Challenges and Opportunities for Computational Social Science

Prof. Dr. Markus Strohmaier

GESIS – Leibniz Institute for the Social Sciences

GESIS Leibniz Institute for the Social Sciences

Five scientific departments:

- Computational Social Science
- Survey Design & Methodology
- Data Archive for the Social Sciences
- Knowledge Technologies for the Social Sciences
- Monitoring Society and Social Change

Social and Computer Sciences (1890)
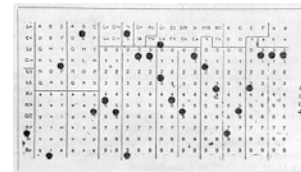# Traditional kinds of data

the 1890 US Census

1ˢᵗ time *Hollerith* machines were used to tabulate US Census data

Herman Hollerith (1860-1929)

- "invented a machine … to tally not only overall numbers but also individual characteristics and even cross-tabulations."

- "invented the **first automatic card-feed mechanism** and the **first key punch**"

PhD thesis "An Electric Tabulating System" (1889)

- developed a crucial foundation for the advancement of **computer science** during the 20ᵗʰ century

3

Social and Computer Science in the 21st century
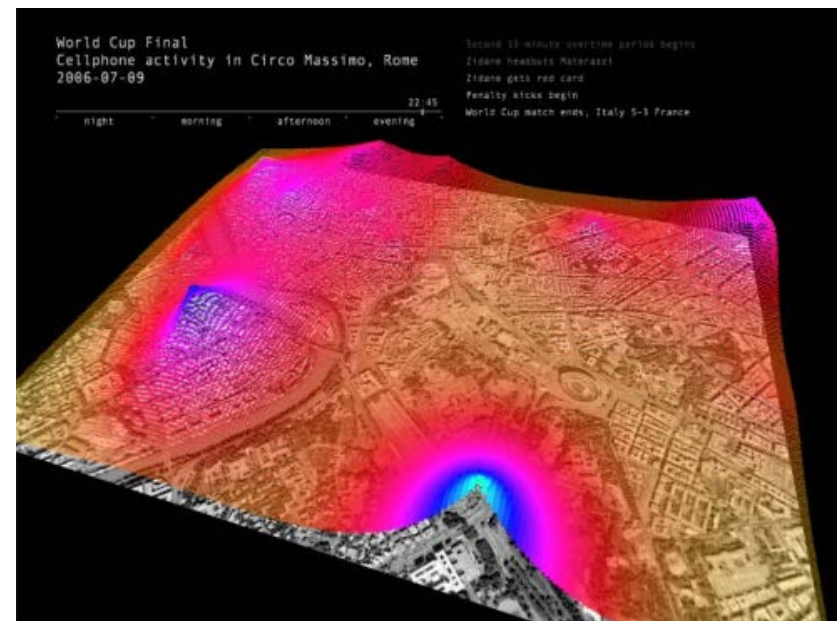# New kinds of data (macro-scale)

Human mobility in societies

Check-ins (Foursqaure, Gowalla, Twitter, …)

Cheng, Zhiyuan, et al. "Exploring Millions of Footprints in Location Sharing Services." *ICWSM* 2011 (2011): 81-88.

Social and Computer Science in the 21st century
New kinds of data (meso-scale)
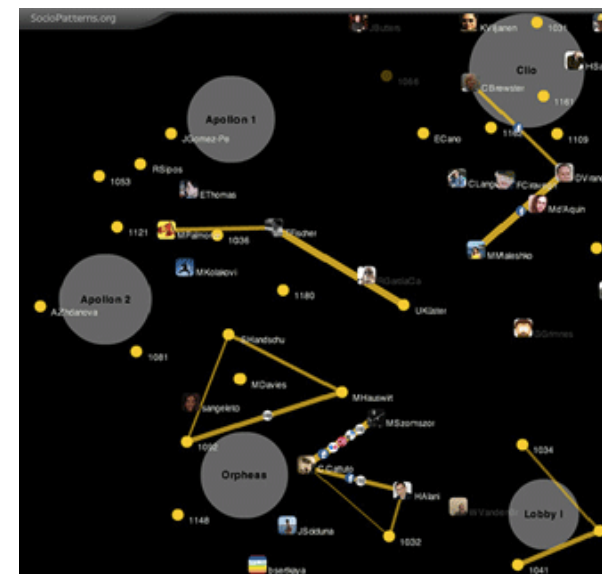
Urban movement analysis from GPS/phone data

