IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON THEORETICAL FOUNDATIONS FOR BIG DATA APPLICATIONS: CHALLENGES AND OPPORTUNITIES

Received March 2, 2016, accepted March 29, 2010, date of publication April 28, 2015, date of current version June 13, 2016.

Digital Object Identifier 10.1109/ACCESS.2016.2559584

Social Set Analysis: A Set Theoretical Approach to Big Data Analytics

RAVI VATRAPU[1,2], RAGHAVA RAO MUKKAMALA[1], ABID HUSSAIN[1], AND BENJAMIN FLESCH[1]
[1]Computational Social Science Laboratory, Copenhagen Business School, Frederiksberg 2000, Denmark
[2]Westerdals Oslo School of Arts, Communication and Technology, Oslo 0178, Norway

Corresponding author: R. R. Mukkamala (rrm.itm@cbs.dk)

The authors were supported by the project Big Social Data Analytics: Branding Algorithms, Predictive Models, and Dashboards funded by Industriens Fond (The Danish Industry Foundation).

ABSTRACT Current analytical approaches in computational social science can be characterized by four dominant paradigms: text analysis (information extraction and classification), social network analysis (graph theory), social complexity analysis (complex systems science), and social simulations (cellular automata and agent-based modeling). However, when it comes to organizational and societal units of analysis, there exists no approach to conceptualize, model, analyze, explain, and predict social media interactions as individuals' associations with ideas, values, identities, and so on. To address this limitation, based on the sociology of associations and the mathematics of set theory, this paper presents a new approach to big data analytics called social set analysis. Social set analysis consists of a generative framework for the philosophies of computational social science, theory of social data, conceptual and formal models of social data, and an analytical framework for combining big social data sets with organizational and societal data sets. Three empirical studies of big social data are presented to illustrate and demonstrate social set analysis in terms of fuzzy set-theoretical sentiment analysis, crisp set-theoretical interaction analysis, and event-studies-oriented set-theoretical visualizations. Implications for big data analytics, current limitations of the set-theoretical approach, and future directions are outlined.

INDEX TERMS Big social data, formal models, social set analysis, big data visual analytics, new computational models for big social data.

I. 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 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 regarding how Big Social Data integrates with the existing datasets of an organization (that is, data from internal enterprise systems) and its relevance to the organisation's key performance indicators. To address these diverse but interrelated issues, this paper presents a novel set-theoretical approach to Big Data Analytics in general and Big Social Data Analytics in particular for Facebook, Twitter and other social media channels.

Specifically, this paper introduces a research program situated in the domains of Data Science [5]–[7] and Computational Social Science [8] with practical applications to Social Media Analytics in organizations [4], [9], [10]. It addresses some of the important theoretical and methodological

limitations in the emerging paradigm of Big Data Analytics of social media data [11]. From an academic research standpoint, Social Set Analysis addresses two major limitations with the current state of the art in Computational Social Science: (i) a vast majority of the extant literature is on twitter datasets with only 5% of the papers analysing Facebook data raising representativeness, validity and methodological concerns [11], and (ii) mathematical modelling of social data hasn't progressed beyond the four dominant approaches [12] of text analysis (information extraction and classification), social network analysis (graph theory), social complexity analysis (complex systems science), social simulations (cellular automata and agent-based modelling).

To put it honestly and provocatively, currently we don't have deep academic knowledge of the most dominant action on social media platforms performed by hundreds of millions of unique users every day: ''like'' on Facebook. In fact, as Claudio Cioffi-Revilla (2013), one of the founding parents of the field of Computational Social Science, astutely observed:

> *Reliance on the same mathematical structure every time (e.g., game theory, as an example), for every research problem, is unfortunately a somewhat common methodological pathology that leads to theoretical decline and a sort of inbreeding visible in some areas of social science research. Dimensional empirical features of social phenomena-such as discreteness-continuity, deterministic-stochastic, fnite-infnite, contiguous-isolated, local-global, long-term vs. short-term, independence-interdependence, synchronic-diachronic, among others-should determine the choice of mathematical structure(s).*

This lack of mathematical imagination coupled with hyperactive boundary-policing of the ''purity of the turf'' of Computational Social Science results in major conceptual and technical limitations when analysing big social data resulting from individuals' and organizations' Facebook and Twitter engagement. There is both a research gap and real-world organisational needs to describe, model, analyse, explain, and predict such interactions as individuals' associations to ideas, values, identities etc [13].

