BIG DATA: NEW CHALLENGES FOR LAW AND ETHICS
International scientific conference
22 - 23 May 2017
Faculty of Law, University of Ljubljana

We don't know what the Questions are, but we know we're gonna find the Answers

Alexander Czadilek
Policy Analyst

epicenter.works
Digital Rights NGO, Vienna
https://www.epicenter.works

Christof Tschohl
Scientific Director

Walter Hötzendorfer
Senior Researcher

Research Institute AG & Co KG
Digital Human Rights Center, Vienna
http://www.researchinstitute.at

RESEARCH INSTITUTE
DIGITAL HUMAN RIGHTS CENTER

- **Research Institute (RI):** Fundamental and applied research at the **intersection of technology, law and society** from a multi- and interdisciplinary perspective
- What do human rights mean in the digital era?
- Portfolio:
  - **Research** on technological and legal aspects of privacy and data protection, information security, cyber security, cybercrime etc.
  - **Smart.Rights.Consulting**: Putting data protection into practice – consulting in organisational and technical implementation of data protection law
  - **Education**: Trainings, seminars and lectures
  - Development of data protection management **software** and tailored software solutions

EPICENTER.WORKS

- 2009 "**AKVorrat**" (Arbeitskreis Vorratsdaten – Working Group on Data Retention Austria)
- April 2014 CJEU: **Annulment of Data Retention Directive**
- June 2014 VfGH (Austrian Constitutional Court): Annulment of Data Retention in Austria
- Since 2016 "**epicenter.works**"
  - Net Neutrality - SaveTheInternet.eu
  - Campaing on "Surviellance Package" of the Austrian federal government

*"**Big data is like teenage sex:***

- *everyone talks about it,*
- *nobody really knows how to do it,*
- *everyone thinks everyone else is doing it,*
- *so everyone claims they are doing it..."*

(Dan Ariely, Duke University)

*"**Big data is like teenage sex:***

- *everyone talks about it,*
- *nobody really knows how to do it,*
- *everyone thinks everyone else is doing it,*
- *so everyone claims they are doing it..."*

(Dan Ariely, Duke University)

…and if you don't be careful, it can go terribly wrong!

Big Data

Volume	• Increasing volume of data (Petabyte) • Challenge: administration and information retrieval
Variety	• Different sources and formats • Complex and often unstructured
Velocity	• Requirement for timely processing • Sensor data und real-time systems
Value	• Use of data and retrieved information • Public, economic and private interests

The Enablers of Big Data

- **Datafication**: Enormous amounts of data are generated every day
  - Smart phones, smart TVs, smart watches etc.
  - Internet use
  - Sensors in cars, machines etc.
  - Internet of Things
  - Surveillance
- **Capacity and methods** for handling and analysing this data

Properties of Big Data

- Analysis of the whole data, not only of a random sample
- „What?“ not „Why?“: correlation not causality
- Possibility to deduce sensitive information from common and trivial data

The definition of Big Data is not clear-cut.

At its core, Big Data starts where we stop asking particular questions and see what algorithms are able to find in the data

research institute

EPICENTER .WORKS for digital rights

algorithm | ˈalgərɪð(ə)m |

noun

a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer: *a basic **algorithm for** division*.

DERIVATIVES

algorithmic | algəˈrɪðmɪk | adjective ,

algorithmically | algəˈrɪðmɪk(ə)li | adverb

ORIGIN

late 17th cent. (denoting the Arabic or decimal notation of numbers): variant (influenced by Greek *arithmos* **'number'**) of Middle English *algorism*, via Old French from medieval Latin *algorismus*. The Arabic source, *al-Ḵwārizmī* **'the man of Ḵwārizm'** (now Khiva), was a name given to the 9th-cent. mathematician Abū Ja'far Muhammad ibn Mūsa, author of widely translated works on algebra and arithmetic.

