



## **Data and youth policy:**

How to use big data and data science to improve (digital) youth work innovatively and successfully?

**Innar Liiv, PhD**

Associate Professor (Data Science)

Tallinn University of Technology

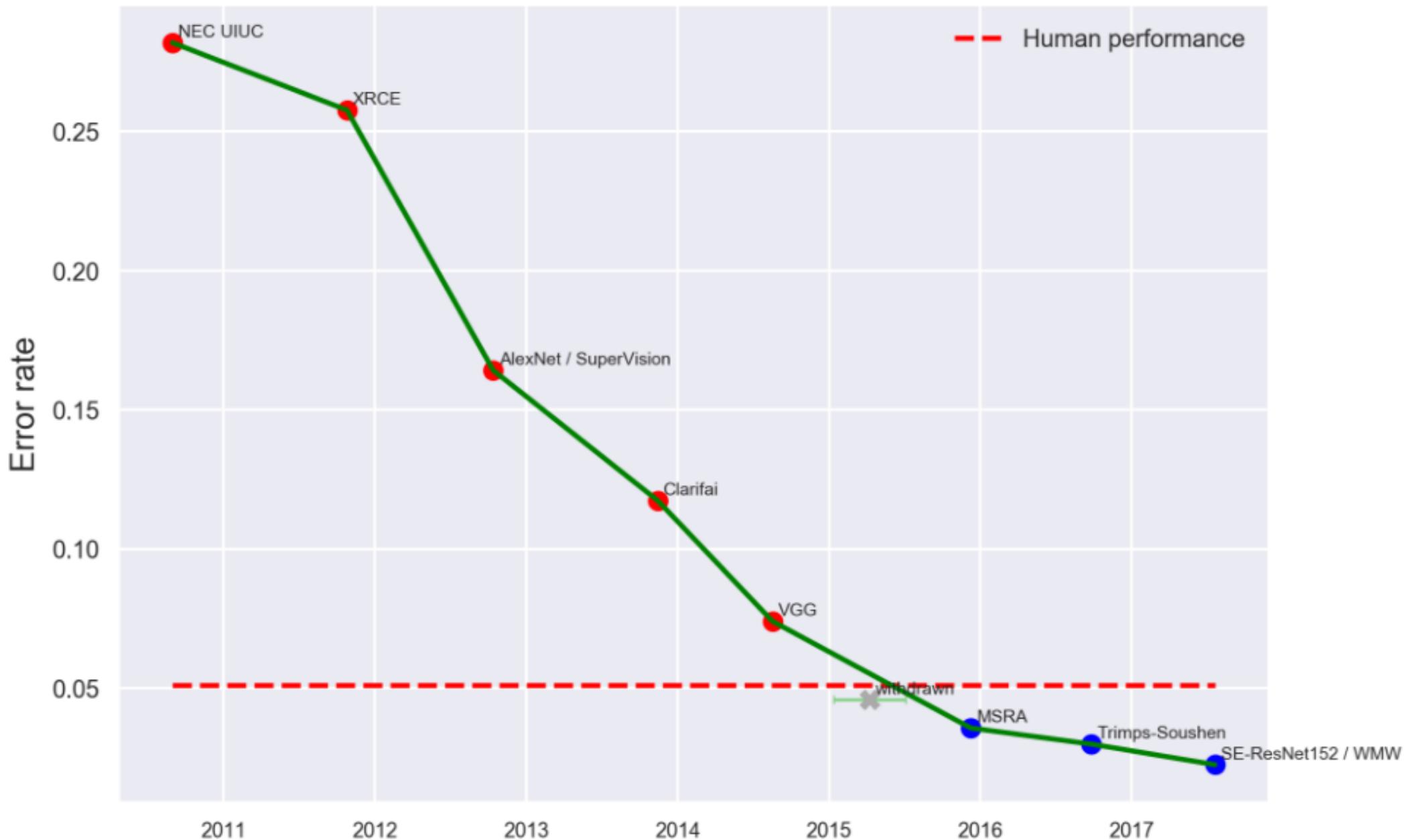
Email: [innar@innar.com](mailto:innar@innar.com) Twitter: [@innarliiv](https://twitter.com/innarliiv)

