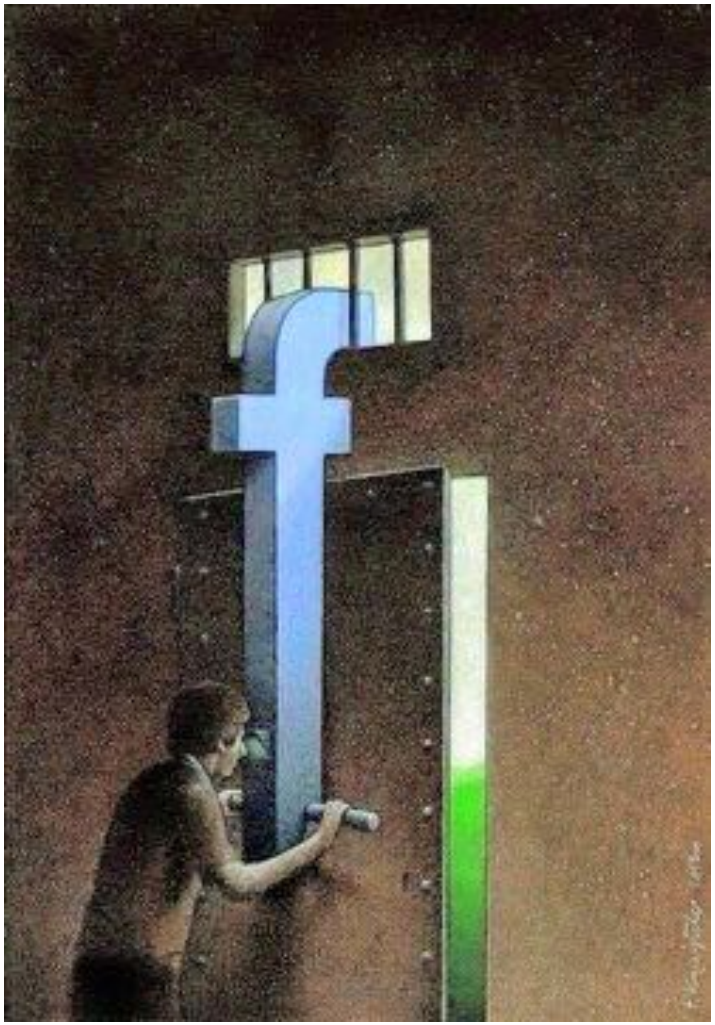
ETH zürich

Build Digital Democracy

By citizens, for citizens

Dr. Evangelos Pournaras

New Emerging Challenges



Big Data

Harvard Business Review

SEARCH

THE MAGAZINE | BLOGS | VIDEO | BOOKS | CASES | WEBINARS | COURS

Guest — Subscribe today and get access to all current articles and HBR online archive.

HBR Blog Network

Big Data's Dangerous New Era of Discrimination

by Michael Schrage | 8:00 AM January 29, 2014

Comments (30)

Congratulations. You bought into Big Data and it's paying off Big Time. You slice, dice, parse and process every screen-stroke, clickstream, Like, tweet and touch point that matters to your enterprise. You now know exactly who your best — and worst — customers, clients, employees and partners are. Knowledge is power. But what kind of power does all that knowledge buy?

Big Data creates Big Dilemmas. Greater knowledge of customers creates new potential and power to discriminate. Big Data — and its associated analytics — dramatically increase both the dimensionality and degrees of freedom for detailed discrimination. So where, in your corporate culture and strategy, does value-added personalization and segmentation end and harmful

ZDNet

White Papers | Hot Topics | Downloads | Reviews | Newsletters

US Edition | Internet of Things | Mobility | Research | Windows | Enterprise Software

ARE YOUR HR & FINANCE SYSTEMS BASED ON TECH OLDER T

MUST READ Hackers jump on the Shellshock Bash bandw

Topic: Big Data — Follow via:

Why big data evangelists should be sent to re-education camps

Summary: *Big data is a dangerous, faith-based ideology. It's fuelled by hubris, it's ignorant of history, and it's trashing decades of progress in social justice.*

By Stilgherrian for The Full Tilt | September 19, 2014 — 07:13 GMT (00:13 PDT)

Follow @stilgherrian — Get the ZDNet Big Data newsletter now

Comments 27 | Votes 7 | Share 458 | Tweet 692 | Share | more +

The last time I wrote about big data, in July, I called it a big, distracting bubble. But it's worse than that. Big data is an ideology. A religion. One of its most important gospels is, of course, at Wired.

In 2008, Chris Anderson talked up a thing called The Petabyte Age in The End of Theory: The Data Deluge Makes the Scientific Method Obsolete.

Has anyone got a pin?

"The new availability of huge amounts of data, along with the statistical tools to crunch these numbers, offers a whole new way of understanding the world. Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all," he wrote.

Declaring the scientific method dead after 2,700 years is quite a claim. Hubris, even. But, Anderson wrote, "There's no reason to cling to our old ways." Oh, OK then.

Now, this isn't the first set of claims that correlation would supersede causation, and that the next iteration of computing practices would "make everything different".

Big Data

Existing **social mining** practices
threaten **social cohesion**

"*surveillance has become increasingly privatized, commercialized and participatory*", Julie E. Cohen

Build Digital Democracy [1]

nature
International weekly journal of science
Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio
Archive > Volume 527 > Issue 7576 > Comment > Article

NATURE | COMMENT

Society: Build digital democracy

Dirk Helbing & Evangelos Pournaras

02 November 2015

Open sharing of data that are collected with smart devices would empower citizens and create jobs, say Dirk Helbing and Evangelos Pournaras.

Subject terms: Computer science · Society

Imaginechina/Corbis

Many choices that people consider their own are already determined by algorithms.

Fridges, coffee machines, toothbrushes, phones and smart devices are all now equipped with communicating sensors. In ten years, 150 billion 'things' will connect with each other and with billions of people. The 'Internet of Things' will generate data volumes that double

Many choices that people consider their own are already determined by algorithms.

Build digital democracy

Open sharing of data that are collected with smart devices would empower citizens and create jobs, say Dirk Helbing and Evangelos Pournaras.

Fridges, coffee machines, toothbrushes, phones and smart devices are all now equipped with communicating sensors. In ten years, 150 billion 'things' will connect with each other and with billions of people. The 'Internet of Things' will generate data volumes that double every 12 hours rather than every 12 months, as is the case now.

Blinded by information, we need 'digital

our decisions, we need information systems that are transparent, trustworthy and user-controlled. Each of us must be able to choose, modify and build our own tools for winnowing information.

With this in mind, our research team at the Swiss Federal Institute of Technology in Zurich (ETH Zurich), alongside international partners, has started to create a distributed,

predictable. Our behaviour is increasingly steered by personalized advertisements and search results, recommendation systems and emotion-tracking technologies. Thousands of pieces of metadata have been collected about every one of us (see go.nature.com/stoqsu). Companies and governments can increasingly manipulate our decisions, behaviour and feelings[1].

An Alternative `Big Data‘ Paradigm

Open & free services – public good

Truly decentralized

How to design & build an alternative “Big Data” paradigm for a sustainable society?

Accurate but privacy-preserving

Participatory

Business opportunities

An Outline

Alternative 'Big Data' System

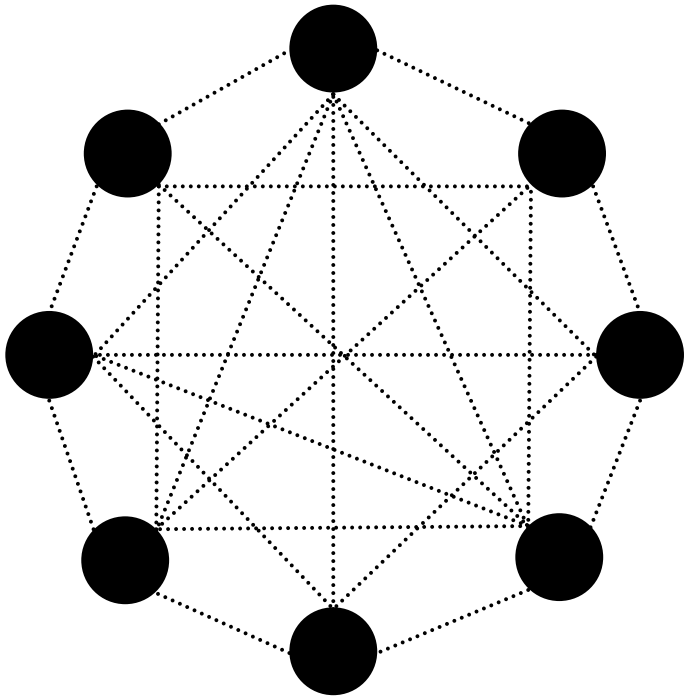
Participatory &
truly decentralized

1. Self-determined Choices

Internet of Things
sensors, controllers, actuators

2. Self-regulatory Information Sharing

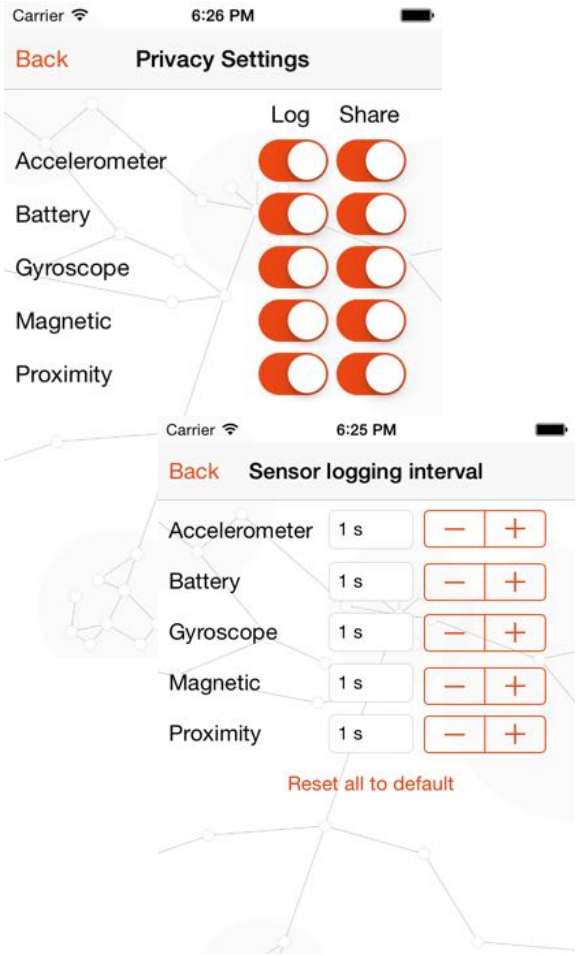
4. Collective Public Good Knowledge

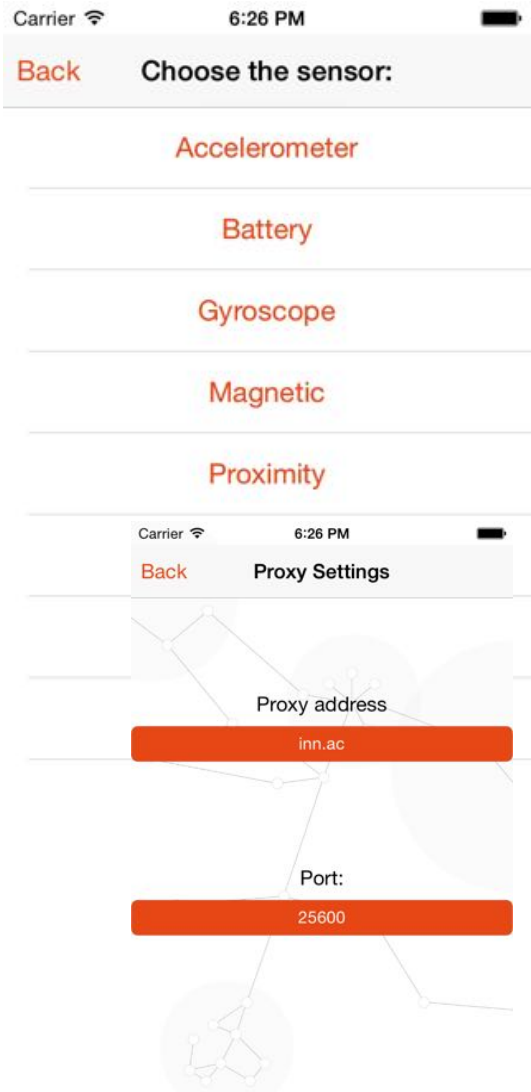


3. Decentralized Data Analytics

1. Self-Determined Choices [2]

Nervousnet 1.0 Beta





Nervousnet 1.0 Beta

Released beginning of June

Android: https://play.google.com/store/apps/details?id=ch.ethz.soms.n

iOS: https://itunes.apple.com/us/app/nervousnet/id1000599804?mt=8

Open & free! Source code in Github
https://github.com/nervousnet

Functionality

Sensor data storage on the phone (local)

