Social Set Analysis: Four Demonstrative Case Studies

Ravi Vatrapu
Computational Social Science Laboratory (CSSL)
Department of IT Management, Copenhagen Business School, Denmark & Westerdals Olso School of Arts, Communication and Technology, Norway
vatrapu@cbs.dk

Abid Hussain
Computational Social Science Laboratory (CSSL)
Department of IT Management, Copenhagen Business School, Denmark
ah.itm@cbs.dk

Niels Buus Lassen
Computational Social Science Laboratory (CSSL)
Department of IT Management, Copenhagen Business School, Denmark
nbl@evalua.dk

Raghava Rao Mukkamala
Computational Social Science Laboratory (CSSL) Department of IT Management, Copenhagen Business School, Denmark
rao@itu.dk

Benjamin Flesch
Computational Social Science Laboratory (CSSL)
Department of IT Management, Copenhagen Business School, Denmark
bfitservice@gmail.com

Rene Madsen
Computational Social Science Laboratory (CSSL)
Department of IT Management, Copenhagen Business School, Denmark
john.rene.madsen@gmail.com

ABSTRACT

This paper argues that the basic premise of Social Network Analysis (SNA) – namely that social reality is constituted by dyadic relations and that social interactions are determined by structural properties of networks-- is neither necessary nor sufficient, for Big Social Data analytics of Facebook or Twitter data. However, there exist no other holistic computational social science approach beyond the relational sociology and graph theory of SNA. To address this limitation, this paper presents an alternative holistic approach to Big Social Data analytics called Social Set Analysis (SSA). Based on the sociology of associations and the mathematics of classical, fuzzy and rough set theories, this paper proposes a research program. The function of which is to design, develop and evaluate social set analytics in terms of fundamentally novel formal models, predictive methods and visual analytics tools for Big Social Data. Four demonstrative case studies, employing SSA, covering the range of descriptive, predictive, visual and prescriptive analytics are presented and briefly discussed.

Categories and Subject Descriptors
Information systems: Information systems applications: Collaborative and social computing systems and tools

General Terms
Formal Methods, Design

Keywords
Social media, Big Social Data analytics, computational social science, social set analysis, social network analysis

1. INTRODUCTION

Social media are fundamentally scalable communications technologies that turn Internet based communications into an interactive dialogue platform [1]. On the "demand-side", users and consumers are increasingly turning to various types of social media to search for information and to make decisions regarding products, politicians, and public services [2]. On the "supply- side", terms such as "Enterprise 2.0" [3] and "social business" [4] are being used to describe the emergence of private enterprises and public institutions that strategically adopt and use social media channels to increase organizational effectiveness, enhance operational efficiencies, empower employees, and co-create with stakeholders. The organizational and societal adoption and use of social media is generating large volumes of unstructured data that is termed 'Big Social Data'. New organizational roles such as *Social Media Manager, Chief Listening Officer, Chief Digital Officer,* and *Chief Data Scientist* have emerged to meet the associated technological developments, organizational changes, market demands, and societal transformations. However, the current state of knowledge and practice regarding social media engagement is rife with numerous technological problems, scientific questions, operational issues, managerial challenges, and training deficiencies. As such, not many Danish organizations are generating competitive advantages by extracting meaningful facts, actionable insights and valuable outcomes from Big Social Data analytics. Moreover, there are critical unsolved problems associated with how Big Social Data integrates with the existing datasets of an organization (that is, data from internal enterprise systems) and its relevance to the key performance indicators for the organization. To address the diverse but interrelated issues associated with Big Social Data for organizations, this paper presents interdisciplinary research project to design, implement and evaluate a set theoretical approach using Big Social Data from Facebook, Twitter and other social media channels.

Specifically, this paper introduces a research programs situated in the domains of Data Science [5,6,7] and Computational Social Science [8] with practical applications to Social Media Analytics in organizations [4, 9, 10]. It addresses one of the important theoretical and methodological limitations in the emerging paradigm of Big Data Analytics of social media data [11]. In particular, it address the major limitation in existing research on Big Social Data analytics that computational methods, formal models and software tools are largely limited to graph theoretical approaches [12], such as SNA [13], and are informed by the social philosophical approach of relational sociology [14]. There are no other unified modeling approaches to social data that integrate the conceptual, formal, software, analytical and empirical realms [15]. This results in a research problem when analyzing Big Social Data from platforms like Facebook and Twitter as such data consists of not only dyadic relations but also individual associations [16]. For Big Social Data analytics of Facebook or Twitter data, the fundamental assumption of SNA that social reality is constituted by dyadic relations and interactions are determined by structural positions of individuals in social networks [17] is neither necessary nor sufficient [18]. To overcome this limitation and address the research problem, this paper proposes an alternative holistic approach to Big Social Data analytics that is based on the sociology of associations and the mathematics of set theory and offers to develop fundamentally new methods and tools for Big Social Data analytics, SSA.

1.1 Overarching Research Question

In order to further research in this area we as ourselves the following research question: How, and to what extent, can methods and tools for SSA derived from the alternative holistic approach to Big Social Data analytics based on the sociology of associations and the mathematics of set theory result in meaningful facts, actionable insights and valuable outcomes?

2. Theoretical Framework

2.1 Philosophy: Computational Social Science

Large-scale and content driven social media platforms such as Facebook are of extreme importance to organizations in terms of marketing communications, corporate social responsibility, democratic deliberation, public dissemination etc. Social media analytics in practice [9, 10, 19] has been based on an implicit, inherent and latent understanding of social associations as expressed by metrics and key performance indicators such as brand sentiment, brand associations, conversation keywords, reach etc.