# Observation: A lot of discussion on digitalization, very little discussion on data analytics.

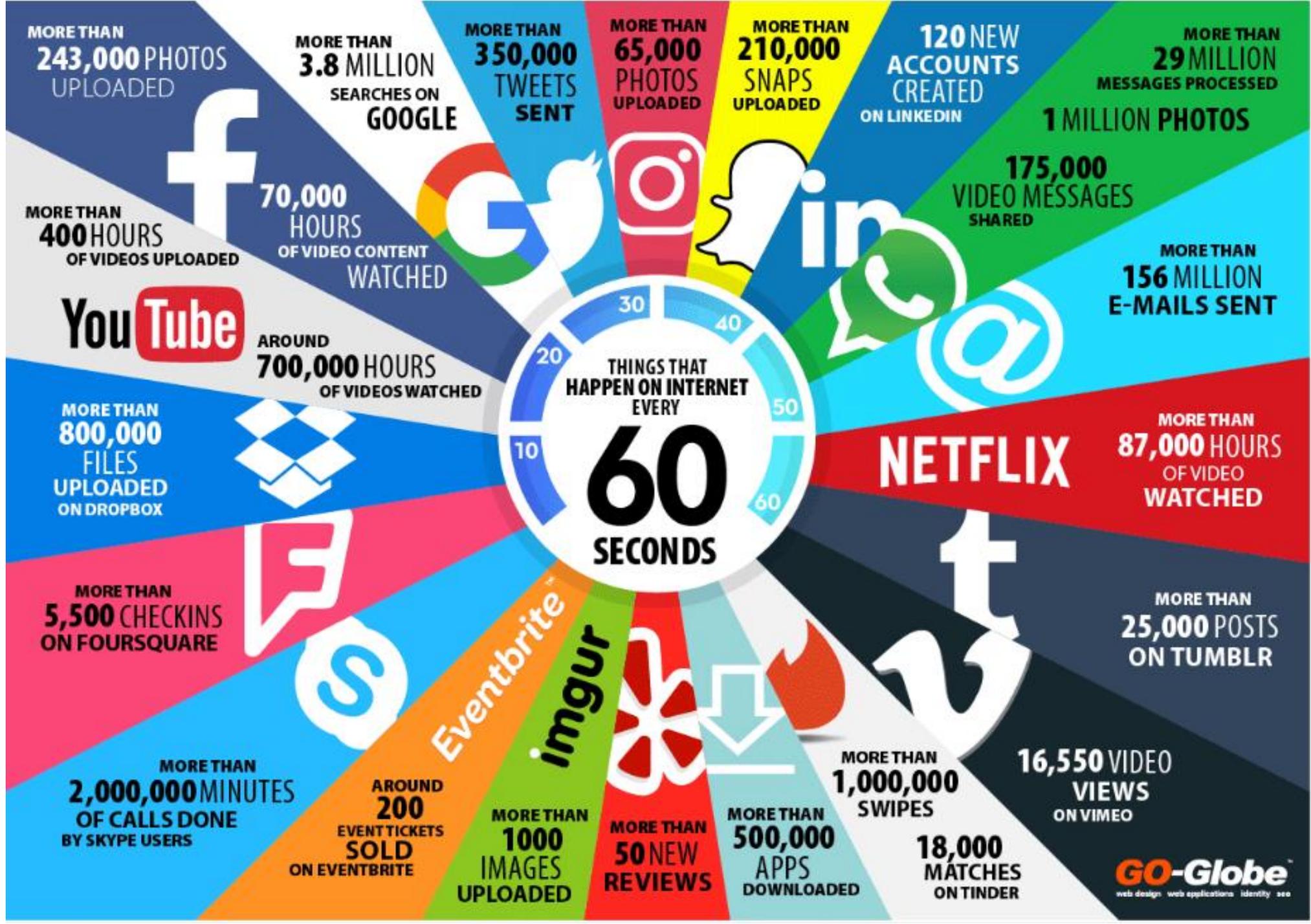
- Access to services, youth-friendly services, services for social inclusion;
- Reaching out to young people and the digital divide;
- Resilience and empowerment for social inclusion;
- Discrimination in the digital space

