

Concepts and Methods of Social Network Analysis

Part 8 (Data Networks)

Berlin-Brandenburgische Akademie der Wissenschaften

Berlin, Oct. 9-11, 2017

Sorry for mixing English and
German slides ...

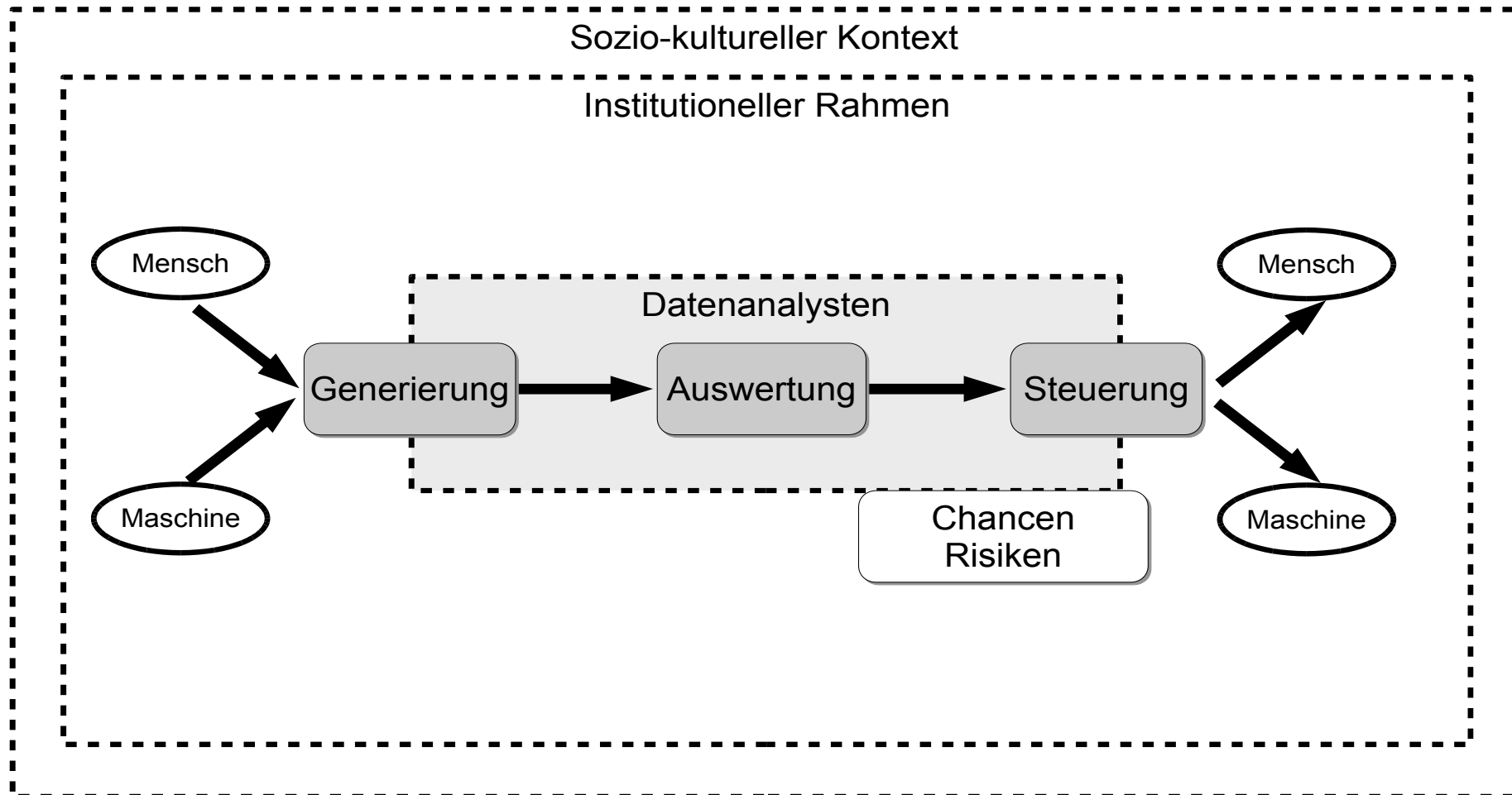
Data Networks in the Era of Big Data



Content

1. Process model
2. Data generation
3. Data analytics
 - Reality mining
 - Page rank
 - Traffic data analysis
 - Methods
4. Governance of complex systems
 - Cases (Flu trends, traffic management, smart grids)
 - Real-time governance
5. Political regulation of Big Data?

1. Big-Data-Prozessmodell



2. Datengenerierung

- Ubiquitäre Datenerfassung
 - durch smarte Geräte
 - automatisiert
 - 3 V's, 3 R's, 13 P's ...
- Algorithmisierung des Sozialen

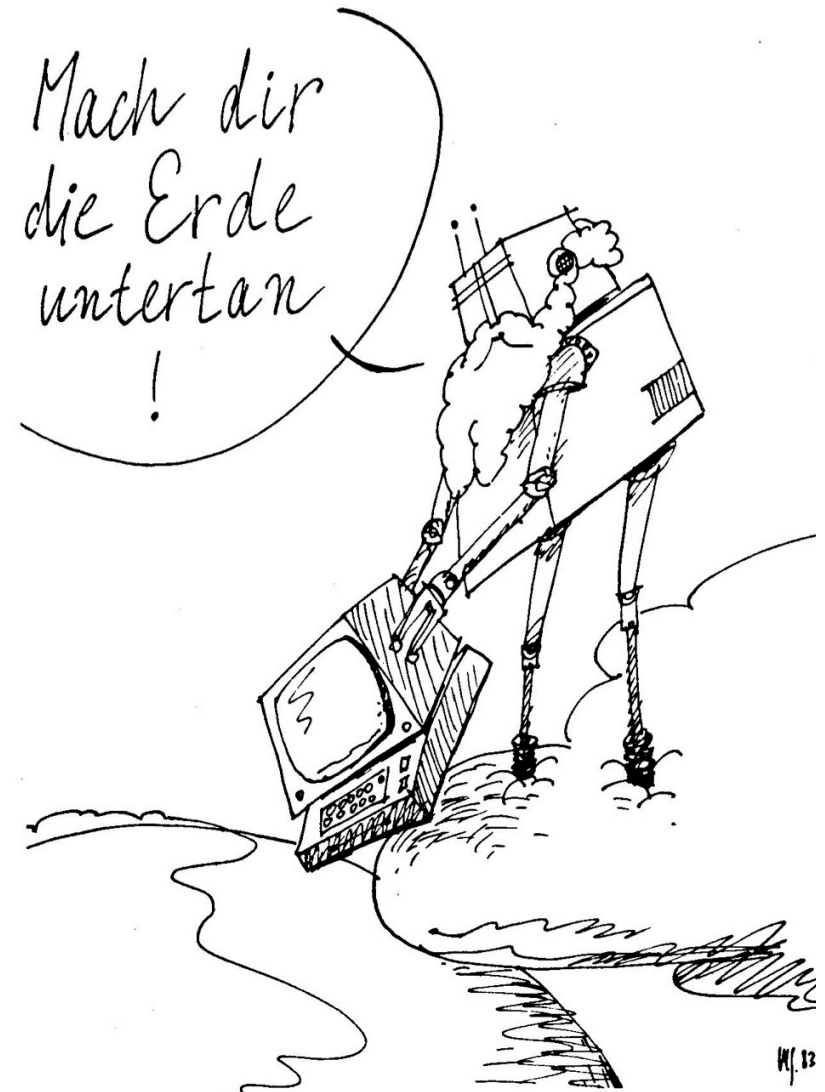
- Datentypen
 - Nutzerdaten
 - Metadaten

2. Selbstvermessung: Praktiken

- Gesundheitsmonitoring
 - Gesunde
 - Körper und Lebensführung optimieren
 - Kranke
 - preiswerte Alternative
 - vereinfachte Protokollierung und Überwachung
- Motivation
 - Sinnstiftung
 - Optimierung
 - Emanzipation
 - neue Normen
 - Gamification