Sensor data storage on Nervousnet proxy (remote)

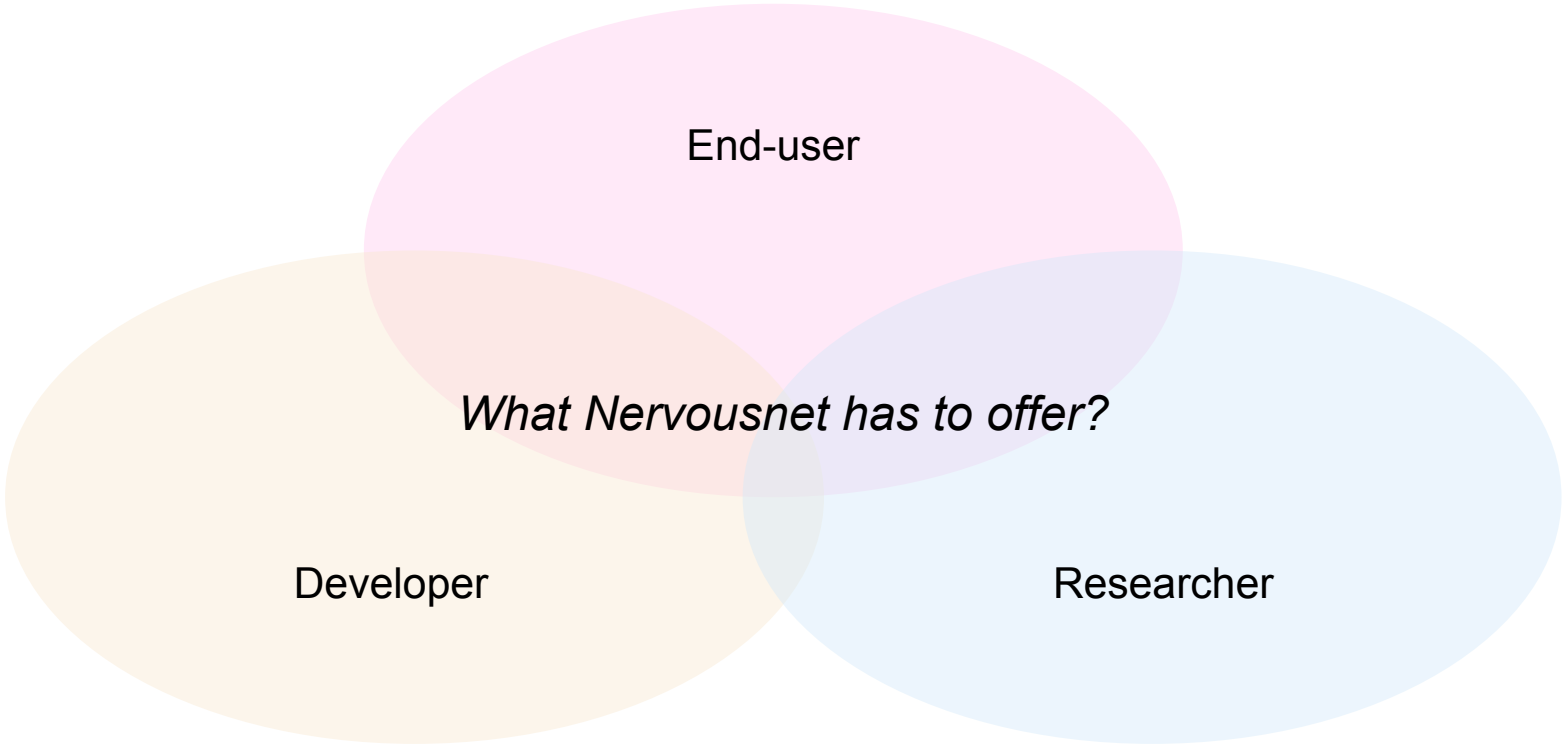
Real-time/query-based data visualization

Controllability of data logging/sharing

Controllability of data sampling

Self-determination of Nervousnet proxy

User Groups


End-user
What Nervousnet has to offer?
Developer
Researcher

Why to use Nervousnet

Flexible infrastructure-independent deployment of social experiments

Researcher

Testbed platform for studying real-world distributed systems

Large-scale spatio-temporal data

Data to validate & test applications

Open-source guiding examples

Developer

Extensible data-driven API for the development of virtual sensors

Data visualization

Awareness of own social behavior – Virtual Sensors

End user

Privacy-preserving access to local sensor data

Analytics on local sensor data

2. Self-Regulatory Information Sharing [3]

A Paradox in Digital Democracy

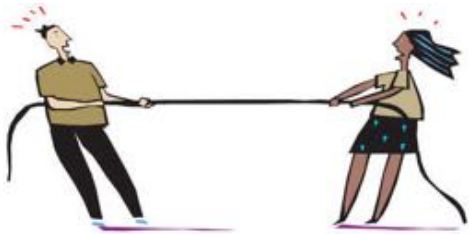
Contradictions

Fragmentation

polarization, deceived citizens, social injustice,

Big Data Analytics vs. Privacy-preservation

More data,
more information, more knowledge, more security, more business opportunities,
more prosperity

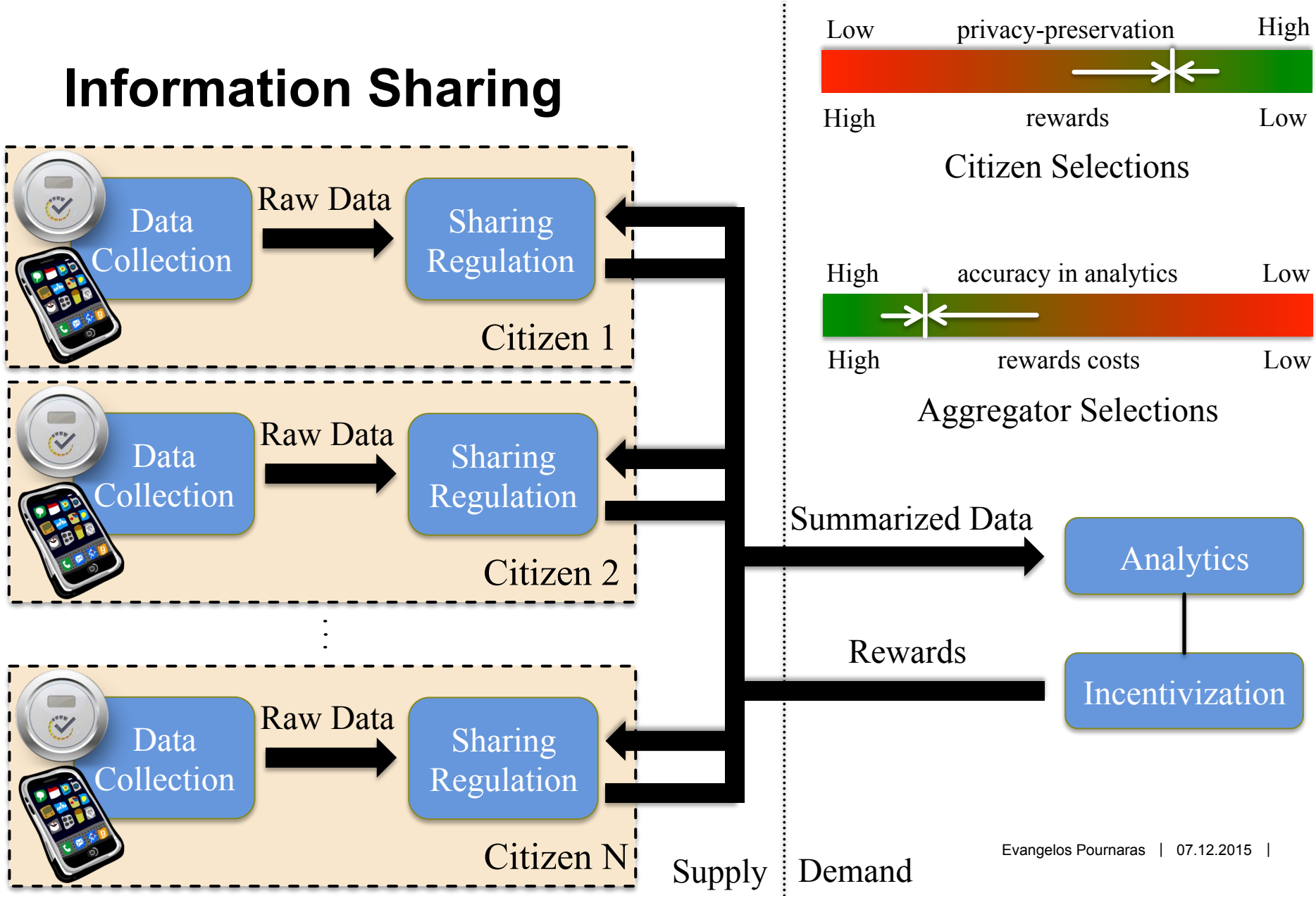


Less data,
less information, less surveillance, less discrimination, more freedom/ justice, more social cohesion,
more prosperity