ALGORITHMS

- We always talk about algorithms and should not forget about the people behind them
- There are also self-learning algorithms where the logic is not designed by humans
- Emerging field of research: understanding the functionality of trained self-learning algorithms

Are we rapidly heading towards a future where we will be surrounded by “smart” machines on which we depend but which we cannot trust because we do not understand them?

Data Protection as a fundamental Right

- Art 8 CFR: Protection of personal data
  - 1. Everyone has the right to the protection of personal data concerning him or her.
  - 2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.
  - 3. Compliance with these rules shall be subject to control by an independent authority.
- Exists independently next to Art 7 CFR (Protection of private and family life), see also Art 8 ECHR
- **Data Protection is not an end in itself** but
  - a catalyst for the exercise of other fundamental rights
  - necessary for the functioning of democracy

Data Protection Legal Framework in Europe

- (EU) **General Data Protection Regulation** (GDPR) 2016/679
  - Replacing Data Protection Directive 95/46/EC from 25 May 2018 onwards
  - Similar aims and principles as Directive 95/46/EC
- Directive (EU) 2016/680 on the protection of personal data processed by competent authorities for the purposes of prevention, investigation, detection or prosecution of criminal offences (**Police Directive**/DPD-PJ)
  - Replacing Framework Decision 2008/977/JHA on the protection of personal data processed in the framework of police and judicial cooperation in criminal matters
  - First harmonised data protection framework in the field of law enforcement and justice (not only cross-border transfer of data but also domestic processing)
  - Must be transposed by Member States until 6 May 2018
- Directive 2002/58/EC on privacy and electronic communications (Commission Proposal: ePrivacy Regulation)

Big Data and the Purpose Limitation Principle

- Purpose limitation is a **fundamental principle of data protection law** (Art 5 (1) (b) GDPR)
- Big Data analysis is often based on data collected for different purposes
- "Ways out:"
  - Compatible purposes (Art 5 (1) (b) and Art 6 (4) GDPR)

ART 6 (4) GDPR: COMPATIBLE PURPOSES

*4. Where the processing for a **purpose other than that** for which the personal data have been **collected** is not based on the data subject's **consent** or on a Union or Member State **law** which constitutes a necessary and proportionate measure in a democratic society to safeguard the objectives referred to in Article 23(1), the controller shall, in order **to ascertain whether processing for another purpose is compatible with the purpose for which the personal data are initially collected**, **take into account**, **inter alia**:*

*(a) any **link between the purposes** for which the personal data have been collected and the purposes of the intended further processing;*

*(b) the **context** in which the personal data have been **collected**, in particular regarding the relationship between data subjects and the controller;*

*(c) the **nature of the personal** data, in particular whether special categories of personal data are processed, pursuant to Article 9, or whether personal data related to criminal convictions and offences are processed, pursuant to Article 10;*

*(d) the possible **consequences** of the intended further processing for data subjects;*

*(e) the existence of **appropriate safeguards**, which may include **encryption** or **pseudonymisation**.*

Big Data and the Purpose Limitation Principle

- Purpose limitation is a **fundamental principle of data protection law** (Art 5 (1) (b) GDPR)
- Big Data analysis is often based on data collected for different purposes
- "Ways out:"
  - Compatible purposes (Art 5 (1) (b) and Art 6 (4) GDPR)
    - Not only comparison of the purposes but processing as a whole
    - Further processing is lawful if there is no significant impact of the further processing on the rights and freedoms the data subjects