2. Selbstvermessung: Risiken

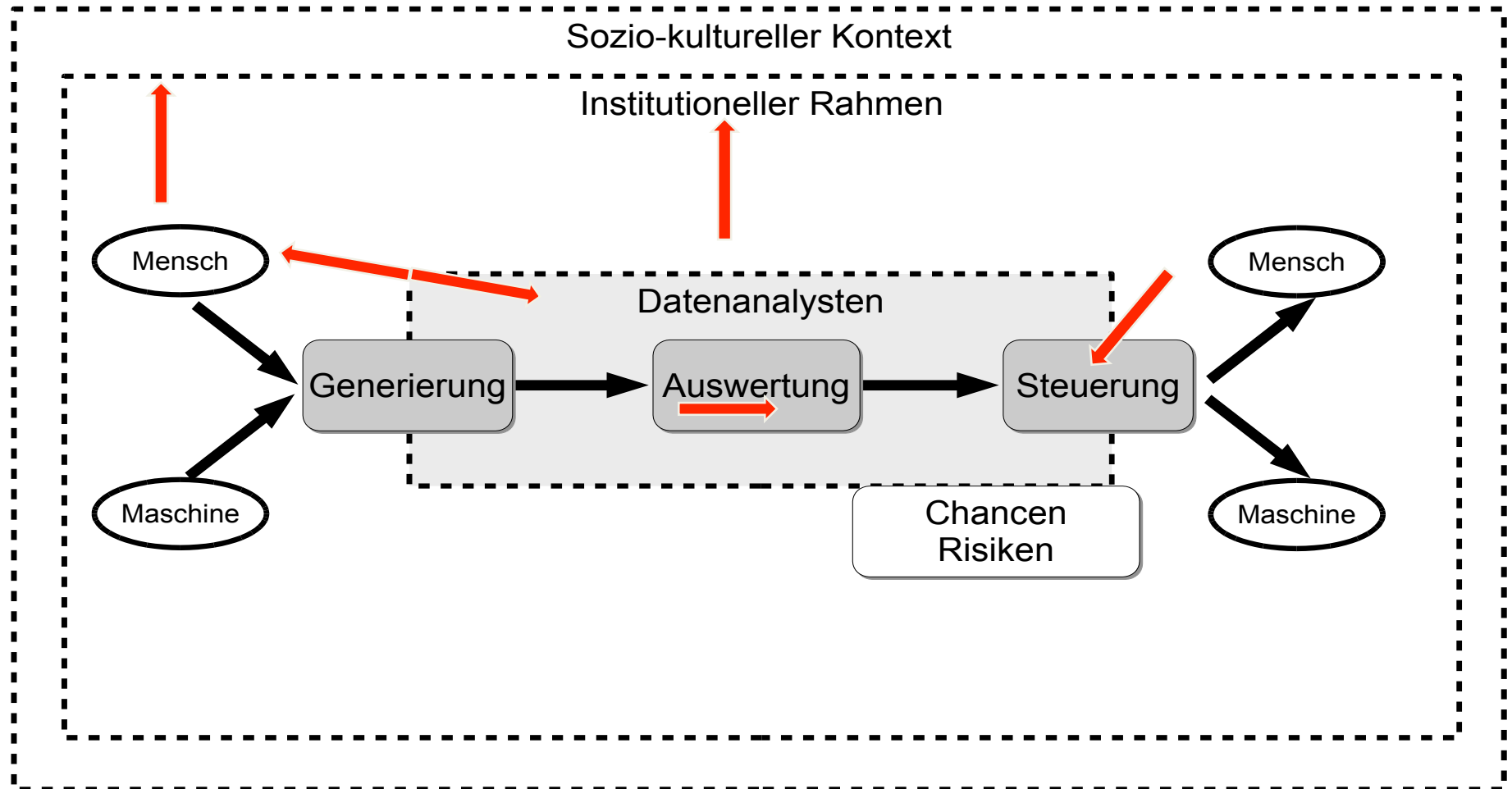
- Konkurrenzkampf
 - Gefährdung von
 - persönlichem Wohlbefinden
 - interpersonellen Beziehungen
- Normierung des Alltagslebens
 - Fokus: Messbarkeit
 - Objektivierungs-Suggestion
 - Deskriptive Daten
 - normative Daten



2. Selbstvermessung: Legitimation

- Weitergabe an Dritte
 - Peers
 - reflektiert und intendiert
 - intermediäre Instanz als Filter
 - Datenanalysten
 - unbemerkt (AGB's)
- Legitimationsstrategien
 - schützenswerte vs. nicht schützenswerte Daten
 - übermächtiges Gegenüber
 - Tauschgeschäft
 - Service gegen Daten / Daten gegen Vergleichswerte

Vertrauen



Content

1. Process model
2. Data generation
3. Data analytics
 - Reality mining
 - Page rank
 - Traffic data analysis
 - Methods
4. Governance of complex systems
 - Cases (Flu trends, traffic management, smart grids)
 - Real-time governance
5. Political regulation of Big Data?

3. Datenverarbeitung

- Massendaten
 - aus Suchmaschinen, Social media etc.
 - vollständige Samples (?)
 - Probleme der Datenreliabilität
- Anwendungsfelder
 - Marketing, Verkehr, Gesundheit etc.
- Data analytics
 - Lagebilder (Makro)
 - Trendprognosen (Makro)
 - Mustererkennung (→ nächste Folie)
 - individuelle Profile (Mikro)
 - Anomalie-Erkennung (Mikro)

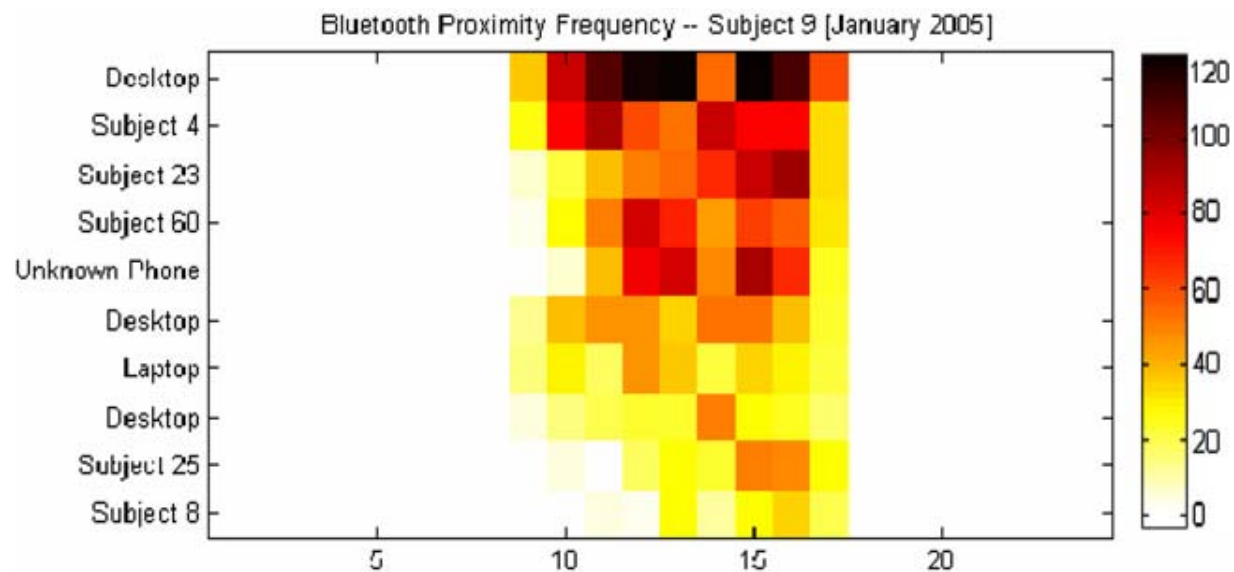
Eagle/Pentland (2006)

Reality mining: sensing complex social systems

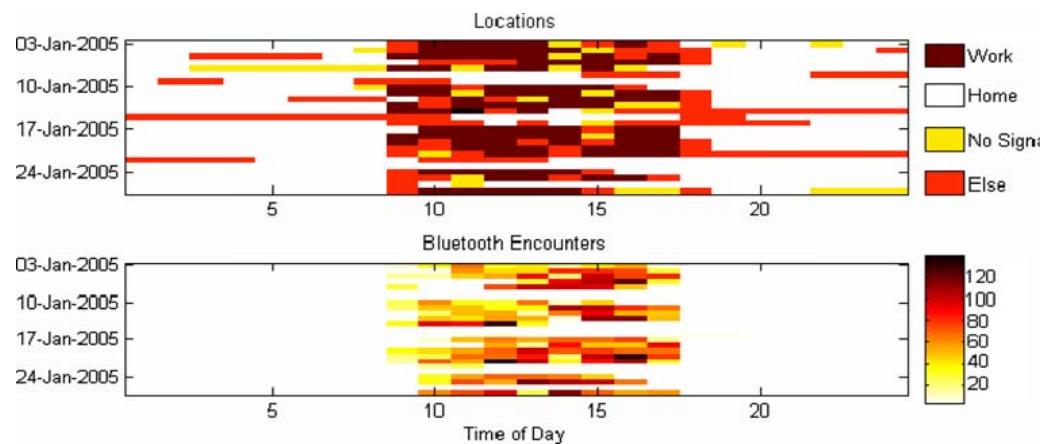
- Analysis of social behavior and social relations
- By means of tracing digital data
- Objectives
 - Identify behavioral patterns
 - Predict future actions
- 100 students and staff from MIT
 - Equipped with Nokia 6600
- 9 months
- 450.000 hours
 - Location, usage, communication



Eagle/Pentland (cont.)