For example, a typical post on F. C. Barcelona's Facebook page generates around 100,000 unique likes, 5,000 comments and 1,000 shares). Facebook users' ''likes'' on any given F. C. Barcelona post could be personal-association to one of the players, identity-association to the Catalan, political-association to pro-independence parties of Catalonia, brand-association to the corporate sponsors etc. The mathematics of set theory is ideally suited to model such associations in the first analysis. Just like graph theory is ideally suited for Social Network Analysis [14] of dyadic relations from the perspective of relational sociology [15], set theory is ideally suited for conceptualising, modelling, and analysing monadic, dyadic, and polyadic human associations to ideas, values and identities [16] from the perspective of the sociology of associations. This is the gist of the set theoretical approach proposed by this paper.

A. OVERARCHING RESEARCH QUESTION

In order to further research in this area we as ourselves the following research question:

How can models, methods and tools for Social Set Analysis 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?

II. CONCEPTUAL FRAMEWORK

A. NEED FOR A PHILOSOPHY OF COMPUTATIONAL SOCIAL SCIENCE

The purpose of this section is to present an argument that we need philosophies of Computational Social Science that explicilty outline and discuss their sociological assumptions, mathematical modelling, computational implementation, and empirical analysis. To the best of our knowledge, no such philosophy of Computational Social Science exists other than Social Network Analysis [17] based on the mathematics of graph theory [18] and the sociology of relations [15]. However, the philosophical assumptions of relational sociology might be not be relevant to all classes of problems in computational social science. For example, for the class of problems that address big social data from the Facebook or Twitter interactions of large brands such as Coca-Cola or a F. C. Barcelona, 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 [19] is neither necessary nor sufficient [13]. Other dominant paradigms of computational social science such as Social Complexity and Social Simulation [12] have varying levels of philosophical and modelling unity and maturity. [12]. Therefore, there is a clear need for a manifest statement and critical examination of philosophical principles that underpin the theoretical, methodological, and analytical aspects of current Computational Social Science approaches.

However, philosophical proposals for Big Data Analytics must avoid the malaise of *over-philosophising* with non-realist ontologies and non-empirical epistemologies (for a precautionary tale from the Humanities and Social Sciences, please cf. [21], [22]) that result in little-to-no methodological innovation in terms of instrumentation, measurement and evaluation of the phenomena of interest. Philosophical framworks for Big Data Analytics should aspire towards positive contributions that go beyond the negative criticisms of assumptions and methods that regularly feature in prominent recent criticisms (for instance, [11] and [23]). We argue that one class of positive contributions would be generative frameworks that provide explicit articulation of philosophical assumptions underlying analytical approaches as well as a *production system* for creating and evaluating

TABLE 1. Five elements of the candidate generative framework for philosophy of computational social science.

Philosophical Dimension	GF-PCSS Element	Key Assumptions
Ontology	Basic Premise	What is social?
		When is it social?
		Being vs. Becoming of social
Epistemology	Social Action	How is it social?
		How does a social entity act and interact?
Methodological	Unit of Analysis	What is the foundational analytical unit?
		What is the minimum viable analytical entity?
Political	Social Structure	What is the social grouping entity?
		What is the social formation unit?
Formal	Mathematics	What is the appropriate mathematical theory for modelling?

new philosophies. To address the analytical limitations identified and to fulfill the critical and generative criteria outlined above, we propose a first version of the generative framework for the philosophy of Computational Social Science.

1) A GENERATIVE FRAMEWORK FOR PHILOSOPHY OF COMPUTATIONAL SOCIAL SCIENCE (GF-PCSS)

The preliminary version of the GF-PCSS comprising of five elements is presented in Table 1 below.

Given the preliminary stage of the GF-PCSS, no claims are made about the exhaustiveness and/or mutual exclusivity of the five elements. We simply claim that the five elements are necessary with no claims made about their sufficiency and orthogonality.

