

The Big Data Paradox and the Network Paradox

Evaluating R&I impacts on sustainable development: what role for big data?
Brussels, 2015-04-18

Christian Gulas
FASresearch, Vienna, Austria



About FASresearch

FASresearch was founded in 1997 and is a social network analytics & strategies company located in Vienna, Austria, with more than 20 years of cross-sector experience in applying the science of networks.

Over the last decade the founder Harald Katzmaier and his international team of scientists and consultants have developed a unique and proprietary set of tools and technologies to empower decision makers in the areas of public affairs, key account management, stakeholder relations, HR and organizational development to navigate through the increasing complexities of our connected world.

What Big Data Analytics and Social Network Analysis have in common

Big Data Characteristics

- Huge volume of data.
- Complexity of data types and structures.
- Speed of new data creation and growth.
- Big data is mostly semi-structured or unstructured.

Big Data Paradox

„The overall working space [of big data] is larger, but the answers lie somewhere in the ‚small‘.“

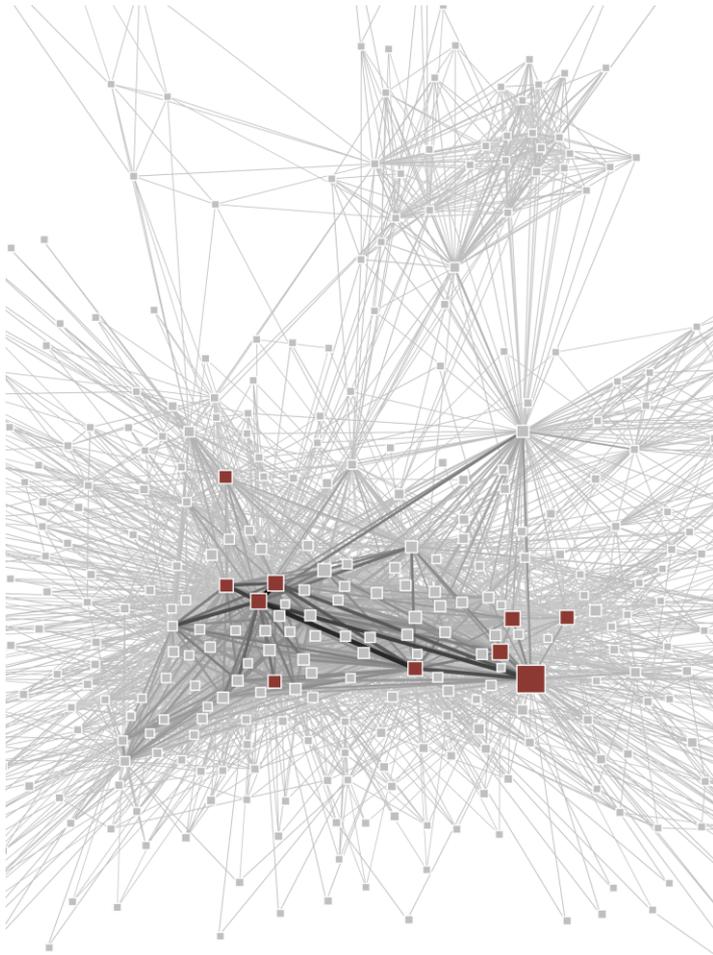
J. Hurwitz et al. (2013).

Big Data Paradox

„The overall working space [of big data] is larger, but the answers lie somewhere in the ‚small“.“

- Big data is about finding the needle in the haystack.
- It is about granularity.
- It is the individual case that matters.

Highly centralized networks



FP7 cooperation network (detail).
Total network: 28.786 institutions, 618.456 links.

- The most central institution (Fraunhofer) controls 11,9% of all network flows.
 - The top 10 institutions control 34,6% of all flows.
-
- Fraunhofer-Gesellschaft e.V.
 - VTT Technical Research Centre of Finland
 - CNR National Research Council, Italy
 - CNRS French National Centre for Scientific Research
 - TNO Netherlands Organisation for Applied Scientific Research
 - CEA Commissariat à l'énergie atomique et aux énergies alternatives
 - Fundacion Tecnalia Research & Innovation
 - CSIC Consejo Superior de Investigaciones Cientificas
 - RIKILT – Institute of Food Safety
 - KU Leuven

Why networks require complete data

1. It is difficult to draw conclusions about networks from samples of relations.

Why networks require complete data

1. It is difficult to draw conclusions about networks from samples of relations.
2. We need to determine precisely the presence or absence of linkages between certain pairs of actors.