2

# Imagenet Image Recognition



Source: <https://www.eff.org/ai/metrics>

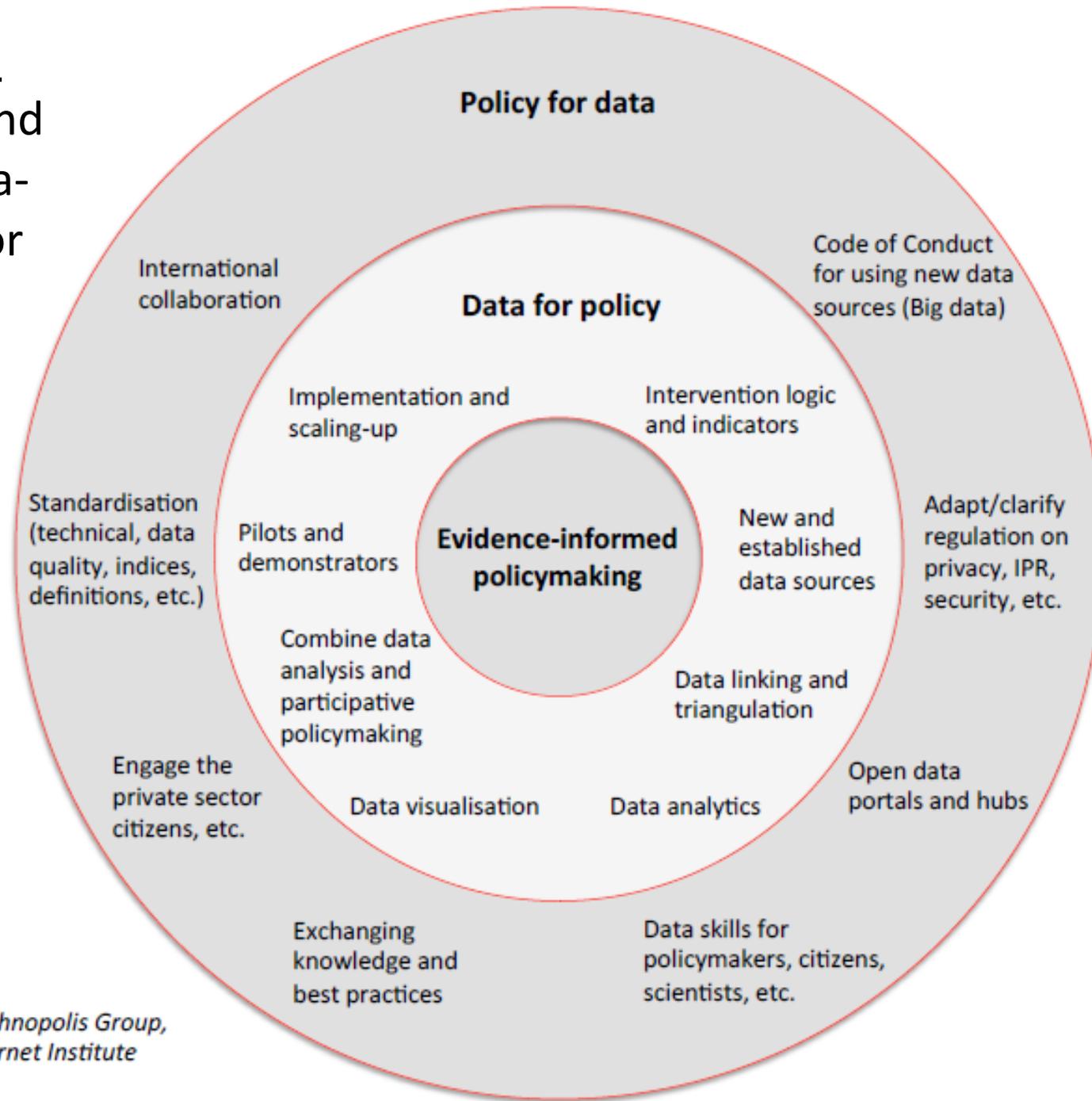


# Data Analytics and Science trends

- Social Data Revolution
- Andreas Weigend,
  - ex-Chief Scientist @ Amazon.com
  - Today a professor @Stanford, UC Berkeley, Tsinghua
- Two data “revolutions”
  - Classic: data is collected on need-basis
  - 1<sup>st</sup> wave: lots of data created everywhere, secondary data, making sense of data
  - 2<sup>nd</sup> wave: people themselves (when motivated!) push new information about their current state/emotions/preferences/needs (statuses, twitter, linkedin, facebook etc)

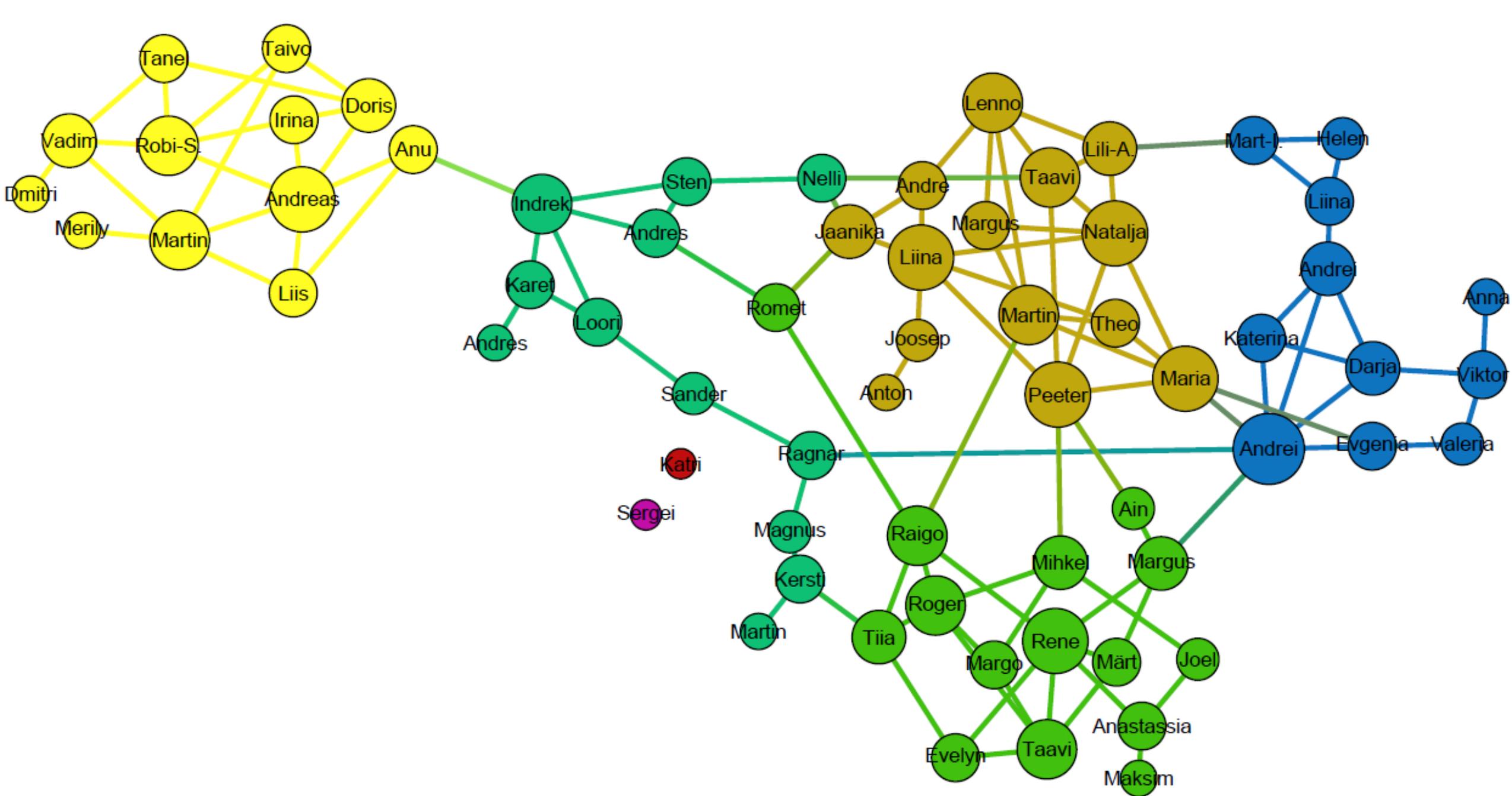
# data4policy.eu:

A study of big data and other innovative data-driven approaches for evidence-informed policy making



The European Commission has commissioned Technopolis Group, Oxford Internet Institute (OII) and the Centre for European Policy Studies (CEPS) to conduct an international study on innovative data-driven approaches to inform policymaking.

