

Analytic consideration of the society by itself

Marilia M. Pisani¹

Renato Fabbri²

¹CCNH/UFABC, ²IFSC/USP

¹marilia.m.pisani@gmail.com, ²renato.fabbri@gmail.com

Abstract: We aim to consider a set of developments on linked social data and analyses performed over the last years, with a focus on the potential use for the civil society and scientific academy. Various conceptualizations have been gathered by interviewing specialists and State authorities and have been validated by them and their institutions. Data from known social networks, such as Facebook, Twitter, IRC and Email, and from more specialized social participation platforms, have been translated to RDF and linked to these conceptualizations in the form of OWL ontologies [1,2]. Social networks were reported as very stable and their language varies with connectivity [3,4]. Also, resources recommendation and experiments have been performed with such data. How are we to articulate the gathering and analysis of such data with the needs of the civil society and the academy? Will it be valuable to deal with the private and State interests that shape our society? Does Anthropological Physics yield reasonable strategies to collect and analyze data from our social structures? How does the global semantic web of linked data [1] relate to linked social data? This document, with the bibliographic items, is available at <https://github.com/ttm/aavo/raw/master/latex/nuvem/nuvem1.pdf>.

1. Introduction

Social analytics is a term associated to Lars-Hendrik Schmidt [5] in philosophy. One might also relate social analytics to social data mining by extension of the terms: visual analytics, data analytics, social media analytics etc, being used nowadays (e.g. look for these terms in the English language Wikipedia). Interestingly, both social analytics have among core practices the reporting of tendencies of the times. And they put emphasis on issues and analytical outcomes that are very controversial and result from recent technological and scientific developments.

On the critical theory tradition, Marcuse have a central role on questioning the conditions imposed for living by technological advances [6–9]. And the consideration of the authoritarian personality traces [10] has gained relevance due to unforeseen social phenomena such as the recent and worldwide rise of the right-winged manifestations and political representation [11, 12].

On physics and computer sciences, both the advent of the complex networks theory and the big data practices had an impact on the utility of considering social systems. One enables forecast of properties and the use of natural laws for deriving understandings. The other provides the data that represents the real systems in scales that make them useful for management and scientific studies (including experiments as in Sections 2.3 and 2.4). Social networking platforms are massively and constantly used and thus fed. They are most often elaborate suites of software with an interface for the user to interact with other users and often makes use of these recent scientific and technological developments. [3]

In this context, we advocate the analytic consideration of the Society by itself. There are many ways to achieve this, and it is performed to some extent [1, 13–16]. Our focus here is on the mining of the data derived from social platforms. It is not very trivial in our context of monetary obsession to achieve engagement in such analytical undertake. Even if there is money making in the process, the citizens already have their livelihood activities. Therefore, some effort has been given to enable this kind of ‘social participation’. Core goals are:

- Develop the critical view, on specialists and layman, of social media and management performed by the State and the private sector.
- Make advances in the data available for consideration and the algorithms for analysis.
- Implement or study the possibility of implementation of software systems.

2. What has been done

On the following sections, we will focus on developments related to social systems and technology. There are methods, software and scientific discoveries that underpins them [3, 4, 17].

2.1. Social participation data

Data from Facebook, Twitter, Email lists, IRC Channels, ParticipaBR, Cidade Democrática and AA have been translated into linked data (RDF) and has been named the LOSD (Linked Open Social Data [1]). Initial OWL ontologies were derived from the data by automated routines. The dataset is oriented towards general scientific research and the development of analytic interfaces [14, 18–21]. Even so, it has been used e.g. by participants of a SESC workshop and has received some attention from the scientific community [3, 22], and mining of the data revealed very stable patterns as predicted by natural laws and strong differentiation of the language of participants across hubs, intermediary and the periphery [3, 4].

2.2. Ontologies

Beyond the ontologies mentioned in the last section, there are social participation ontologies that describe social participation instances and mechanisms [14]. Such instances and mechanisms are: Conference (Conferência), Forum (Fórum), Committee (Comitê), Council (Conselho), Ombudsman (Ouvidoria), Public Consultation (Consulta Pública), Dialog Table (Mesa de diálogo), Monitoring Table (Mesa de Monitoramento), Intercouncil Forum (Fóruns Interconselhos), Audience (Audiência), Virtual Environment (Ambiente Virtual). They received additional ontologies relating them to documents and larger scopes.

These ontologies are contextualized by the whole semantic web (or Giant Global Graph [1]) and by other social participation ontologies: the OPa (Ontology of ParticipaBR [14]), the OPS (Ontology of Social Participation [23]), the OCD (Ontology of Cidade Democrática [14]), the OAA (Ontology of AA [14]). There are minor ontologies, such as the one about the Magic Box (Caixa Mágica [24], a social participation wifi gadget). These OWL ontologies (and the SKOS vocabularies which most of them have, following a pattern suggested in [25]) enable sound conceptual discussions, navigation by humans and machine (data discovery), automated inference; and data organization, integration and linkage by means of conceptualizations. It seems reasonable to use these ontologies to link the social participation data and enable a scalable participatory legacy of machine- and human- friendly representation of data.