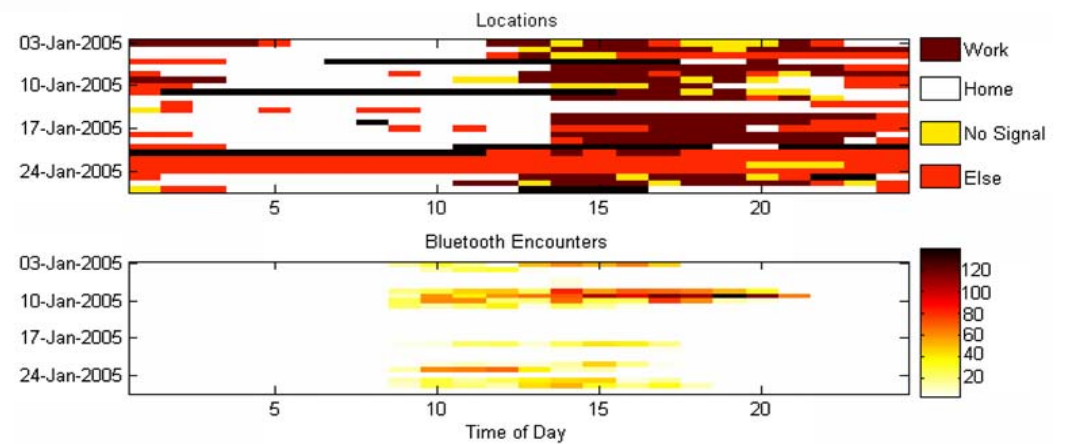


Eagle/Pentland (cont.)

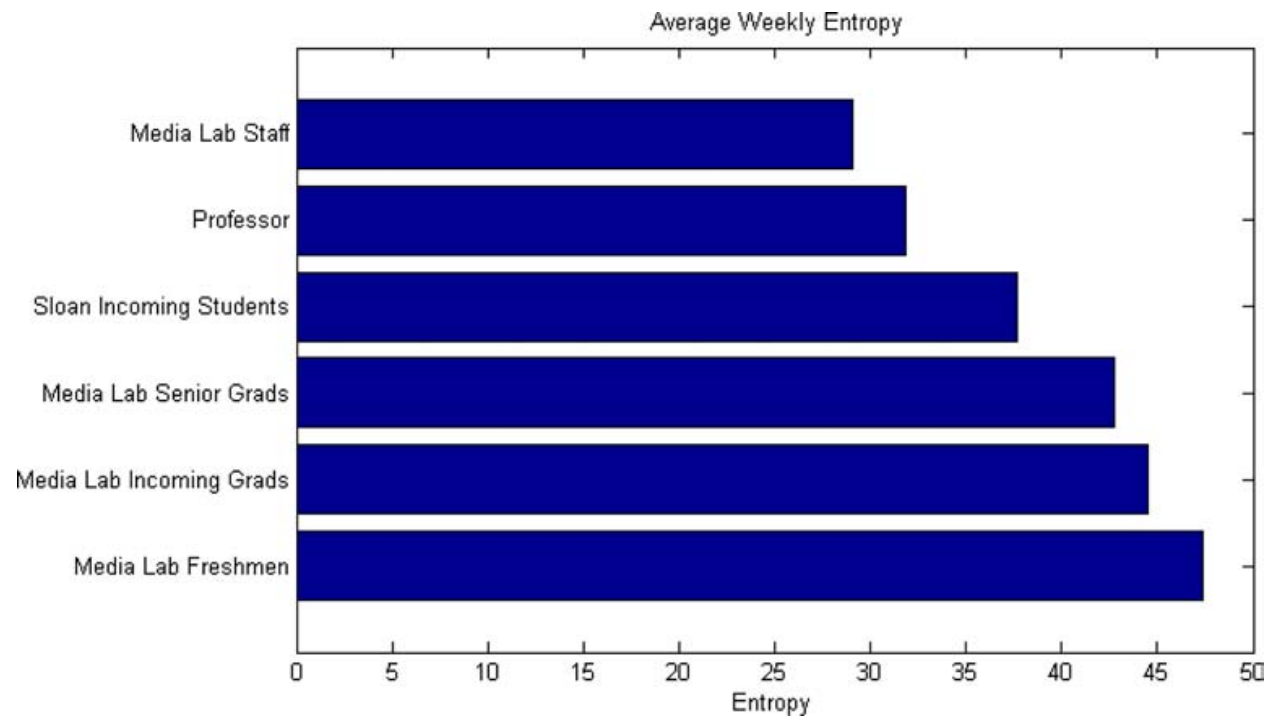


low entropy

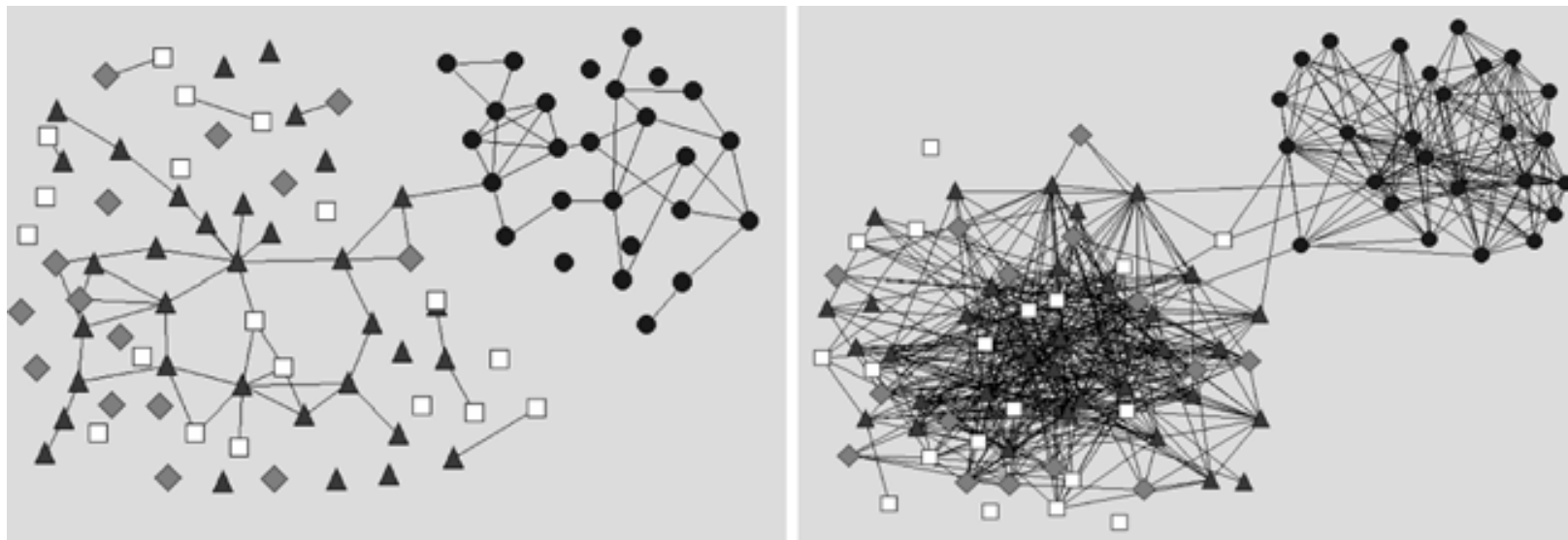
high entropy







Eagle/Pentland (cont.)



Eagle/Pentland (cont.)