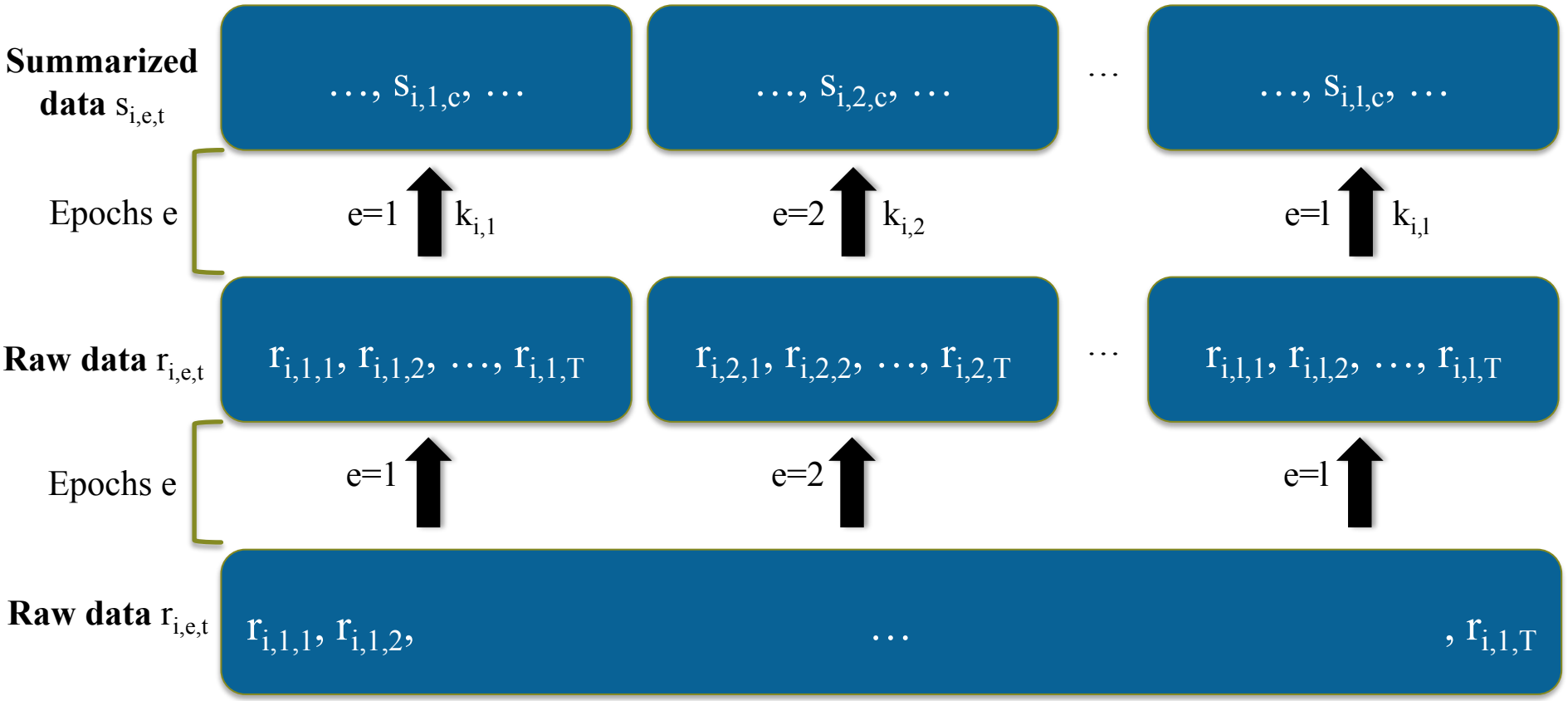
Research communities cancel out their contributions

How to cancel out their paradox?

Information Sharing

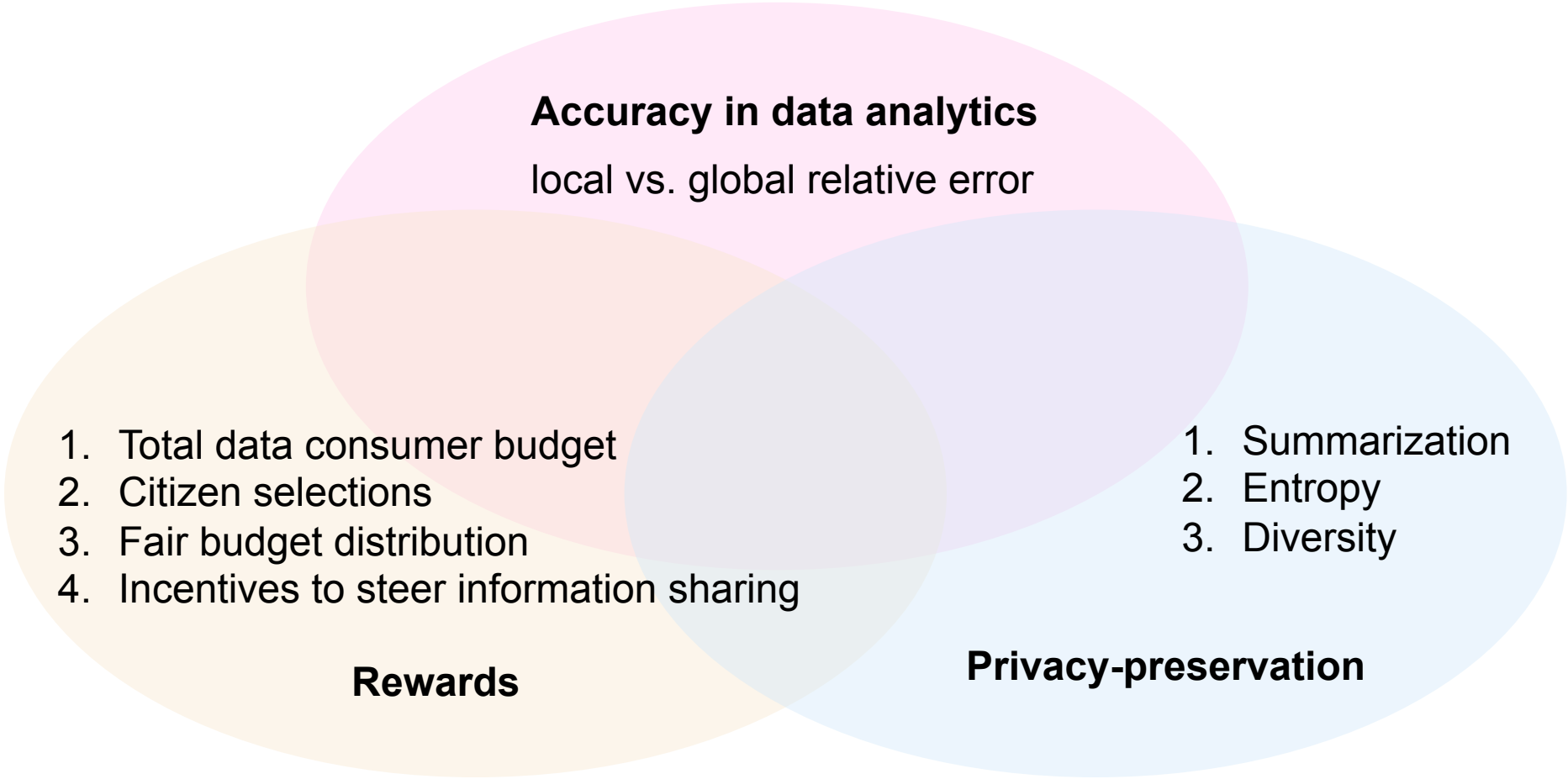


Privacy-preserving Information Sharing



Agent i
e.g. 30 mins data for a year
T=17520

Trade-off Measurements



Implementation

Unsupervised learning

Several implementation algorithms

Clustering

Semi-automated – Empirical

Fully-automated, data-driven

Customizable – number of clusters

Manual - unbounded

Smart Grid – 6435 participants – 1 sensor

Datasets

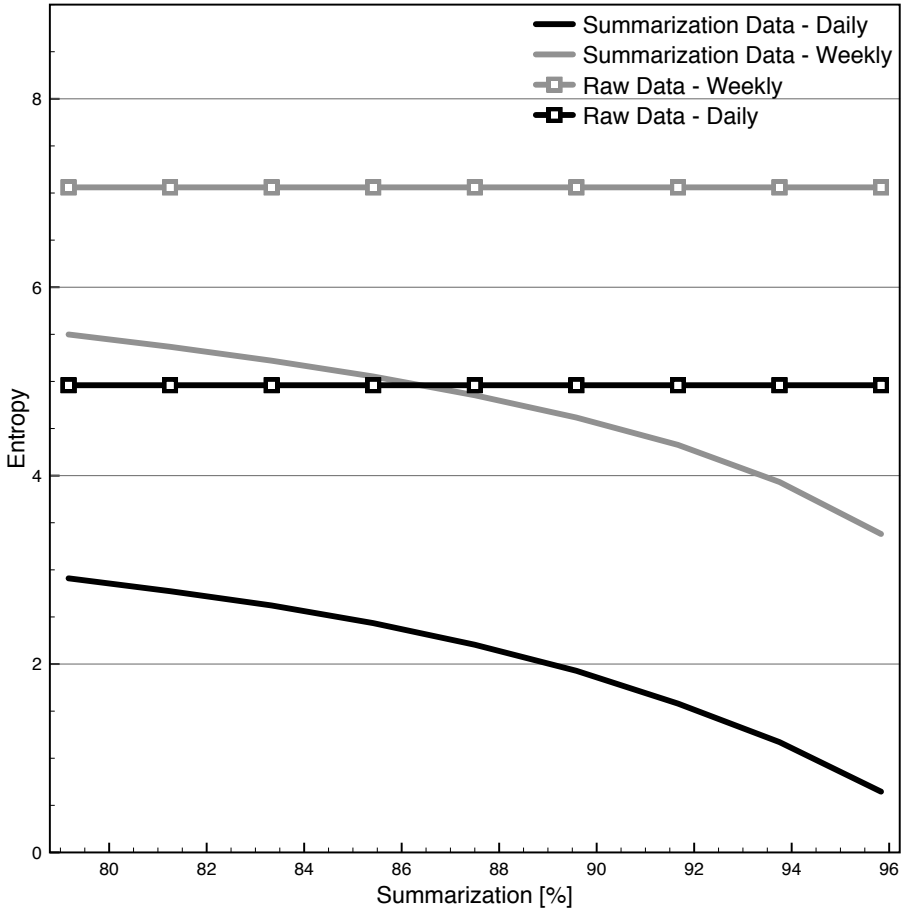
Nervousnet– 154 participants – several sensors