Source: Technopolis Group,  
Oxford Internet Institute  
and CEPS





## Mailis Reps

@Mailis\_Reps

Mailis Reps, Estonian politician, Estonian Centre Party Deputy Chairman and member of the board, and from 23 November 2016, Minister of Education and Research.

Tweets  
**67**

Following  
**84**

Followers  
**168**

Likes  
**28**

**Tweets**

**Tweets & replies**

**Media**

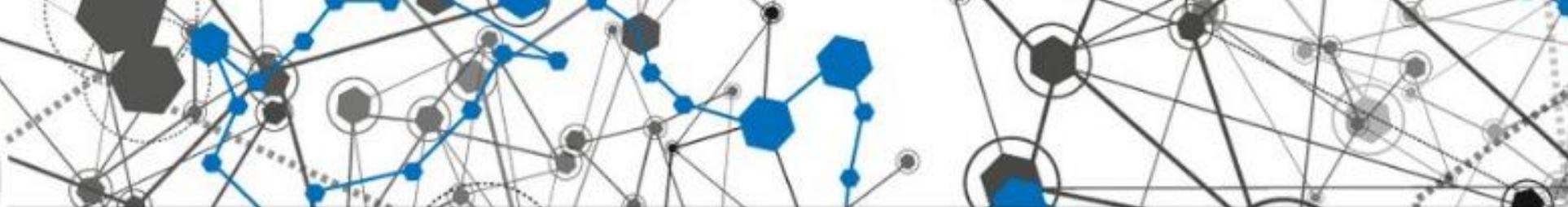


**Mailis Reps** @Mailis\_Reps · Jun 18

Õpetajad ootavad kooliit sama, mida teistelt tööandjatelt: karjäärivõimalusi, pайдlikkust, head töökeskkonda ja panusele vastavat palka. Viimasest alustades saan öelda, et järgmisel aastal tõuseb õpetajate keskmine palk 1500 euroni!

 Translate Tweet





Tweets **67** Following **84** Followers **168** Likes **28**

[Follow](#)

## Mailis Reps

@Mailis\_Reps

Mailis Reps, Estonian politician, Estonian Centre Party Deputy Chairman and member of the board, and from 23 November 2016, Minister of Education and Research.

[facebook.com/repmailis/](https://facebook.com/repmailis/)

Joined March 2017

[Tweet to Mailis Reps](#)

11 Photos and videos




**Mailis Ostra**  
@mailisostra

[Follow](#)



**CyprusInno**  
@CyprusInno

The first & largest bi-zonal, intercommunal digital platform connecting entrepreneurs in divided #Cyprus. Join us on our journey to...

[Follow](#)



**Grazia Cannarsa**  
@GraziaCannarsa

[Follow](#)



**Sylwia Radomska**  
@RadomskaSylwia

#Nowoczesna / Individual Member of @ALDEParty

[Follow](#)



**Ott Lumi**  
@ottlumi

Black belt in complicated questions of life and science [fipra.com](http://fipra.com), [metaadvisory.ee](http://metaadvisory.ee)

[Follow](#)



**keskerakond.eu**  
@keskerakond\_eu

[Follow](#)