Why networks require complete data

1. It is difficult to draw conclusions about networks from samples of relations.
2. We need to determine precisely the presence or absence of linkages between certain pairs of actors.
3. Innovation comes from deviation. For example, new knowledge, cultural change etc. emerge at the periphery of networks.

Network embeddendness of CSOs

FP7

Decile	Embeddedness	HES	REC	PRC	PUB	OTH	CSO1	CSO2	CSO3	CSO4	Total	
1	*****	37,7	23,9	4,4	11,6	4,7	2,3	3,1	5,8	5,6	10,0	Core
2	*****	13,3	14,6	8,4	13,6	8,3	7,9	6,2	11,1	11,2	10,0	
3	*****	9,2	10,7	10,2	9,8	5,7	5,6	6,5	7,7	9,4	10,0	
4	*****	6,7	9,0	10,8	10,0	7,3	10,3	8,5	12,0	6,8	10,0	Semi-periphery
5	*****	5,1	6,5	10,8	11,9	10,4	15,9	15,8	13,5	10,7	10,0	
6	*****	5,7	7,5	10,4	11,2	18,2	15,4	11,6	10,1	16,4	10,0	
7	****	5,7	6,7	10,7	10,4	15,1	11,2	15,2	14,9	13,6	10,0	
8	***	4,5	7,8	11,0	10,4	7,8	14,0	11,9	8,2	9,4	10,0	Periphery
9	**	4,4	5,6	11,8	5,3	8,9	11,2	13,7	11,1	11,5	10,0	
10	*	7,7	7,7	11,5	5,9	13,5	6,1	7,5	5,8	5,4	10,0	
Total		100										

Interim Conclusion

- Innovations – e.g. the idea of cultural change – particularly emerge at the periphery of networks.
- Strengthening the ties between core and periphery of networks is necessary.
- The concept of sustainability may be also regarded as something that has its origins at the periphery of networks.
- Therefore identifying network actors which carry the idea of sustainability is a typical big data problem. It depends on the individual case.