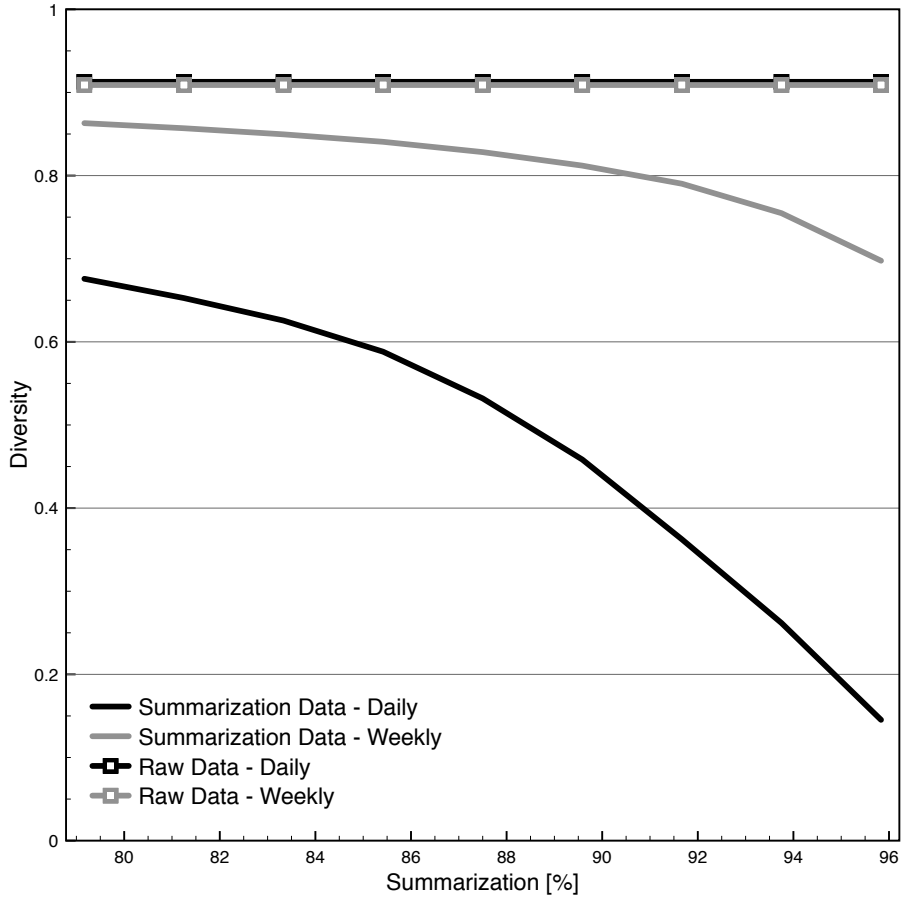
Survey questions

Privacy preferences

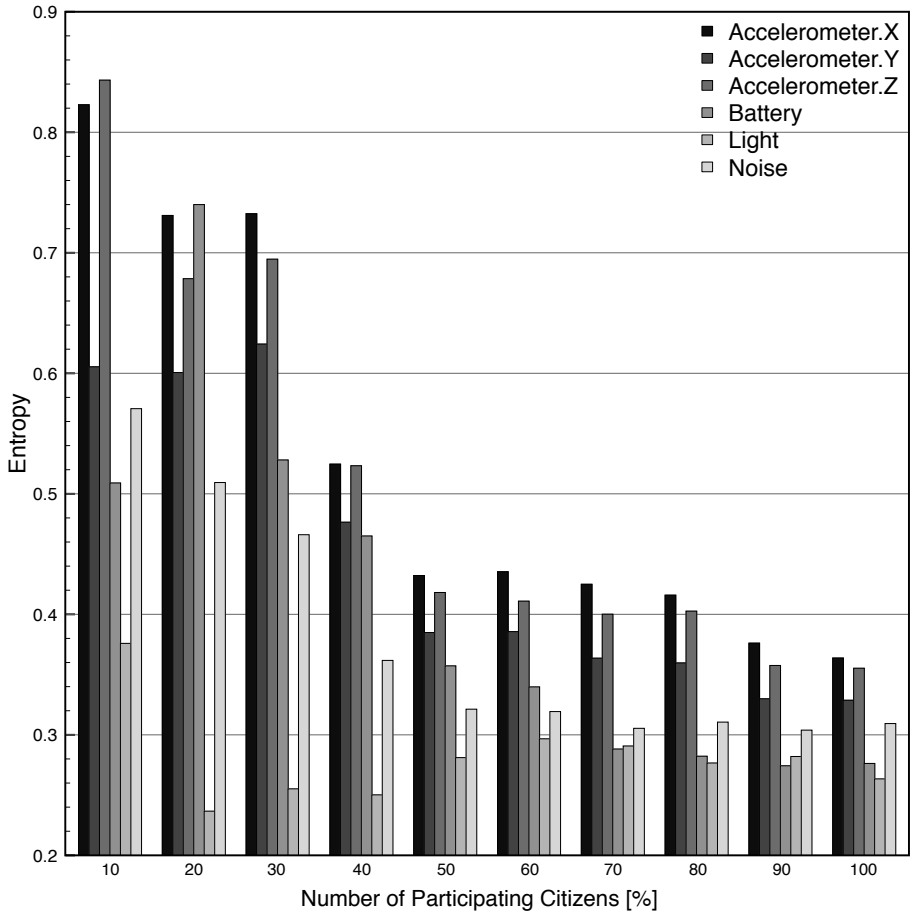
Mapping survey answers to a summarization range