**Appearance** x

Nodes Edges

Unique Attribute

#c0c0c0

Apply

**Layout** x

---Choose a layout

Run

Threads

Threads number 3

Behavior Alternatives

Dissuade Hubs

LinLog mode

Prevent Overlap

Edge Weight Influence 1.0

Tuning

Scaling 10.0

Stronger Gravity

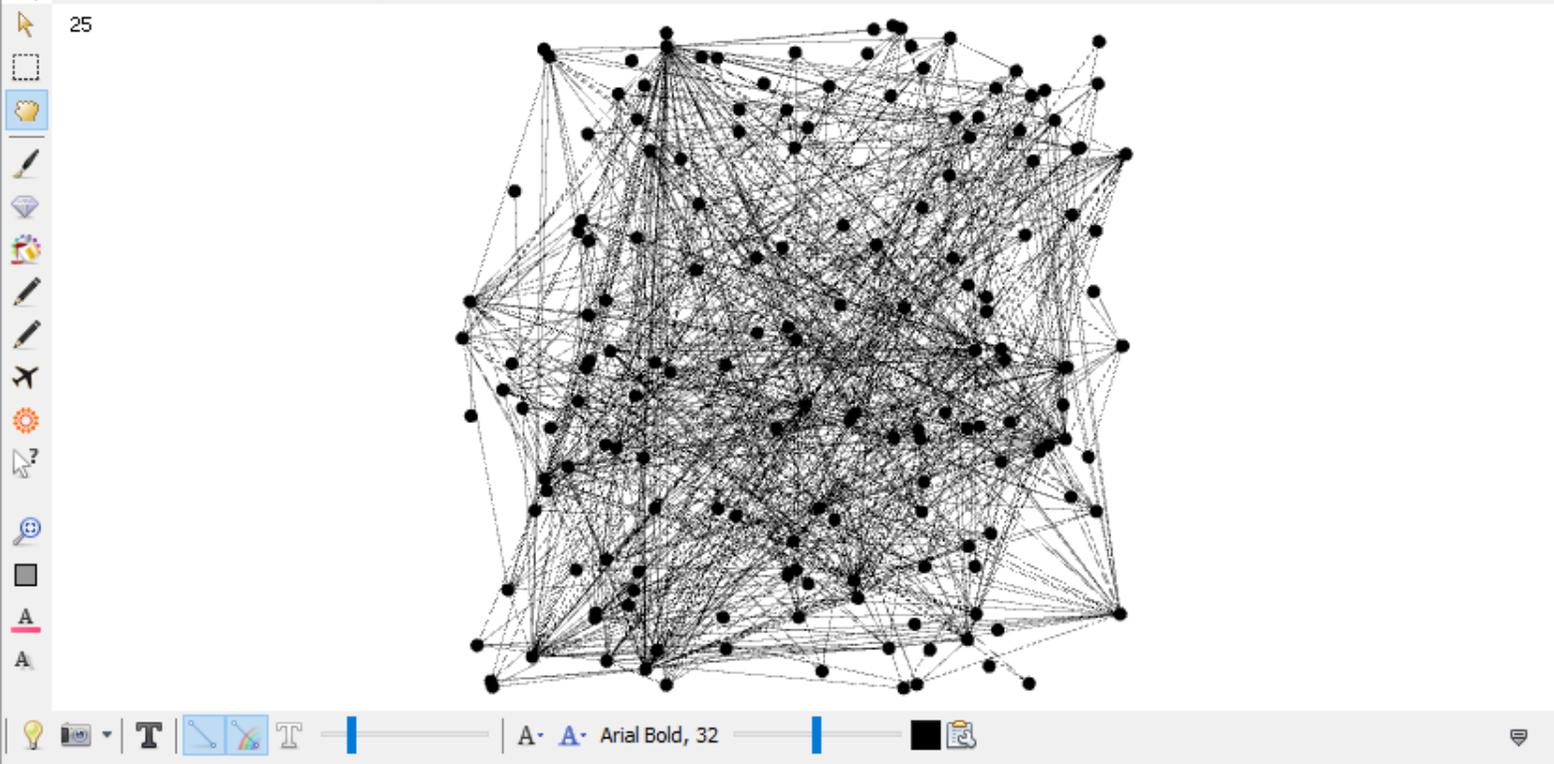
**ForceAtlas 2**

Presets... Reset

**Graph** x

Dragging (Configure)

25



Global Edges Labels

Background color:  Zoom Highlight selection

Autoselect neighbor

Arial Bold, 32

**Context** x

**Nodes:** 168

**Edges:** 754

Undirected Graph

**Filters** **Statistics** x

Settings

**Network Overview**

Average Degree Run

Avg. Weighted Degree Run

Network Diameter Run

Graph Density Run

Modularity Run

PageRank Run

Connected Components Run

**Node Overview**

Avg. Clustering Coefficient Run

Eigenvector Centrality Run

**Edge Overview**

Avg. Path Length Run

**Dynamic**

**Appearance**

Nodes Edges

Unique Attribute

Betweenness Centrality

Min size: 10 Max size: 40

Spline...

Apply

**Layout**

ForceAtlas 2

Run

Threads

Threads number: 3

Behavior Alternatives

Dissuade Hubs:

LinLog mode:

Prevent Overlap:

Edge Weight Influence: 1.0

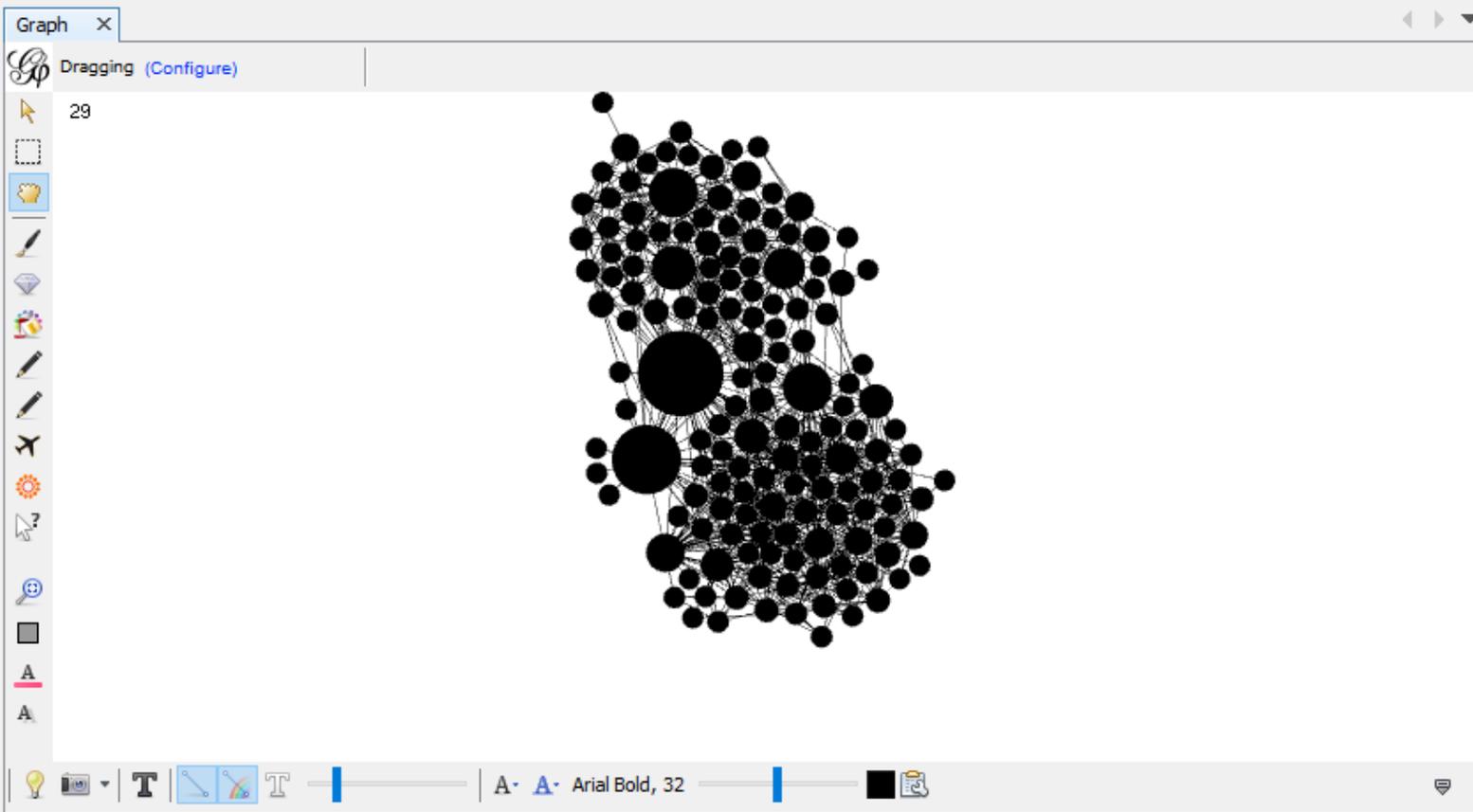
Tuning

Scaling: 2.0

Stronger Gravity:

ForceAtlas 2

Presets... Reset



**Context**

Nodes: 149

Edges: 754

Undirected Graph

**Filters Statistics**

Settings

**Network Overview**

Average Degree	8.976	Run
Avg. Weighted Degree	8.976	Run
Network Diameter	5	Run
Graph Density	0.054	Run
Modularity	0.372	Run
PageRank		Run
Connected Components	20	Run

**Node Overview**

Avg. Clustering Coefficient		Run
Eigenvector Centrality		Run

**Edge Overview**

Avg. Path Length	2.586	Run
------------------	-------	-----

**Dynamic**

**Appearance**

Nodes Edges

Unique Attribute

Modularity Class

- 3 (48.99%)
- 0 (31.54%)
- 13 (19.46%)

Apply

**Layout**

Contraction

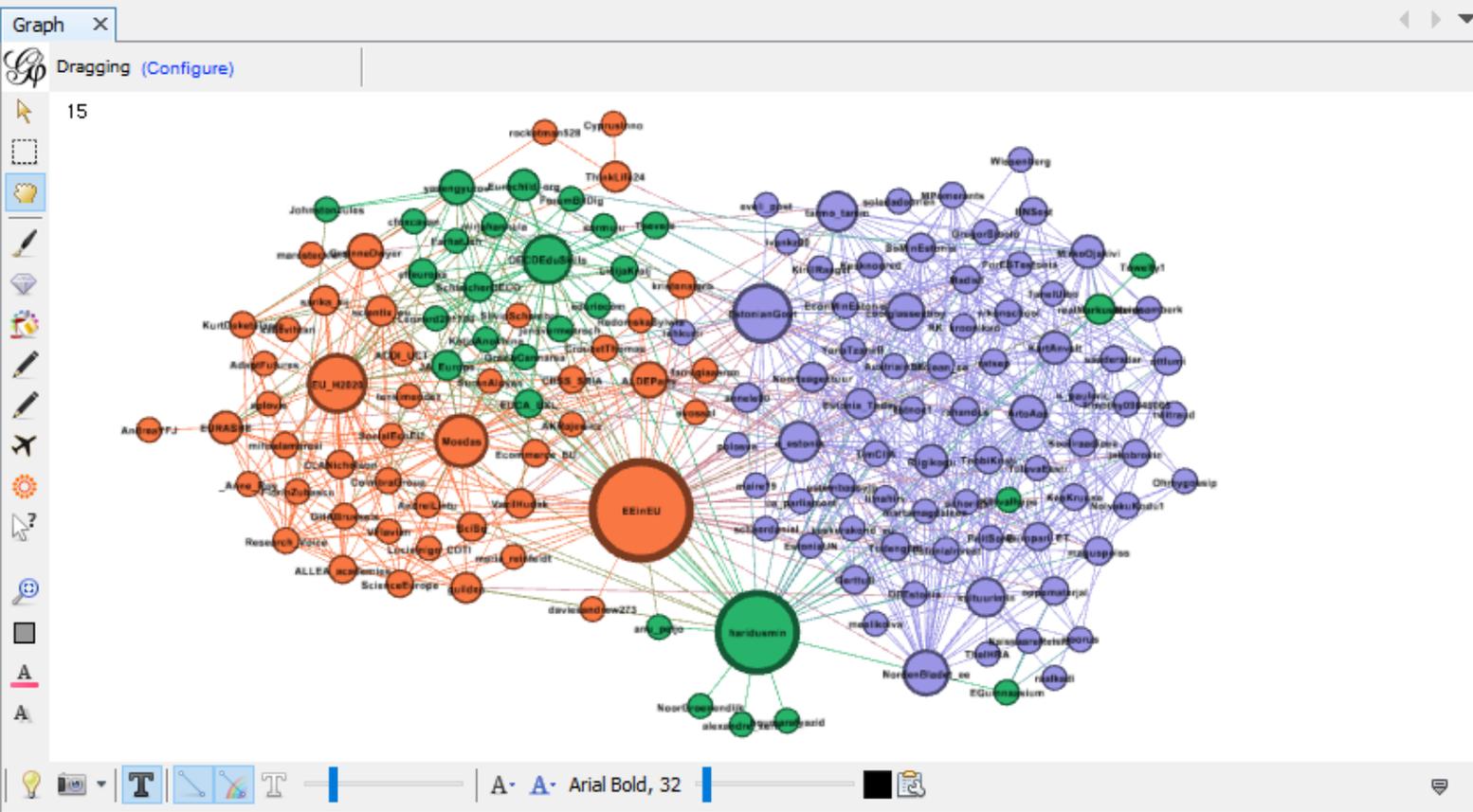
Run

properties

Scale factor 0.8

Contraction

Presets... Reset



**Context**

Nodes: 149

Edges: 754

Undirected Graph

**Filters Statistics**

Settings

**Network Overview**

Average Degree	8.976	Run
Avg. Weighted Degree	8.976	Run
Network Diameter	5	Run
Graph Density	0.054	Run
Modularity	0.372	Run
PageRank		Run
Connected Components	20	Run