Table 2 below seeks to illustrate the positive contribution of the GF-PCSS. First, the framework is used to explicitly state the latent philosophical assumptions of one dominant traditional approach in Computational Social Science, Social Network Analysis. Second, the framework is used to better understand the limitations of Social Network Analysis with respect to large-scale social media platforms that are increasingly content driven. Social Network Analysis is primarily concerned with how social actors relate to each other and not so much with how content is generated, interacted and circulated in terms of ideas, aspirations, values, and identities. However, 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], [24] has been based on an implicit, inherent and latent understanding of human associations as expressed by metrics and key performance indicators such as brand sentiment, brand associations, conversation keywords, reach etc. Further, Social Network Analysis assumes *homophiliy* rather than explaining

TABLE 2. Contrasting philosophies of computational social science.

	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	Molecular Relations	Atomic Actions
Unit of Analysis	Dyadic	Monadic, Dyadic & Polyadic
Social Configuration	Networks	Sets
Social Explanation	Structural	Agentic
Mathematics	Graph Theory	Set Theory

the agentic mechanisms constituting it. Third and last, GF-PCSS is used to generate a new holistic approach termed Social Set Analysis and make a positive contribution. Social Set Analysis is based on the philosophical principles derived from ecological psychology, micro sociology, associational sociology [25], and the mathematics of the set theory (crisp sets, fuzzy sets, rough sets, and random sets) [26].

To be clear, our argument is not that current approaches in Computational Social Science such as Social Network Analysis (based on relational sociology, graph theory, and network analysis) are invalid or ineffective. Instead, our argument, as articulated and illustrated in Tables 1 & 2, is that a generative framework of the philosophy can be used to make a fundamental change in the foundational mathematical logic of the formal model from graphs to sets which can yield new analytical insights for a new class of problems (in our case, organizational use of social media).

B. SET THEORETICAL BIG SOCIAL DATA ANALYTICS

As articulated in [27], based on Smithson and Verkuilen [28] there are five advantages to applying classical set theory [29] in general and fuzzy set theory [26] 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.).

We now present a theory of social data based on the philosophical framework for Social Set Analysis discussed above.

C. THEORY OF SOCIAL DATA

For the purposes of systematically collecting and analysing big social data, we argue that any candidate theory of social data must support conceptual and mathematical modelling of data at the software log level. After all, it is a fact that the outcomes from big social data collection from modern web service calls or historic web crawling methods are nothing more than digital trace records and software log entries. As such, an appropriate theory of social data would be operational at the micro-genetic level of social media interactions as they unfold in the real-time and in the actual-space of a computer screen of some kind (desktop monitor, laptop display or the mobile phone screen). For Social Set Analysis, we have selected the theory of socio-technical interactions by Vatrapu [30]–[32] as it conceptualises perception of and interaction on the screen in real-time and actual-space. The theory of socio-technical interations [30]–[32] is derived from the following sources:

1) the ecological approach to perception and action [33]
2) the enactive approach to the philosophy of mind [34]
3) the phenomenological approach to sociology [35], [36]

A more detailed exposition of the theory of socio-technical interactions regarding its ontological and epistemological assumptions and principles, is beyond the scope of this paper but for a consise overview, please confer [32].

We use the theory of socio-technical interactions [30]–[32] to describe how individual data items (or trace records) such as Facebook posts, likes, comments etc. come into being. In other words, we use the theory of socio-technical interactions to describe the phenomenon of big social data generation from its constituent individual interactions of Facebook posts, comments, likes etc. That said, the scope and extent of the theory of social data are restricted to providing phenomenological grounding for modelling of the social data retrieved from social media platforms such as Facebook. The theory of social data is outlined the next subsection (II-D).

As already mentioned, the theory of social data is drawn from the theory of socio-technical interactions [30]–[32]. 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 and there are two types of socio-technical interactions:

1) Interacting with the technology: An example could be using the Facebook app on the user's smartphone.
2) Interacting with others socially using the technology: An example could be liking a picture posted by a friend in the Facebook app on the user's smartphone.

These socio-technical interactions are theoretically conceived as

1) Perception and appropriation of socio-technical affordances
2) 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) [30]–[32] refers to a technology supported, interactional social relationship between two or more actors.

Socio-technical interactions as described above result in electronic trace data that is termed *social data*. In case of the previously mentioned example where a Facebook user liking a picture posted by a friend 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 which is the analytical focus of this paper. Our argument is not that there exists only one set of candidates for the theory of social data, conceptual model of social data and the formal model of social data as proposed in this paper. Instead, our argument is that a theoretically informed and empirically oriented research project in big social data analytics must incoporate these components (theory, conceptual and formal models of social data) and computationally realise each of them within IT-Artefacts.