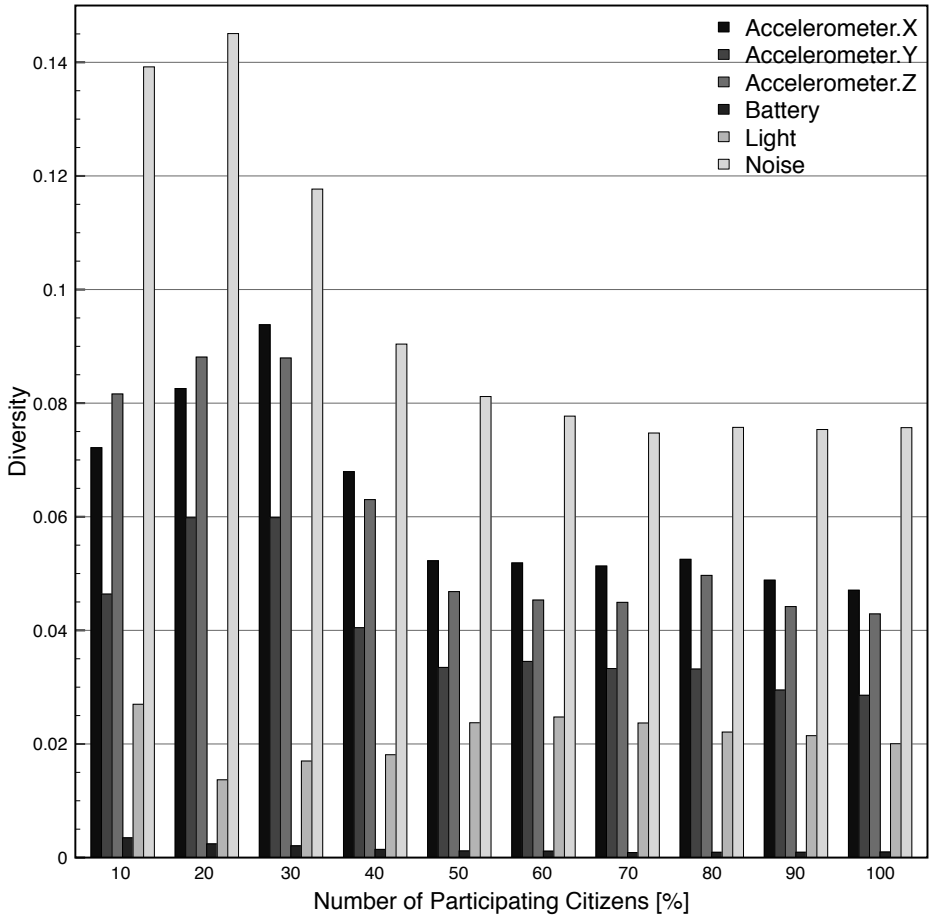
Privacy-preservation – Smart Grid



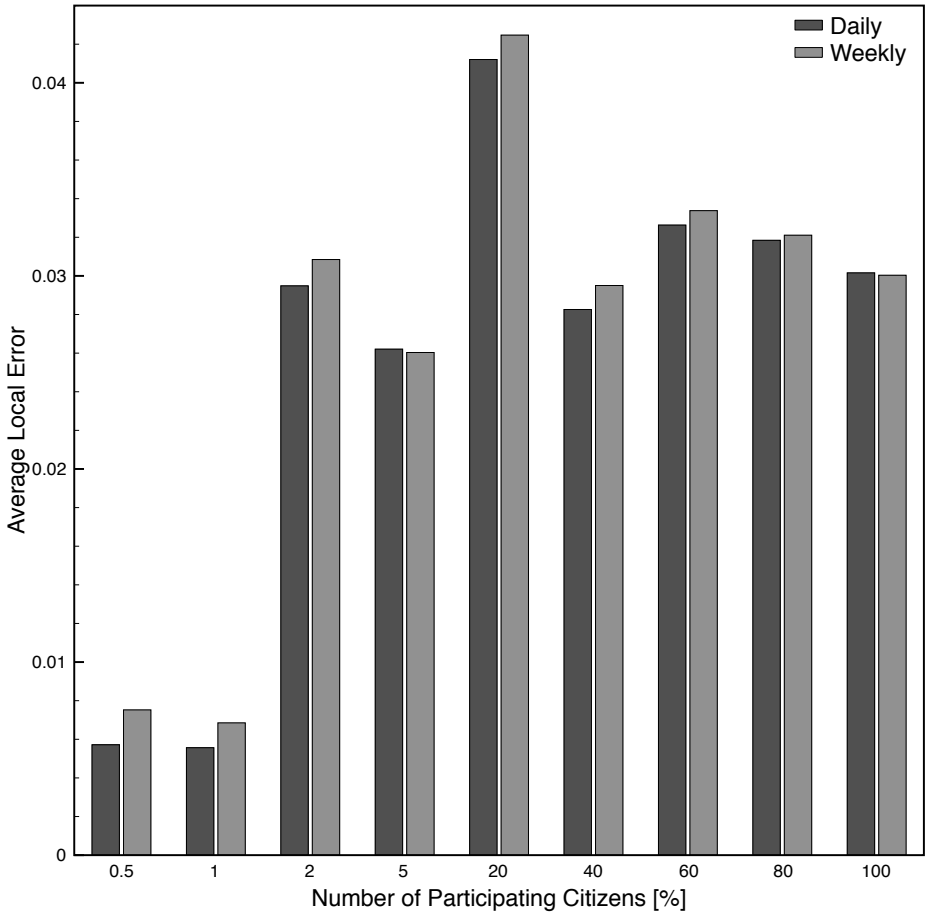


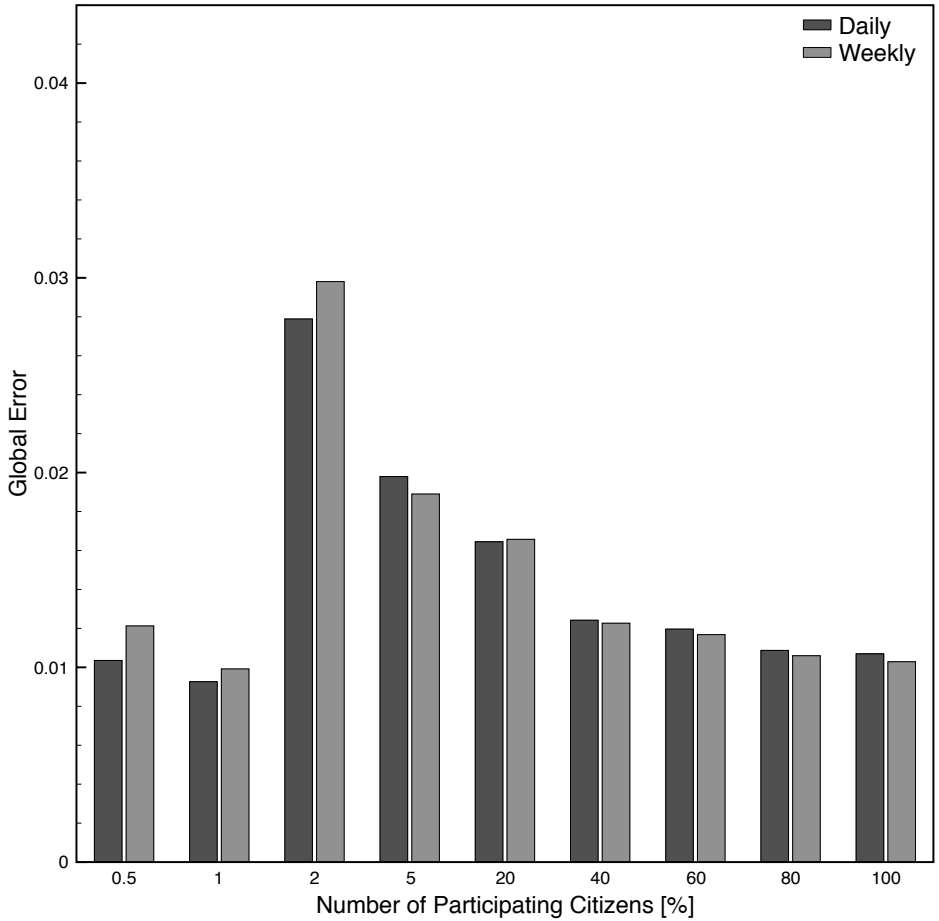
Privacy-preservation - Nervousnet



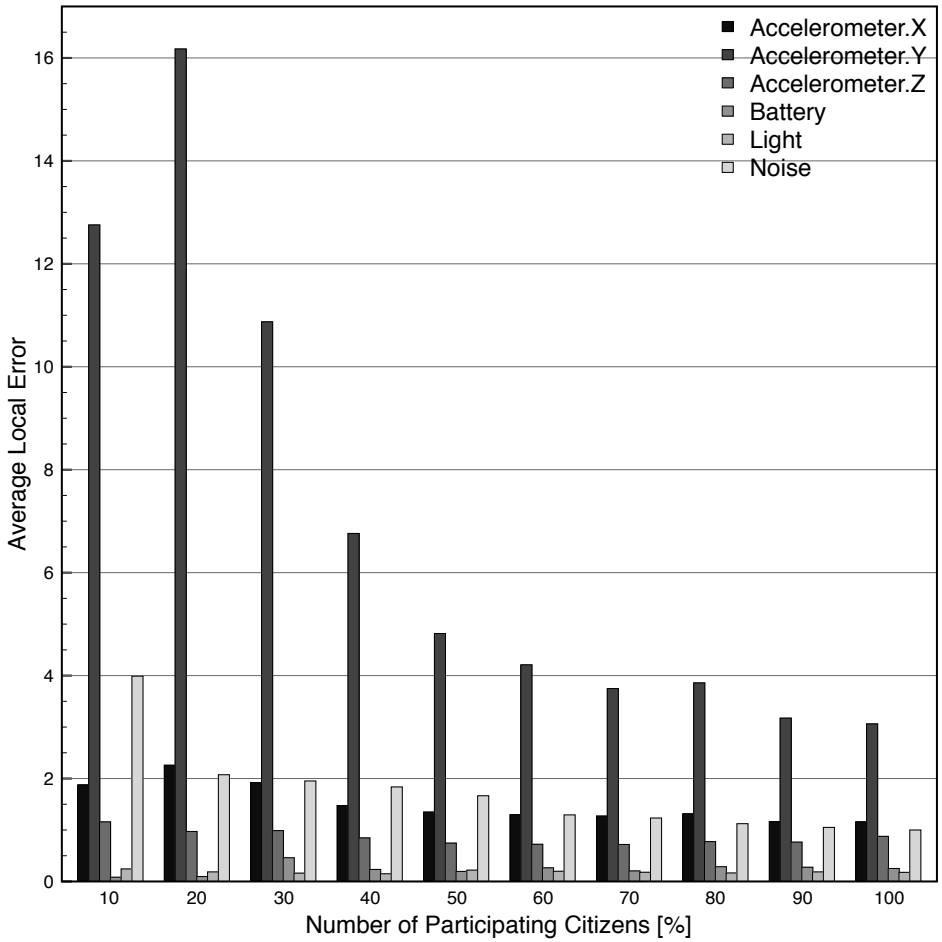


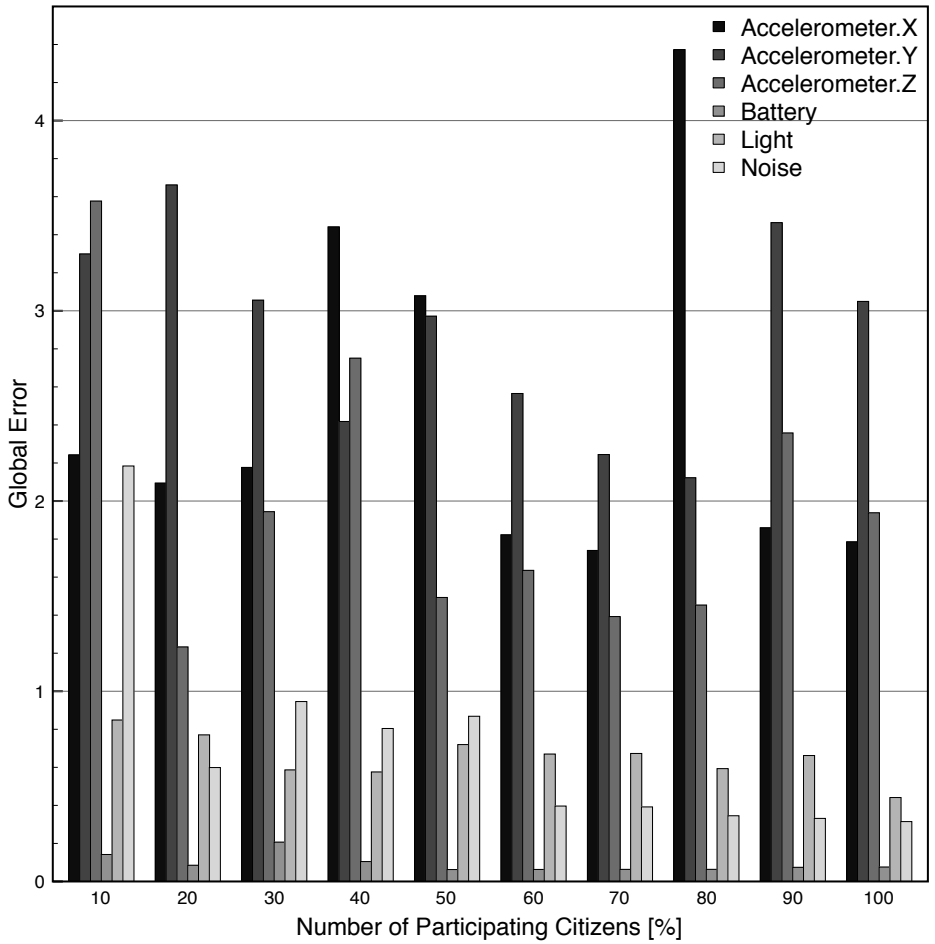
Data Analytics Accuracy – Smart Grid





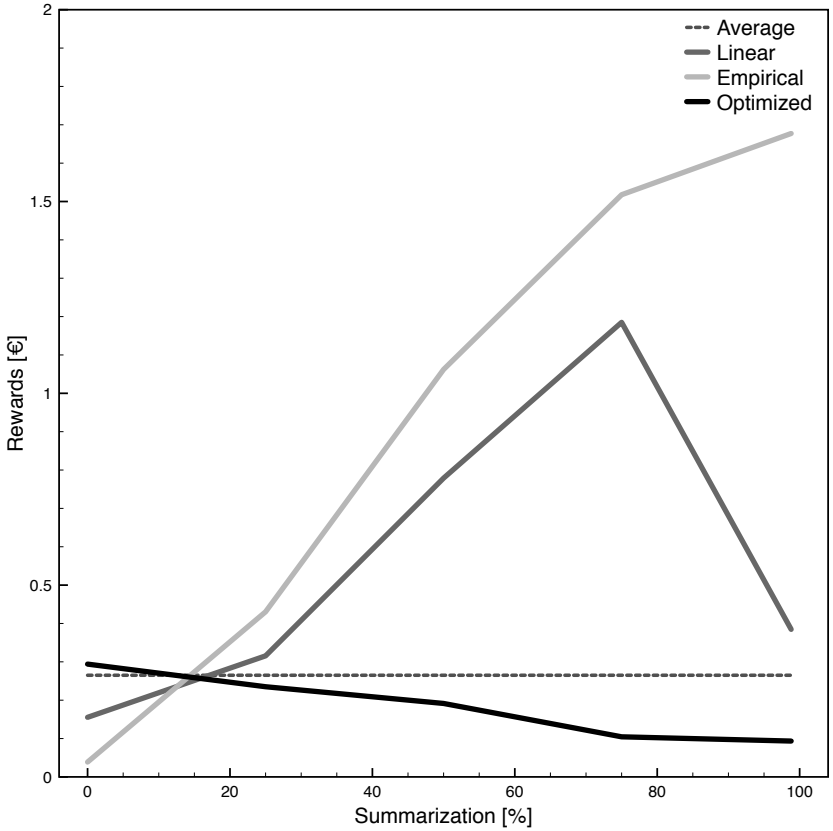
Data Analytics Accuracy – Nervousnet





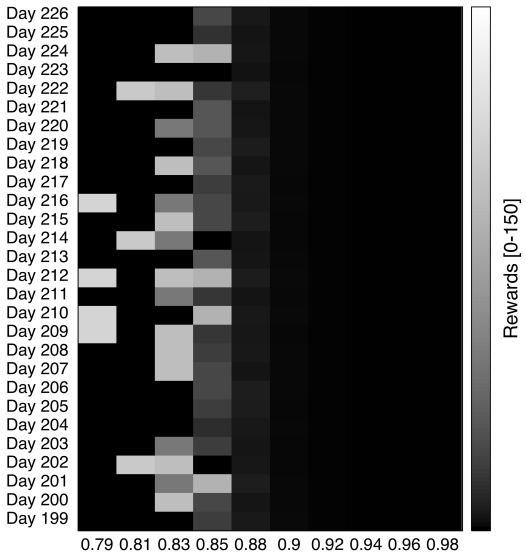
Rewards – Smart Grid

Semi-automated

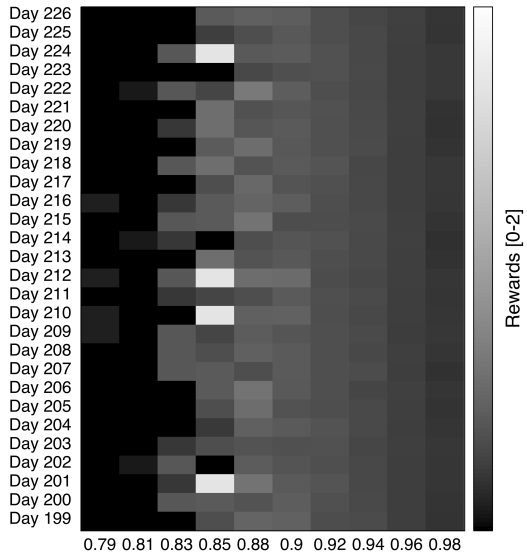


Fully-automated – data driven

Linear



Optimized



Summarization

Conclusions

Higher summarization, higher privacy-preservation

More participants, higher privacy-preservation

From canceling out research contributions
to canceling out their paradox

Sensor types influence privacy-preservation & accuracy

Local errors cancel out resulting in low global errors

Incentivization can be optimized to be fair

An Outline

Alternative 'Big Data' System

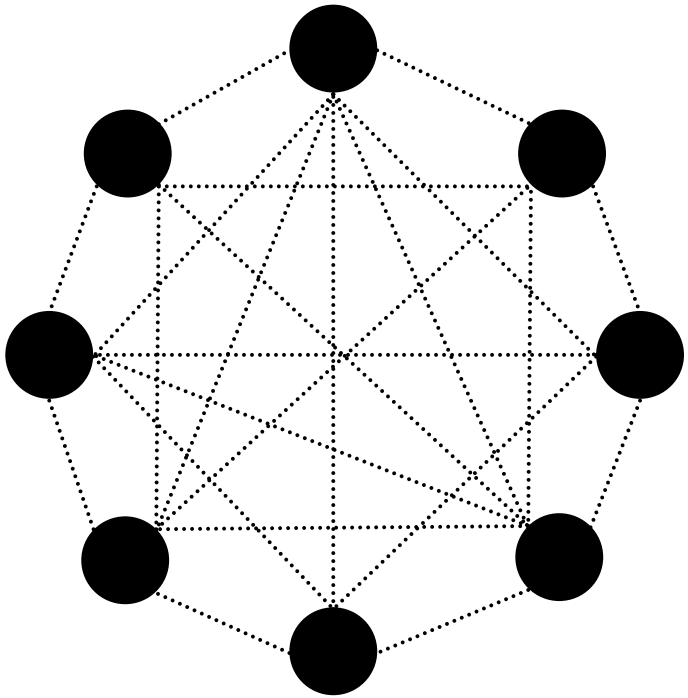
Participatory &
truly decentralized

1. Self-determined Choices

Internet of Things
