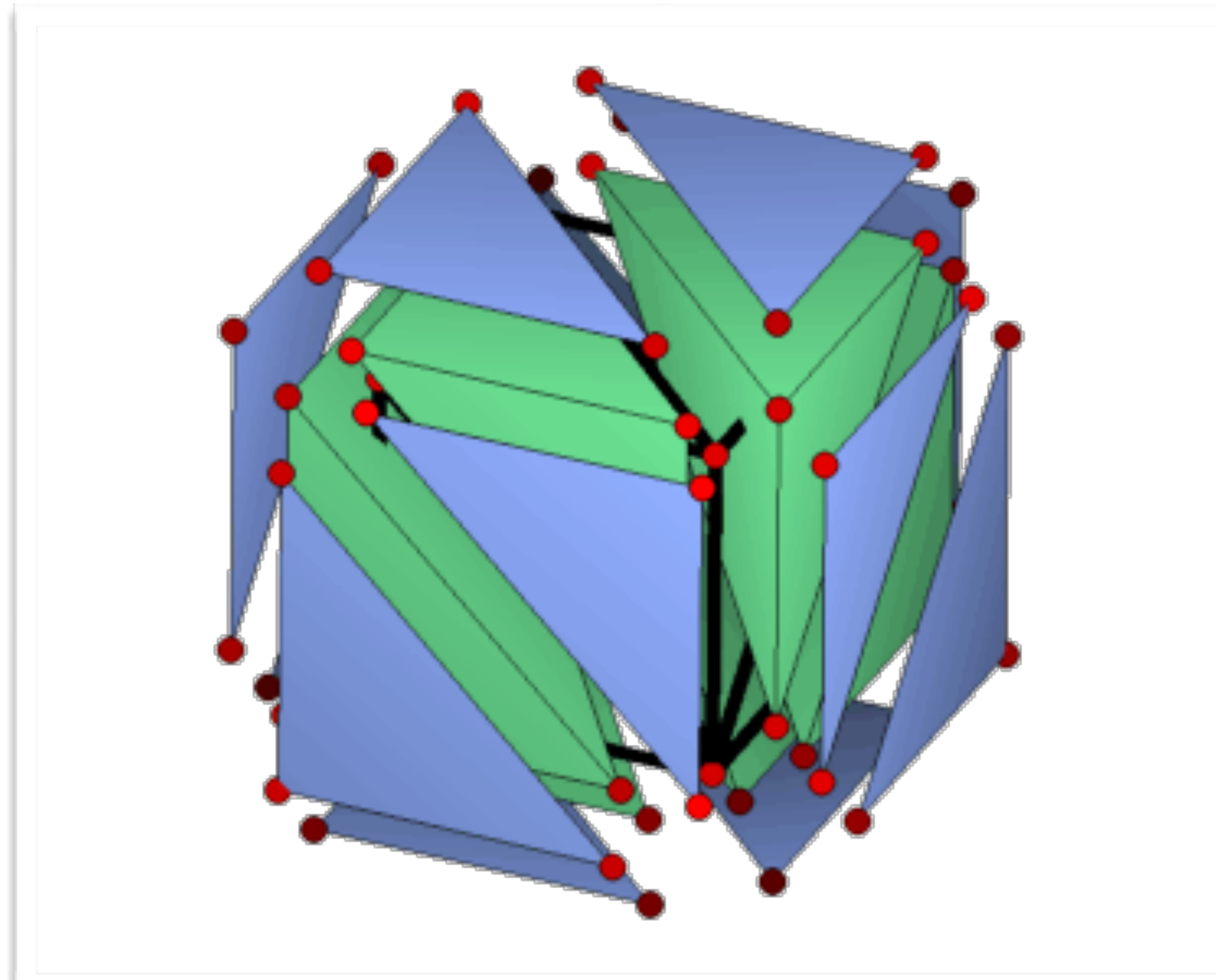
Simplicial Complex





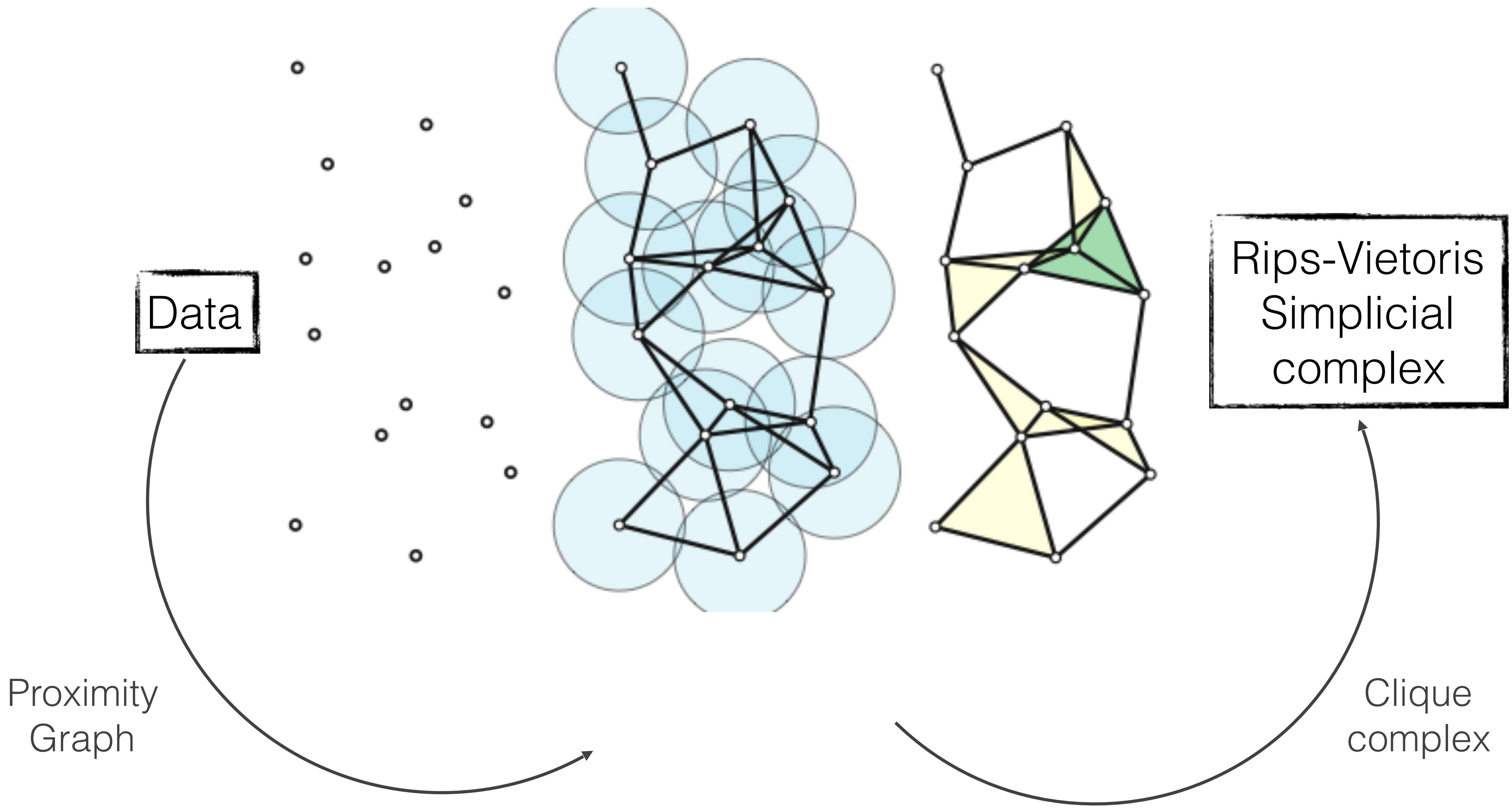
Simplicial Complex



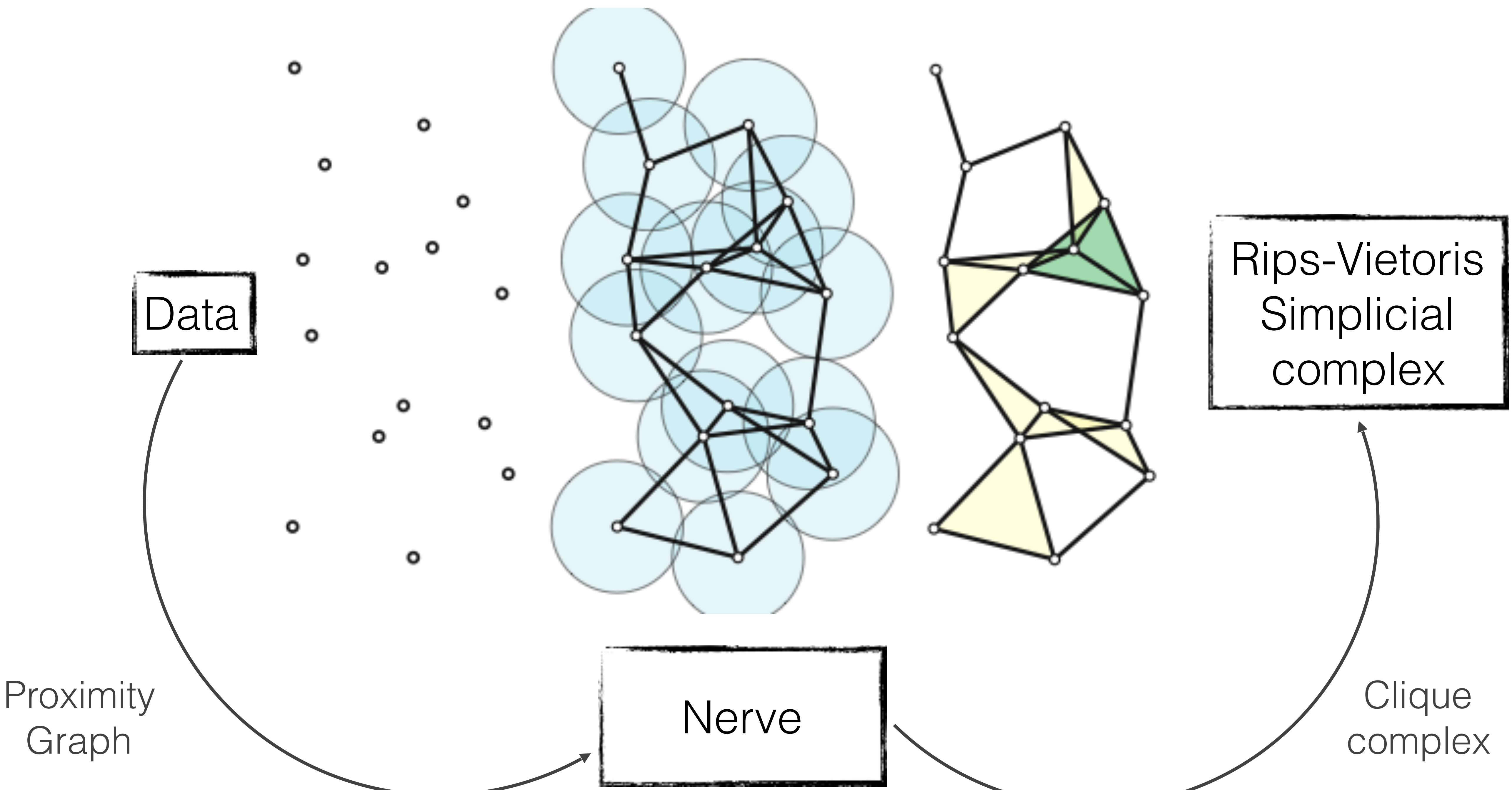


An abstract simplicial complex on a set of vertices V
is collection S of subsets of V such that,
if A belongs to V and B is a subset of A, then also B is in S

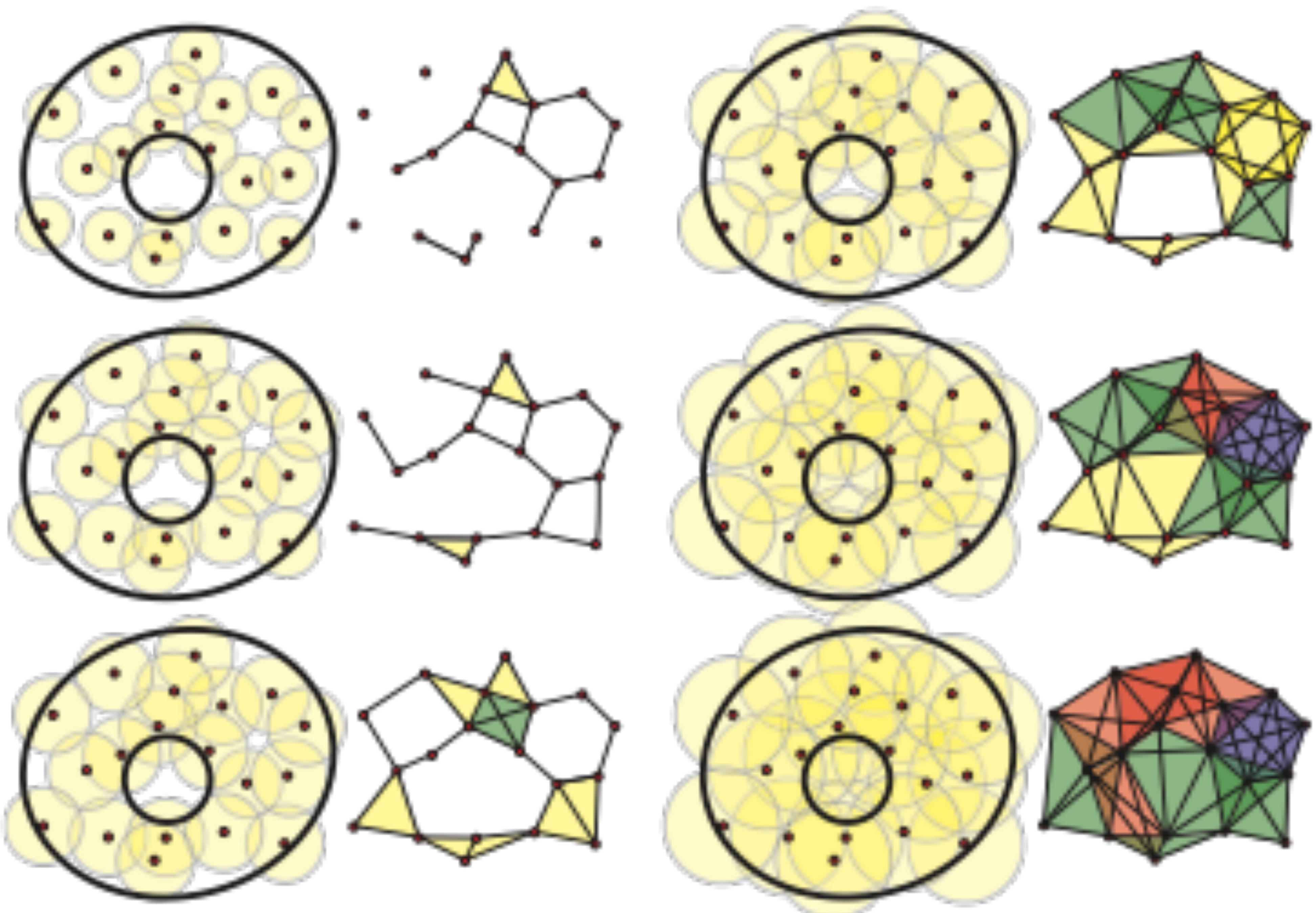
Data to Complex: metric case



Data to Complex: metric case



Across Scales



Ghrist, R. (2008). Barcodes: The persistent topology of data. *Bulletin-American Mathematical Society*, *45*(1), 61.

fMRI of altered states

fMRI of altered states

JOURNAL OF THE ROYAL SOCIETY Interface

rsif.royalsocietypublishing.org

Research

Cite this article: Petri G, Expert P, Turkheimer F, Carhart-Harris R, Nutt D, Hellyer PJ, Vaccarino F. 2014 Homological scaffolds of brain functional networks. *J. R. Soc. Interface* **11**: 20140873.
http://dx.doi.org/10.1098/rsif.2014.0873

Received: 5 August 2014
Accepted: 3 October 2014

Subject Areas:
computational biology

Keywords:
brain functional networks, fMRI, persistent homology, psilocybin

Author for correspondence:
P. Expert
e-mail: paul.expert@kcl.ac.uk

Homological scaffolds of brain functional networks

G. Petri[1], P. Expert[2], F. Turkheimer[2], R. Carhart-Harris[3], D. Nutt[3], P. J. Hellyer[4] and F. Vaccarino[1,5]

[1]ISI Foundation, Via Alassio 11/c, 10126 Torino, Italy
[2]Centre for Neuroimaging Sciences, Institute of Psychiatry, Kings College London, De Crespigny Park, London SE5 8AF, UK
[3]Centre for Neuropsychopharmacology, Imperial College London, London W12 0NN, UK
[4]Computational, Cognitive and Clinical Neuroimaging Laboratory, Division of Brain Sciences, Imperial College London, London W12 0NN, UK