The theoretical aim of this project is to make a positive contribution in terms of an associational sociological approach to Big Social Data analytics in order to address the twin problems of (a) largely absent academic research, and (b) mostly latent industry practice on social media analytics from a sociology of associations in general and SSA in particular.

As such, the primary scientific objectives of this project are to theoretically formulate, mathematically model and empirically investigate an alternate holistic approach based on associational sociology [20], set theory and fuzzy set theory [21], and SSA [22]. To achieve these objectives, the theory of social data is developed and discussed next. Table 1 below shows the philosophical comparison between the traditional approach of SNA and the proposed new approach of SSA. Our criticism of the limitations of the relational sociology assumptions of SNA are with respect to large-scale social media platforms that are increasingly social content driven.

Table 1: Two Philosophies of Computational Social Sciences

	Social Network Analysis	**Social Set Analysis**
Basic Premise	There exists a relation between social actor A and social actor B	There exists an association by actor A with some entity E which can be an actor or an artifact
Social Action	Interpersonal Relations	Individual Actions
Unit of Analysis	Dyadic	Monadic, Dyadic & Polyadic
Social Structure	Networks	Sets
Mathematics	Graph Theory	Set Theory

2.2 Set Theoretical Big Social Data Analytics

As articulated in [22], based on Smithson and Verkuilen [23] there are five advantages to applying classical set theory [24] in general and fuzzy set theory [21] in particular to computational social sciences:

(1) Set-theoretical ontology is well suited to conceptualize vagueness, which is a central aspect of social science constructs. For example, in the social science domain of marketing, concepts such as brand loyalty, brand sentiment and customer satisfaction are vague.

(2) Set-theoretical epistemology is well suited for analysis of social science constructs that are both categorical and dimensional. That is, set-theoretical approach is well suited for dealing with different and degrees of a particular type on construct. For example, social science constructs such as culture, personality, and emotion are all both categorical and dimensional. A set-theoretical approach can help conceptualize their inherent duality.

(3) Set-theoretical methodology can help analyze multivariate associations beyond the conditional means and the general linear model. In addition, set theoretical approaches analyze human associations prior to relations and this allows for both quantitative variable centered analytical methods as well as qualitative case study methods.

(4) Set-theoretical analysis has high theoretical fidelity with most social science theories, which are usually expressed logically in set-terms. For example, theories on market segmentation and political preferences are logically articulated as categorical inclusions and exclusions that natively lend themselves to set theoretical formalization and analytics.

(5) Set-theoretical approach systematically combines set-wise logical formulation of social science theories and empirical analysis using statistical models for continuous variables. For example, in the case of predictive analytics, it is possible to employ set and fuzzy theory to dynamically construct data points for independent variables such as brand sentiment (polarity, subjectivity, etc.).

2.3 Theory of Social Data

The theory of social data is drawn from the theory of socio-technical interactions [25,26,27]. Social media platforms such as Facebook and Twitter, at the highest level of abstraction, involve individuals interacting with (a) technologies and (b) other individuals. These interactions are termed *socio-technical interactions*. There are two types of socio- technical interactions: 1) interacting with the technology (for example, using the Facebook app on the user's smartphone and 2) interacting with others socially using the technology (for example, liking a picture of a friend in the Facebook app on the user's smartphone). These socio-technical interactions are theoretically conceived as (a) perception and appropriation of socio-technical affordances, and (b) structures and functions of technological intersubjectivity. Briefly, socio-technical affordances are action-taking possibilities and meaning-making opportunities in an actor-environment system bound by the cultural-cognitive competencies of the actor and the technical capabilities of the environment. Technological intersubjectivity (TI) refers to a technology supported, interactional social relationship between two or more actors. A more detailed explanation of the theoretical framework regarding its ontological and epistemological assumptions and principles, is beyond the scope of this paper but for details, please consult Vatrapu (2010).

Socio-technical interactions as described above result in electronic trace data that is termed "social data". For the example discussed of a Facebook user liking a friend's picture on their smartphone app, the social data is not only rendered in the different 'timelines' of the user's social network but it is available via the Facebook graph API. Large volumes of such micro-interactions constitute the macro world of Big Social Data that is the analytical focus of this paper. Based on the theory of social data described above, the conceptual model of social data is presented below.

2.4 Conceptual Model of Social Data

Figure 1 presents the conceptual model of social data.

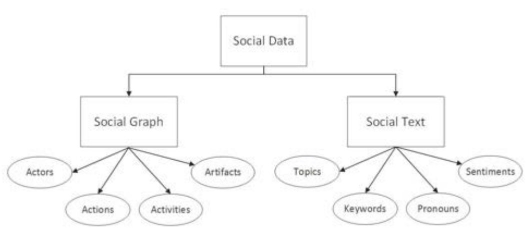

Figure 1. Social Data Model

Social data consists of two types: *Social Graph* and *Social Text*. Social Graph relates to the first aspect of socio-technical interactions that involve perception and appropriation of affordances (which users/actors act up on which technological features to interact with what other social actors in the systems). Social Text relates to the second aspect of socio-technical interactions: structures and functions of technological intersubjectivity (what the users/actors are trying to communicate to each other and how they are trying to influence each other through language). Social Graph consists of the structure of the relationships emerging from the appropriation of social media affordances such as posting, linking, tagging, sharing, liking etc. It focuses on identifying the **actors** involved, the **actions** they take, the **activities** they undertake, and the **artifacts** they create and interact with. Social Text consists of the communicative and linguistic aspects of the social media interaction such as the **topics** discussed, **keywords** mentioned, **pronouns** used and **sentiments** expressed. Figure 1 presents the descriptive model of social data.