sensors, controllers, actuators

2. Self-regulatory Information Sharing

4. Collective Public Good Knowledge



3. Decentralized Data Analytics

3. Decentralized Data Analytics [4]

Data Analytics

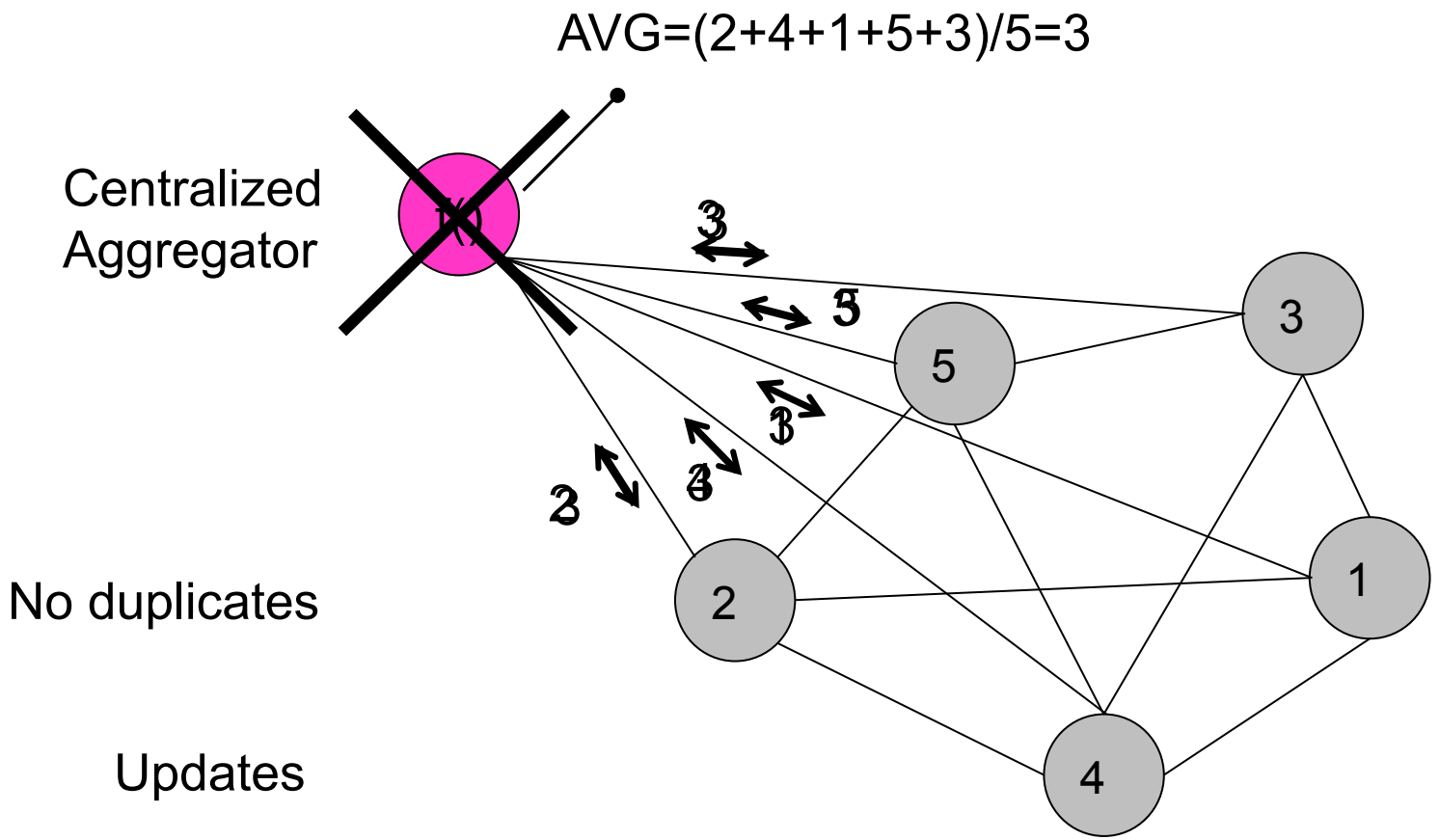
Common light-weight process for large-scale real-time analytics

Aggregation functions, SUM, AVG, MAX, MIN, STDEV, COUNT, etc.

Main component of the MapReduce model

Privacy-preserving

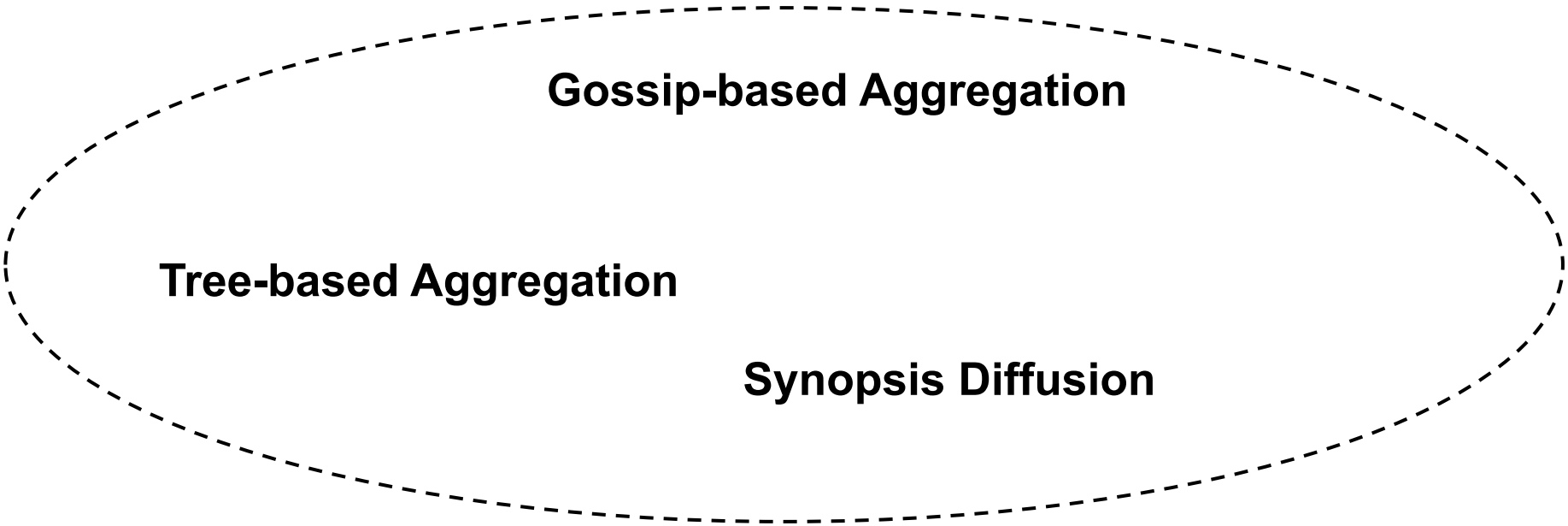
The Aggregation Problem



Dynamic Decentralized Aggregation

Static Values

Aggregation-function dependent



Routing-dependent

Inaccuracies: Duplicate & outdated values

Research Question

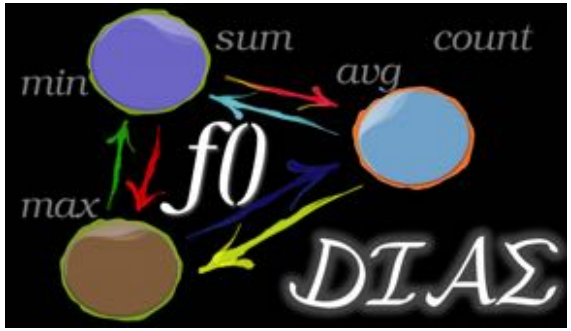
How to accurately compute in a decentralized fashion aggregation functions under a varying input sample?

DIAS

Computes almost any aggregation function – *without changing a single line of code*

Gossip-based routing decoupled from aggregation

DIAS – The Dynamic Intelligent Aggregation Service

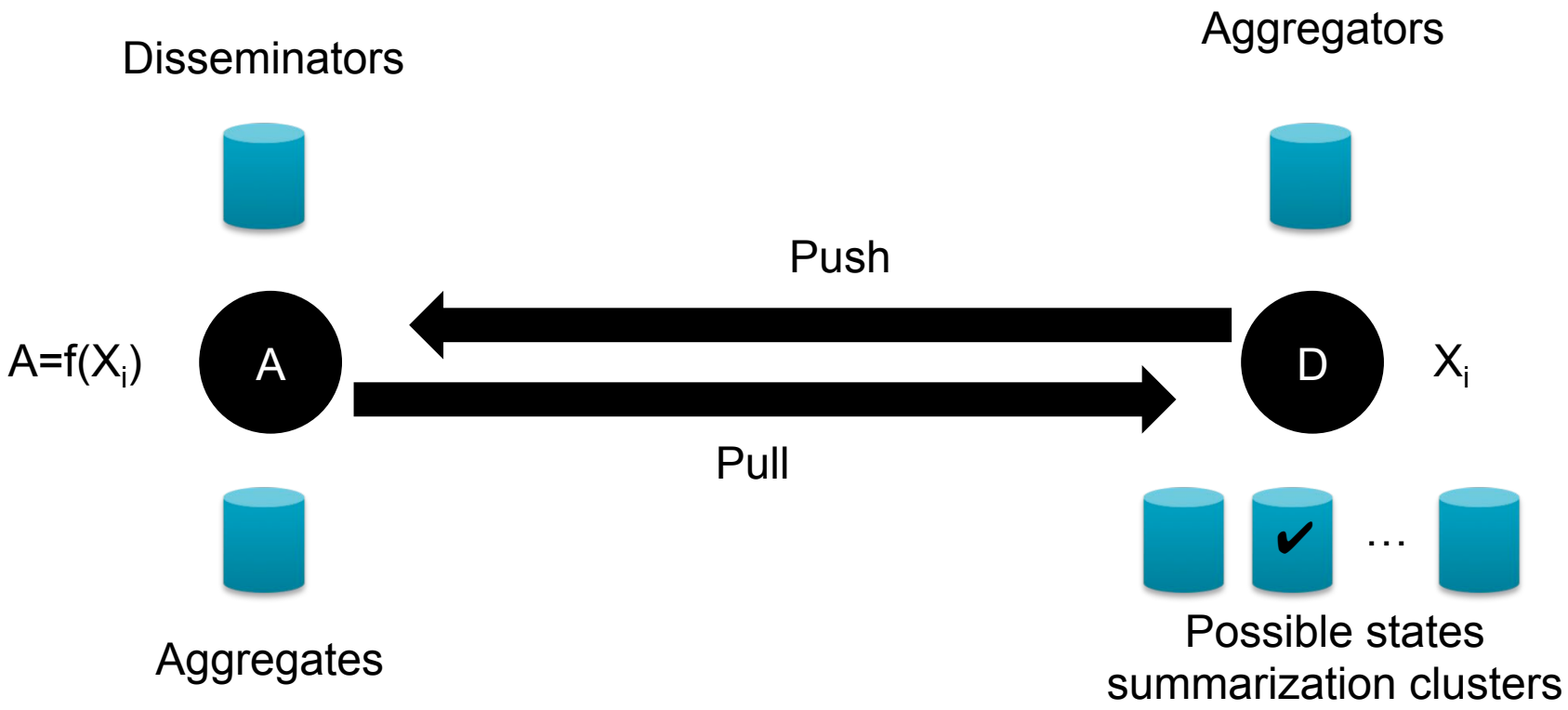


In-network aggregation
Every participating node **contributes input** & **receives output**