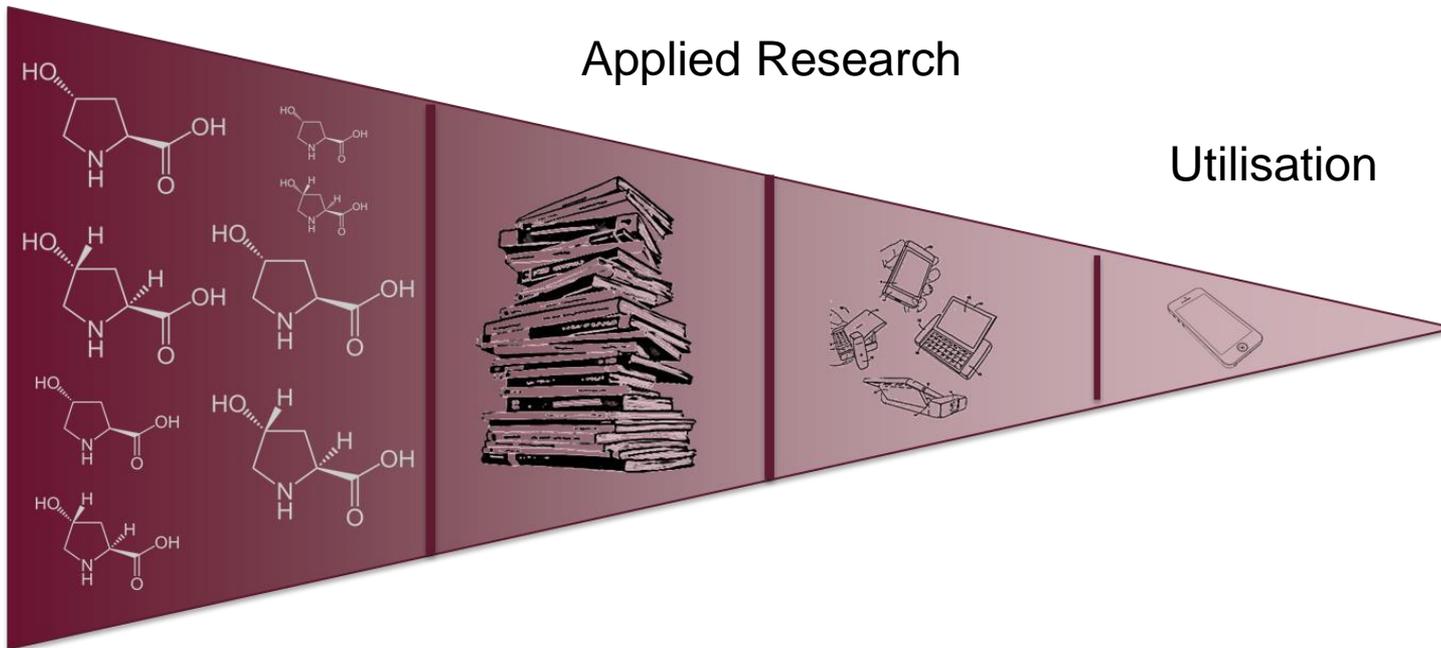
Challenges

Value Creation Ecology

Basic Research

Applied Research

Utilisation



Projects

Publications

Patents

Comparing Publications and Patents

- We wanted to find out which science fields generate a high amount of publications.
- Then we determined in which technology fields many patents are applied.
- Then we wanted to compare research and patent activity.
- Therefore we needed a common classification of science fields for both publications and patents.

Naïve Bayes Classifier

ID	PMID 2081898
Titel	Carbamazepine monotherapy in the treatment of alcohol withdrawal
Abstract	<p>More than 135 different strategies for medical treatment have been described for the treatment of alcohol withdrawal syndromes the substances used most frequently benzodiazepines, barbiturates, or clomethiazol themselves pose some risk for abuse or addiction anticonvulsants, especially carbamazepine CBZ, have been discussed for the treatment of alcohol withdrawal since the early seventies various studies report favourable results with cbz, usually combined with sedative agents nineteen out-patients and 19 in-patients took part in an open study of cbz in alcohol withdrawal the dose of cbz was adjusted individually and ranged from a mean dose of 761 mg on day 1 to 616 mg on day 3 and to 388 mg on day 7 in the group of out-patients, and from 789 mg on day 1, 694 mg on day 3 to 562 mg on day 7 in the sample of in-patients the objective clinical scale in assessment and measurement of alcohol withdrawal ocsamaw was used for treatment evaluation statistical analysis showed a significant improvement on the 5%-level in both groups four in-patients needed concomitant treatment with oxazepam nausea and pruritus were the most common side-effects of CBZ treatment</p>

ID	PMID 19546922
Titel	Q-switching of a fiber laser with a single crystal photo-elastic modulator
Abstract	<p>A study of using a single crystal photo-elastic modulator for active q-switching of a fiber laser is presented the modulator, which oscillates in a longitudinal eigenmode, was realized with litao3 this induces due to the photo-elastic effect a modulated artificial birefringence which modulates the polarization of passing light when used together with a polarizer inside a laser cavity the laser photon life time is strongly modulated and the laser may start to emit laser pulses we realized this with a fiber laser based on a 5m long double clad nd-doped fiber the pulse repetition frequency was 400 khz and the pulse duration 300ns</p>

IPC Patent Class A61K 31

Medicinal preparations containing organic active ingredients.