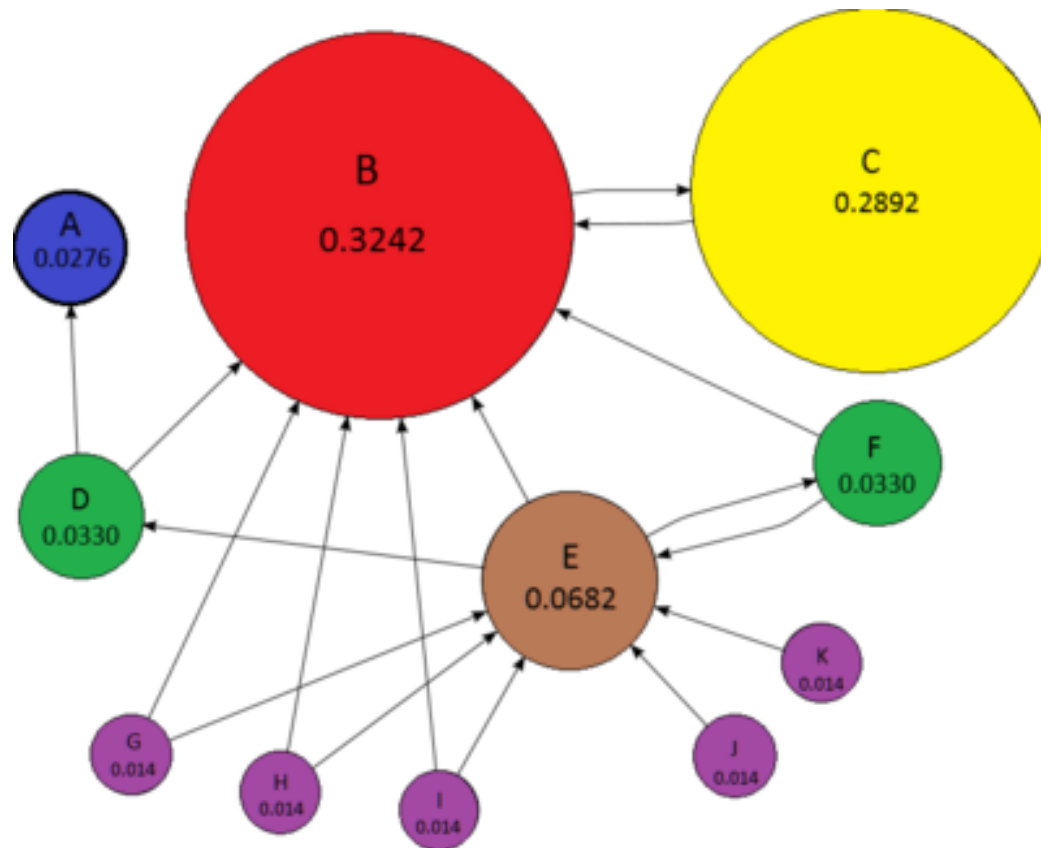


friendship (left) and proximity networks (right)

-  = Media Lab Freshman und Mitarbeiter
-  = Media Lab Neue Masterstudenten
-  = Media Lab Senior Masterstudenten
-  = Sloan business school Studenten

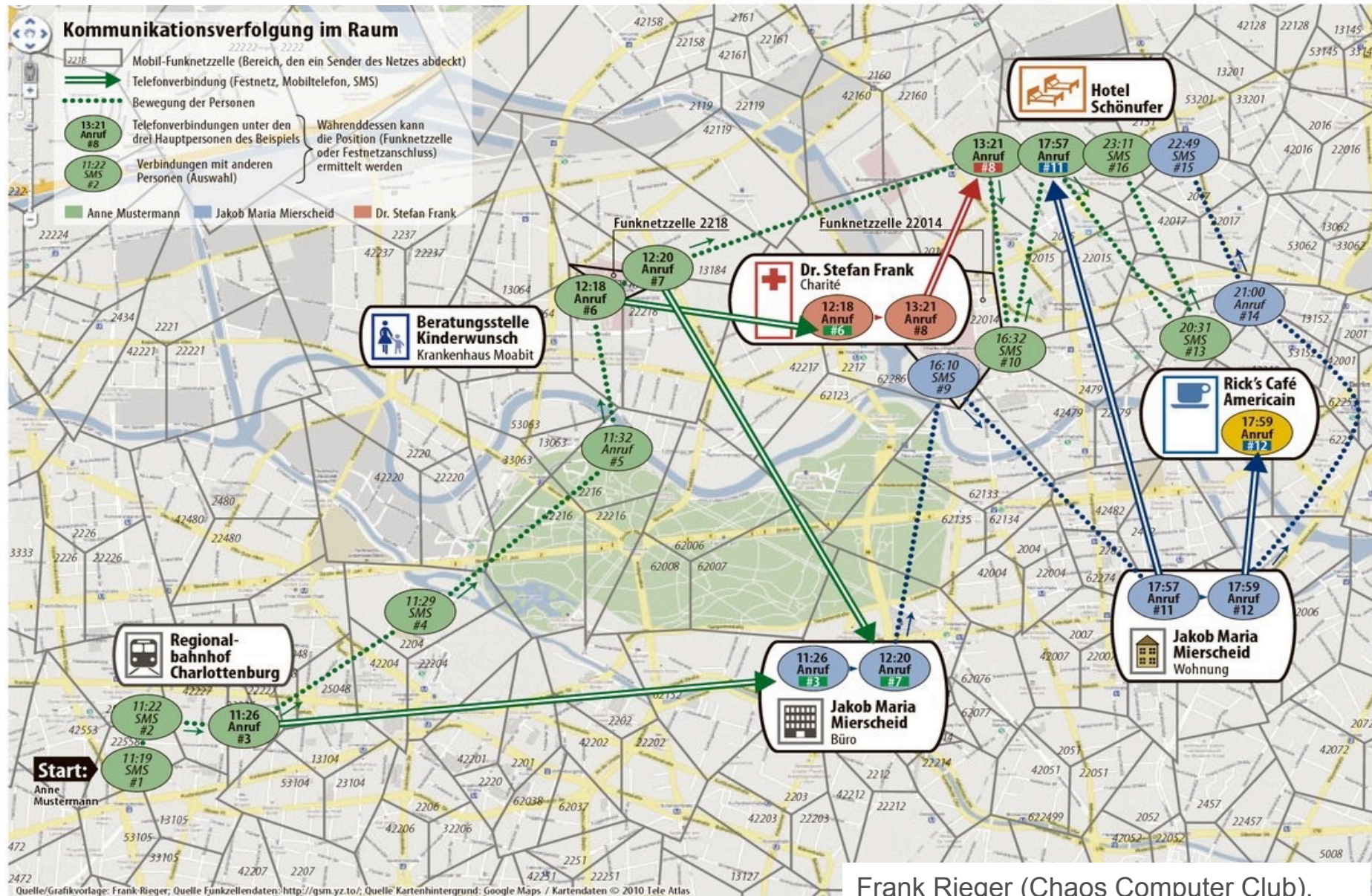
3. Page rank (Google)

- See NetLogo



Verkehrsdatenanalyse

Dr. Weyer



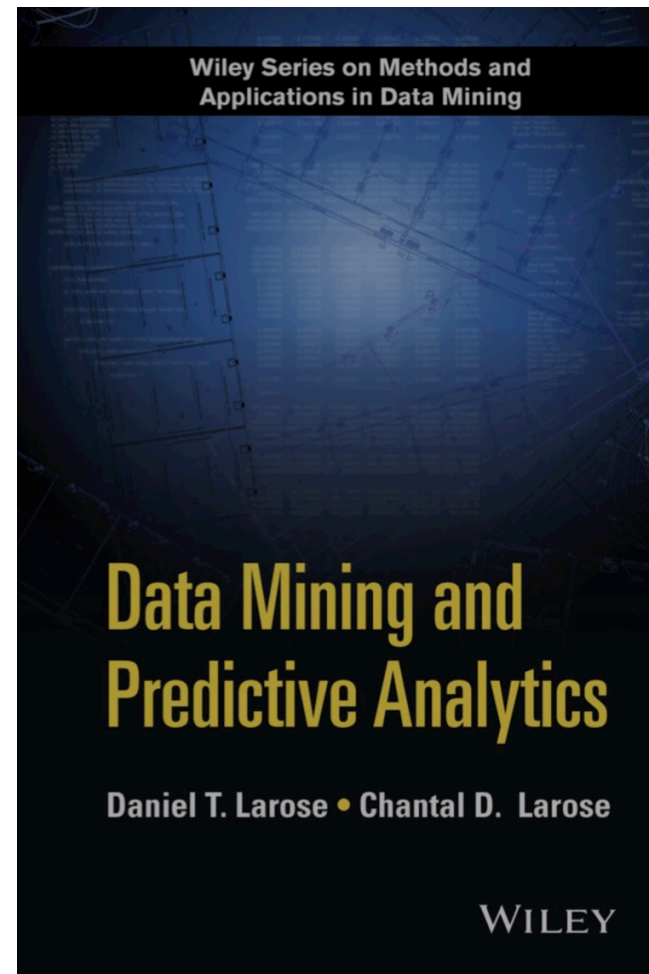
Frank Rieger (Chaos Computer Club),
Gutachten für BVerfG, in FAZ 20.02.2010

3. Datenverarbeitung

- Massendaten
 - aus Suchmaschinen, Social media etc.
 - vollständige Samples (?)
 - Probleme der Datenreliabilität
- Anwendungsfelder
 - Marketing, Verkehr, Gesundheit etc.
- Data analytics
 - Lagebilder (Makro)
 - Trendprognosen (Makro)
 - Mustererkennung
 - individuelle Profile (Mikro)
 - Anomalie-Erkennung (Mikro)

3. Datenverarbeitung: Methoden

- traditionelle Verfahren
 - Statistik
 - Netzwerkanalyse



[3.9 Deriving New Variables: Flag Variables](#)