Aggregates are accurately updated
(i) under a varying input sample or
(ii) when nodes join/leave the network

Double-counting is prevented

Aggregation Model



Privacy by Design

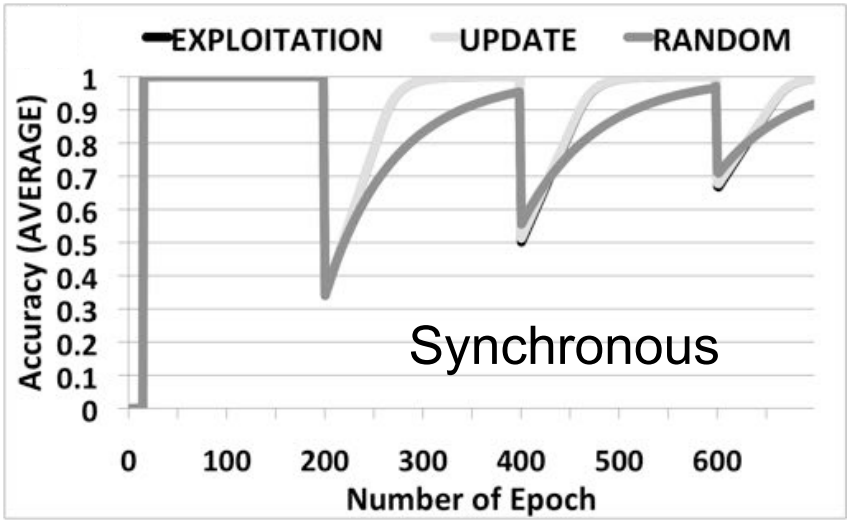
Aggregated possible states **are summarized data instead of raw data**

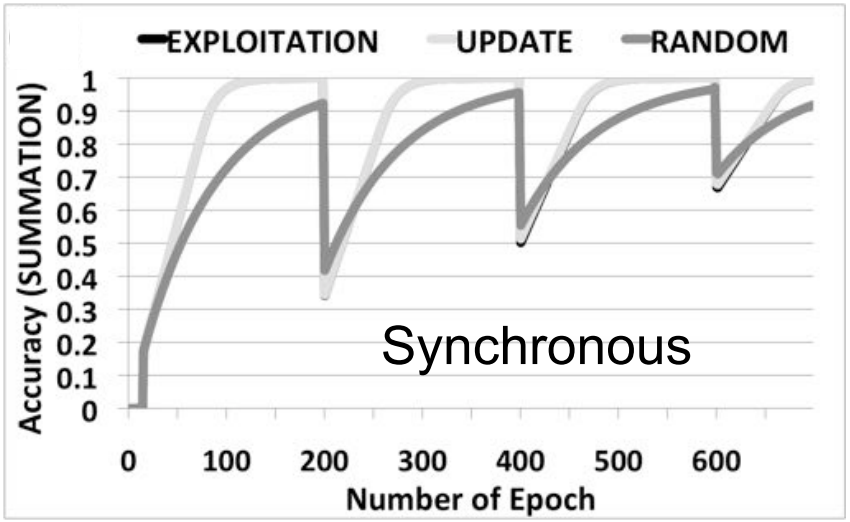
Privacy for Free!

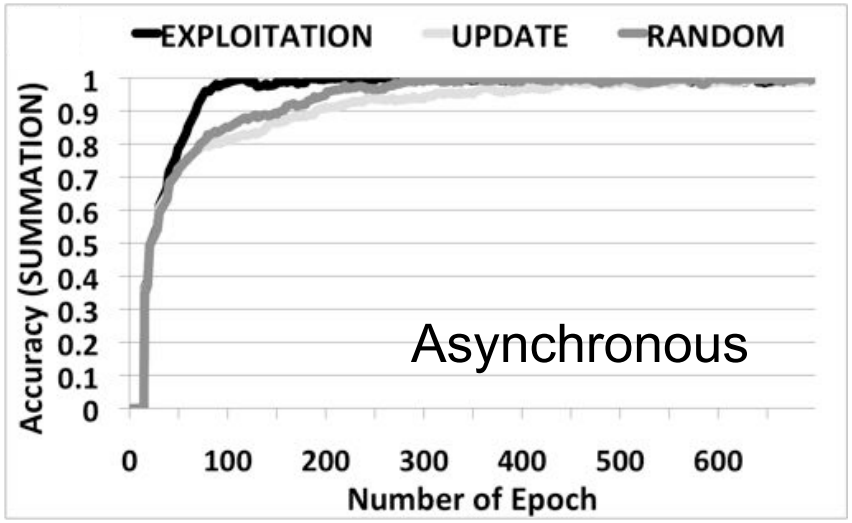
Bloom filter information **is implicit and not explicit** - membership

Join & leaves are handled **by only using bloom filters – no other information leak**