[5]Dipartimento di Scienze Matematiche, Politecnico di Torino, C.so Duca degli Abruzzi no 24, Torino 10129, Italy

Networks, as efficient representations of complex systems, have appealed to scientists for a long time and now permeate many areas of science, including neuroimaging (Bullmore and Sporns 2009 *Nat. Rev. Neurosci.* **10**, 186–198. (doi:10.1038/nrn2618)). Traditionally, the structure of complex networks has been studied through their statistical properties and metrics concerned with node and link properties, e.g. degree-distribution, node centrality and modularity. Here, we study the characteristics of functional brain networks at the mesoscopic level from a novel perspective that highlights the role of inhomogeneities in the fabric of functional connections. This can be done by focusing on the features of a set of topological objects—*homological cycles*—associated with the weighted functional network. We leverage the detected topological information to define the *homological scaffolds*, a new set of objects designed to represent compactly the homological features of the correlation network and simultaneously make their homological properties amenable to networks theoretical methods. As a proof of principle, we apply these tools to compare resting-state functional brain activity in 15 healthy volunteers after intravenous infusion of placebo and psilocybin—the main psychoactive component of magic mushrooms. The results show that the homological structure of the brain's functional patterns undergoes a dramatic change post-psilocybin, characterized by the appearance of many transient structures of low stability and of a small number of persistent ones that are not observed in the case of placebo.