[3.10 Deriving New Variables: Numerical Variables](#)

[3.11 Using EDA to Investigate Correlated Predictor Variables](#)

[3.12 Summary of Our EDA](#)

[The R Zone](#)

[R References](#)

[Exercises](#)

[Chapter 4: Dimension-Reduction Methods](#)

[4.1 Need for Dimension-Reduction in Data Mining](#)

[4.2 Principal Components Analysis](#)

[4.3 Applying PCA to the *Houses* Data Set](#)

[4.4 How Many Components Should We Extract?](#)

[4.5 Profiling the Principal Components](#)

[4.6 Communalities](#)

[4.7 Validation of the Principal Components](#)

[4.8 Factor Analysis](#)

[4.9 Applying Factor Analysis to the *Adult* Data Set](#)

[4.10 Factor Rotation](#)

[4.11 User-Defined Composites](#)

[4.12 An Example of a User-Defined Composite](#)

[The R Zone](#)

[R References](#)

[Exercises](#)

[Part II: Statistical Analysis](#)

[Chapter 5: Univariate Statistical Analysis](#)

[5.1 Data Mining Tasks in Discovering Knowledge in Data](#)

[5.2 Statistical Approaches to Estimation and Prediction](#)

[5.3 Statistical Inference](#)

[5.4 How Confident are We in Our Estimates?](#)

[5.5 Confidence Interval Estimation of the Mean](#)

[5.6 How to Reduce the Margin of Error](#)

[5.7 Confidence Interval Estimation of the Proportion](#)

[5.8 Hypothesis Testing for the Mean](#)

[5.9 Assessing The Strength of Evidence Against The Null Hypothesis](#)

[5.10 Using Confidence Intervals to Perform Hypothesis Tests](#)[5.11 Hypothesis Testing for The Proportion](#)[Reference](#)[The R Zone](#)[R Reference](#)[Exercises](#)[Chapter 6: Multivariate Statistics](#)[6.1 Two-Sample \$t\$ -Test for Difference in Means](#)[6.2 Two-Sample Z-Test for Difference in Proportions](#)[6.3 Test for the Homogeneity of Proportions](#)[6.4 Chi-Square Test for Goodness of Fit of Multinomial Data](#)[6.5 Analysis of Variance](#)[Reference](#)[The R Zone](#)[R Reference](#)[Exercises](#)[Chapter 7: Preparing to Model the Data](#)[7.1 Supervised Versus Unsupervised Methods](#)[7.2 Statistical Methodology and Data Mining Methodology](#)[7.3 Cross-Validation](#)[7.4 Overfitting](#)[7.5 Bias-Variance Trade-Off](#)[7.6 Balancing The Training Data Set](#)[7.7 Establishing Baseline Performance](#)[The R Zone](#)[R Reference](#)[Exercises](#)[Chapter 8: Simple Linear Regression](#)[8.1 An Example of Simple Linear Regression](#)[8.2 Dangers of Extrapolation](#)[8.3 How Useful is the Regression? The Coefficient of Determination, \$r^2\$](#) [8.4 Standard Error of the Estimate,](#)[8.5 Correlation Coefficient](#)[8.6 Anova Table for Simple Linear Regression](#)[8.7 Outliers, High Leverage Points, and Influential Observations](#)[8.8 Population Regression Equation](#)

3. Datenverarbeitung: Methoden

- traditionelle Verfahren
 - Statistik
 - Netzwerkanalyse
- neuartige Methoden
 - data mining, reality mining usw.
 - machine learning, statistical relational learning
- Datenverarbeitung in Echtzeit
- Algokratie

3. Datenverarbeitung: Soziologie

- Soziologie und Big Data
 - Internet-Soziologie
 - nicht-responsive Verhaltensdaten
 - Massen-Experimente
- Computational social science
 - text mining
 - ...

Content

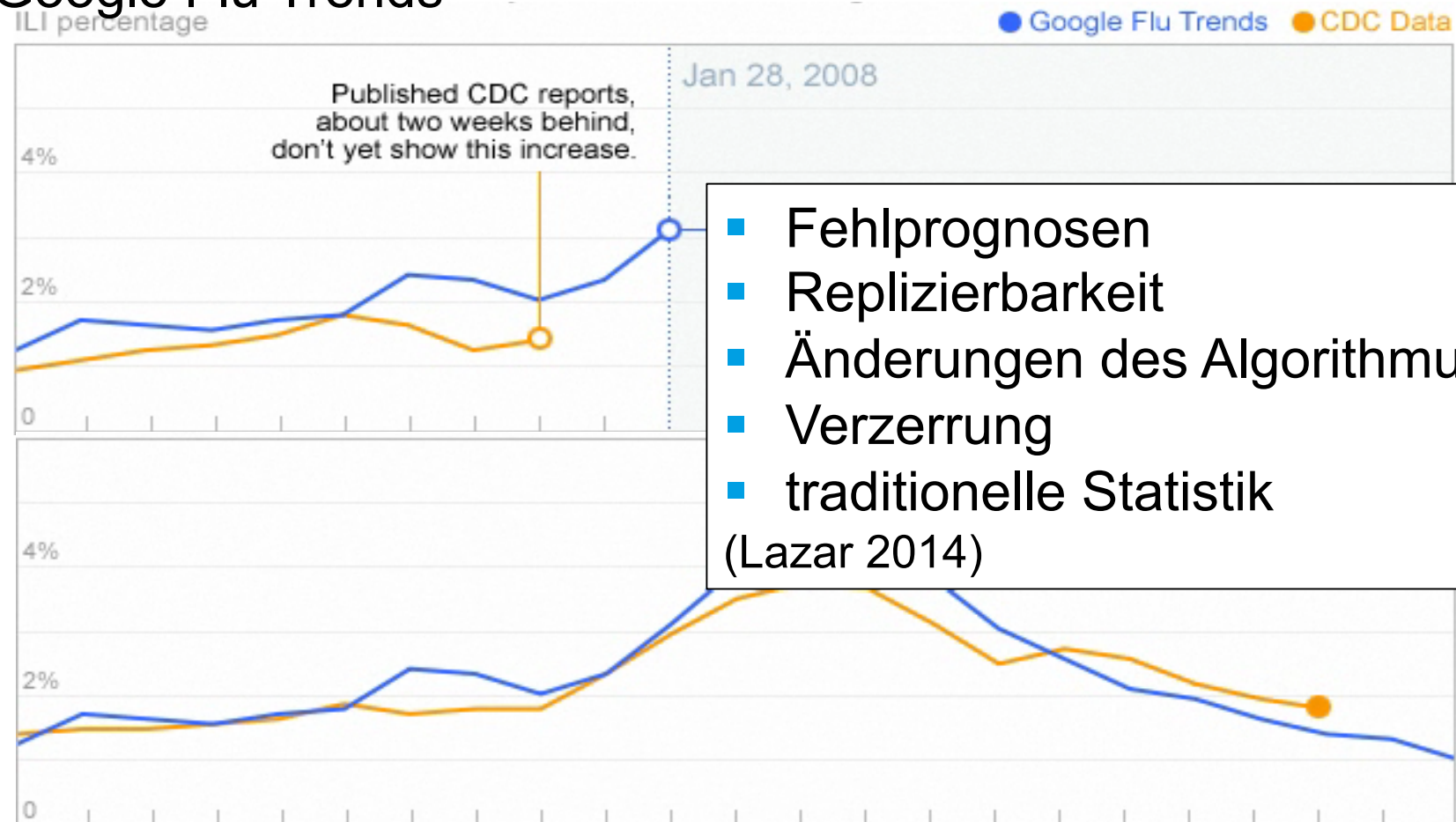
1. Process model
2. Data generation
3. Data analytics
 - Reality mining
 - Page rank
 - Traffic data analysis
 - Methods
4. Governance of complex systems
 - Cases (Flu trends, traffic management, smart grids)
 - Real-time governance
5. Political regulation of Big Data?