D. CONCEPTUAL MODEL OF SOCIAL DATA

Based on the theory of social data described above, we present the conceptual model of social data below.

In general, *Social data* consists of two types: (a) Interactions (what is being done) and Conversations (what is being said). Interactions refer to the first aspect of socio-technical interactions constituted by the perception and appropriation of affordances (which users/actors perceive which socio-techncial affordances to interact with what other social

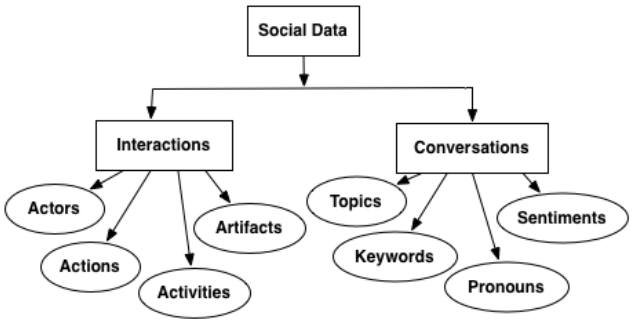

FIGURE 1. Conceptual Model of Social Data [37].

actors on social media platforms). Conversations relates to the second aspect of socio-technical interactions: structures and functions of technological intersubjectivity (what the users/actors are communicating to each other and how they are influencing each other through both natural language as well as design language of the social media plaforms). Interactions 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. Conversations consists of the communicative and linguistic aspects of the social media interaction such as the topics discussed, keywords mentioned, pronouns used and emotions expressed. Figure 1 presents the conceptual model of social data.

E. ILLUSTRATIVE EXAMPLE OF SOCIAL DATA

Let us say that the research domain is Corporate Social Responsbility (CSR) and the research question is to what extent do Facebook walls function as online public spheres with regard to CSR in terms of marketing campaigns as well as crises. 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.

Focusing on the interactional aspect of big social data allows the examination of the breadth of engagement 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' Interactional analysis of big social data can help reveal the patterns and dynamics of actors' mobility across space (multiple facebook walls) and time (before, during and after campaigns/crises).

Focusing on the conversational aspect of big social data allows the examination of 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 four key aspects of the posts and comments: topics, keywords and emotions. As with interactional analysis, conversational analysis of big social data can help reveal the patterns and dynamics of actors' conversational genres across space (multiple facebook walls) and time (before, during and after campaigns/crises).

III. RESEARCH METHODOLOGY

Our research methodology is shown in Fig. 2 and is described below:

1) Systematically collect big social data about organisations from Facebook, Twitter etc using the Social Data Analytics Tool [37], [38] developed in the Computational Social Science Laboratory (http://cssl.cbs.dk) and other research and commercial tools.
2) Technically combine organisational process data with business social data so that the resulting dataset legally compliant, ethically correct, privacy adherent, and data security ensured
3) Big Social Data Analytics: Phase One: Adopt current methods, techniques and tools from Computational Social Science to model and analyse
   a) Interaction Analysis: Social Network Analysis, Complex Systems Dynamics, Event Study Methodology from Finance, Data Mining from Computer Science
      - Who is doing what, when, where, how and with whom?
      - Social media users and organisational stakeholders (like consumers) liking pictures of cute puppies posted by Walmart on its official Facebook wall every third Sunday according to its social media marketing strategy.
   b) Conversation Analysis: Computational Linguistics & Machine Learning
      - What are the things human actors (and fraudulent accounts/robots) saying?
      - Social media users and organisational stakeholders (like consumers) commenting on those pictures of cute puppies by discussing/mentioning various topics/keywords of organisational/societal relevance/irrelevance and expressing their subjective feelings etc.
4) Applying set theoretical methods and techniques drawn from crisp sets, fuzzy sets, rough sets and random sets [26], [28], [29], [39].
5) Software realisation of the empirical findings from traditional and novel (set theoretical) approaches to Computational Social Science as a tools for Organisations.
6) Publication of research findings in peer-reviewed conferences, journals and edited books.
7) Generation of instrumental benefits for Organisations in terms of meaningful facts (sensible data), actionable insights (applicable information), valuable outcomes (constructive knowledge) and sustainable impacts (wisdom)