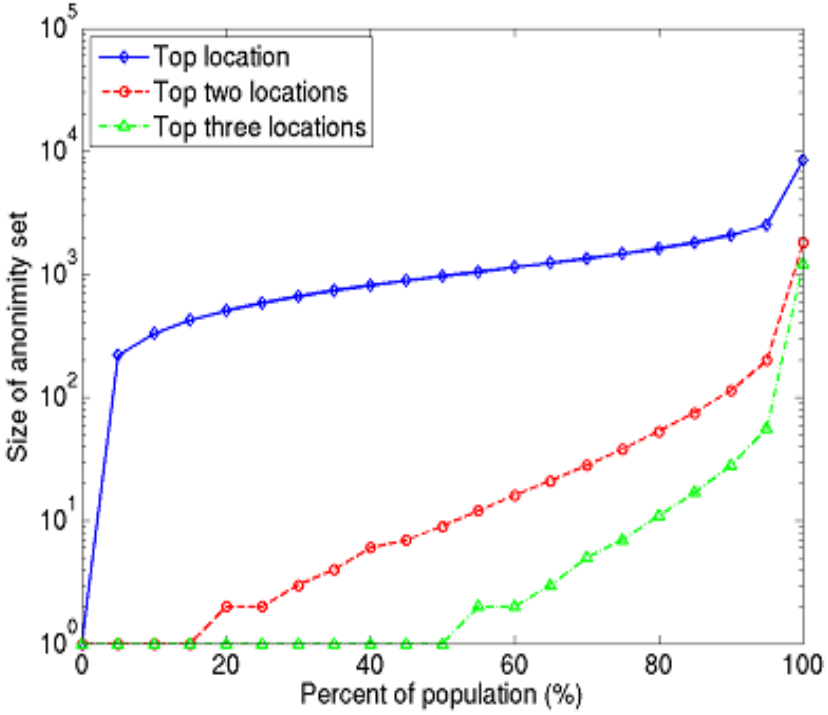
Big Data and the Purpose Limitation Principle

- Purpose limitation is a **fundamental principle of data protection law** (Art 5 (1) (b) GDPR)
- Big Data analysis is often based on data collected for different purposes
- "Ways out:"
  - Compatible purposes (Art 5 (1) (b) and Art 6 (4) GDPR)
    - Not only comparison of the purposes but processing as a whole
    - Further processing is lawful if there is no significant impact of the further processing on the rights and freedoms the data subjects
  - Scientific research (Art 5 (1) (b) and Art 89 (1) GDPR): appropriate safeguards, pseudonymisation (if purposes can still be fulfilled)
  - Consent of the data subjects (freely given, specific, informed): very unlikely
  - Anonymisation: often very difficult
- In (almost) any case: Privacy Impact Assessment necessary

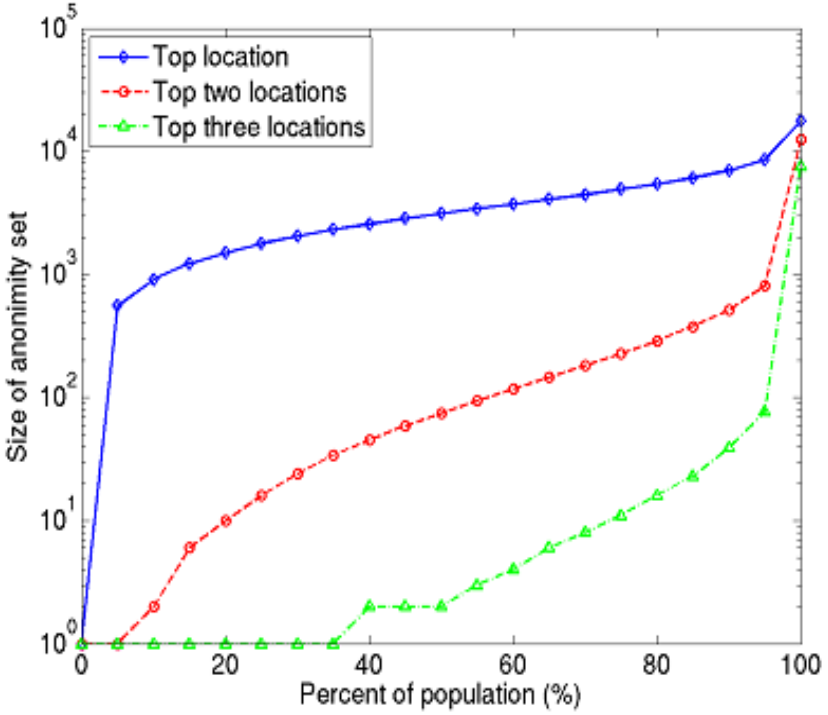
ANONYMISATION IS DIFFICULT

- …and Big Data makes it even harder:
  - Big Data increases the possibility of identifying persons
  - Public availability of a lot of personal data (on the internet)
- Erasure of obvious identifying attributes (like the name) will likely not be enough
- Individual combinations of other attributes can be very revealing
- Measures of Anonymisation:
  - k-Anonymity: At least k people have exactly the same combination of attributes in a dataset
  - l-diversity
  - t-closeness