**Node Overview**

Avg. Clustering Coefficient	Run
Eigenvector Centrality	Run

**Edge Overview**

Avg. Path Length	2.586	Run
------------------	-------	-----

**Dynamic**



# UN Global Pulse taxonomy

of novel big data sources

- A. Data exhaust** – passively collected data from people’s use of digital services ([mobile phones](#), credit card, other logs)
- B. Online activity** – Twitter, Search Engines, blogs, sentiments
- C. Sensing Technologies** – satellite or infrared of changing landscapes, traffic patterns, light emissions, urban development (remote and personal sensing)
- D. Citizen Reported or Crowd-sourcing** – humanitarian emergencies, mobile phone-based surveys, [user generated maps](#)



# UNITED NATIONS GLOBAL PULSE

Harnessing big data for development and humanitarian action

Search  SEARCH



ABOUT

PROJECTS

LABS

NEWS

CHALLENGES

PRIVACY

PARTNERSHIPS

CONTACT

HOME

## Projects

Welcome to the repository of Global Pulse's projects. Find out more about collaborative research, prototypes and experiments analyzing digital data to support global development and humanitarian action.



Monitoring In Real Time The Implementation Of HIV Mother-To-Child Prevention Programme



Understanding Immunisation Awareness And Sentiment Through Analysis Of Social Media And News Content (2015)



Measuring Poverty With Machine Roof Counting



Feasibility Study: Crowdsourcing High-Frequency Food Price Data In Rural Indonesia

## BROWSE BY LAB

Jakarta Kampala New York

## BROWSE BY PROGRAMME

- Climate & Resilience
- Data Privacy & Protection
- Economic Well-being
- Food & Agriculture
- Gender
- Humanitarian Action
- Post-2015
- Public Health
- Real-time Evaluation

## BROWSE BY REGION

SUBSCRIBE TO OUR NEWSLETTER

Photo



Satellite image



# ~2.000.000.000 CDRs/day

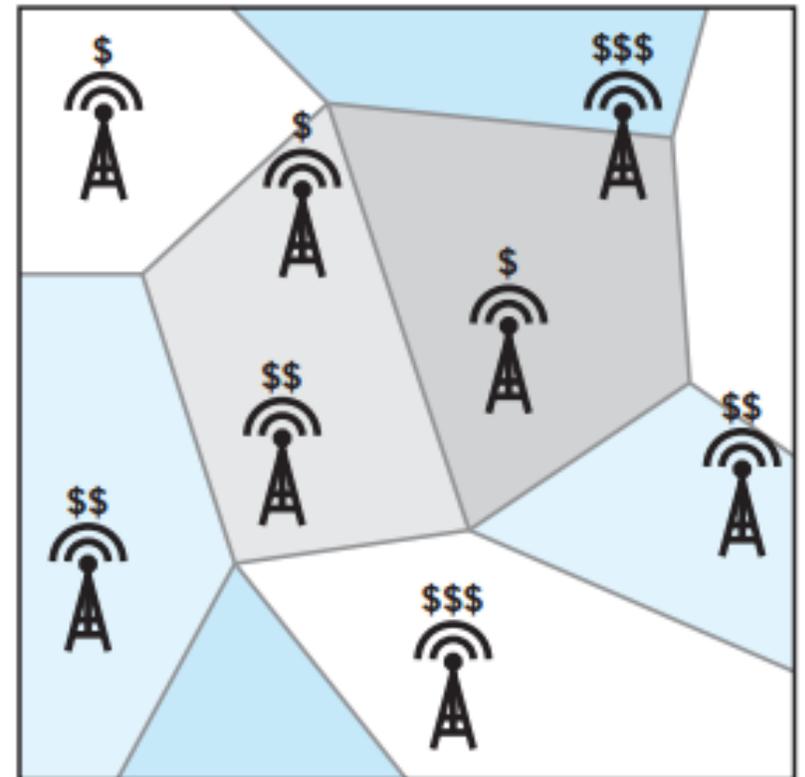
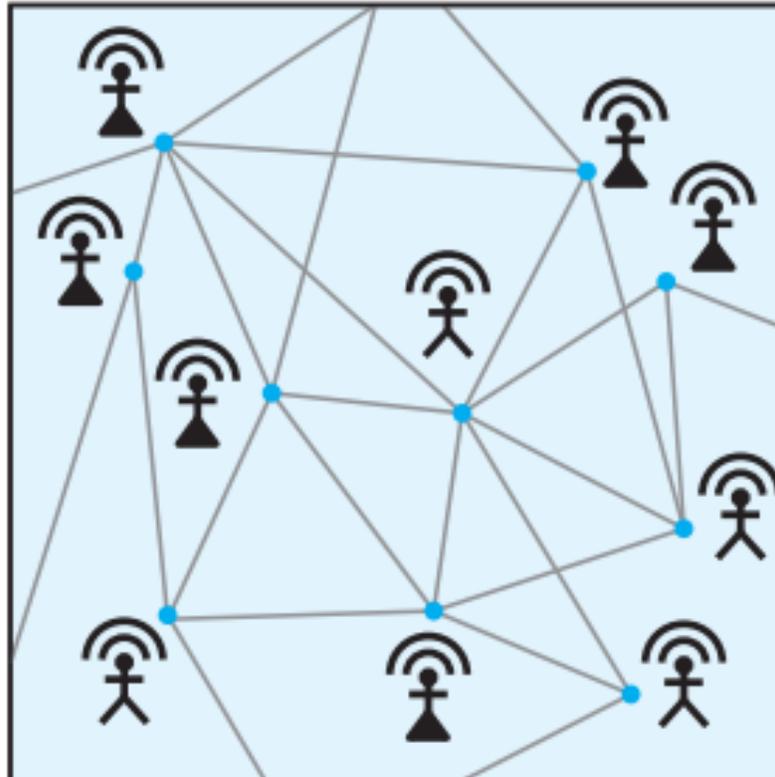
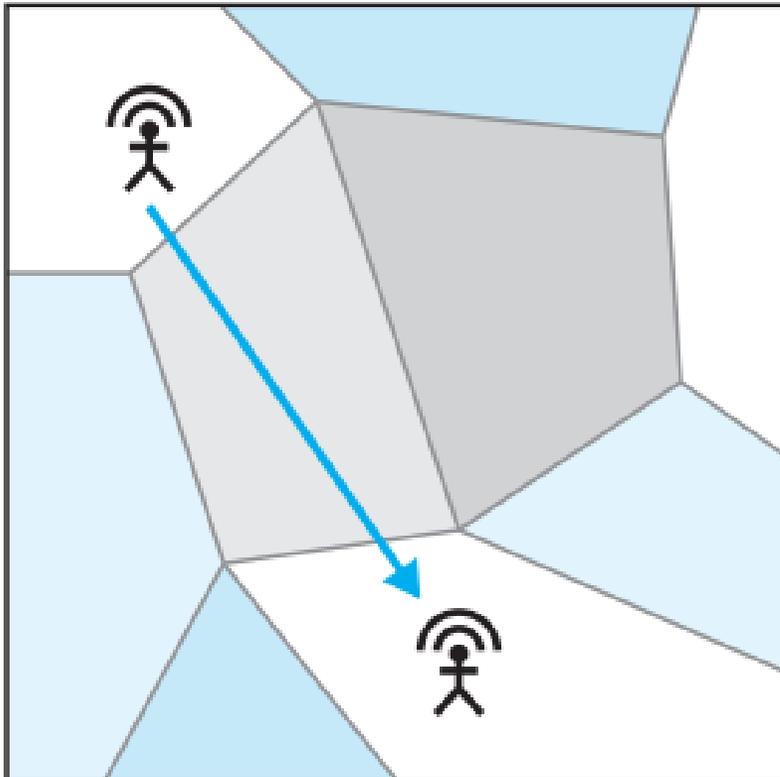