4. Steuerung komplexer Systeme

- Steuerung individuellen Verhaltens
- Echtzeitsteuerung komplexer Systeme
 - Prognosen

4. Steuerung komplexer Systeme

Google Flu Trends



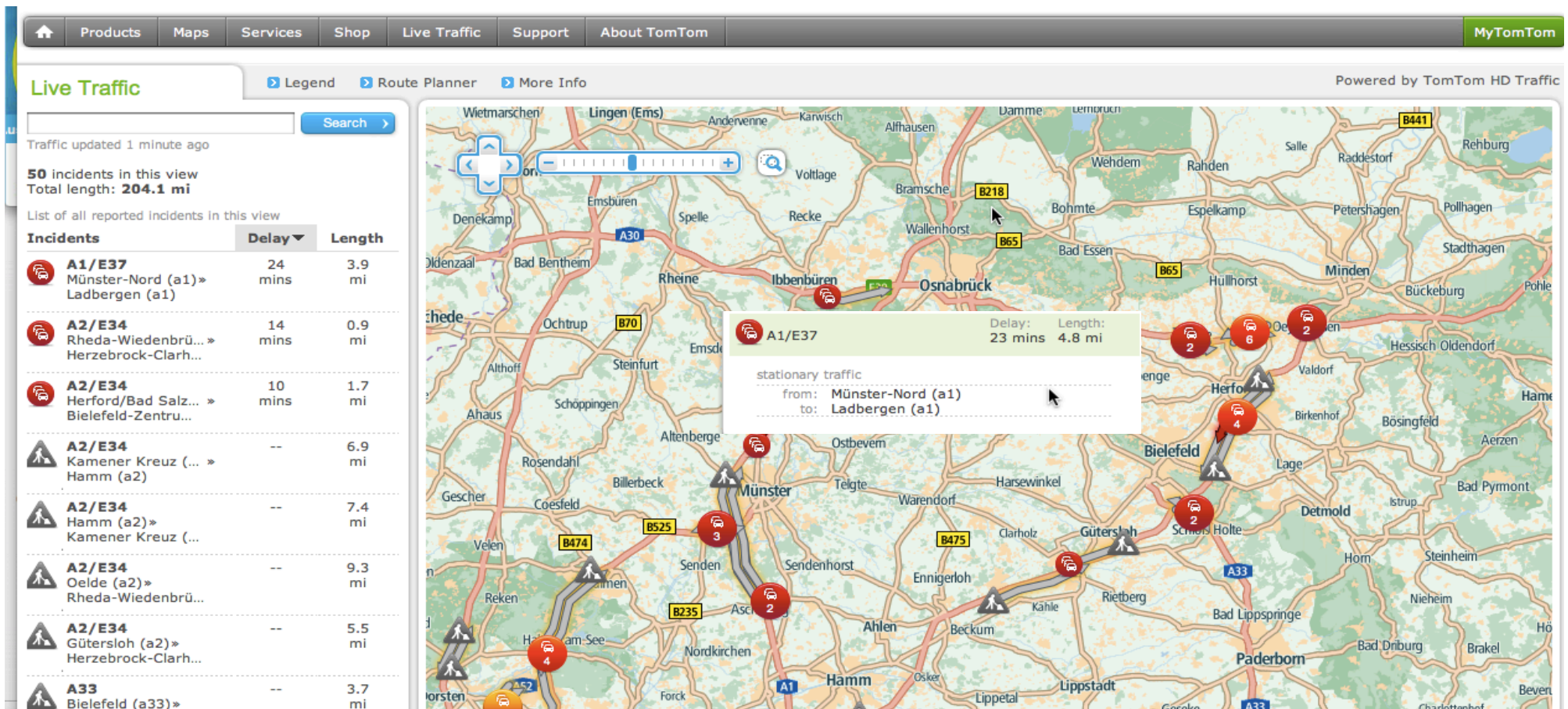
- Fehlprognosen
 - Replizierbarkeit
 - Änderungen des Algorithmus
 - Verzerrung
 - traditionelle Statistik
- (Lazar 2014)

Quellen: www.google.org/flutrends/about/how.html, Ginsberg 2009

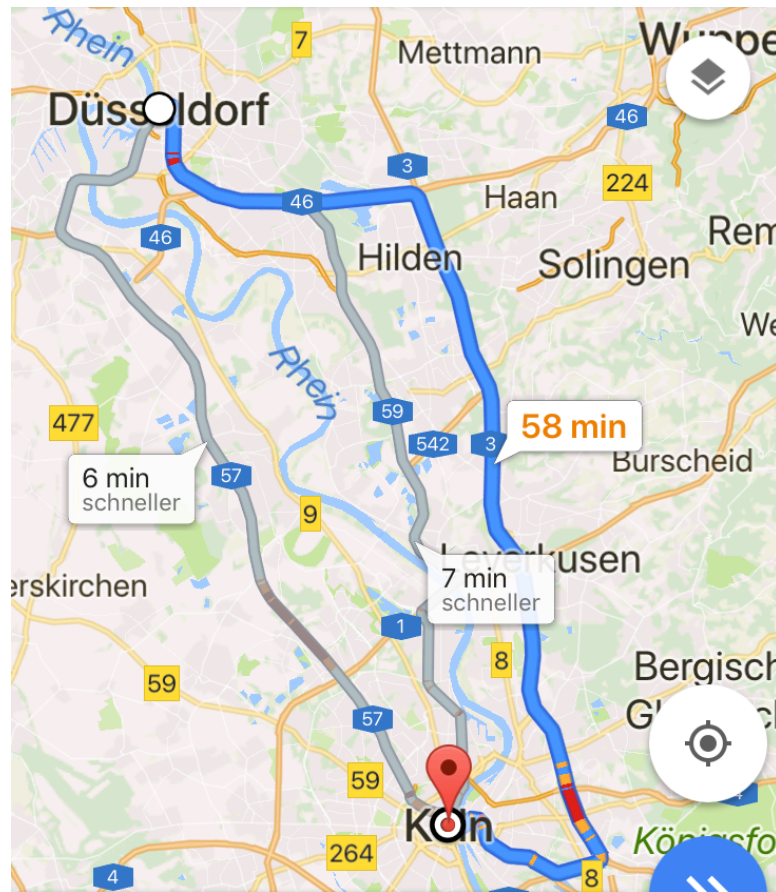
4. Steuerung komplexer Systeme

- Steuerung individuellen Verhaltens
- Echtzeitsteuerung komplexer Systeme
 - Prognosen
 - intelligente Netze (Verkehr, smart grid)
- Predictive policing
- Macht im Netz
- Politische Regulierung