The Amsterdam Real Time Project

Calabrese, F., Colonna, M., Lovisolo, P., Parata, D., & Ratti, C. (2011). Real-time urban monitoring using cell phones: A case study in Rome, *IEEE Transactions on Intelligent Transportation Systems, 12*(1), 141-151.

Social and Computer Science in the 21st century
New kinds of data (micro-scale)
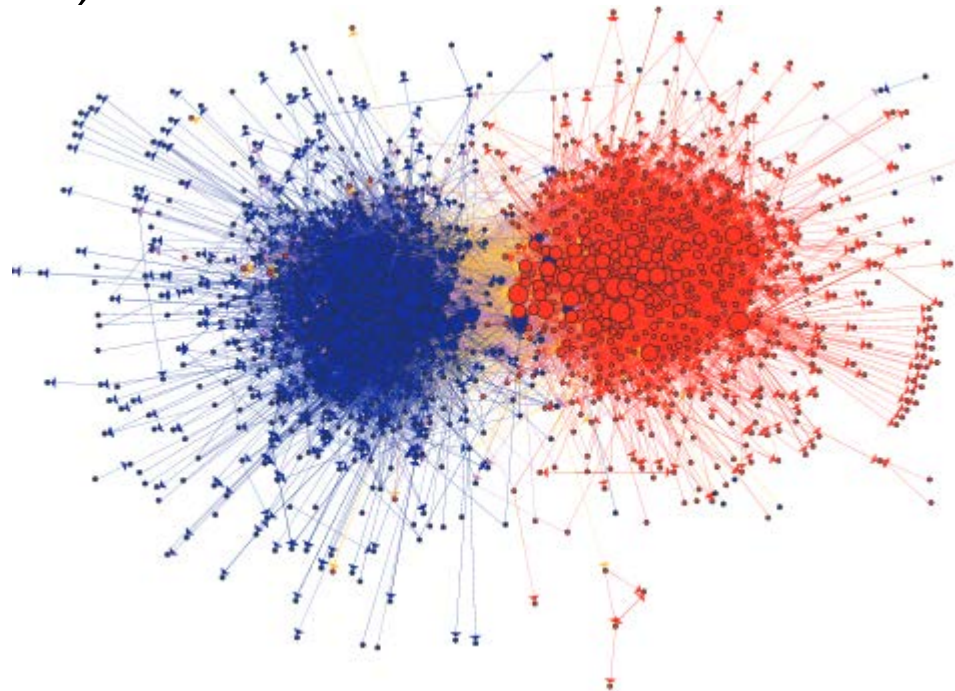
Social Sensing via RFID

Cattuto, C., Van den Broeck, W., Barrat, A., Colizza, V., Pinton, J. F., & Vespignani, A. (2010). Dynamics of person-to-person interactions from distributed RFID sensor networks. *PloS one*, *5*(7), e11596.

Social and Computer Science in the 21st century
New kinds of data

Analysis of political conversations in the blogosphere
(US election 2004)

Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., ... & Van Alstyne, M. (2009). Life in the network: the coming age of computational social science. *Science (New York, NY)*, *323*(5915), 721.

