Ethnography and Big Data

Ethnography Workshop #1

16.6.2014, Doctoral Training Center, Department Of Computer Science, University Of Trento, Italy

Outline

1. Introductions
2. Outline of the workshop
3. Definitions of Big Data
4. Possible benefits of such a rapprochement, but also some problems it would face
5. Recent developments relevant to chances for rapprochement
6. Some links of relevance
7. Exercise: Take Steps toward an Ethnography of Big Data
8. Report back and summary

Definition of Ethnography

- Ethnography: As fieldwork/participant observation that is
  - experiential,
  - embodied,
  - not only qualitative, and
  - not only descriptive

Definitions of Big Data

"Big data are high volume, high velocity, and/or high variety information assets that require new forms of processing to enable enhanced decision making, insight discovery and process optimization." - 2012, Gartner Group (An American research/advisory firm)

" *Big Data is* no more exact a notion than Big Hair. Nothing magic happens when you get to the 18th or 19th zero...What's new is the way data is generated and processed. It's *like dust in that regard, too. We kick up clouds of it wherever we go*. Cellphones and cable boxes; Google and Amazon, Facebook and Twitter; cable boxes and the cameras at stoplights; the bar codes on milk cartons; and the RFID chip that whips you through the toll plaza — each of them captures a sliver of what we're doing, and nowadays they're all calling home. *It's only when all those little chunks are aggregated that they turn into Big Data; then the software called analytics can scour it for patterns*." - Geoffrey Nunberg

Additional Definitions

"*Big Data is notable not because of its size, but because of its relationality to other data. Due to efforts to mine and aggregate data, Big Data is fundamentally networked. Its value comes from the patterns that can be derived by making connections between pieces of data,* about an individual, about individuals in relation to others, about groups of people, or simply about the structure of information itself." boyd and Crawford

"The first is to make explicit how *"Big Data" are intimately associated with computing; indeed, the notion that they are a separate species of data is connected to the idea that they are generated more or less "automatically," as traces normally a part of mediation by computing. Such data are "big" in the sense that they are generated at a much higher rate than are those of large-scale, purpose-collected sets*" - Hakken

Rapprochement between Ethnography and Big Data: Benefits

- To Ethnography, from Big Data
  - Big Data can be studied as an important phenomenon/aspect of contemporary culture
  - Big Data techniques can be used as a source of information regarding other topics--e.g., digital technology and social change—as part of methodological triangulation
  - A possible focus for contemporary cultural critique: (Crawford and boyd piece)
- To Big Data, from Ethnography
  - Behavioral meaning of scraped data is fraught (see Marres/Weltrevede piece)
  - How much and in what ways to discount technical affordances
  - The impact of analytic tools on data is also difficult to assess
  - Main problem is interpretation, finding meaning in the data
  - All of these ethnography can help with (see Gross, et. al. piece)

School of Informatics and Computing
Colloquium Series

Alex Vespignani

Sternberg University Distinguished Professor, Northeastern University

Tuesday, April 30, 2013

3:00pm

IMU Georgian Room

Modeling and forecast of socio-technical systems in the data-science age

Abstract: The big data revolution and the ever increasing computational power are providing the necessary data, numerical experiments and validation tests which are finally adding an "applied" dimension to complex networks and systems science. Data-driven computational models are finally providing a quantitative understanding of the emerging tipping points and nonlinear properties that often underpin the most interesting characteristics of human collective behavior and networks. These models are at the center of data gathering and computational modeling infrastructures that offer predictive analytic tools for processes as diverse as epidemic outbreaks, information spreading, knowledge diffusion, and Internet packet routing. The field of complexity science has thus entered a new stage of its life in which applications and quantitative results allow us to deal with problems which have a huge impact on our lives such as pandemic emergencies, traffic congestion, systemic risks, and the emergence of social collective behavior.

Here I review some of the recent progress toward the developing of infrastructures for the analysis and forecast of socio-technical systems that integrates the complex features and heterogeneities of real-world systems.