1. Motivation

The understanding of global brain organization and its large-scale integration remains a challenge for modern neurosciences. Network theory is an elegant framework to approach these questions, thanks to its simplicity and versatility [1]. Indeed,

fMRI of altered states

JOURNAL OF THE ROYAL SOCIETY Interface

rsif.royalsocietypublishing.org

Research

Cite this article: Petri G, Expert P, Turkheimer F, Carhart-Harris R, Nutt D, Hellyer PJ, Vaccarino F. 2014 Homological scaffolds of brain functional networks. *J. R. Soc. Interface* **11**: 20140873.
http://dx.doi.org/10.1098/rsif.2014.0873

Received: 5 August 2014
Accepted: 3 October 2014

Subject Areas:
computational biology

Keywords:
brain functional networks, fMRI, persistent homology, psilocybin

Author for correspondence:
P. Expert
e-mail: paul.expert@kcl.ac.uk

Homological scaffolds of brain functional networks

G. Petri[1], P. Expert[2], F. Turkheimer[2], R. Carhart-Harris[3], D. Nutt[3], P. J. Hellyer[4] and F. Vaccarino[1,5]

[1]ISI Foundation, Via Alassio 11/c, 10126 Torino, Italy
[2]Centre for Neuroimaging Sciences, Institute of Psychiatry, Kings College London, De Crespigny Park, London SE5 8AF, UK
[3]Centre for Neuropsychopharmacology, Imperial College London, London W12 0NN, UK
[4]Computational, Cognitive and Clinical Neuroimaging Laboratory, Division of Brain Sciences, Imperial College London, London W12 0NN, UK
[5]Dipartimento di Scienze Matematiche, Politecnico di Torino, C.so Duca degli Abruzzi no 24, Torino 10129, Italy

Networks, as efficient representations of complex systems, have appealed to scientists for a long time and now permeate many areas of science, including neuroimaging (Bullmore and Sporns 2009 *Nat. Rev. Neurosci.* **10**, 186–198. (doi:10.1038/nrn2618)). Traditionally, the structure of complex networks has been studied through their statistical properties and metrics concerned with node and link properties, e.g. degree-distribution, node centrality and modularity. Here, we study the characteristics of functional brain networks at the mesoscopic level from a novel perspective that highlights the role of inhomogeneities in the fabric of functional connections. This can be done by focusing on the features of a set of topological objects—*homological cycles*—associated with the weighted functional network. We leverage the detected topological information to define the *homological scaffolds*, a new set of objects designed to represent compactly the homological features of the correlation network and simultaneously make their homological properties amenable to networks theoretical methods. As a proof of principle, we apply these tools to compare resting-state functional brain activity in 15 healthy volunteers after intravenous infusion of placebo and psilocybin—the main psychoactive component of magic mushrooms. The results show that the homological structure of the brain's functional patterns undergoes a dramatic change post-psilocybin, characterized by the appearance of many transient structures of low stability and of a small number of persistent ones that are not observed in the case of placebo.

1. Motivation

The understanding of global brain organization and its large-scale integration remains a challenge for modern neurosciences. Network theory is an elegant framework to approach these questions, thanks to its simplicity and versatility [1]. Indeed, in recent years, networks have become a prominent tool to analyse and understand

frontiers

Insights into brain architectures from the homological scaffolds of functional connectivity networks

Lord, L.D. [1], **Expert, P.** [2], **Fernandes, H.M.** [1,3] **Petri, G.** [4], **Van Hartevelt, T.J.** [1,3], **Vaccarino, F.** [5], **Deco, G.** [6,7], **Turkheimer, F.E.** [2], **Kringelbach, M.L.** [1,3]

[1] *Department of Psychiatry, University of Oxford, United Kingdom*
[2] *Institute of Psychiatry, Kings College London, United Kingdom*
[3] *Center of Functionally Integrative Neuroscience, Aarhus University, Denmark*
[4] *Institute for Scientific Interchange (ISI Foundation), Torino, Italy*
[5] *Department of Mathematical Sciences, Politecnico di Torino, Torino, Italy*
[6] *Center for Brain and Cognition, Universitat Pompeu Fabra, Barcelona, Spain*
[7] *Instituci Catalana de la Recerca i Estudis Avanats (ICREA), Universitat Pompeu Fabra, Spain*

Correspondence*:
Prof. Morten L Kringelbach
Hedonia Research Group
Department of Psychiatry, University of Oxford,
morten.kringelbach@queens.ox.ac.uk

ABSTRACT

In recent years, the application of network analysis to neuroimaging data has provided useful insights about the brain's functional and structural organization in both health and disease. This has proven a significant paradigm shift from the study of individual brain regions in isolation. Graph-based models of the brain consist of vertices, which represent distinct brain areas, and edges which encode the presence (or absence) of a structural or functional relationship between each pair of vertices. By definition, any graph metric will be defined upon this dyadic representation of the brain activity. It is however unclear to what extent these dyadic relationships can capture the brain's complex functional architecture and the encoding of information in distributed networks. Moreover, because network representations of global brain activity are derived from measures that have a continuous response (i.e. interregional BOLD signals), it is