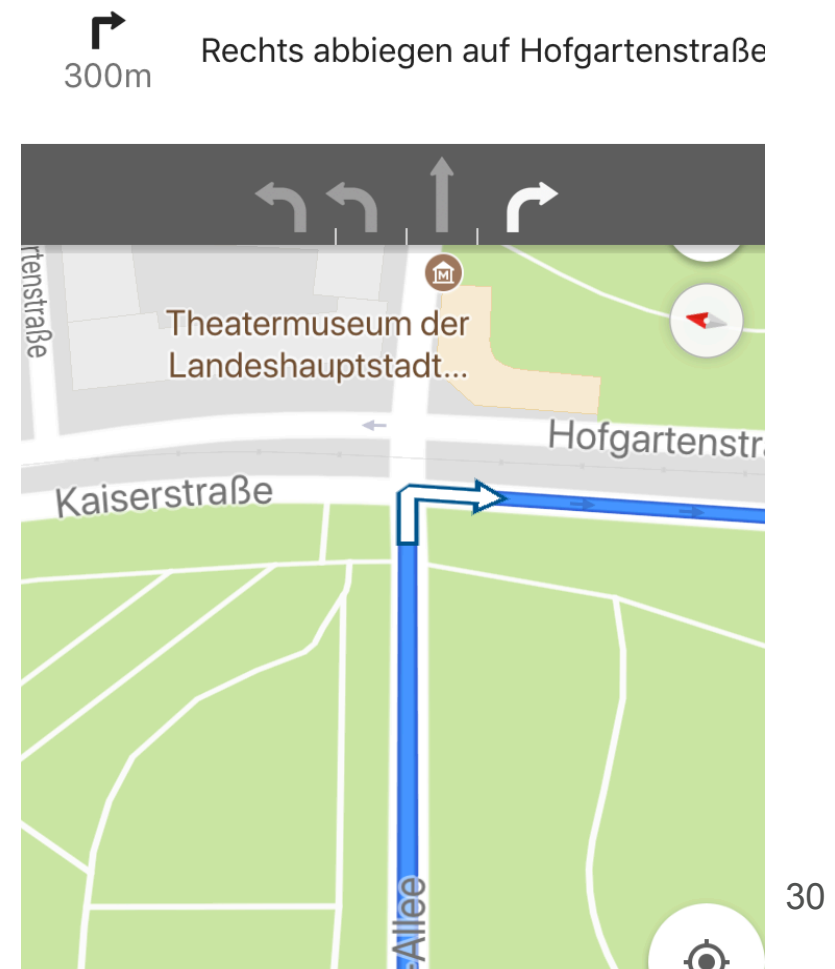
Praxisbeispiel: TomTom HD traffic



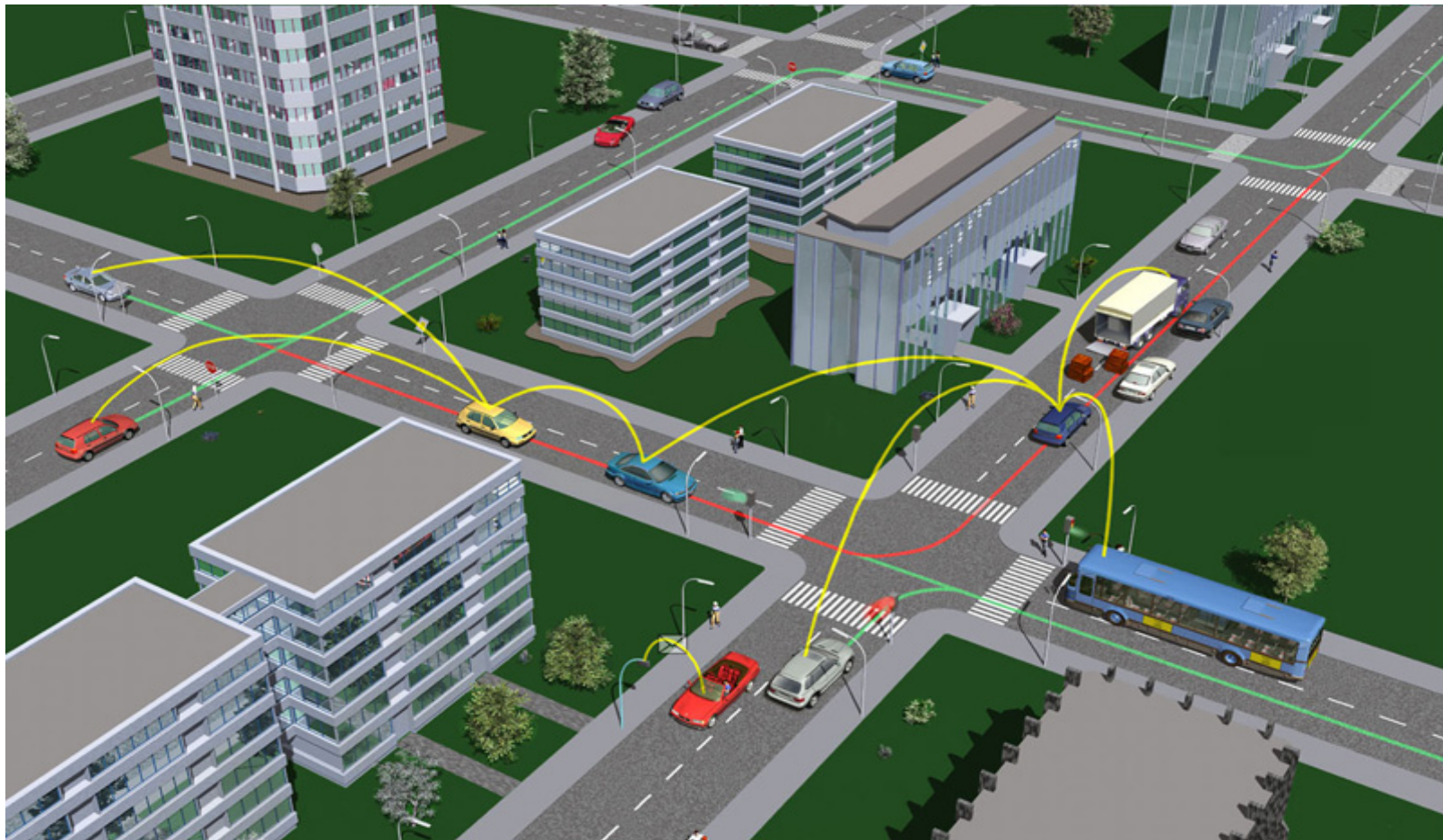
Echtzeit-Steuerung komplexer Systeme



| 2017 |



Car2Car Communication

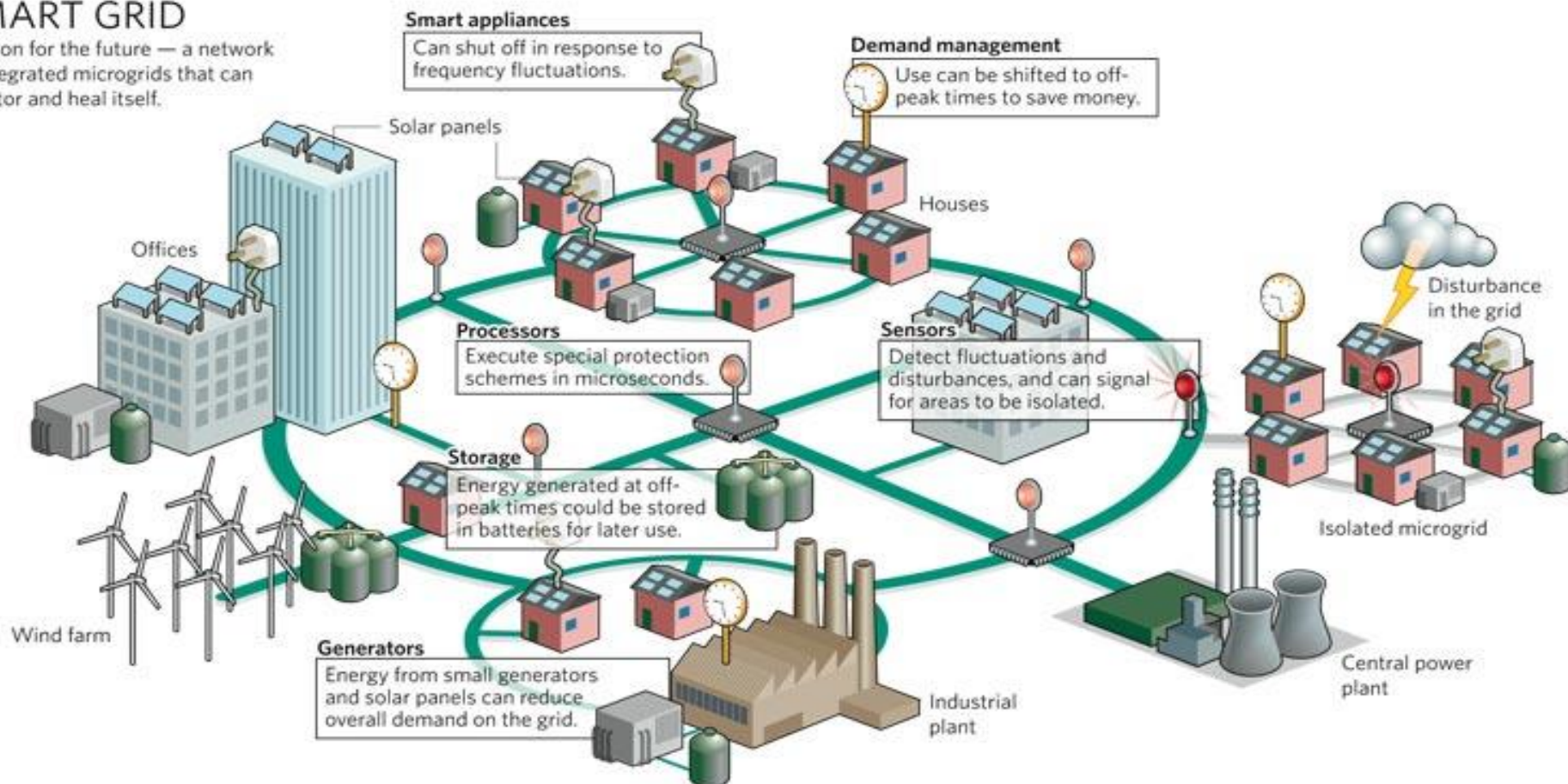


<http://www.car-2-car.org/index.php?id=131>

Smart grid

SMART GRID

A vision for the future — a network of integrated microgrids that can monitor and heal itself.



Content

1. Process model
2. Data generation
3. Data analytics
 - Reality mining
 - Page rank
 - Traffic data analysis
 - Methods
4. Governance of complex systems
 - Cases (Flu trends, traffic management, smart grids)
 - Real-time governance
5. Political regulation of Big Data?

Exkurs

- Grenzen traditioneller Governance
 - hierarchische Steuerung (top-down)
 - dysfunktionale Effekte; Ziele nicht erreicht
 - dezentrale Koordination (bottom-up)
 - ungesteuerte Selbstorganisation riskant

- Suche nach neuen Formen intelligenter Steuerung
 - Widerspruch?
 - Verknüpfung von Unvereinbarem?
 - Steuerung **UND** Selbstorganisation?

Schwarmintelligenz in natürlichen Systemen

- erstaunliche Leistungsfähigkeit
- dezentrale, lokale Koordination
- einfache Regeln; kein „leader“
- Beispiele
 - Vogelschwarm
 - Fischschwarm
- Anwendungen in
 - Robotik
 - VKI/MAS-Forschung
 - Logistik u.a.m.