2.5 Illustrative Example of Social Data

Let us say that the research domain is online politics and the research question is to what extent do Facebook walls function as online public spheres [28]?. Then, the set theoretical approach to computational social science can be employed to specify measures of the extent to which the Facebook Walls are serving as online public spheres as discussed below.

The Social Graph aspect of social data allow us to examine the breadth of the public sphere by reporting the overall number of posts made (artifacts), which of the Facebook walls received most posts and whether they linked out to other sources of information. In addition to looking at the posts in the aggregate we also can look at them individually and map linkage across walls. Was the posting entirely independent in that individuals (actors) only posted (action) to one wall or did they post more widely on two or three walls?

The Social Text aspect of data allow us to examine the depth of the engagement taking place through the Facebook walls and thus whether walls are acting as an online public space. In particular we can look at three key aspects of the posts: their length, their focus in terms of the use of pronouns in the posts - categorizing them as inward (use of ***I***) or outward (use of ***you*** and ***they***); and the direction of sentiment being either positive or negative.

3. Research Objectives and Hypotheses

Objective#1 (O1): Formulate robust computational methods for Social Set Analytics

Objective#2 (O2): Build and Evaluate Predictive Models for Social Set Analytics

Objective#3 (O3): Design and Evaluate Dashboards of Visualization Techniques for Social Set Analytics

Formalism is at the heart of computation and O1 is aimed at specification and verification of formal models for social data. This is the fundamental first step in the development and evaluation of the new set theoretical approach to big data proposed. The set theoretical formalism of Big Social Data in O1 then informs the development and evaluation of predictive methods and models in O2. Together, O1 and O2 will inform the design, development and evaluation of new visual analytics techniques in dashboards for researchers and practitioners in the real world. Three research hypotheses corresponding to the three scientific objectives are advanced below:

Hypothesis#1 (H1): Formal models of social data based on set theory in general and classic sets, fuzzy sets and rough sets in particular will outperform formal models based on graph theory in solving content related Big Social Data analytics.

Hypothesis#2 (H2): Predictive methods based on social data measures derived from domain-specific and set-theoretical formal models will outperform existing predictive models in terms of accuracy, goodness of fit, and explanatory power.

Hypothesis#3 (H3): SSA dashboards that implement set theoretical formal models and predictive methods will be more efficient and effective compared to existing social network visualizations for interactional dynamics tasks such as identification of user affinity and loyalty, evolution of sentiment, detection of bursts, and formation of communities.

All three hypotheses are theoretically informed by associational sociology [20, 29]. H1 is a computational hypothesis that is also informed by well-known data mining techniques drawn from Fuzzy logic [21] and rough sets [30] that have not yet been extensively applied to social data analytics, despite social science applications. H2 is a statistical hypothesis that is further informed by prior work in social psychology [31] and consumer psychology [32] that argue for the primacy of individual psychological associations at micro-level of social interactions for outcomes at the meso-level of organizations and the macro-level of markets, communities and societies. H3 is a design hypothesis that is additionally informed by the new principle of visual analytics to "analyze first and visual next" [33] which explores the dashboard design space with the set theoretical formal models and predictive methods.

In the next section, three demonstrative case studies are presented to illustrate the feasibility of the SSA approach and to discuss the empirical findings obtained.

4. Case Study #1: Predictive Analytics

Recent research in the field of computational social science have shown how data resulting from the widespread adoption and use of social media channels such as Twitter can be used to predict outcomes such as movie revenues, localized moods, and epidemic outbreaks. Underlying assumptions for this research stem from predictive analytics in that social media actions such as tweeting, liking, commenting and rating are proxies for user/consumer's attention to a particular object/product and that the shared digital artifact that is persistent can create social influence. In this paper, we demonstrate how social media data from Twitter and Facebook can be used to predict the quarterly sales of iPhones and revenues of clothing retailer, H&M, respectively. Based on a conceptual model of social data consisting of Social Graph (actors, actions, activities, and artifacts) and Social Text (topics, keywords, pronouns, and sentiments), we develop and evaluate linear regression models that transform (a) iPhone tweets into a prediction of the quarterly iPhone sales with an average error close to the established prediction models from investment banks [34] and (b) Facebook likes into a prediction of the global revenue of the fast fashion company, H&M. We discuss the findings and conclude with implications for predictive analytics with Big Social Data.

Our basic premise is that social media actions can serve as proxies for user's attention and as such have predictive power. Our central research question for this demonstrative case study is: *To what extent can Big Social Data predict real-world outcomes such as sales and revenues?*

We adhered to the methodological schematic recommended by Shmueli and Koppius [35] for building empirical predictive models. Table 2 below presents the dataset collected for predictive analytics purposes of this paper.

We built on and extended the predictive analytics method of Asur and Huberman [36] and examine if the same principles for predicting movie revenue with Twitter data can be used to predict iPhone sales and H&M revenues for Facebook data. That is, if a tweet/like can serve as a proxy for a user's attention towards a product and an underlying intention to purchase and/or recommend it. We extend Asur and Huberman (2010) in three important ways: (a) addition of Facebook social data, (b) theoretically informed time lagging of the independent variable, social media actions, and (c) domain-specific seasonal weighting of the dependent variable, sales/revenues.

In summary, drawing from the theoretical framework of Awareness, Interest, Desire and Action (AIDA) and Hierarchy of Effects models in marketing [37], combined with the assumption that social media actions such as tweeting, liking, commenting and rating are proxies for user/consumer's attention to a particular object/product, we demonstrated how social media data from Twitter and Facebook can be used to predict real-world outcomes such as sales and revenues. Figure 2 and figure 3 present the predicted vs. actual charts.