IPC Patent Class G02 B6

Light guides; Structural details of arrangements comprising light guides and other optical elements, e.g. couplings

Prospect

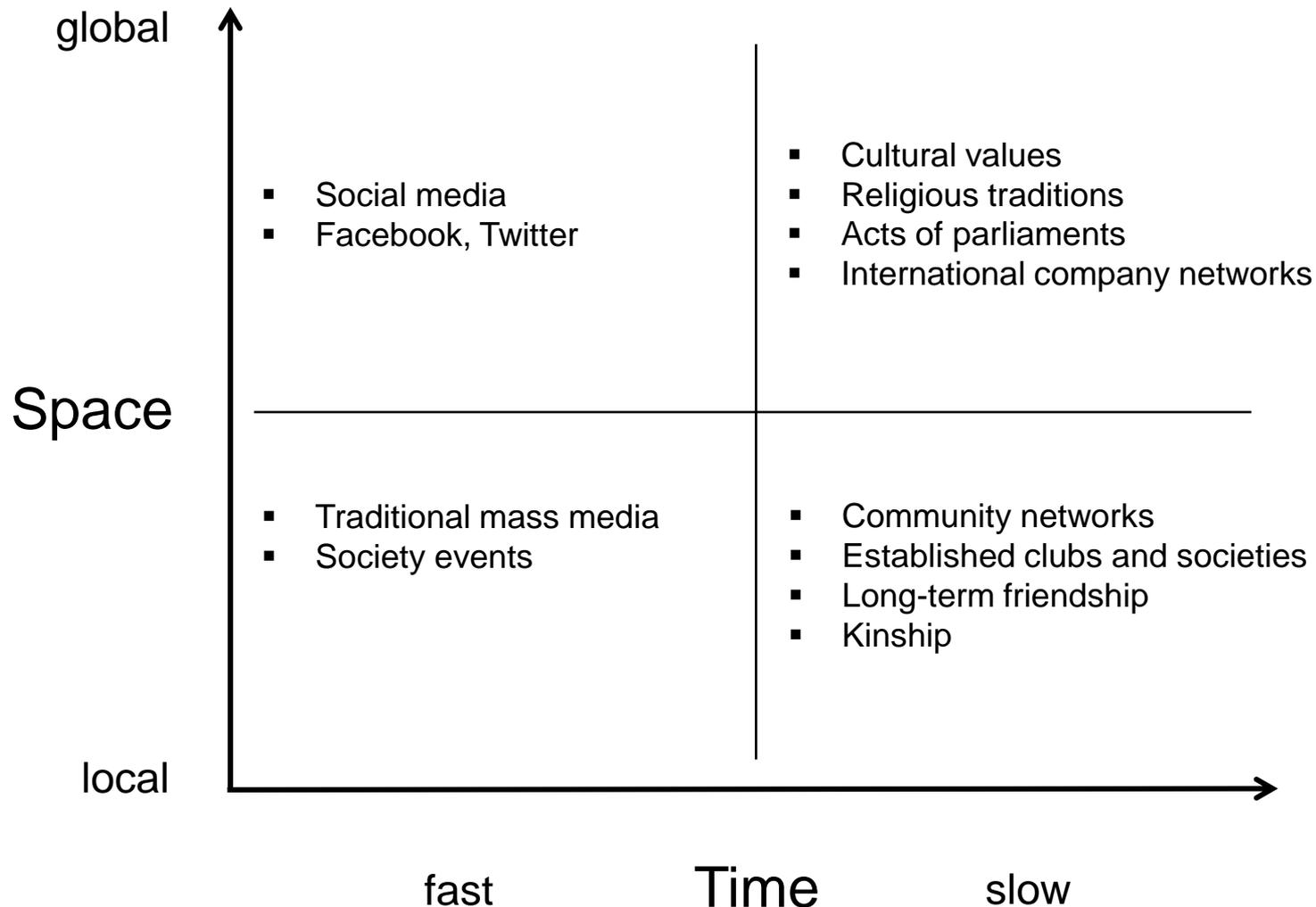
Importance of models

- We often do not understand the meaning of the data because we lack context.
- To understand the meaning of big data and the meaning of networks we need theory based models.
- We do not find these models in the data.