Schwarmintelligenz in sozialen Systemen

- Surowiecki (2005)
 - hohe Intelligenz von Gruppen
- Open Innovation, Crowdsourcing
 - InnoCentive
 - Tchibo ideas
 - Page Rank (Google)
- natürliche versus soziale Systeme
 - Ameisen: trieb-/instinktgesteuert
 - Menschen: subjektiv rationale Alternativ-Entscheidungen

Grenzen der kollektiven Intelligenz

- Surowiecki (2005)
 - Aggregationsmechanismus nötig
 - Selbstorganisation funktioniert nicht von allein

- Was ist, wenn der Schwarm in die falsche Richtung fliegt?
 - Beispiel Love Parade
 - Beispiel Finanzmärkte
 - Sind Verkehrsstaus intelligent?

(Apropos: Wer weiß, was die richtige Richtung ist?)

Notwendigkeit von Steuerung ...

ergibt sich aus ...

1. nicht-intendierten Effekten „ungezügelter“ Selbstorganisation
 - vgl. Willke, Beck, Loorbach u.a.m.

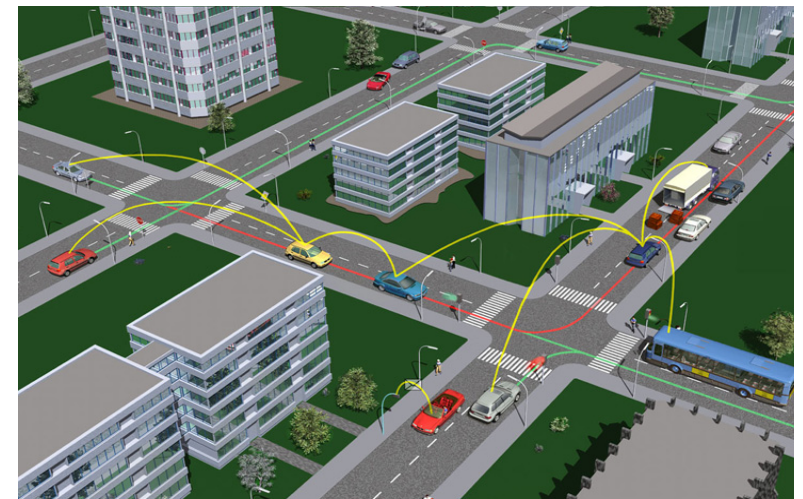
2. politisch konsentierten Zielen
 - per Selbstorganisation schwer erreichbar
 - Versagen interventionistischer Steuerung

Die Praxis der Steuerung verteilter Systeme

- dumme Netze
 - produzieren Datenmüll, Blackouts etc.
- intelligente Netze
 - bewältigen Probleme der Koordination und Aggregation
- „Working in practise, but not in theory“
 - Soziologie hat bislang keine Theorie verteilter Systeme.

Verkehrstelematik

- Echtzeit-Kommunikation
 - bi-direktional
 - Fahrzeuge sind „Knoten im Netz“
- Echtzeit-Steuerung
 - dezentral generierte Daten
 - zentral hinterlegte Algorithmen
 - dezentrale Entscheidungen
- Adaptive Steuerung
 - Anpassung des Systems
 - Anpassung der Akteure



Echtzeit-Steuerung komplexer Systeme

- dezentrale Strukturen **UND** zentrale Steuerung
 - New („mixed“) mode of governance
- Planung **UND** Selbstorganisation
 - „Ad-hoc-Planung“ (situatives Handeln)
 - Konsequenzen für Handlungsfähigkeit der Akteure?
- hochgradig automatisiert
- zeitlich verdichtet
 - simultan statt sequenziell

Paradigm-shift(s) in management (Rochlin 1998)

1. hierarchical, centralized control; rational planning (1950s/60s)
 - assembly line
2. decentralized, flexible, participatory self-organization (1970/80s)
 - personal computer
3. central control of decentralized structures (micro-management)
 - Internet

„Computer trap“ (Rochlin 1998)

- unintended consequences of computerization
 - striving for improvement of efficiency
 - Losses of autonomy and slack
 - of learning capacities
- growing dependency, increasing vulnerability of society
- transformation of society
 - deconstruction of social institutions
 - irreversal process
 - (large-scale) experiments

4. Governance of complex systems

- Governance *by* algorithms? (Saurwein et al. 2015)
 - O'Reilly 2013
 - Lessig 2000, 2006: Code is law

- Governance *of* algorithms? (Just et al. 2016)
 - Smart regulation
 - Role of the state?

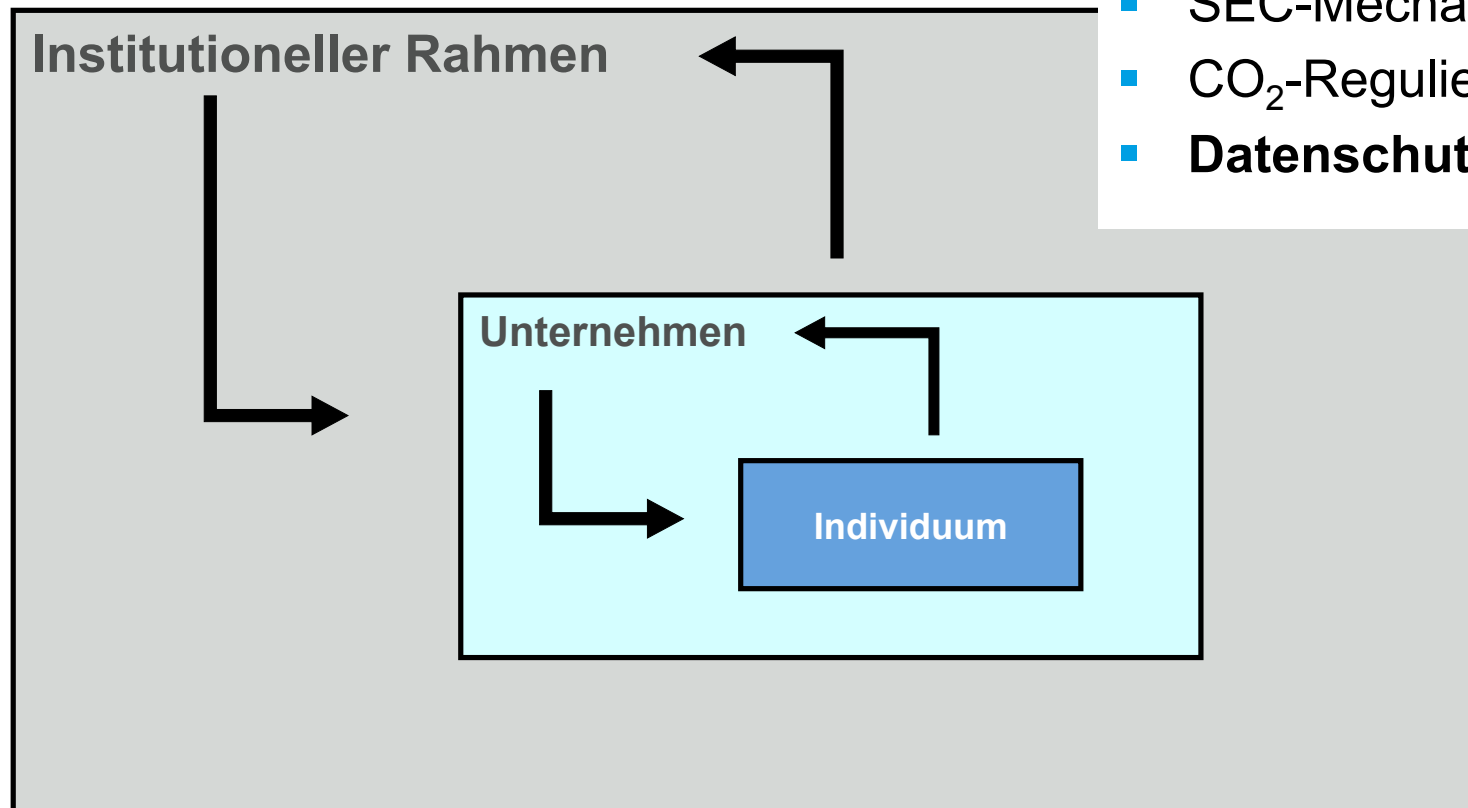
Content

1. Process model
2. Data generation
3. Data analytics
 - Reality mining
 - Page rank
 - Traffic data analysis
 - Methods
4. Governance of complex systems
 - Cases (Flu trends, traffic management, smart grids)
 - Real-time governance
5. Political regulation of Big Data?

5. Institutionelle Regulierung

Beispiele

- SEC-Mechanismus
- CO₂-Regulierung
- **Datenschutz**



Nach dem Datenskandal

Kurs der Google-Aktie bricht dramatisch ein. (FAZ 14.03.2018)



Nach dem Organskandal

So wenig Organspenden wie noch nie. (statista.de)