Table 2: Overview of Dataset

Company	Data Source	Time Period	Size of Dataset
Apple	Twitter	2007 → October 12, 2014	500 million+ tweets containing "iPhone"
H & M	Facebook	January 01, 2009 → October 12, 2014	~15 million Facebook events

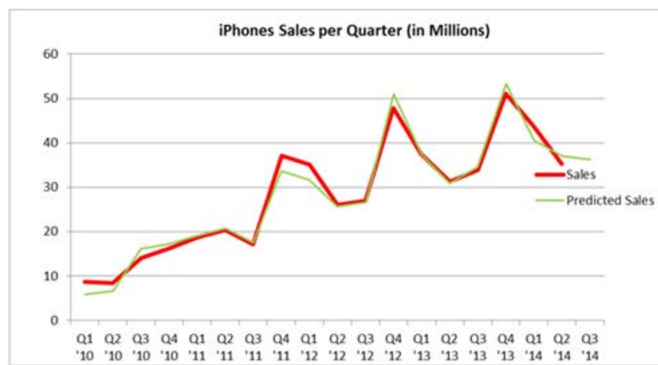

Figure 2: Predicted vs. Actual Sales of iPhones from Tweets

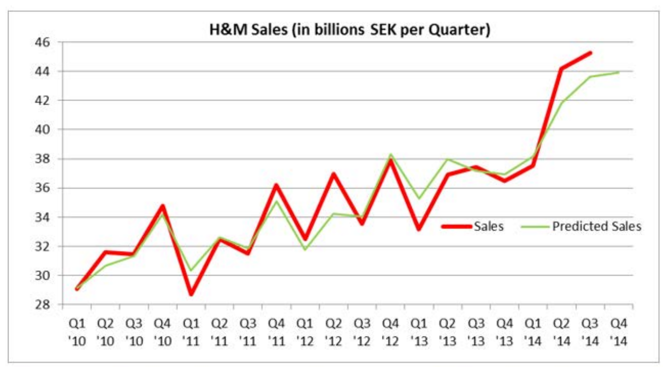

Figure 3: Predicted vs. Actual Revenues of H&M from facebook likes

5. Case Study #2: Visual Analytics

This case study presents a state-of-the art visual analytics dashboard of approximately 180 million Facebook actions from 11 different companies that have been mentioned in the traditional media in relation to the garment factory accidents in Bangladesh.

The application domain for the dashboard is Corporate Social Responsibility (CSR) and the targeted end-users are CSR researchers and practitioners. The design of the dashboard was based on the SSA approach to computational social science [18]. The development of the dashboard involved cutting-edge open source visual analytics libraries from D3.js and creation of new visualizations (actor mobility across time, conversational comets etc.). Evaluation of the dashboard consisted of technical testing, usability testing, and domain-specific testing with CSR students and yielded positive results.

The garment industry in Bangladesh is the second-largest exporter of clothing after China, and employs more than 3 million - mainly female - workers. The garment industry in Bangladesh has rapidly grown during the past 20 years while approving of lax safety regulations and frequent accidents [38]. According to [39]:

"*Bangladesh's garment sector […] employs forty percent of industrial workers and earns eighty percent of export revenue. Yet, the majority of workers are women. They earn among the lowest wages in the world and work in appalling conditions. Trade unions and associations face brutal conditions as labour regulations are openly flouted*."

On April 24, 2013, factory disasters in the Bangladeshi garment sector culminated in the largest textile industry tragedy to date with the collapse of Rana Plaza, a factory building in an industrial suburb of Bangladesh's capital Dhaka in which more than 1100 garment workers died during the factory's collapse and fires. This event has been reported by media outlets all over the world and deeply shocked many consumers of clothing products originating from Bangladesh. In various research publications, the safety and struggles of workers in the Bangladesh garment industry have been widely discussed [39], including ongoing protests [40], globalization-related problems [41] and ethical aspects of the factory disasters [42]. Nevertheless, the lack of publicly shown empathy by many major textile industry companies created a public outcry against perceived unethical behaviour in textile industry supply chains. In many cases, this public outcry was expressed by consumers and directly addressed to the respective clothing brands, which were in the consumers' immediate reach through means of social media channels such as Facebook. The factory disasters in Bangladesh prompted major textile industry brands like H&M and Walmart to join campaigns supporting textile workers' rights in Bangladesh. The introduction of better methods of supply chain management, such as social contracts in supply chains, showed a more sustainable, but lagging response [43]. Our dashboard is a computational artifact that is situated in this particular nexus of social science research and practice.

Our research methodology consisted of seven steps. First, we assembled a list of real-world events with respect to the Bangladesh factory accidents. Second, we created a list of the traditional news media (print newspapers, TV and radio) reports of the real-world factory accidents in Bangladesh. Third, we reviewed the media reports and extracted a list of 11 multi-national companies that have been frequently mentioned in the traditional media reports in relation to the Bangladesh garment factory accidents. Fourth, since Strategic Corporate Social Responsibility communication is conducted by companies on their Facebook pages, we extracted the full archive of the social data from the Facebook walls of the 11 companies using the Social Data Analytics Tool (SODATO) [44]. Fifth, we designed, developed and evaluated the Social Set Visualizer dashboard of this Facebook corpus of approximately 180 million data points. Sixth, we addressed and answered a set of research questions (see below) using the dashboard. Seventh and finally, we deployed the dashboard (with access control) to support ongoing research by CSR researchers and practitioners. Figure 4 presents the dashboard.