Privacy by Design







Towards an Operational System

Real-world prototype in Euler HPC at ETH

An Outline

Alternative 'Big Data' System

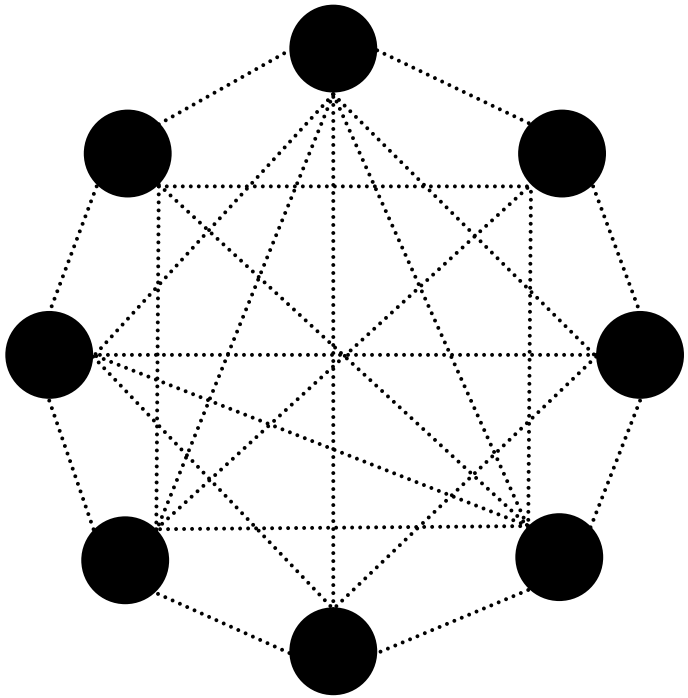
Participatory &
truly decentralized

1. Self-determined Choices

Internet of Things
sensors, controllers, actuators

2. Self-regulatory Information Sharing

4. Collective Public Good Knowledge



3. Decentralized Data Analytics

4. Collective Public Good Knowledge

Applications

Privacy-preserving localization & navigation

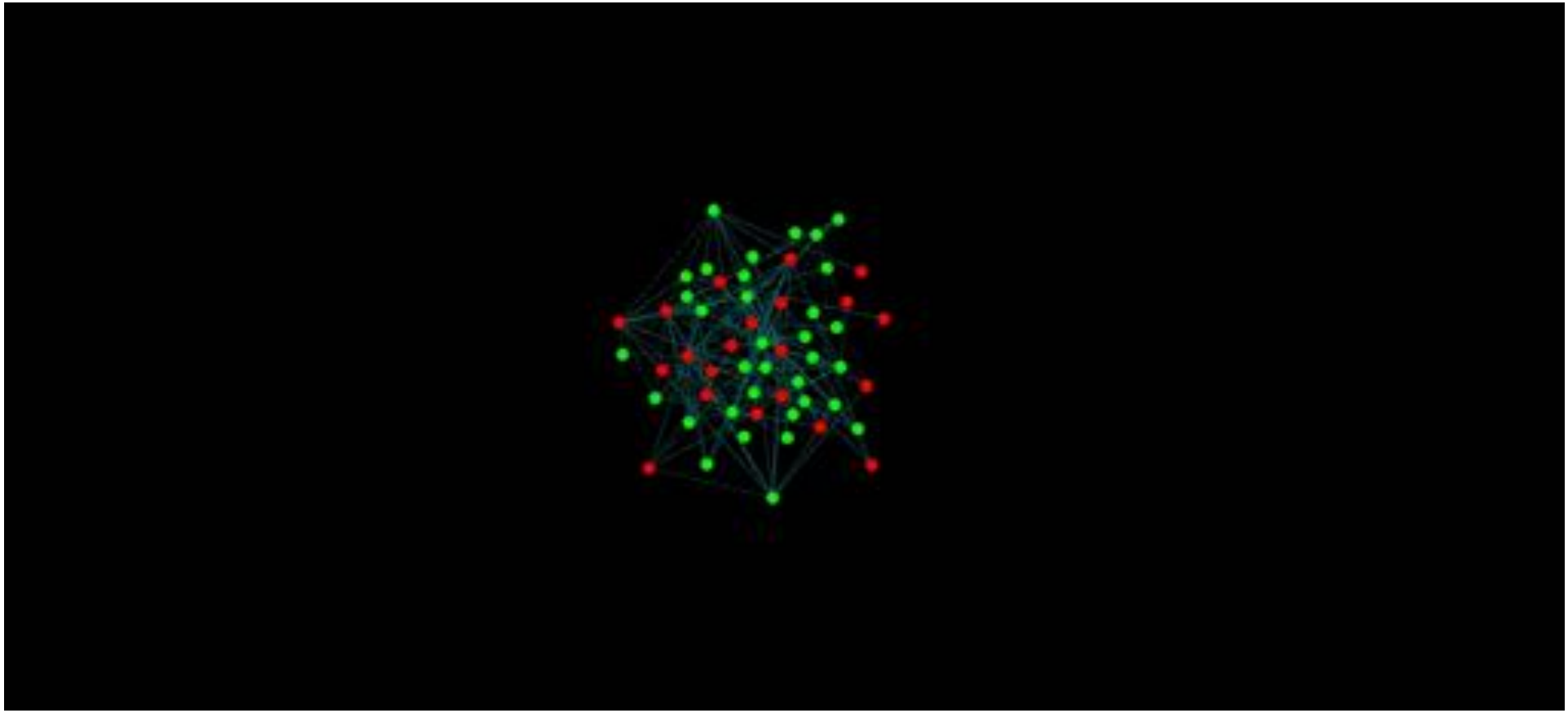
Mobility pattern recognition

Earthquake detection

Evacuation systems

Ambient assisting living

Privacy-preserving Localization – CCC 2014



Final Conclusion

An alternative more ethical & democratic Big Data paradigm is required

Unleashing force: *systems truly decentralized & privacy-preserving by design*

Long-term sustainable participation, social justice, social cohesion

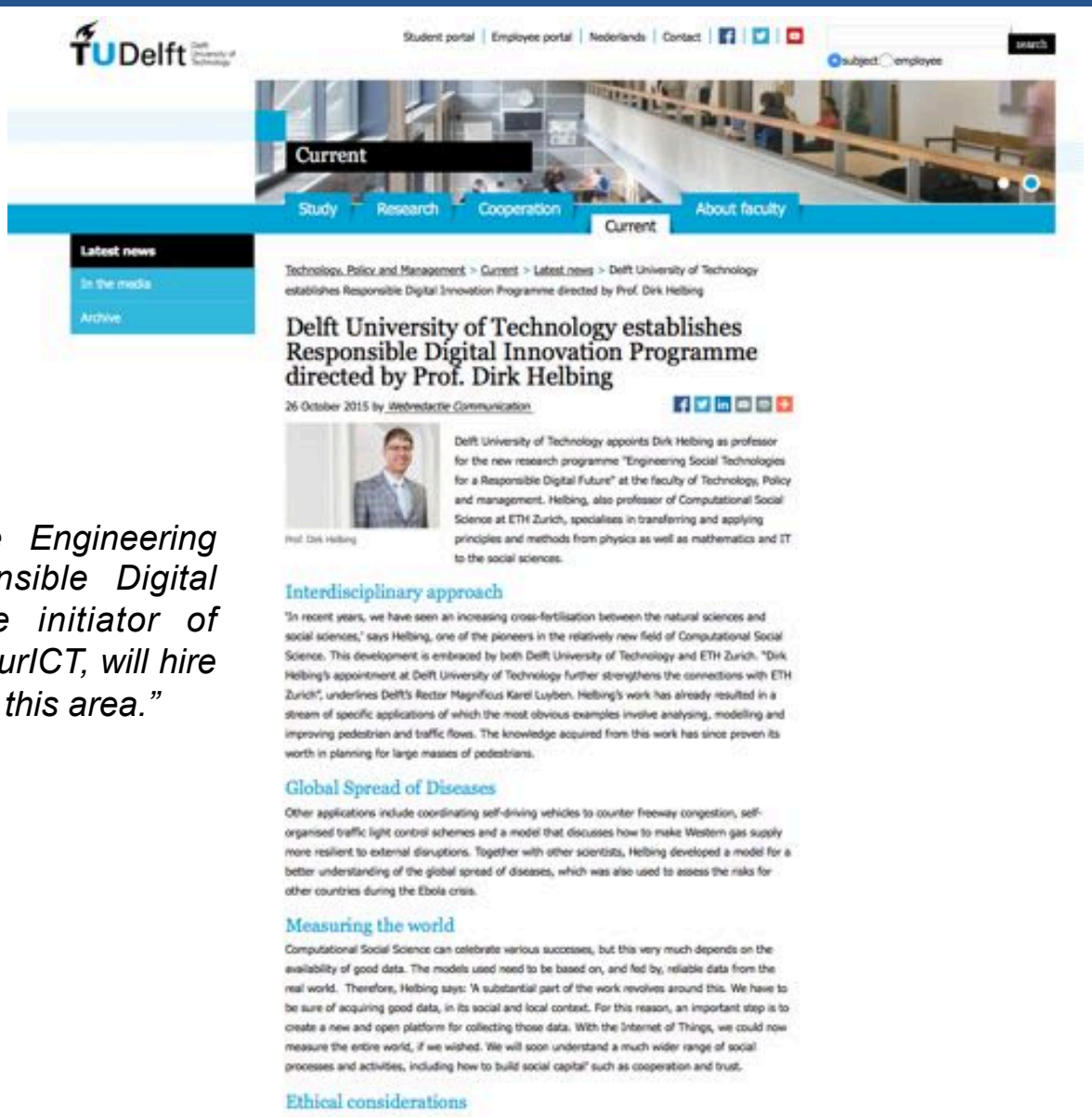
Citizen in the loop: Digital democracy by citizens for citizens

ETH zürich

Announcements & Remarks

Research Hubs

"As coordinator of the programme Engineering Social Technologies for a Responsible Digital Future, Helbing, who is also the initiator of Nervousnet and the coordinator of FuturICT, will hire ten PhD candidates to work actively in this area."



TU Delft

Student portal | Employee portal | Nederlands | Contact

search

subject employee

Current

Study | Research | Cooperation | About faculty

Current

Latest news

In the media

Archive

Technology, Policy and Management > Current > Latest news > Delft University of Technology establishes Responsible Digital Innovation Programme directed by Prof. Dirk Helbing

Delft University of Technology establishes Responsible Digital Innovation Programme directed by Prof. Dirk Helbing

26 October 2015 by Webredactie Communication

Prof. Dirk Helbing

Delft University of Technology appoints Dirk Helbing as professor for the new research programme "Engineering Social Technologies for a Responsible Digital Future" at the faculty of Technology, Policy and management. Helbing, also professor of Computational Social Science at ETH Zurich, specialises in transferring and applying principles and methods from physics as well as mathematics and IT to the social sciences.

Interdisciplinary approach

'In recent years, we have seen an increasing cross-fertilisation between the natural sciences and social sciences,' says Helbing, one of the pioneers in the relatively new field of Computational Social Science. This development is embraced by both Delft University of Technology and ETH Zurich. "Dirk Helbing's appointment at Delft University of Technology further strengthens the connections with ETH Zurich", underlines Delft's Rector Magnificus Karel Luyben. Helbing's work has already resulted in a stream of specific applications of which the most obvious examples involve analysing, modelling and improving pedestrian and traffic flows. The knowledge acquired from this work has since proven its worth in planning for large masses of pedestrians.

Global Spread of Diseases

Other applications include coordinating self-driving vehicles to counter freeway congestion, self-organised traffic light control schemes and a model that discusses how to make Western gas supply more resilient to external disruptions. Together with other scientists, Helbing developed a model for a better understanding of the global spread of diseases, which was also used to assess the risks for other countries during the Ebola crisis.

Measuring the world

Computational Social Science can celebrate various successes, but this very much depends on the availability of good data. The models used need to be based on, and fed by, reliable data from the real world. Therefore, Helbing says: 'A substantial part of the work revolves around this. We have to be sure of acquiring good data, in its social and local context. For this reason, an important step is to create a new and open platform for collecting those data. With the Internet of Things, we could now measure the entire world, if we wished. We will soon understand a much wider range of social processes and activities, including how to build social capital' such as cooperation and trust.

Ethical considerations

http://www.tbm.tudelft.nl/en/current/latest-news/article/detail/tu-delft-introduceert-engineering-social-sciences-for-a-responsible-digital-future-programma-o-1/

DSS Workshop

2nd International Workshop on Data-driven Self-regulating Systems

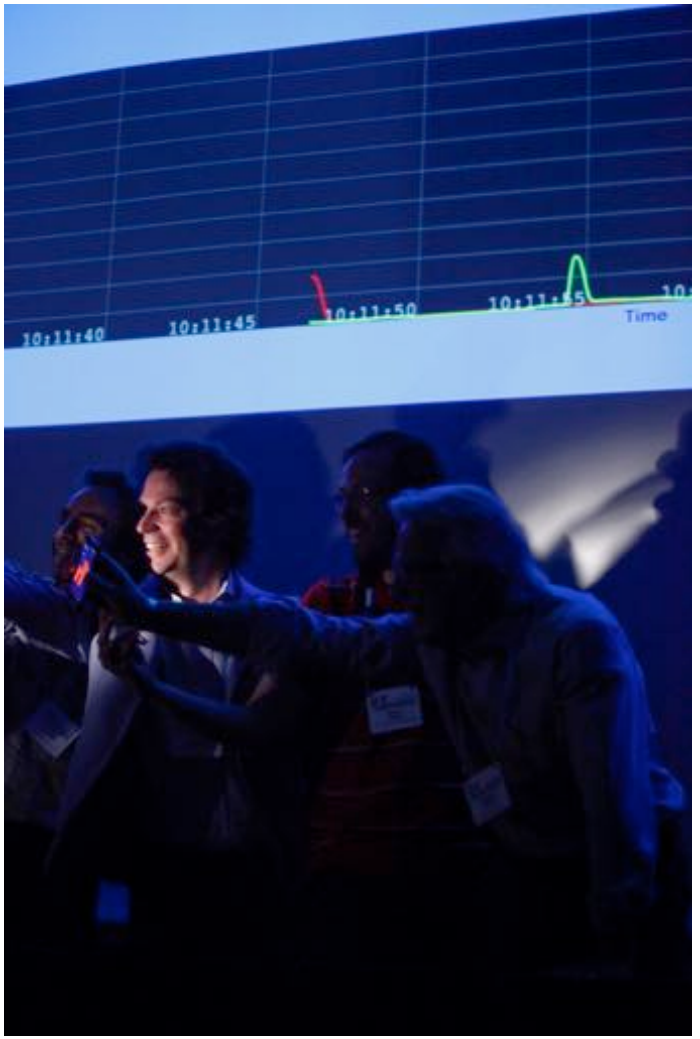
http://dss2015.inn.ac

Demos

Swarmpulse

https://www.youtube.com/watch?v=IU_3XKqWesE&feature=youtu.be

Nervousnet Competition Game

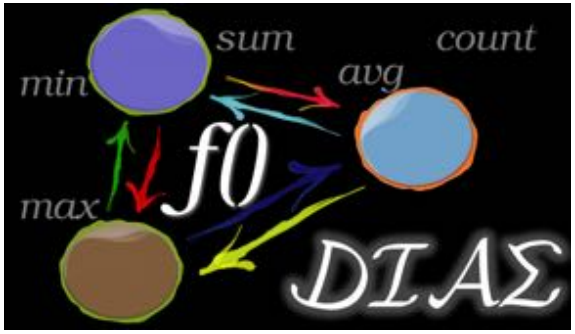


Questions?

nervousnet.ethz.ch

http://www.sobigdata.eu

http://www.swarmpulse.net



[1] Dirk Helbing and Evangelos Pournaras, *Build Digital Democracy*, Nature, Vol. 527, pp. 33-34, 2015 © Macmillan Publishers Limited

[2] Evangelos Pournaras, Izabela Moise and Dirk Helbing, *Privacy-preserving Ubiquitous Social Mining via Modular and Compositional Virtual Sensors*, in the proceedings of the 29th IEEE International Conference on Advanced Information Networking and Applications-AINA-2015, pages 332-338, Gwangju, South Korea, March 2015. © IEEE

[3] Evangelos Pournaras, Jovan Nikolic, Pablo Velasquez, Marcello Trovati, Nik Bessis and Dirk Helbing, *Self-regulatory Information Sharing in Participatory Social Sensing, submitted*

[4] Evangelos Pournaras, Martijn Warnier and Frances M.T. Brazier, *A Generic and Adaptive Aggregation Service for Large-scale Decentralized Networks*, Complex Adaptive Systems Modeling, 1:19, 2013 © SpringerOpen

epournaras@ethz.ch www.evangelospournaras.com