2.3. Critical theory and anthropological physics

The potential for prejudice and the study of the self were considered in [3, 4, 26]. In summary, the hubs, intermediary and periphery sectorialization of a network might be achieved (or imposed) in a social network. It yields a classification of the participants by the comparison of the degree (or strength) distribution against that of an Erdős-Rényi network. Hubs are valued by literature (although the intermediary are also reported as authorities and structurally most important agents) [27], which favors the potential for prejudice of a quantitative classification (or typology) of human beings. At the same time, the classification of the participants (hub, intermediary or peripheral) vary in time constantly [3, 4] and across scales and networks, which minimized the potential for prejudice of such classification.

Violence, prejudice and authoritarianism, and the authoritarian personality and technology, are frequently co-occurring themes in the two dossier of the Nexos Networks for Critical Theory and Interdisciplinary Research [28]. We also considered the cold-blooded reception of the reality enabled technological devices such as drones [22].

The Anthropological Physics is concerned with the observation of natural phenomena in the social structures of the observer and has been considered in [3, 15, 16]. Comte talked about anthropological physics, but it had the sense which physical anthropology has, which is a branch of anthropology (the other is sociocultural anthropology). Anthropological physics, in the sense we use and understand correct, is physics (research on the natural phenomena mainly of complex networks) with anthropological aspects. The writing and study of diaries is a traditional ethnographic asset, which resembles the collection and study of one's data by the person itself, social data. The current guidelines are centered on keeping the process open, using free and open source tools and file formats, and publicly accessible repositories of software, gadgets, data, writings, etc; while gathering and studying the social data related to the researcher (and other data if needed and afterwards). [3, 15, 16, 26]

2.4. Self-transparency, AA and the fundamental cycle

A system for sharing work, development and research processes (actually, any dedication) is described in [29, 30]. It is the Algorithmic Autoregulation (AA) methodology and it has received a number of software implementations. It is

based on voluntary logging of messages (called shouts) about what is being done (reading, coding, etc). The Shouts may be linked to sessions and screencasts or be blind-reviewed by other users. The methodology and software have initiated discussions e.g. about documenting academic dedication not only through traditional documents (e.g. articles, books) and participation (e.g. conferences, thesis defense), as currently performed e.g. by Lattes Curriculum. Such documentation (log of dedications) would enable one to dedicate more time to individual tasks and make more paced and relevant contributions.

The concept has reached other initiatives, such as the Brazilian federal portal of social participation ParticipaBR. And have incited discussions that yield interesting concepts, such as the fundamental cycle, which is an idealized interaction between Society and a management sector of it, such as the State. The individuals that feel the need to change the way living is imposed by the State should be able to act on the State, have results and get paid for the time they put into performing the social participation. The existence of such a cycle would eradicate extreme poverty, as the poor would be paid to act and change State measures for the poor or a community. Thus, the ideal fundamental cycle is a social participation model which is very shabbily practiced and is defined by giving the one in need good conditions to change the context which puts he/she in need. The mechanisms or institutions that give the individual the conditions and means to act is then called the State or Government or, more accurately nowadays, the Institutionalized Power. Notice that, as such, the institutionalized power have the goal to allow the interference of the individual in its management and that this is extended to the private initiatives. Social participation in the private sector seems to be very incipient.

2.5. *Audiovisual Analytics platform*

Described in [31], it is an envisioned audiovisual analytics software for social data, with focus on network and textual data. It will also be sent to the I Workshop @NUVEM, together with this document, for separate appreciation.

3. **What shall we do?**

Most certainly we will make the LOSD and social participation ontologies (Sections 2.1 and 2.2) available e.g. on DataHub and Data.World. In the Nexos research network, we considered a number of times to study authoritarian personality traces [10] on this data. And to further consider the Anthropological Physics aspects that arise from studying our own social structures [15,16]. The self-transparency is having some attention in scientific writings and hacks [29,30] but seems to be lacking user bases. The developments for Audiovisual Analytics are incipient and have the development focus of only one researcher and developer. The Nexos research network is very active nation wide. It has an overall tendency to tackle subjects within the critical theory tradition and thus considers the society and potential enhancements through criticism of its organization, which is valuable for the technological developments being made [28].

Research groups @Nuvem of Cloud Computing and Intelligent Societies might have further context to share about these issues. For example, are there SparQL endpoints with (Brazilian) social participation (including self-transparency) data? Is it necessary to contact lawyers to better know the limits of our possibilities to gather and research our social data within the Anthropological Physics perspective and are there well known guidelines? Is there more participatory linked data available in Brazil? Ontologies? We might benefit from directions on better linking LOSD to the semantic web (DBpedia, other participatory data, etc), and for a reasonable way to keep the data online (through DataHub or Data.World or both? A Pubby-like interface?), and to develop an audiovisual analytics software (persistence, web?, analysis methods, audiovisual rendering, etc, as described in [31]).

Other questions that might initiate nice discussions or entail collaboration are: how to enable a self-transparency user base? Where to keep the linked data? How to manage the ontologies and keep their constant development (as needed and predicted in the literature)? How to achieve a reasonable use of our social (open linked) data? Is it possible to have an audiovisual analytics platform with our own (scraped) social data, with facilities for media rendering and interaction experiments (collection and diffusion of information), that enables the user base and interested parties to collectively gather social data, analyses and conceptualizations? Is the societal consideration of our social data relevant for equilibrium with the State and private sectors? Is it possible to achieve social participation in the private sector, e.g. to have civil representation in the management of companies, to regulate matters such as YouTube adds in Google, nutrient standards in McDonald's, and other things in other major and minor institutions?

3.1. *Acknowledgments*

FAPESP (project 2017/05838-3); ICMC/USP and UFABC researchers; the open source and free software communities that made these developments possible.