Due to space restrictions, we cannot report all the findings from the visual analytics case study. That said, selected findings are that (a) the global supply chain concerns with regard to Bangladesh garment factories have been expressed by Facebook users from as far back as 2009, (b) there are many instances of authentic displays of support and expressions of empathy from Facebook users as well as robotic incidents of slactivism, (c) many of the use of the word

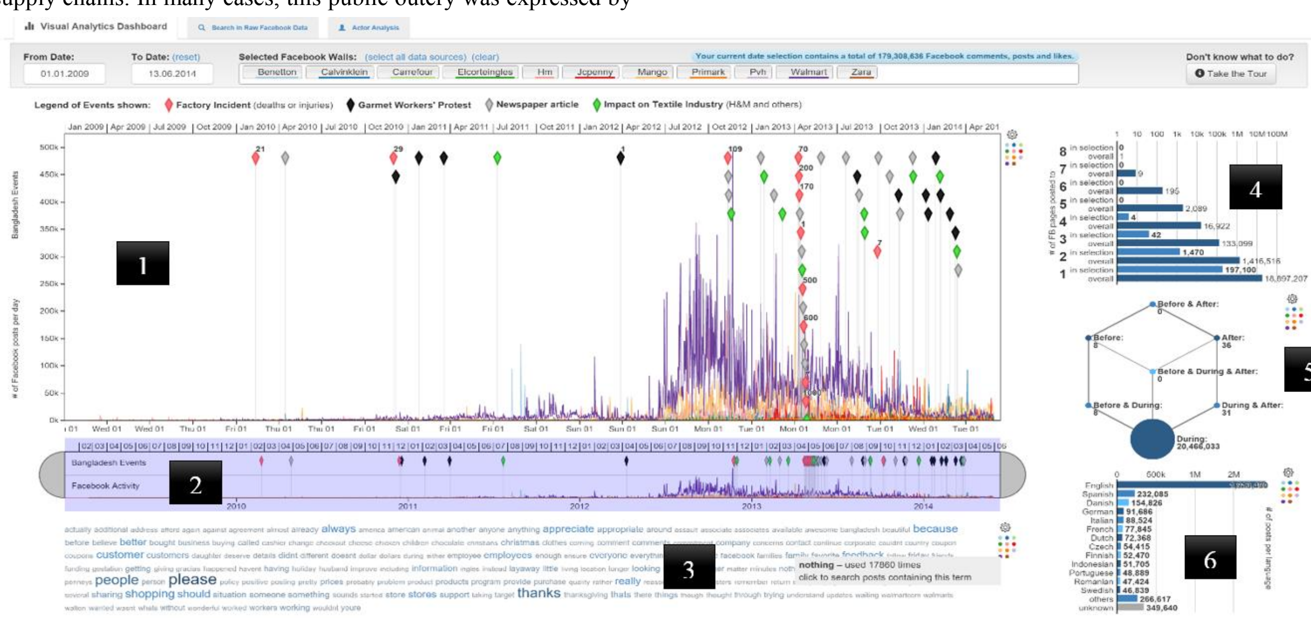

Figure 4: Social Set Visualizer: For the selected time period (see date range fields in top-left) and selected facebook walls (see colour coded selection bubble chart next to visualizations) → [1] Facebook Activity Chart; [2] Timeline of Bangladesh Factory Accidents & Facebook Actions; [3] Word Cloud of Text from Posts and Comments; [4] Actor Mobility across Space (facebook walls); [5] Actor Mobility across Time (before, during, after time-period of selection and combinations of them); and [6] Language Distribution

"please" with respect to opening of new stores in the case of H&M, (d) protestors and activists employed different social media strategies on the different Facebook walls of companies but with little evidence for social influence (in terms of the number of likes and comments on their posts), (e) similarly, companies followed not only different CSR strategies but also different social media strategies before, during and after the Bangladesh garment factory accidents, (f) for almost all of the accidents, a majority of the users posting during the news-cycle of it (that is, traditional media coverage of the event) did not return to the Facebook walls (that is, social media engagement during factory accidents is episodic with little overlap to the "business-as-usual" period before or after the accident).

6. Case Study # 3: Descriptive Analytics

This case study describes and demonstrates the analytical framework and computational aspects of SODATO in the domain of political science. The method has been previously applied to the U.S. elections of 2008 [45,46,47]. Here we replicated and extended the analysis to Danish Elections in 2011. Our research question was to measure the extent to which Facebook walls function as online public spheres. To do so we extracted the Facebook walls of three prominent candidates in the 2011 Danish general election.

These data were then linked with unique keys for Facebook wall id and poster ids. This meant that we could track the extent to which individuals appeared in all three walls or just one or two. We also aggregated individual data by politicians so we had a record of all their posts/status updates and comments. For the Social Graph analysis, we compiled a list of unique posters to the three Facebook walls being studied. In order to count the links and videos we combined all the links from the free text of the posts and comments as well as dedicated link type of posts. String matching database queries were applied to count number of total links and links to particular web host, for example YouTube.

For the Social Text analysis, we calculated the number of words within each post used by defining word boundaries through the use of space separator character. First we parsed the text to identify the pronouns used and then applied analysis to identify how positive or negative it was to the candidate. For this task, we created a sentiment analysis tool specific to our project. The sentiment analyzer was originally built for and trained on the Danish language user reviews collected from the Internet [see 48] and then applied to the Facebook. The analyzer classifies texts into three categories: Positive, Negative and Neutral. The posts were divided into monthly time slices, giving 13 slices for each of the three candidates across the period of September 2010 to September 2011. Each post in the time-period was labeled with a sentiment of positive, negative or neutral by the sentiment analysis tool. Then for each month, the ratio of Positive to Negative postings was automatically computed. For one candidate, Pia, there was little activity until November 2010, so we use 11 time slices for this candidate. There were a total of 25,987 unique posters in the data corpus. Of this total the vast majority (22740) posted on only one wall (Pia: 2157, Lars: 9315, and Helle: 11268). A much smaller proportion, around one in twenty individuals (1,439) posted at least once on two Facebook walls (Pia+Helle: 242, Helle+Lars: 1071, and Lars+Pia: 495). Finally, a small number of users posted at least once on all three Facebook walls. The largest number of wall-crossing users was 1071 for the Helle and Lars combination. 495 and 242 used crossed the partisan barrier to post on both Pia's Facebook wall and on one or both of the other politicians' walls of Lars and Helle respectively.