Biography: Alessandro Vespignani is currently Sternberg University Distinguished Professor at Northeastern University in Boston, where he leads the Laboratory for the Modeling of Biological and Socio-technical Systems. He is fellow of the American Physical Society, member of the Academy of Europe, and fellow of the Institute for Quantitative Social Sciences at Harvard University. He is also serving in the board/leadership of several international scientific journals and the Institute for Scientific Interchange Foundation. He is president elected of the Complex Systems Society. Vespignani is focusing his research activity in modeling diffusion phenomena in complex systems, including data-driven computational approaches to infectious diseases spread.

SCHOOL OF INFORMATICS AND COMPUTING
INDIANA UNIVERSITY
Bloomington

Rapprochement: Problems in Getting it

1. Some BD trivializes all previous study of the social as "pre-scientific"
2. Some Ethnography is mindlessly critical of Big Data for, e.g., positivism, hype,
3. Some BD not well informed about previous study of the social
4. Some Ethnography not sufficiently concerned with problems of generalization
5. Some BD interested only in correlation, not explanation
6. Some Ethnography ignorant of BD (although not as much)

The Problem in a Nutshell: Mutual Hostility, or Mutual Ignorance?

- David: More hostility, although more from BD than from Ethnography
- Kalpana: More ignorance, although more from BD than from Ethnography
- Essential to decide which if to develop a program to pursue such a rapprochement
- Some documents which display each

That was then, but this is now

- Last year, spent rest of the workshop spent identifying resources that could be supportive of such a rapprochement.
- Several developments since last year shift the ground on which a rapprochement between Ethnography and Big Data would have to be constructed
- These developments lead us to shift focus, from Ethnography and Big Data—How can they be pursued together—to the Ethnography of Big Data—what kind of a thing is it?
- Place in context of historic dynamic of Computer Science

Some recent developments/events that may make a Rapprochement easier/more necessary in the Long Run (but not in the Short)

- Some spectacular failures, such as Google Flu debacle
- Emergence of a new journal, *Big Data and Society*
- Pushback from other fields: Burrell, boyd
- New users of Big Data: Data Journalists, Citizen Hackers,
- Admission by some that they just don't know what to do with BD, beyond predicting consumption behavior of those already consuming
- Possibly Fewer assertions that only BD is science?
- Associated bridge-building or other issues at least being raised: e.g., ethics

Developments Decreasing Rapprochement Likelihood

- Emergence of new BD structures; such as academic programs in, e.g., "Data Science" that ignore social science
- Powerful extensions of BD rhetorics: e.g., Data-driven Policy, D-d Science, White House Science Policy
- Considerable, even increased funding for BD through, e.g., NSF, European Commission/Horizon 2020
- Data "haves" and "have nots" issue; few limits to the power of those with the most BD and in best position to "keep the gate" of access to it: e.g., Google, Facebook, Amazon, Twitter

Why Chances for Rapprochement have declined

- Were those promoting rapprochement stronger, then we should continue pursuing it;
- But were those decreasing stronger, then we should stop trying to access field in this way.
- We think second set, those discouraging rapprochement, more powerful
- In ethnography, when we can't gain access to the field in one way, we step back and try to find another way.
- So, make BD the subject of ethnographic enquiry, develop an ethnography of BD

Task today: Develop A Preliminary Ethnography of Big Data

- "Hang out" with Data: Use the provided links, and anything else you want, to answer:
  - Who are the BD "natives," and what do they think (and say) BD is;
  - How could one study BD practices, and how do these practices relate to what the natives think; and
  - What do we think we know, if only provisionally, about what BD is, and what else do we need to know?

Topics Relevant to Big Data Today

Big Data in the News

Education and data science

Data driven journalism

Citizens and Big Data

Big Data Hackathons

Opinions from Ethnography

(Access via:
http://disi.unitn.it/~dandrea/workshop/