Social and Computer Science in the 21st century New kinds of data

Analysis of political conversations on Twitter (Egyptian revolution 2011)

23.Jän	24.Jän	25.Jän
qaeda-linked	#jan25	#jan25
blames	braced	blocked
militant	protests	tahrir
palestinian	bust	protest
al-qaida-linked	revolution	protest
group	anti-government	tear
army	berlin	thousands
bombing	prepares	#25jan
korean	nefertiti	demonstrations
#jan25	braces	violent
interior	#25jan	#tahrir
qaeda	demonstrations	anti-government
ceo	#arabprotest	square
al-qaida-linked	activists	armored
church	ceo	riot
#palestinepaper	return	tiananmen
accusations	#j	demonstrators
attacked	kate	protestors
conclusive	tunisian-inspired	fired

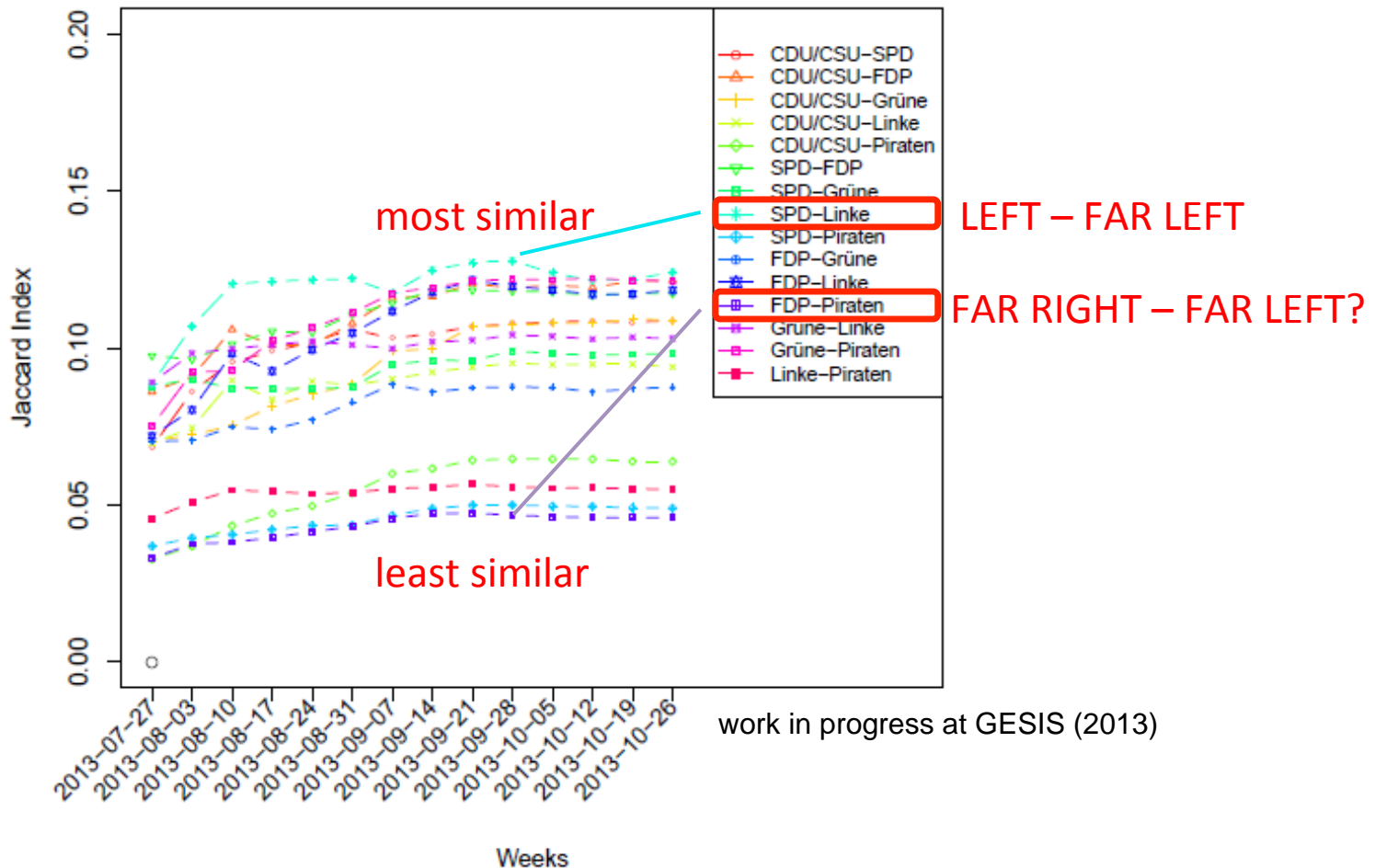1. day of the protests

Based on an analysis of about 100 mio tweets on egypt

Days on which the internet was shut down

8

Markus Strohmaier, work at (XEROX) PARC in collaboration with Lichan Hong (then at PARC, now at Google)

German Parliamentary Elections 2013
Similarity between parties as measured by hashtags

work in progress at GESIS (2013)

Observations

The **digital world is tracking the social world more and more closely**.

This enables us to **use computation** to discover patterns, build models, validate social theories and learn about societies.

Computational Social Science

Computational Social Science:
"The science that investigates social phenomena through the medium of computing and algorithmic data processing."
[adapted from CSSSA]

CSSSA: http://computationalsocialscience.org/

SOCIAL SCIENCE

Computational Social Science

David Lazer,[1] Alex Pentland,[2] Lada Adamic,[3] Sinan Aral,[2,4] Albert-László Barabási,[5]
Devon Brewer,[6] Nicholas Christakis,[1] Noshir Contractor,[7] James Fowler,[8] Myron Gutmann,[3]
Tony Jebara,[9] Gary King,[1] Michael Macy,[10] Deb Roy,[2] Marshall Van Alstyne,[2,11]

A field is emerging that leverages the capacity to collect and analyze data at a scale that may reveal patterns of individual and group behaviors.