From a public perspective, the results suggest a "partisan sphere" [49] rather than inclusive public space of individuals engaging on multiple platforms and considering multiple-view points. Figure 5 presents the wall-crossing results. Such figures do not suggest a wide breadth of engagement and are not strongly supportive of the emergence of Facebook as a public sphere in the election. There seemed to be little cross-party debate occurring with individuals using the technology to engage with other points of view. Within this small sub-population of the Danish public, most people focused on one candidate and did not seek to engage with supporters of other candidates.

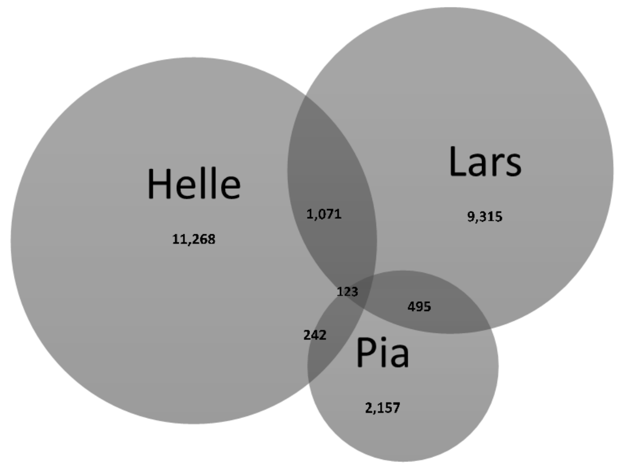

Figure 5: Overview of Facebook wall crossing

7. Case Study # 4: Prescriptive Analytics

Copenhagen Zoo experienced a social media crisis, which started on February 8, 2014, due to an impending euthanizing of a young giraffe, called Marius, and lasted until February 13, 2014 [50]. Major international media also participated in the case of Marius. *British BBC* and *The Guardian* newspaper has also referred to the euthanizing, *CNN* followed the case on both network and TV, and *The New York Times* had also written about Marius' death.

The analytical objective for conducting SSA was to identify the structural properties of social media crises with reference to the domain-specific theories of crisis communication and management discusses in the theoretical framework section. Specifically, we were interested in the three time-periods of before, during and after crisis. We conducted SSA across the three time-periods for (a) overall distribution of user actions (Figure 6), (b) distribution of likes by Facebook users on the artifacts (posts and comments) created by the company (Figure 7), and (c) distribution of comments by Facebook users on posts created by the Copenhagen Zoo (Figure 8).

As can be seen in Figure 6, a disproportionately high percentage of Facebook users only interacted with the Facebook wall of Copenhagen Zoo (86%), the proportion of users interacting during the crises is much higher compared to the total time-period. To put it differently, SSA of actors across the time-periods of before, during and after crises confirms not only the operational definition of a social media crises but also reveals the voluminous and transient nature of user attention. That is, there were many more actors interacting during the crises but they stop interacting after the crises had passed. How this change in user behaviour occurs could be a function of not only the type of social media crises it is but also the type of social media crisis communication and management strategies employed by the companies.

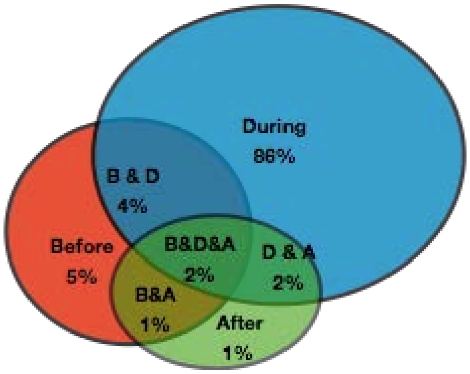

Figure 6. Social Set Analysis of Zoo Actors during crisis

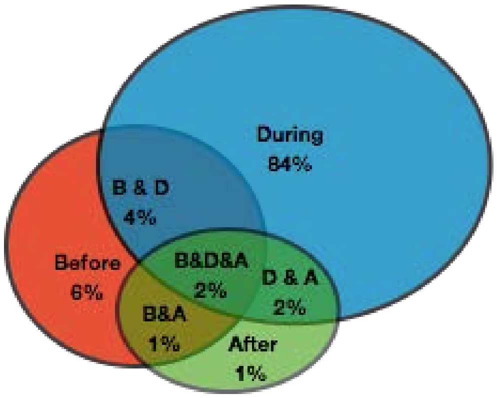

Figure 7. SSA of Actors who Liked Zoo wall Admin Posts

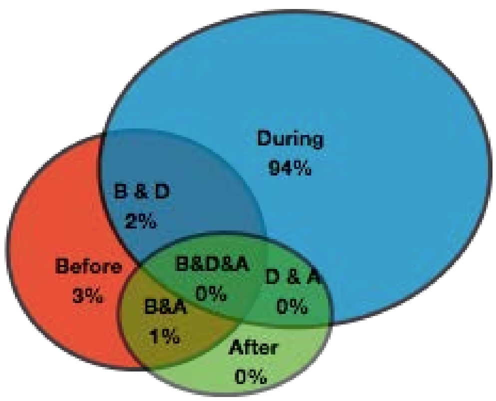

Figure 8. SSA of actors who commented on Zoo admin posts

Figure 7 shows the temporal distribution of Facebook users' likes to the artifacts (posts and comments) created by the company (Facebook wall administrator). Based on associational sociology and social influence theories in social psychology, we conceptualize the action of a 'Facebook like' as a positive association with the artifact (Facebook post or comment) and/or actor (Facebook user). This type of SSA reveals the positive endorsement of the company's communication actions before, during and after the crises. As can be seen from Figure 7, surprisingly high proportion of total likes were received during the crises for the Copenhagen Zoo (84%). This can be a structural indicator that the social media crisis might actually be a net positive for the companies concerned in terms of customer loyalty and brand parameters.