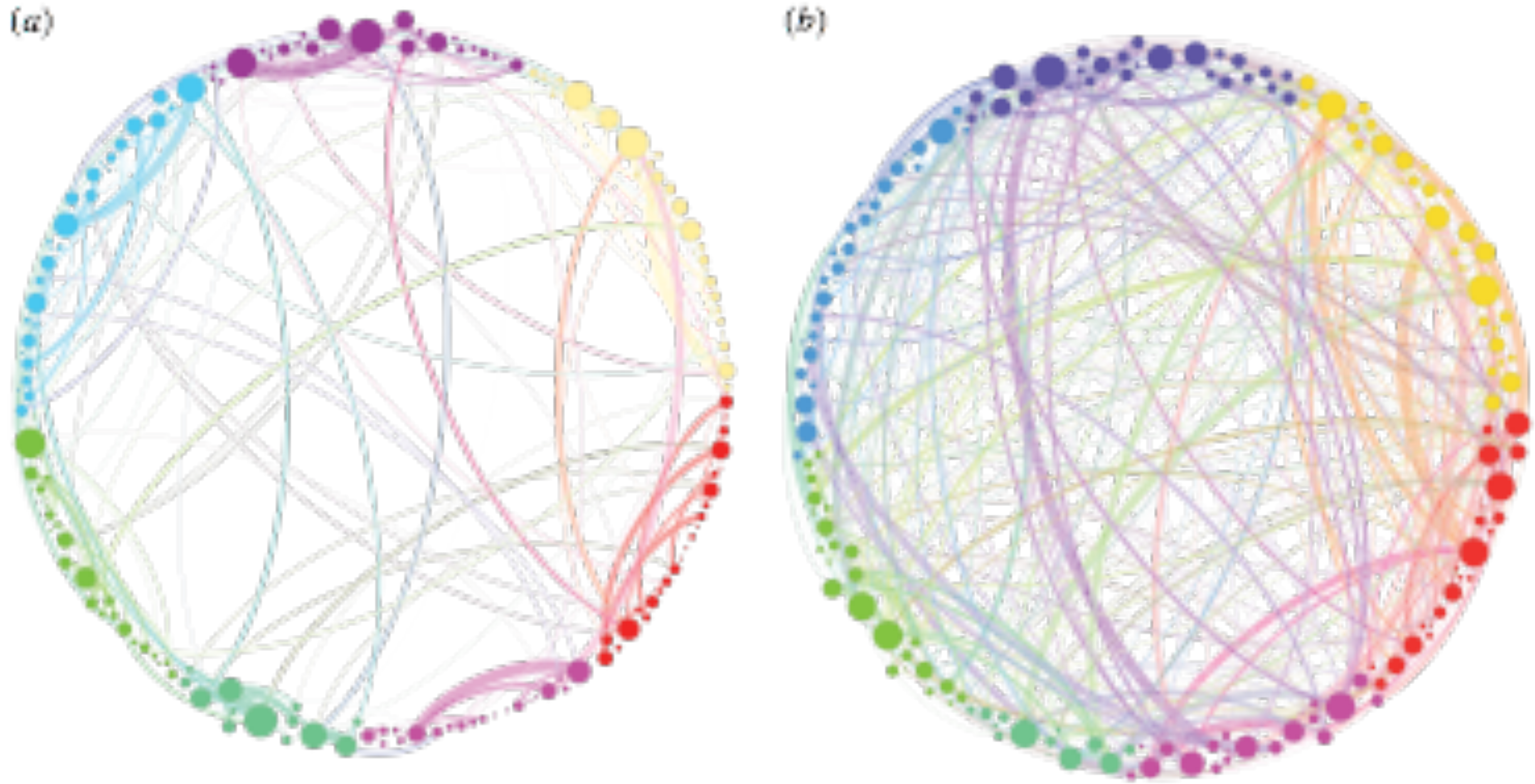
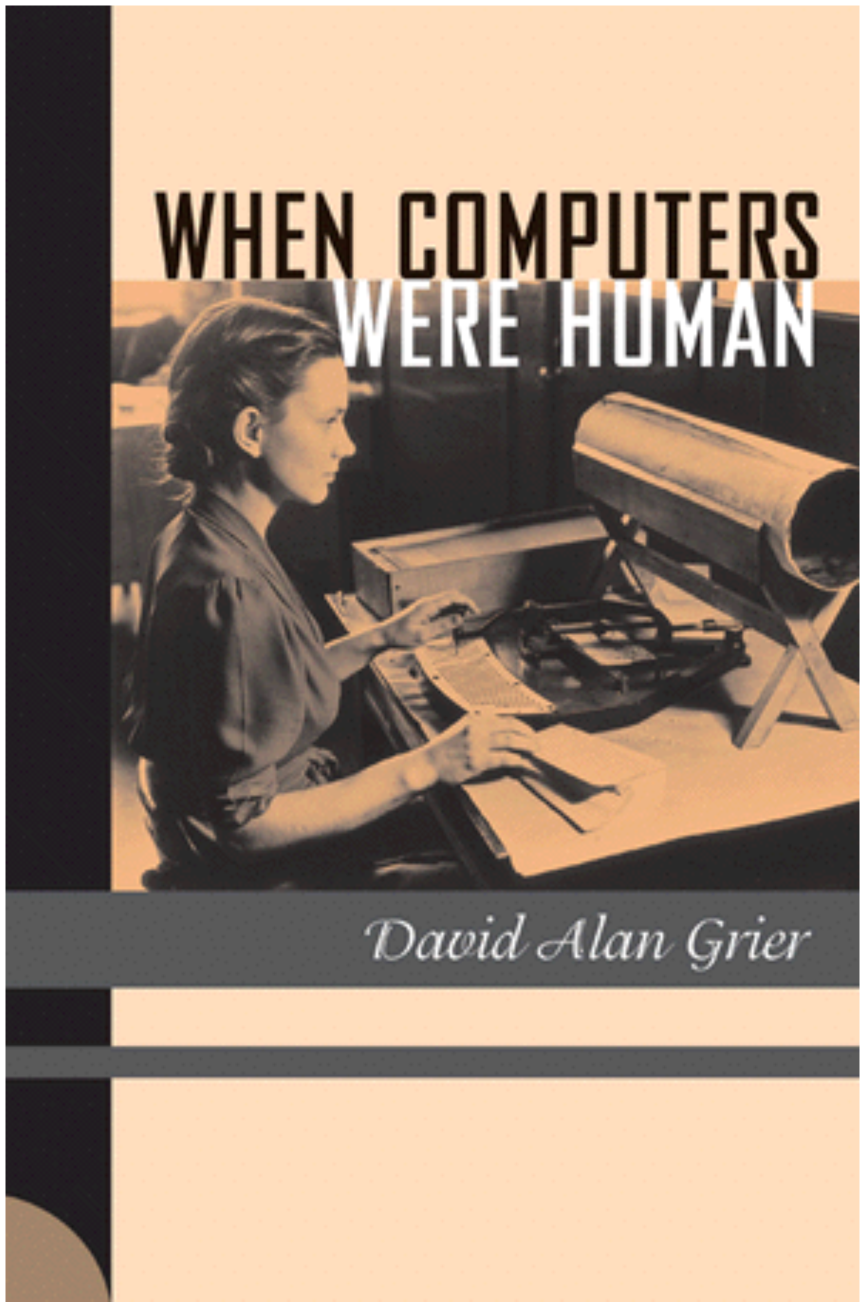


Figure 6. Simplified visualization of the persistence homological scaffolds. The persistence homological scaffolds $\mathscr{H}^{p}_{\mathrm{pla}}$ (*a*) and $\mathscr{H}^{p}_{\mathrm{psi}}$ (*b*) are shown for comparison. For ease of visualization, only the links heavier than 80 (the weight at which the distributions in figure 5*a* bifurcate) are shown. This value is slightly smaller than the bifurcation point of the weights distributions in figure 5*a*. In both networks, colours represent communities obtained by modularity [49] optimization on the placebo persistence scaffold using the Louvain method [50] and are used to show the departure of the psilocybin connectivity structure from the placebo baseline. The width of the links is proportional to their weight and the size of the nodes is proportional to their strength. Note that the proportion of heavy links between communities is much higher (and very different) in the psilocybin group, suggesting greater integration. A labelled version of the two scaffolds is available as GEXF graph files as the electronic supplementary material. (Online version in colour.)

Social computation

Populations of users facing collectively difficult problems using a small cognitive overhead

The human computer



"Before the computers were machines they were persons"

D.A. Grier

easy for humans - hard for bots

CAPTCHA

Completely **A**utomated **P**ublic **T**uring test to tell **C**omputers and **H**umans **A**part

A. Broder (1997), L. von Ahn e M. Blum (2000)

easy for humans - hard for bots

CAPTCHA

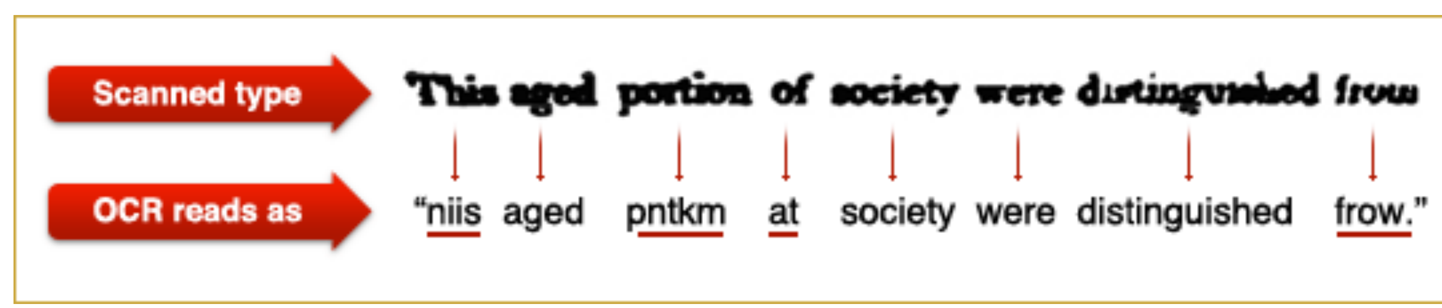
Completely **A**utomated **P**ublic **T**uring test to tell **C**omputers and **H**umans **A**part

A. Broder (1997), L. von Ahn e M. Blum (2000)

easy for humans - hard for bots

reCAPTCHA

L. von Ahn (2006)



Every time our CAPTCHAs are solved, that human effort helps digitize text, annotate images, and build machine learning datasets. This in turn helps preserve books, improve maps, and solve hard AI problems.