EXCAMPLE: LOCATION ANONYMITY



(b) Cell



(c) Zip code

Hui Zang and Jean Bolot. 2011. Anonymization of location data does not work: a large-scale measurement study. In *Proceedings of the 17th annual international conference on Mobile computing and networking* (MobiCom '11). ACM, New York, NY, USA, 145-156.

Privacy by Design

- „Data protection by design“ as a new principle laid down in Art 25 GDPR / Art 20 Police Directive
- Systematic approach for compliance with data protection law and for mitigation/elimination of risks
- Duty to “implement appropriate technical and organisational measures“, in particular during design and implementation in order to
  - implement data-protection principles, such as data minimisation, in an effective manner and to integrate the necessary safeguards into the processing
  - preclude the privacy-infringing use of a system by technical and organisational means
- Most essential measure: **data minimisation**

Privacy by Design in Practice: Example of Mobile Phone Location Data

- AGETOR Project: Monitoring of the movements of crowds at large events through anonymised mobile phone location data
- Privacy by Design Measures:
  - Reduction of accuracy
    - location accuracy
    - time accuracy (reduction of precision of the value or of number of values)
  - Reduction of the monitoring period
  - Cancellation of runaway values (difficult)
  - Adding Noise/Dummy Data (very difficult)
  - Aggregation
- The measures must be very specific to the purpose of the processing, otherwise usefulness of the data decreases faster than the likelihood of re-identification

Privacy Impact Assessment
Art 35 GDPR

- Necessary prior to processing where a type of processing in particular **using new technologies**, and taking into account the **nature, scope, context** and **purposes** of the processing, **is likely to result in a high risk to the rights and freedoms of natural persons**.
- In particular (not exhaustively):
  - Automated systematic and extensive evaluation of personal aspects based on automated processing, including profiling, having legal or similarly significantly effects concerning natural persons
  - processing on a large scale of special categories of personal data relating to criminal convictions
  - Large-scale systematic monitoring of a publicly accessible areas
- Big Data applications will often fulfil these criteria

AUTOMATED INDIVIDUAL DECISION-MAKING, INCLUDING PROFILING

- Art 22 GDPR, Art 11 Police Directive, Art 15 DPD
- Decisions are prohibited which
  - are based solely on automated processing, including profiling and
  - produce legal similarly significant effects concerning the data subject
- Exceptions:
  - necessary for entering into, or performance of, a contract between the data subject and a data controller
  - authorised by Union or Member State law which lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests
  - explicit consent of the data subject
- Art 22 (4) GDPR: Not based on special categories of personal data
- Art 11 Police Directive: Not resulting in discrimination based on special categories of personal data

Open Source Intelligence (OSINT) And Predictive Policing

- Analysis of data in the public domain e.g. in criminal investigations or terrorism prevention
- Right to privacy also extends to data in public domain
  - ECtHR 3.5.2000, Rotaru v. Romania (28341/95)
  - CJEU 13.5.2014, Google Spain (C-131/12)
- Data quality (!)
- Art 10 Police Directive: Lawful where the data was manifestly made public by the data subject
- E.g. Germany is testing a software for automatically predicting the potential threat emanating form a person
- Who controls the algorithms? Usually OSINT software is developed by private companies
- Classical investigatory police work vs. software that is not fully understood
- Factual psychological effect of over-confidence in technology
- Privacy by Design and Privacy Impact Assessment