Figure 8 shows the temporal distribution of Facebook users' comments to the posts created by the company. We find that the proportion of comments before and during the crises for Copenhagen Zoo (3% and 94%) have highly skewed distribution of comments during the before and during periods of the social media crises. Since Facebook does not have a 'dislike' button, comments are the only method for users to express negative associations, sentiments and expressions (positive sentiments and expressions can also be expressed this way). Given the distribution of likes for Copenhagen Zoo's posts and comments, the SSA of comments reveals an interesting pattern of higher likes for the company's artifacts as well as higher number of comments.

SSA results suggest that the type crisis as well as crisis communication and management strategies employed might be different across the four cases. In order to uncover the substantive nature of the interactional patterns revealed by SSA, we conducted qualitative content analysis of the Big Social Data corpus using two methods: (a) netnographic analysis of the Facebook walls before, during and after the crises and (b) manual sentiment analysis and topic analysis of posts and comments during the crises. These analyses helped shed light on the nature of the crises and the crises communication and management strategies, employed by the companies.

8. Conclusion and Outlook

Computational social science research has reached a point where social media activity is ubiquitous, yet hard to collect and analyze in domain-specific ways (with the notable exception of epidemiology). In conjunction with complex event timelines as depicted by the Bangladesh garment factory disasters; tweets about iPhone; and user actions on the Copenhagen Zoo and H&M Facebook walls, Big Social Data presents numerous opportunities for attaining deep insights. As illustrated by the four case studies above, SSA covers the range of predictive, visual, descriptive and prescriptive analytics presents the means of reaching important insights. Taken together, the four demonstrative case studies offers preliminary evidence that SSA as a new, viable, holistic approach to computational social science in general and social data analytics in particular. Based on associational sociology and set theory, SSA can yield complementary insights to the dominant trifecta of relational sociology, graph theory and SNA.

REFERENCES

[1] Montalvo, R.E. 2011. Social Media Management. *International Journal of Management & Information Systems (IJMIS)* 15, 3, 91-96.
[2] Vollmer, C. and Precourt, G. 2008. *Always on: Advertising, marketing and media in an era of consumer control*. McGraw-Hill Professional.
[3] McAfee, A. 2009. *Enterprise 2.0: New collaborative tools for your organization's toughest challenges*. Harvard Business School Press.
[4] Vatrapu, R. 2013. *Understanding Social Business*. Springer, City, 2013 *Incomplete citation.
[5] Cleveland, W.S. 2001. Data science: an action plan for expanding the technical areas of the field of statistics. *International statistical review* 69, 1, 21-26.
[6] Loukides, M. 2012. *What is data science?* O'Reilly Media.
[7] Ohsumi, N. 2000. From data analysis to data science. *Studies in Classification, Data Analysis and Knowledge Organization*, 329-334.
[8] Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., and Van Alstyne, M. .2009. Computational Social Science. *Science* 323, 5915, 721-723.