CAPTCHA

Completely **A**utomated **P**ublic **T**uring test to t… …d **H**umans **A**part

A. Broder (1997), L. von Ahn e M. Blum (…

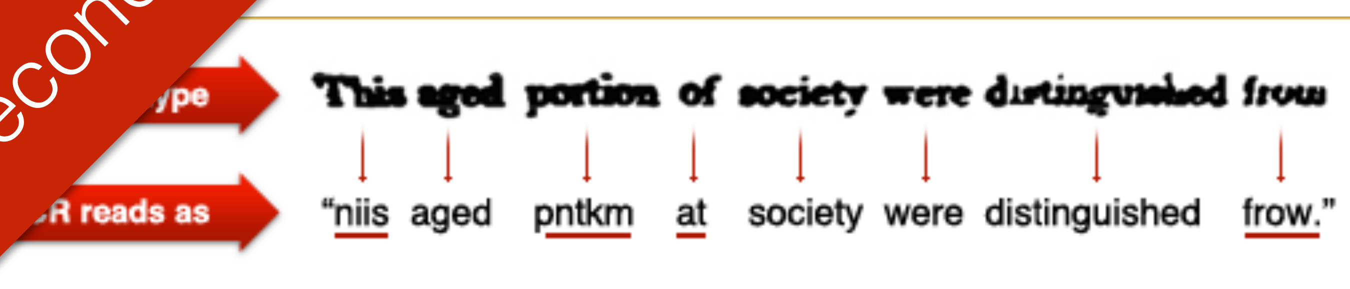
easy for hum… …d for bots

more than 200.000.000 every day !!

9 seconds each !

reCAPTC…

L. von Ahn (2006)



Every time our CAPTCHAs are solved, that human effort helps digitize text, annotate images, and build machine learning datasets. This in turn helps preserve books, improve maps, and solve hard AI problems.


The ESP Game
The images shown during the game may be subject to copyright.
The ESP Game
Ready!
Click to start the game
© 2005 Carnegie Mellon University, all rights reserved. Patent Pending.


The ESP Game
The images shown during the game
may be subject to copyright.
The ESP Game
Ready!
Click to start the game
© 2005 Carnegie Mellon University, all rights reserved. Patent Pending.

Citizen Science

...individual volunteers or networks of volunteers, many of whom may have no specific scientific training, perform or manage research-related tasks such as observation, measurement or computation.

Games and experiments

Gioco di Shannon

Gioco di Shannon

“Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

Gioco di Shannon

"Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

p

Gioco di Shannon

"Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

po

Gioco di Shannon

“Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

pot

Gioco di Shannon

"Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

pote

Gioco di Shannon

“Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

potes

Gioco di Shannon

"Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

potess

Gioco di Shannon

“Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

potesse

Gioco di Shannon

“Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

potesse_

Gioco di Shannon

“Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

potesse_e

Gioco di Shannon

“Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

potesse_es

Gioco di Shannon

“Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

potesse_ess

Gioco di Shannon

“Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

potesse_esse

Gioco di Shannon

“Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

potesse_esser

Gioco di Shannon

“Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

potesse_essere

Gioco di Shannon

“Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

potesse_essere_

Gioco di Shannon

“Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

potesse_essere_l

Gioco di Shannon

“Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

potesse_essere_ lo

Gioco di Shannon

“Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

potesse_essere_lo_

Gioco di Shannon

"Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

potesse_essere_lo_s

Gioco di Shannon

“Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

potesse_essere_lo_sp

Gioco di Shannon

“Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

potesse_essere_lo_spa

Gioco di Shannon

“Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

potesse_essere_lo_spaz

Gioco di Shannon

"Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

potesse_essere_ lo_spazi

Gioco di Shannon

“Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

potesse_essere_lo_spazio

Gioco di Shannon

“Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

potesse_essere_lo_spazio,

Gioco di Shannon

"Si capisce che si stava tutti lì, - fece il vecchio Qfwfq, - e dove altrimenti? Che ci...

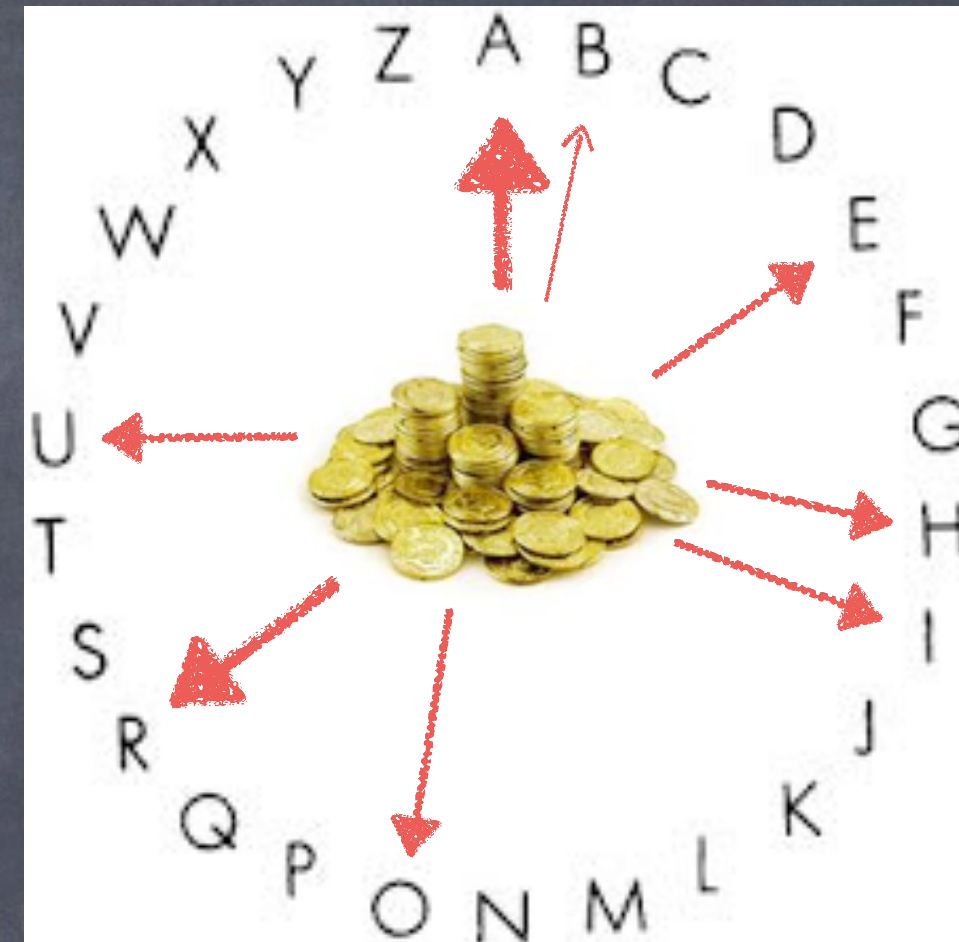
potesse_essere_ lo_spazio,

...nessun ancora lo sapeva. E il tempo, idem: cosa volete che ne facessimo, del tempo, stando lì pigiati come acciughe"

Gambling and Complexity

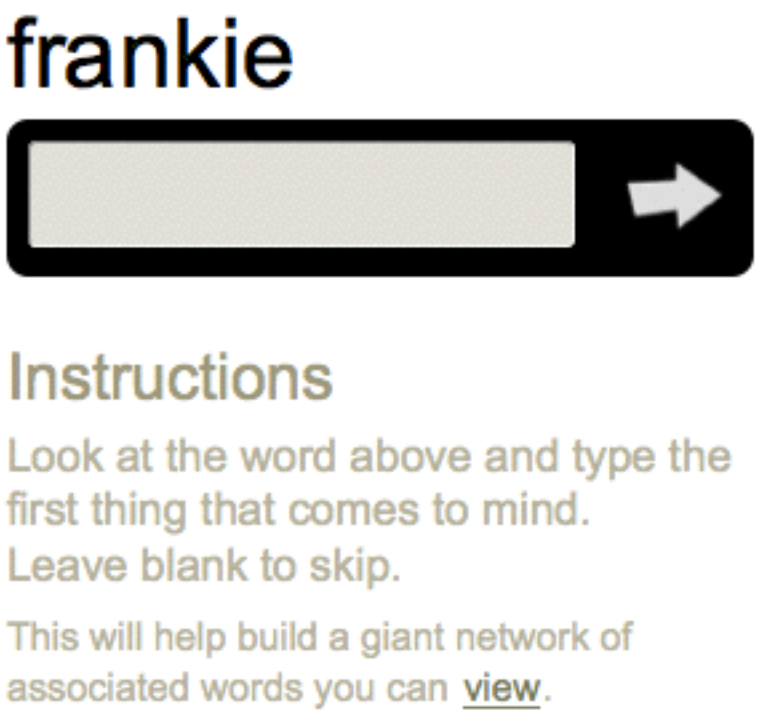
guessing the next outcome with a proportional gambling