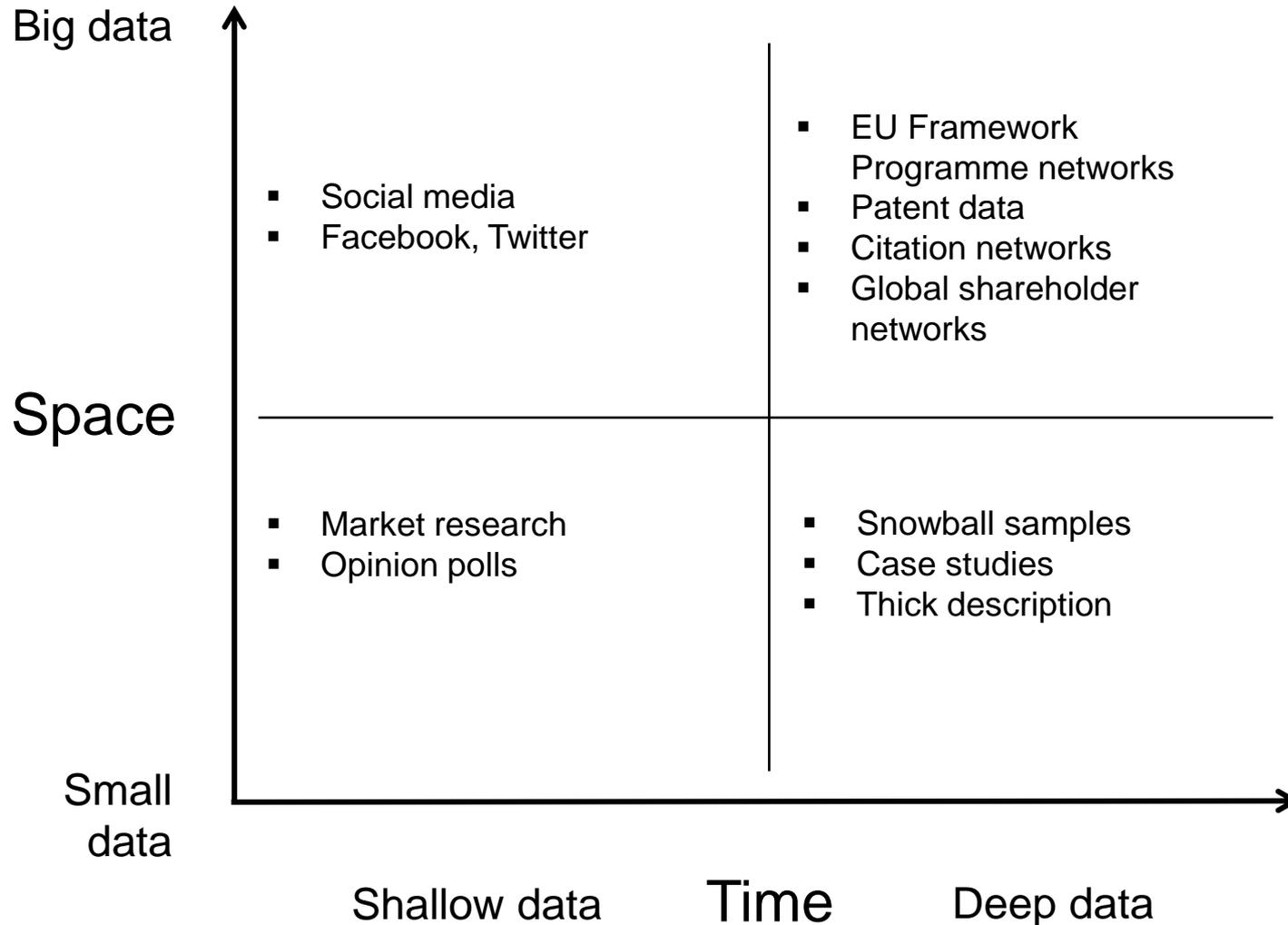
Big data without Context



Classification of Social Entities



Classification of Data



Thank you very much for your attention!

Evaluating R&I impacts on sustainable development: what role for big data?
Brussels, 2015-11-23

Christian Gulas
FASresearch, Vienna, Austria

christian.gulas@fas.at
www.fas-research.com