We live life in the network. We check our e-mails regularly, make mobile phone calls from almost any location, swipe transit cards to use public transportation, and make purchases with credit cards. Our movements in public places may be captured by video cameras, and our medical records stored as digital files. We may post blog entries accessible to anyone, or maintain friendships through online social networks. Each of these transactions leaves digital traces that can be compiled into comprehensive pictures of both individual and group behavior, with the potential to transform our understanding of our lives, organizations, and societies.

ment agencies such as the U.S. National Security Agency. Computational social science could become the exclusive domain of private companies and government agencies. Alternatively, there might emerge a privileged set of academic researchers presiding over private data from which they produce papers that cannot be

critiqued or replicated. Neither scenario will serve the long-term public interest of accumulating, verifying, and disseminating knowledge.

What value might a computational social science—based in an open academic environment—offer society, by enhancing understanding of individuals and collectives? What are the

Computational Social Science

- is an **instrument-enabled scientific discipline**
- similar to microbiology, radio astronomy, or nanoscience —new scientific fields […] that were enabled by the **microscope**, **radar**, and **electron microscope**
- [where] the instrument **drives the development of theory and understanding**…"

Cioffi-Revilla, Claudio. "Computational social science." *Wiley Interdisciplinary Reviews: Computational Statistics* 2.3 (2010): 259-271.

Challenges for Research Infrastructures

computational infrastructures for dealing with:

- **More data**: analyzing large amounts of data

- **Fuzzy data**: cleaning up inprecise and noisy data

- **New kinds of data:** processing real-time sensor streams and web data

- **Correlations**: understanding What (in addition to Why)

Modern Hollerith Machines (2013) Challenges for Computer Science

Figure 1: Taxonomy of machine learning tools.