THE FACEBOOK ALGORITHMIC FACTORY

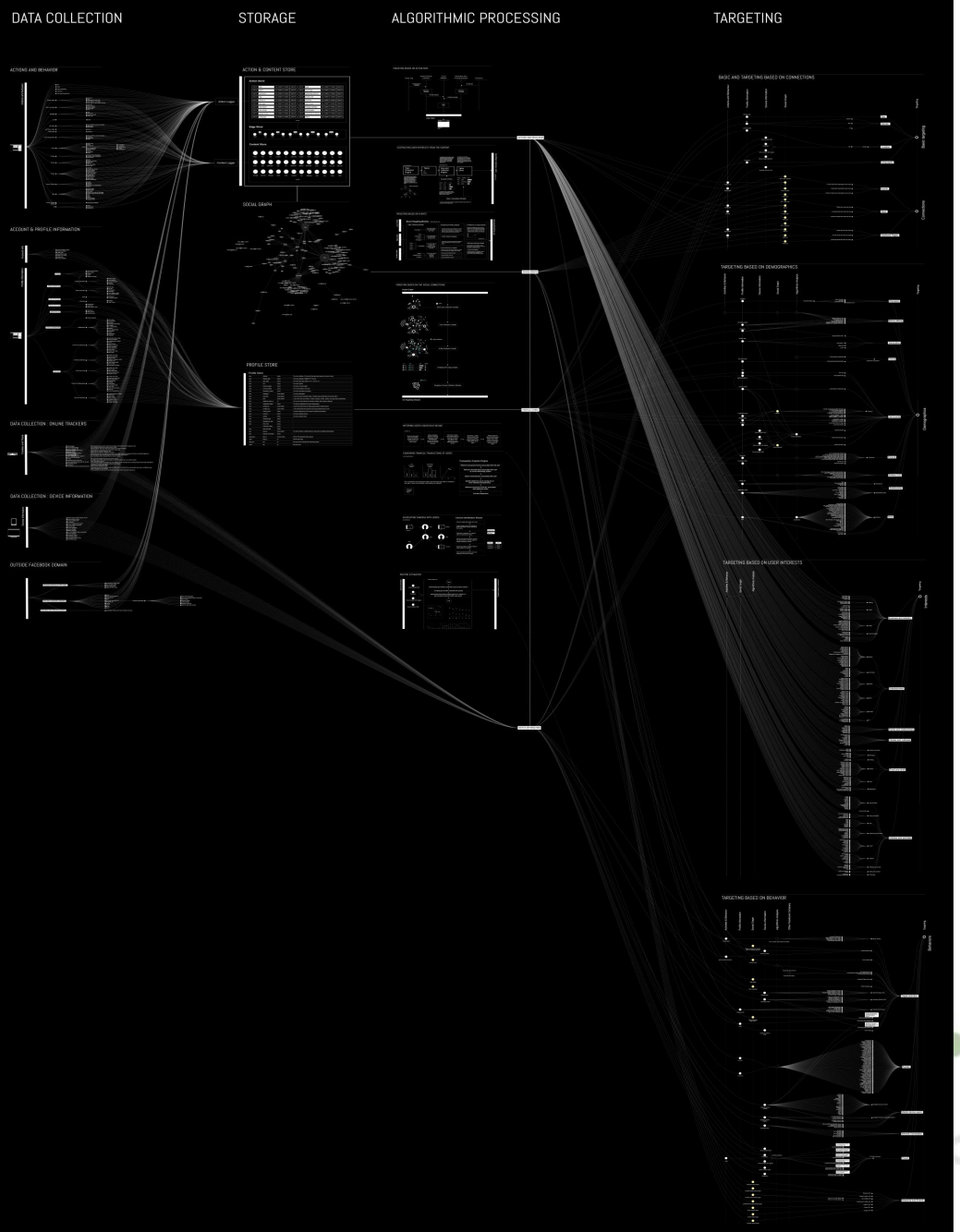
- Research of Share Labs, Belgrade https://labs.rs/en/
- Based on publicly available data (> 8000 patents and Graph API)
- Analysis of the Facebook Social Graph
  - Meta structure connecting all data in one structure
  - Objects (nodes) – connections (edges)
  - Action store – content store – edge store – profile store