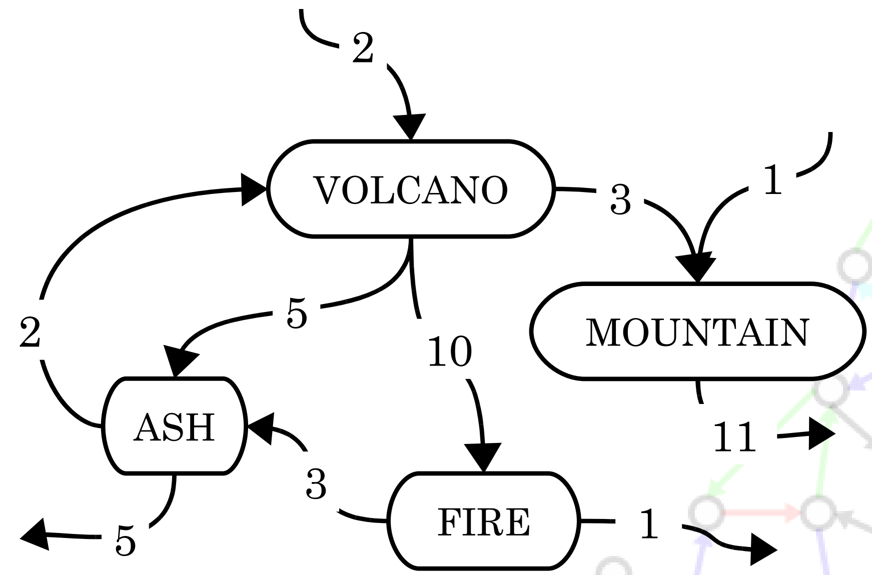
p**r**ofond

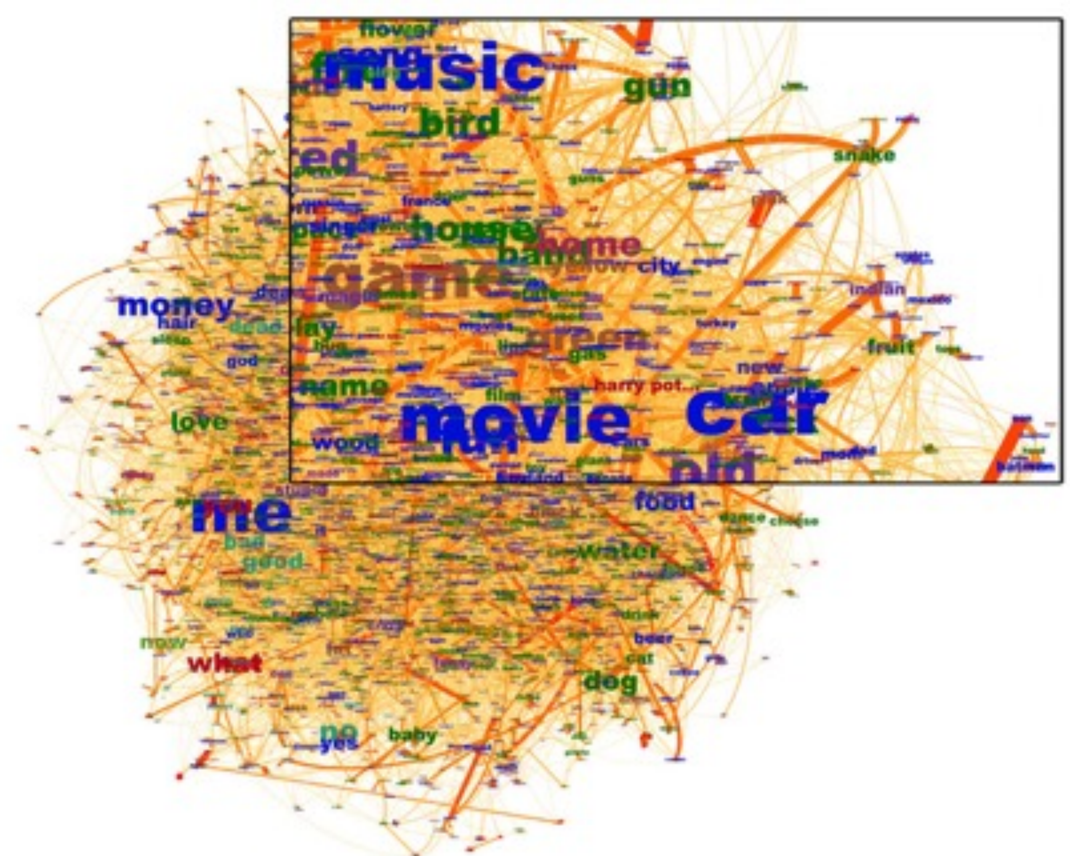


Final capital ⟷ Complexity of the phenomenon

Web as a laboratory for social sciences





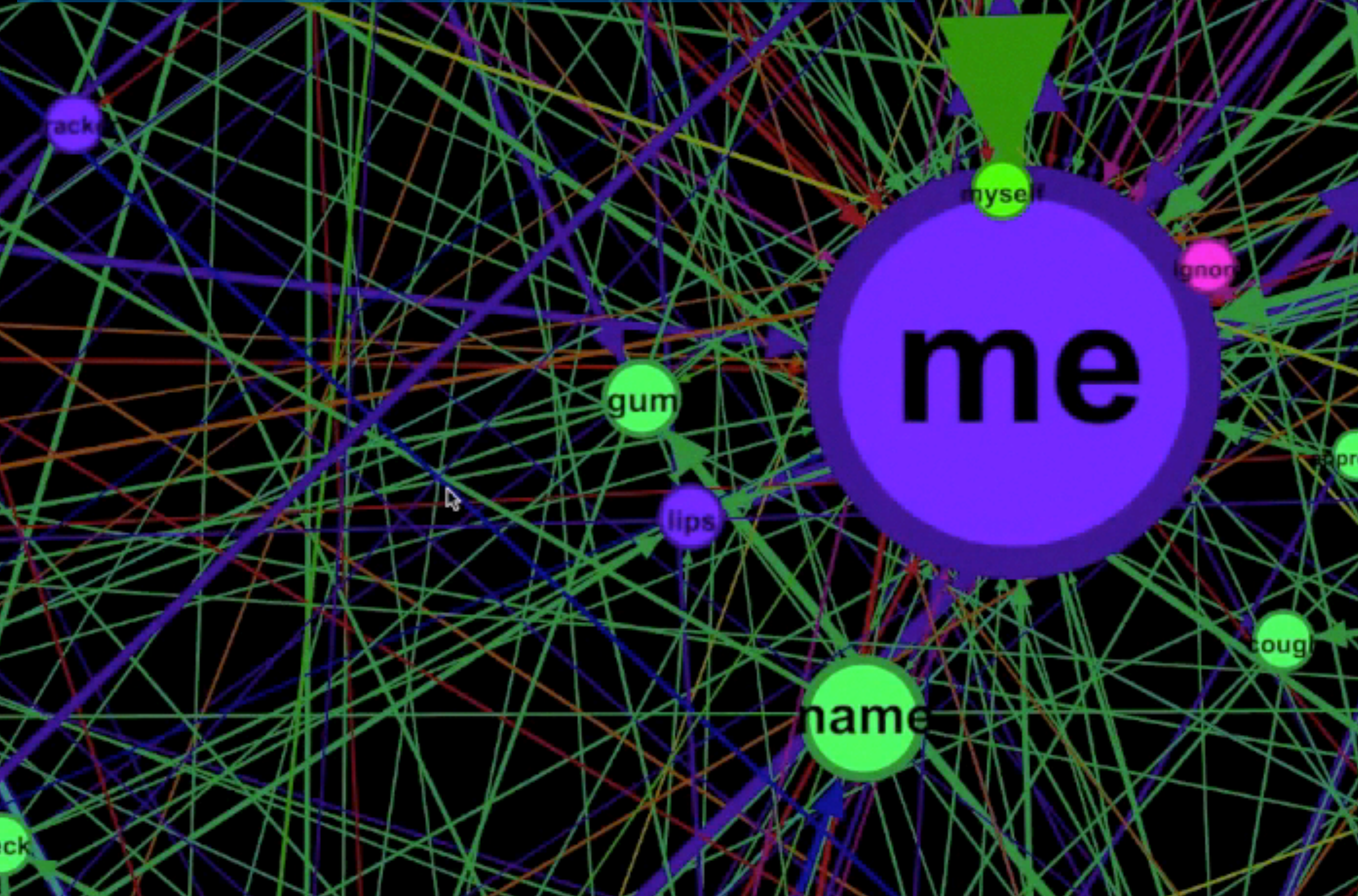


In less than a year:

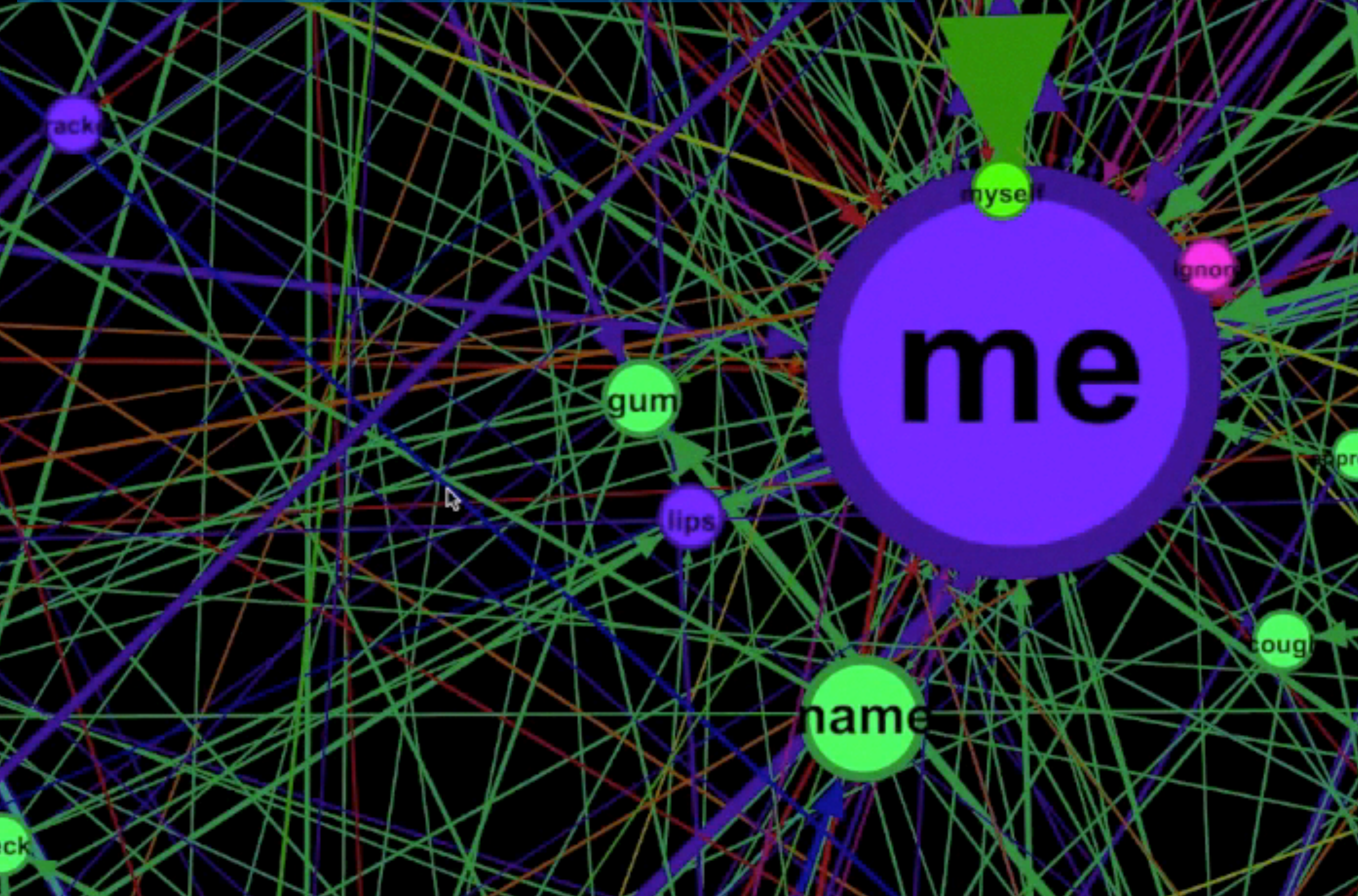
- 150 000 players
- 600 000 words
- 7 millions associations

http://www.humanbraincloud.com/

Conceptual network



Conceptual network



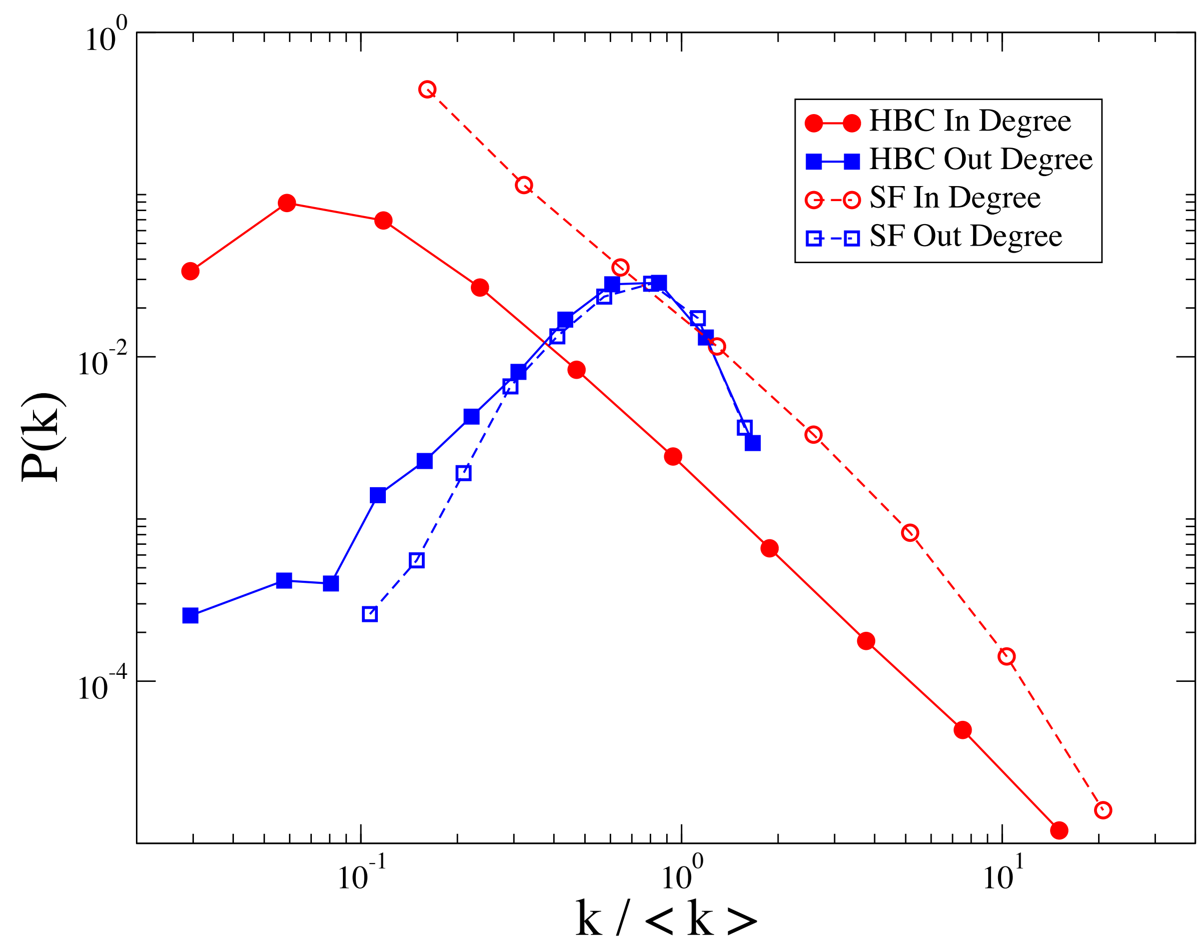
Comparison with a classical experiment: South Florida Free Association Norms (SF)