Data processing
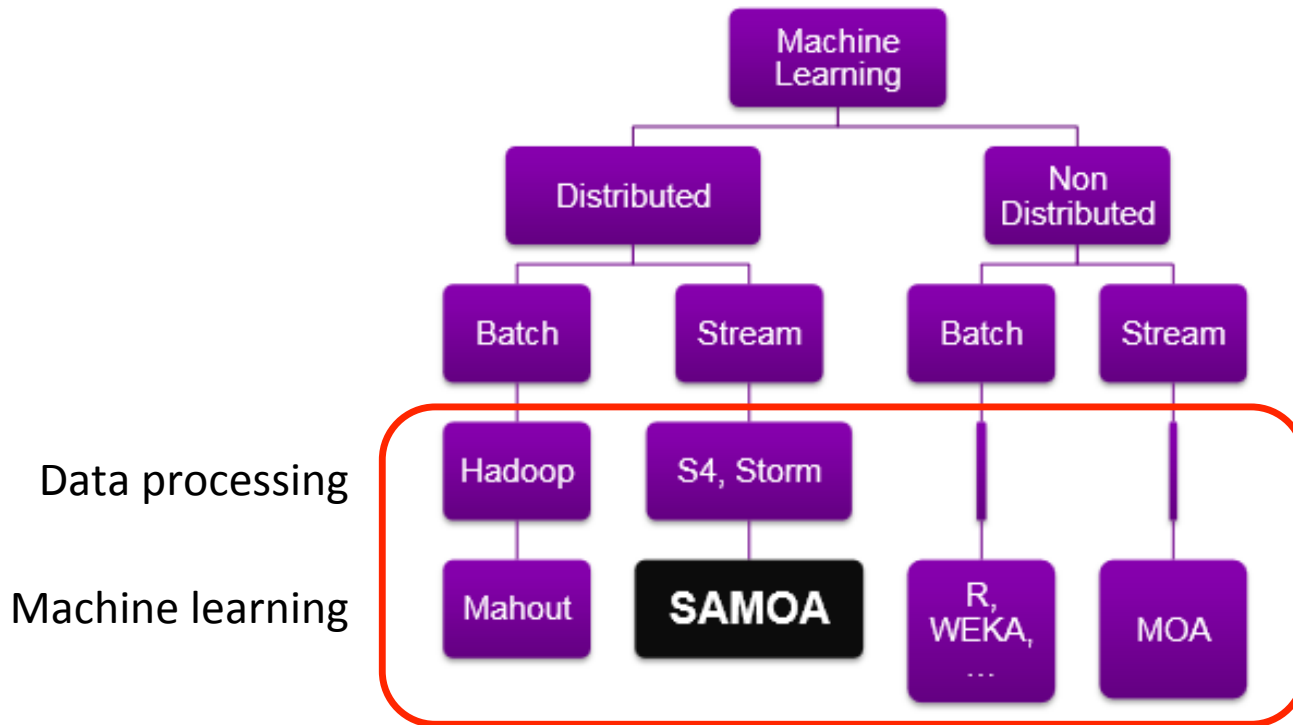
Machine learning

Social-focused algorithms and techniques

Algorithms need to expand the focus from space and time complexity to include social complexity.

De Francisci Morales, Gianmarco. "SAMOA: a platform for mining big data streams." *Proceedings of the 22nd international conference on World Wide Web companion*. International World Wide Web Conference, 2013.

A coming crisis (2007)
Challenges for Social Sciences

"... both the sample survey and the in-depth interview are increasingly dated research methods, which are unlikely to provide a robust base for the jurisdiction of empirical sociologists in coming decades."

Mike Savage and Roger Burrows. "The coming crisis of empirical sociology." *Sociology* 41.5 (2007): 885-899.

- The problem of „found data"
- Data quality & data collection
- single channel and self selection biases
- computation-focused social theories

Outlook

Data mgt. issues

Computational Social Science

Legal issues

Archival issues

Computer Science

Social Sciences

Ability to process large datasets, algorithms, data mining

Knowledge about social theories, methods, and issues

Herman Hollerith (1860-1929)

16

Vielen Dank.

Markus Strohmaier