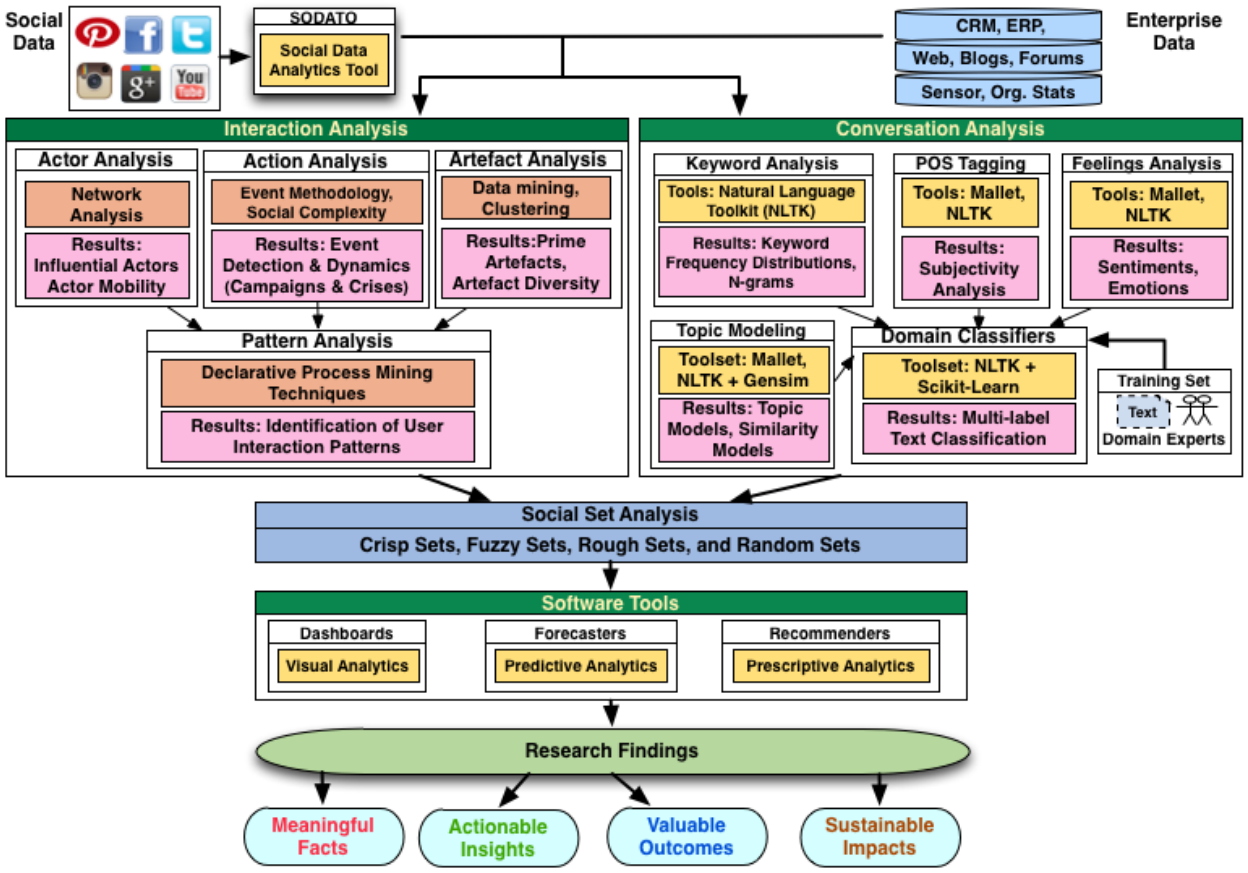

FIGURE 2. Research framework for set-theoretical big social data analytics.

IV. RELATED WORK

A. SOCIAL NETWORK ANALYSIS

Social Network Analysis can be traced back to 1979, where Tichy et.al. [40] used it as a method to examine the relationships and organisational social structures. In the later years, cognitive social structures as a solution for social network related problems was proposed by Krackhardt [41]. The field of social computing attracted many researchers due to the latest developments of online social media since last decade. Even though It is not possible to refer to an extensive list of research articles in this emerging area, however we refer some of the latest and important works here.