research institute

EPICENTER .WORKS for digital rights

FACEBOOK RESEARCH
SHARE LAB – WWW.LABS.RS
2018

FACEBOOK ALGORITHMIC FACTORY

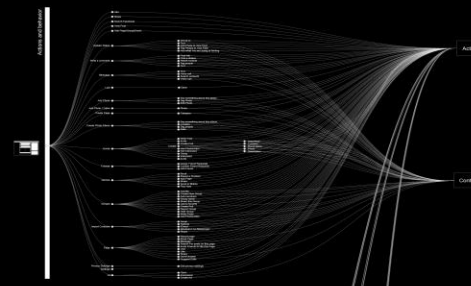


FACEBOOK RESEARCH
SHARE LAB - WWW.LABS.RS
2016

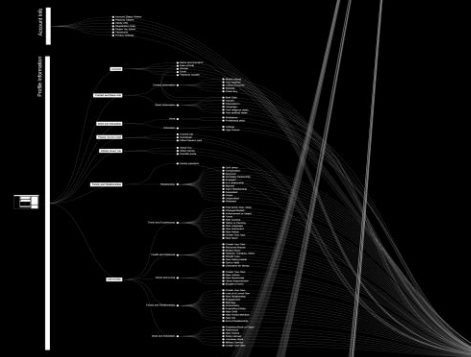
FACEBOOK ALGORITHMIC FACTORY

DATA COLLECTION

ACTIONS AND BEHAVIOR



ACCOUNT & PROFILE INFORMATION



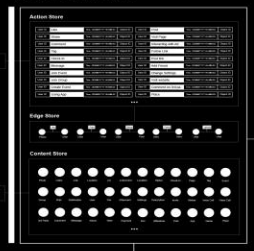
DATA COLLECTION : ONLINE TRACKERS

DATA COLLECTION : DEVICE INFORMATION

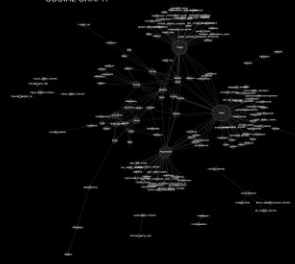
OUTSIDE FACEBOOK DOMAIN

STORAGE

ACTION & CONTENT STORE



SOCIAL GRAPH

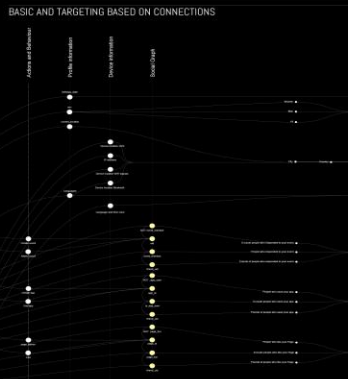


PROFILE STORE

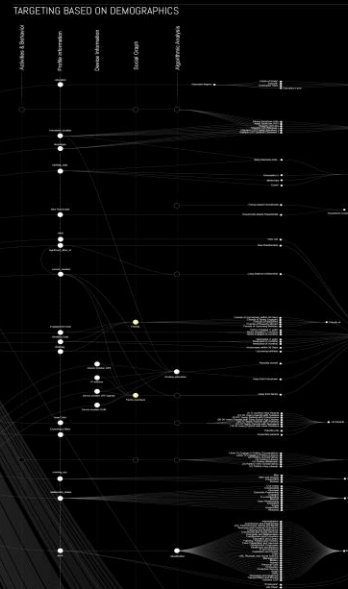
ALGORITHMIC PROCESSING

TARGETING

BASIC AND TARGETING BASED ON CONNECTIONS



TARGETING BASED ON DEMOGRAPHICS



TARGETING BASED ON USER INTERESTS

THE FACEBOOK ALGORITHMIC FACTORY

- Research of Share Labs, Belgrade https://labs.rs/en/
- Based on publicly available data (> 8000 patents and Graph API)
- Analysis of the Facebook Social Graph
  - Meta structure connecting all data in one structure
  - Objects (nodes) – connections (edges)
  - Action store – content store – edge store – profile store
- Data collection – Data storage/analysis – targeting/advertising
- Facebook as an algorithmic machine