[9] Sterne, J. 2010. *Social media metrics: How to measure and optimize your marketing investment*. John Wiley & Sons.
[10] Sponder, M. 2011. *Social Media Analytics: Effective Tools for Building, Intrepreting, and Using Metrics*. McGraw-Hill Professional.
[11] Tufekci, Z. 2014. Big Questions for Social Media Big Data: Representativeness, Validity and Other Methodological Pitfalls. In *Proceedings of the 8th International Aaai Conference on Weblogs and Social Media.* ICWSM. (Ann Arbor, USA, June 2-4, 2014) .
[12] Gross, J. L. and Yellen, J. 2005. *Graph theory and its applications*. CRC press.
[13] Borgatti, S. P., Mehra, A., Brass, D. J. and Labianca, G. 2009. Network analysis in the social sciences. *Science*, 323, 5916, 892-895.
[14] Emirbayer, M. 1997. Manifesto for a relational sociology. *The American Journal of Sociology* 103, 2, 281-317.
[15] Mukkamala, R., Hussain, A. and Vatrapu, R. Towards a Formal Model of Social Data. n.d. *IT University Technical Report Series,*TR-2013-1692013), https://pure.itu.dk/ws/files/54477234/ITU_TR_54472013_54477169.pdf.
[16] Vatrapu, R., Mukkamala, R., and Hussain, A. 2014. *Towards a Set Theoretical Approach to Big Social Data Analytics: Concepts, Methods, Tools, and Empirical Findings*.
[17] Mizruchi, M. S. 1994. Social network analysis: Recent achievements and current controversies. *Acta sociologica* 37, 4, 329-343.
[18] Vatrapu, R., Mukkamala, R., and Hussain, A. 2014. *A Set Theoretical Approach to Big Social Data Analytics: Concepts, Methods, Tools, and Findings*.
[19] Lovett, J. 2011. *Social Media Metrics Secrets*. Wiley.
[20] Latour, B. 2005. *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford University Press.
[21] Ragin, C. C. 2000. *Fuzzy-set social science*. University of Chicago Press.
[22] Mukkamala, R., Hussain, A., and Vatrapu, R. 2014. Towards a Set Theoretical Approach to Big Data Analytics. In *Proceedings of the 3rd IEEE International Congress on Big Data 2014, (*Anchorage, United States).
[23] Smithson, M. and Verkuilen, J. 2006. *Fuzzy set theory: applications in the social sciences*. Sage.
[24] Kechris, A. S. and Kechris, A. S. 1995. *Classical descriptive set theory*. Springer-Verlag New York.
[25] Vatrapu, R. 2007. *Technological Intersubjectivity and Appropriation of Affordances in Computer Supported Collaboration*. Doctoral Dissertation, University of Hawaii at Manoa, Honolulu: Available at http://www.itu.dk/people/rkva/docs/Vatrapu-Dissertation.pdf.
[26] Vatrapu, R. 2009. *Toward a Theory of Socio-Technical Interactions in Technology Enhanced Learning Environments*. Springer-Verlag.
[27] Vatrapu, R. 2010. Explaining culture: an outline of a theory of socio-technical interactions. In *Proceedings of the 3rd ACM International Conference on Intercultural Collaboration.* ICIC 2010*,* 111-120.
[28] Robertson, S. and Vatrapu, R. *Digital Government*. 2010.
[29] Garfinkel, H. 1967. *Studies in Ethnomethodology*. Prentice-Hall, Englewood Cliffs, NJ.
[30] Kryszkiewicz, M. 1998. Rough set approach to incomplete information systems. *Information sciences* 112, 1, 39-49.
[31] Cialdini, R. B. and Goldstein, N. J. 2004. Social influence: Compliance and conformity. *Annu. Rev. Psychol.* 55, 591-621.
[32] Haugtvedt, C. P. and Herr, P. M. 2008. *Handbook of consumer psychology*. Lawrence Erlbaum Associates, New York, NY.
[33] Keim, D., Andrienko, G., Fekete, J.D., Görg, C., Kohlhammer, J., and Melançon, G. 2008. *Visual analytics: Definition, process, and challenges*. Springer.
[34] Lassen, N., Madsen, R. and Vatrapu, R. 2014. Predicting iPhone Sales from iPhone Tweets. In *Proceedings of IEEE 18th International Enterprise Distributed Object Computing Conference* (Ulm, Germany, 2014). EDOC 2014 , 81-90, ISBN: 1541-7719/1514, DOI: 1510.1109/EDOC.2014.1520.
[35] Shmueli, G. and Koppius, O. R. 2011. Predictive analytics in information systems research. *Mis Quarterly* 35, 3, 553-572.
[36] Asur, S. and Huberman, B. A. 2010. *Predicting the future with social media*. IEEE.
[37] Belch, G. E., Belch, M. A., Kerr, G. F., and Powell, I. 2008. *Advertising and promotion: An integrated marketing communications perspective*. Mcgraw-Hill.
[38] Sato, H. 2014. Cournot Competition and Reduction of Corruption to Prevent Garment Factory Fires in Bangladesh. *Advances in Management and Applied Economics* 4, 4, 17-20.
[39] Khanna, P. 2011. Making labour voices heard during an industrial crisis: workers' struggles in the Bangladesh garment industry. *TRAVAIL, capital et société* 44, 2.
[40] Himi, S. A. and Rahman, A. 2013. Workers Unrest in Garment Industries in Bangladesh: An Exploratory Study. *Journal of Organization and Human Behaviour* 2, 3, 49-55.
[41] Rahman, S. 2013. *Broken Promises of Globalization: The Case of the Bangladesh Garment Industry*. Lexington Books.
[42] Stewart, K. L. 2013. An ethical analysis of the high cost of low-priced clothing. *proceedings of Northeastern Association of Business, Economics, and Technology*. NABET. 128.
[43] Islam, M. A., Deegan, C., and Gray, R. Social Audits and Multinational Company Supply Chain: A Study of Rituals of Social Audits in the Bangladesh Garment Industry. *Available at SSRN 2466129*2014).
[44] Hussain, A. and Vatrapu, R. 2014. *Social Data Analytics Tool (SODATO)*. Springer International Publishing.
[45] Robertson, S. 2011. *Changes in referents and emotions over time in election-related social networking dialog*. IEEE.
[46] Robertson, S., Vatrapu, R., and Medina, R. 2010. Online Video "Friends" Social Networking: Overlapping Online Public Spheres in the 2008 U.S. Presidential Election. *Journal of Information Technology & Politics* 7, 2-3, 182-201.
[47] Robertson, S., Vatrapu, R. and Medina, R. 2010. Off the wall political discourse: Facebook use in the 2008 U.S. presidential election. *Information Polity* 15, 1-2, 11-31.
[48] Hardt, D. and Wulff, J. 2012. What is the Meaning of 5*'s? An Investigation of the Expression and Rating of Sentiment. *Empirical Methods on Natural Language Processing.* In *Proceedings of the Conference on Natural Language Processing*, 319-326.
[49] Vatrapu, R., Robertson, S., and Dissanayake, W. 2008. Are Political Weblogs Public Spheres or Partisan Spheres? *International Reports on Socio-Informatics* 5, 1, 7-26.
[50] Zimmerman, C., Chen, Y., Hardt, D., and Vatrapu, R. 2014. Marius, the giraffe: a comparative informatics case study of linguistic features of the social media discourse. In *Proceedings of the Proceedings of the 5th ACM international conference on Collaboration across boundaries: culture, distance & technology* (Kyoto, Japan, 2014). ACM.