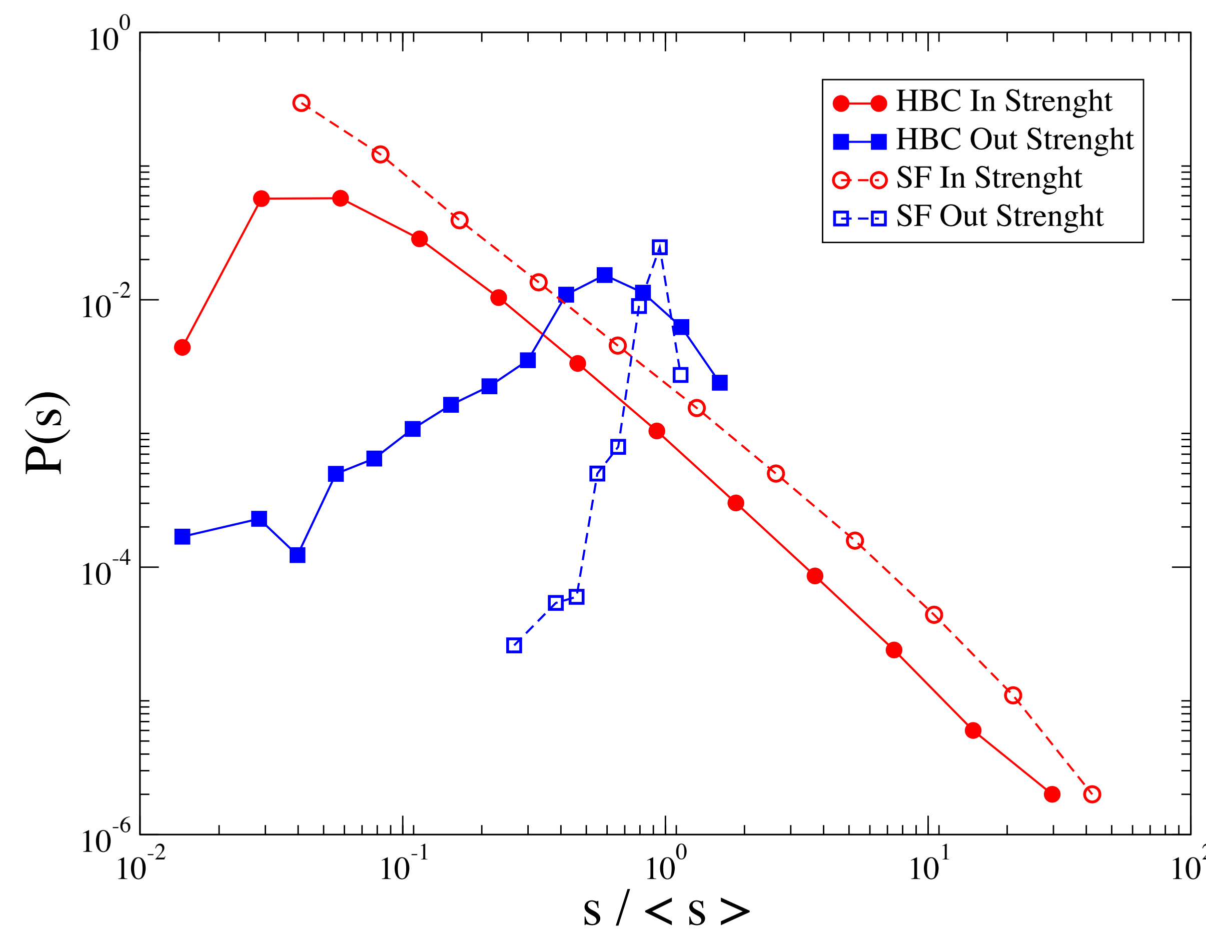
Nelson, McEvoy

5 000 words

60000 associations

Degree (k) distribution for HBC and SF

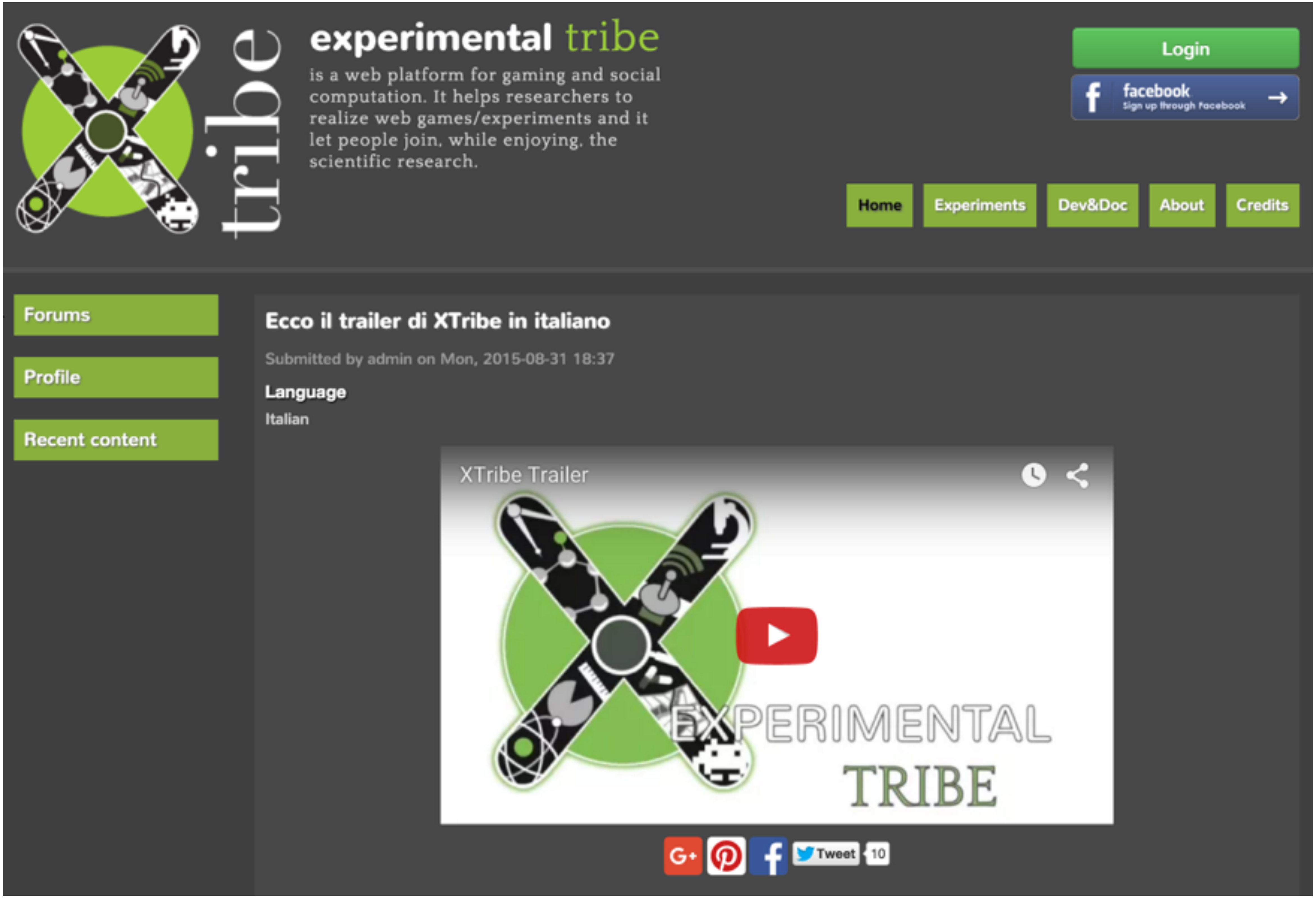




Cosine similarity between HBC and SF links is 0.85 (0.48 reshuffled)

$$CS_i = \frac{\sum_j l_{ij}^{HBC} \cdot l_{ij}^{SF}}{\sqrt{\sum_j (l_{ij}^{HBC})^2 \cdot \sum_j (l_{ij}^{SF})^2}},$$

The Xtribe platform for web-gaming



http://www.xtribe.eu/

tribe

- it will allow virtually any researcher to realize his own experiment with minimal effort, paving the way of the use of the web as a standard "laboratory" for social sciences.

- it will allow virtually any researcher to realize his own experiment with minimal effort, paving the way of the use of the web as a standard "laboratory" for social sciences.
    - it can be a strong "basin of attraction" for people willing to participate to experiments, making in this way recruitment much easier than for single-experiment platforms.

- it will allow virtually any researcher to realize his own experiment with minimal effort, paving the way of the use of the web as a standard "laboratory" for social sciences.
    - it can be a strong "basin of attraction" for people willing to participate to experiments, making in this way recruitment much easier than for single-experiment platforms.
- research areas: opinion and language dynamics, decision making, game-theory, geography, human mobility, economics, psychology, etc...

cum grano salis

In nearly every detective novel since the admirable stories of Conan Doyle there comes a time when the investigator has collected all the facts he needs for at least some phase of his problem. These facts often seem quite strange, incoherent, and wholly unrelated. The **great detective**, however, realizes that no further investigation is needed at the moment, and that **only pure thinking will lead to a correlation of the facts collected**. So he plays his violin, or lounges in his armchair enjoying a pipe, when suddenly, by Jove, he has it! Not only does he have an explanation for the clues at hand but he knows that certain other events must have happened. Since he now knows exactly where to look for it, he may go out, if he likes, to collect further confirmation for his theory.

cum grano salis

In nearly every detective novel since the admirable stories of Conan Doyle there comes a time when the investigator has collected all the facts he needs for at least some phase of his problem. These facts often seem quite strange, incoherent, and wholly unrelated. The **great detective**, however, realizes that no further investigation is needed at the moment, and that **only pure thinking will lead to a correlation of the facts collected**. So he plays his violin, or lounges in his armchair enjoying a pipe, when suddenly, by Jove, he has it! Not only does he have an explanation for the clues at hand but he knows that certain other events must have happened. Since he now knows exactly where to look for it, he may go out, if he likes, to collect further confirmation for his theory.

Albert Einstein and Leopold Infeld

The evolution of physics (1938)
Cambridge University Press.

Open questions

Data-driven or **questions-**driven research?

The new role of data in science
…and in policy making

The role of modelling and theory in science
…and in policy making