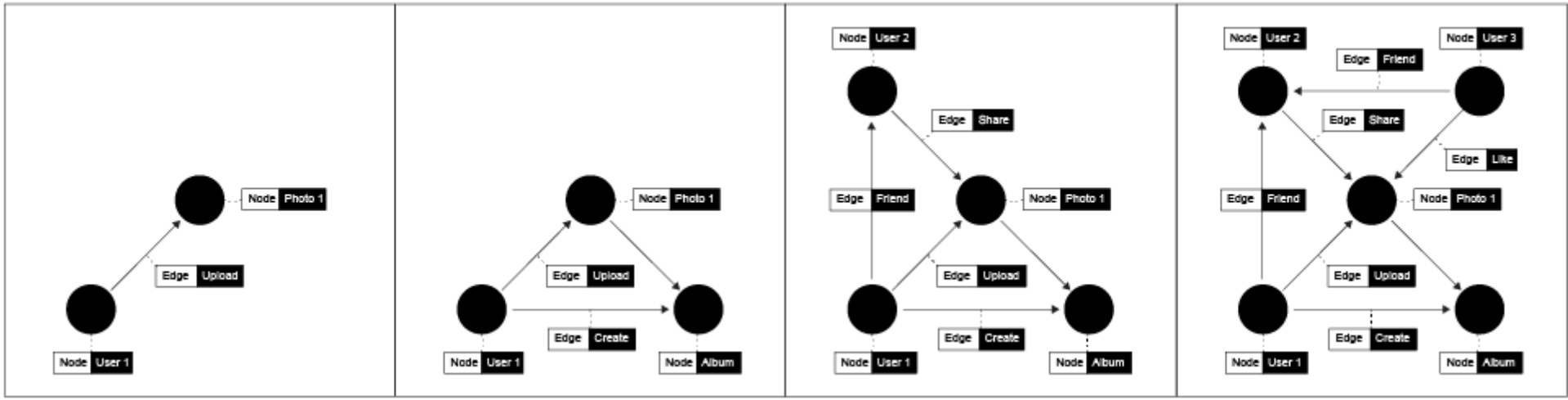
research institute

SHARE LABS, BELGRADE

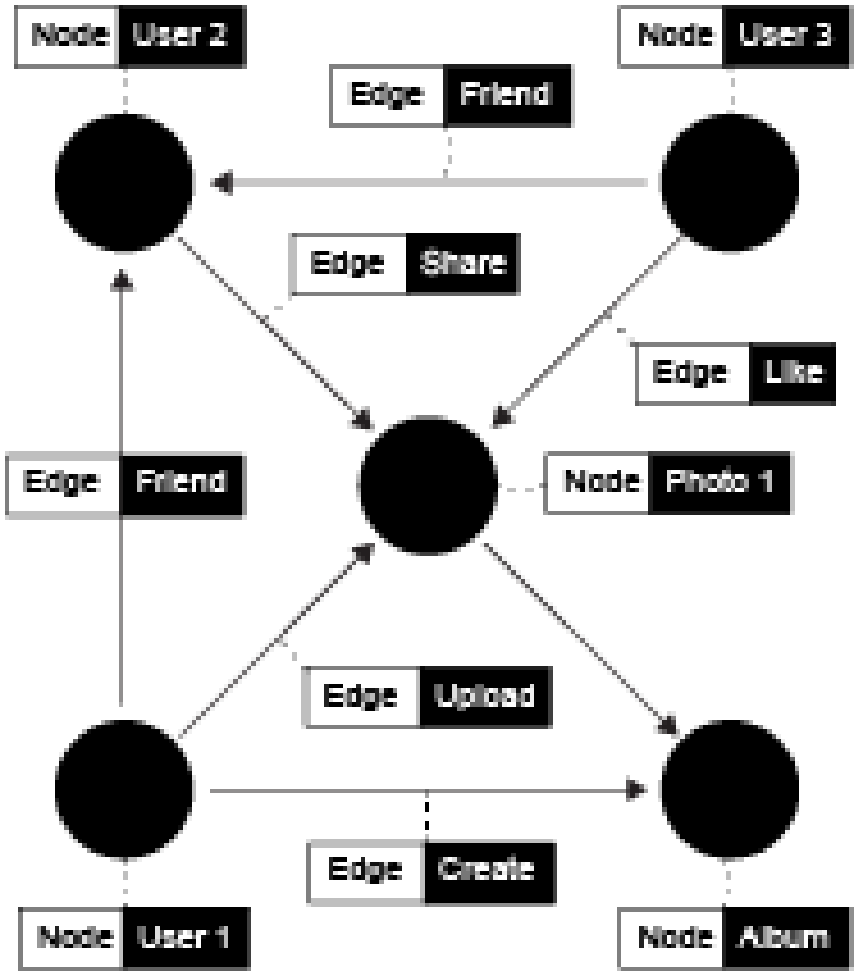
EPICENTER .WORKS for digital rights

1400 MILLIONS	**890** MILLIONS	**300** PETABYTES
MONTHLY ACTIVE USERS	PEOPLE LOG INTO FACEBOOK DAILY	OF USER DATA
1.3 TRILLION	**4300** MILLIONS	**17** BILLION
LIKES SINCE 2004	LIKES EVERY DAY	LOCATION-TAGED POSTS
10 BILLIONS	**350** MILLION	**4.75** BILLION
MESSAGES SENT DAILY	UPLOADED PHOTOS DAILY	SHARES DAILY

SHARE LABS, BELGRADE



SHARE LABS, BELGRADE



THE FACEBOOK ALGORITHMIC FACTORY

- Research of Share Labs, Belgrade https://labs.rs/en/
- Based on publicly available data (> 8000 patents and Graph API)
- Analysis of the Facebook Social Graph
  - Meta structure connecting all data in one structure
  - Objects (nodes) – connections (edges)
  - Action store – content store – edge store – profile store
- Data collection – Data storage/analysis – targeting/advertising
- Facebook as an algorithmic machine
- Huge, structured data set of user activities which is very attractive for targeting mechanisms (ads, political influence)
- Algorithms are transforming behavioural data into a final product
- Research shows that what kind of data Facebook stores is beyond our control

SHARE LABS, BELGRADE