~300 GBs/day  
~160 bytes/each



Source: [iol.telecomitalia.com/](http://iol.telecomitalia.com/)

CALLER ID	CALLER CELL TOWER LOCATION	RECIPIENT PHONE NUMBER	RECIPIENT CELL TOWER LOCATION	CALL TIME	CALL DURATION
X76VG588RLPQ	2°24' 22.14", 35°49' 56.54"	A81UTC93KK52	3°26' 30.47", 31°12' 18.01"	2013-11-07T15:15:00	01:12:02



# WHAT CAN YOUR PHONE TELL ABOUT YOU?

Yves-Alexandre de Montjoye, Jordi Quoidbach,  
Florent Robic, Alex "Sandy" Pentland



**H**ow much can someone know about your personality just by looking at the way you use your phone? We provide the first evidence that personality can be reliably predicted from standard mobile phone logs. Using a set of novel psychology-informed indicators that can be computed from data available to all carriers, we were able to predict users' personalities. As mobile cellular subscriptions have hit six billion throughout the world, our method enables cost-effective, questionnaire-free investigation of personality-related questions at a large-scale, opening exciting avenues for the use of personality to better understand customers.



## Personality: The Big 5



**Neuroticism** is the tendency to experience unpleasant emotions easily.



**Openness** is the tendency to be intellectually curious, creative, and open to feelings.



**Extraversion** is the tendency to seek stimulation in the company of others, to be outgoing and energetic.



**Agreeableness** is the tendency to be warm, compassionate, and cooperative.



**Conscientiousness** is the tendency to show self-discipline, be organized, and aim for achievement.

36 indicators  
5 categories



Location



Basic  
Phone Use



Diversity



Regularity



Active User  
Behaviors

## Neuroticism



Daily distance  
travelled

## Extraversion



The richness and  
evenness of a  
user's text contacts

## Openness



Average time be-  
tween sent and  
received texts

## Agreeableness



The richness and  
evenness of a  
user's text contacts

## Conscientiousness



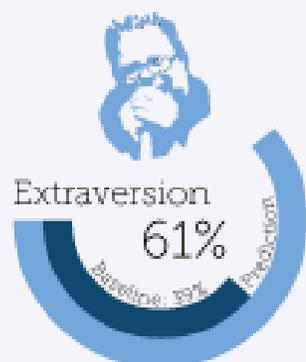
Time regularity be-  
tween phone calls

## How Well We Predict

**W**e predict if a user has a low, medium, or high score for each of the dimensions. The percentages are the likelihood of us predicting correctly a user's personality trait using our psychology-based indicators. The baseline is how well we'd be doing at random.



1.7x better  
than random



1.6x better  
than random



1.4x better  
than random



1.4x better  
than random



1.3x better  
than random

# Questions for brainstorming

- What are the relevant new data sources?
- How can we use them?
- What should we do with the information? Who cares? Which political decisions need faster information from novel sources? Do we need faster information? Does it come with unanticipated risks?

virtual youth work, internet-based youth work, cyber-exclusion, cyber-bullying