- “The ultimate product of Facebook’s surveillance economy is a deep insight into your interests and behaviour patterns, exact **knowledge** who you really are and prediction how you will eventually behave in the future, packed in user **profiles**.”

- Targeting based on connections, demographics, user interests, behaviour,…

- Algorithmic black-box vs. algorithmic transparency?

Concluding Remarks

- We are calling for a very deliberate use of Big Data
- Awareness of the consequences of one's actions
  - Privacy Impact Assessment
  - „Social Impact Assessment"
  - **Public Accountability Test**: To imagine how it would feel like to announce and explain the Big Data activities to the public
  - General Call for Transparency
- Awareness of what algorithms actually do
- Specific Privacy by Design measures
- Use pseudonymisation and anonymization – but do not be overconfident of these measures

BIG DATA: NEW CHALLENGES FOR LAW AND ETHICS
International scientific conference
22 - 23 May 2017
Faculty of Law, University of Ljubljana

We don't know what the Questions are, but we know we're gonna find the Answers

Alexander Czadilek
Policy Analyst | alexander.czadilek@epicenter.works

epicenter.works
Digital Rights NGO, Vienna
https://www.epicenter.works

Christof Tschohl
Scientific Director | christof.tschohl@researchinstitute.at

Walter Hötzendorfer
Senior Researcher | walter.hoetzendorfer@researchinstitute.at

Research Institute AG & Co KG
Digital Human Rights Center, Vienna
http://www.researchinstitute.at