In their research article, Zhan and Fang in [42] provided an detailed overview about research on social networking analysis, human behavioural modelling and security aspects in the context of social networks. Social network analysis based on measuring social relations using multiple data sets has been explored in [43]. In the context of multi-agent systems using social network analysis, a framework for calculating reputations has been proposed by [44]. An algorithm to find overlapping communities via social network analysis was explored in [45]. Moreover, analysis of subgraphs in the social network based on the characteristic features: leadership, bonding, and diversity was studied by the authors in [46]. All these works focussed on using social network analysis and other graph related formalisms as main tools for analysis of social media where the primary focus is on the structural aspects of social data. On the other hand, our work primarily focussed on using set theory and fuzzy logics for analysis of both structure and content of social media data. Therefore we are not only interested in analyzing the structural aspects of social data (as networks or sets) but also in understanding the substantive aspects of social data (as sentiments, topics, keywords, pronouns).

B. SOCIAL TEXT ANALYSIS

A comprehensive state-of-the-art review of computational linguistics is provided by Pang and Lee [47]. They provided approaches to analyse natural language texts, and identify three different technical terms: opinions, sentiments, and subjectivity. In this paper, we adopt Pang and Lee's [47] technical interpretation that opinion mining and sentiment analysis can be treated as identical and conduct sentence level rather than sub-sentence level sentiment analysis as discussed in [48]. Other methods and techniques for sentiment analysis are presented and discussed in [47]–[52]. Below is a selected

listing of related work in sentiment analysis of social data ranging over a variety of methods, techniques, and tools.

Prior work has shown sentiment analysis of social data can be used to predict movie revenues [53], correlate with contemporaneous and subsequent stock returns [54], exploring cultural and linguistic differences in ratings and reviews [55], sentiment evolution in political deliberation on social media channels [56], assess sentiment towards a new vaccine [57], and explore semantic-level precedence relationships between participants in a blog network [58]. To briefly expand, [58] proposed a methodology for the detection of bursts of activity at the semantic level using linguistic tagging, term filtering and term merging, where a probabilistic approach was used to estimate temporal relationships between the blogs. Asur and Huberman [53] showed that sentiment analysis on Twitter's content urls, retweets and their hourly rates can predict box-office movies revenues to a high degree of precision.

In contrast to the existing approaches, we used Set and Fuzzy Set Theory for the formal modelling of associations between actors, actions, artifacts, topics and sentiments in order to provide a systemic treatment of relationship, vagueness and uncertainty in the social data. The existing sentiment analysis techniques (as cited above) use only the classification of individual artifacts (such as either positive or negative or neutral), but not the probabilities associated with the classification labels returned by the sentiment analysis method and/or tool. In contract, our approach uses fuzzy sets to represent artifact sentiment with classification along with their probabilities (e.g. positive: 0.20, negative: 0.65, neutral: 0.15) as explained later.

V. FORMAL MODEL OF THE CONCEPTUAL SOCIAL DATA

In this section, we will provide formal semantics for the concepts of social data, which is based on social data model that was initially presented in [27] and [59], but refined according to the changes in the conceptual model of social data presented in Sec. II-C.

Notation: For a set A we write $\mathcal{P}(A)$ for the power set of A (i.e. set of all subsets of A) and $\mathcal{P}_{disj}(A)$ for the set of mutually disjoint subsets of A. The cardinality or number of elements in a set A is represented as $|A|$. Furthermore, we write a relation R from set A to set B as $R \subseteq A \times B$. A function f defined from a set A to set B is written as $f : A \to B$, where a if f is a partial function then it is written as $f : A \rightharpoonup B$.

First, we define type of artifacts in a socio-technical system as shown in Def. 1.

Definition 1: We define $\mathbb{R}$ as a set of all artifact types as $\mathbb{R} = \{\text{status, comment, link, photo, video}\}$.

Definition 2: We define $\mathbb{A}_{\mathbb{CT}}$ as a set of actions that can be performed as $\mathbb{A}_{\mathbb{CT}} = \{post, comment, share, like, tagging\}$.

As explained in the conceptual model (Sec. II-C), the *Social Data* model contains *Interactions* (what is being done) and *Conversations* (what is being said). The formally *Social Data* is defined in Def. 3 as follows,

Definition 3: Formally, Social Data is defined as a tuple $\mathsf{D} = (\mathsf{I}, \mathsf{C})$ where

i) I is the *Interactions* representing the structural aspects of social data as defined further in Def. 4
ii) C is the *Conversations* representing the content of social data and is further defined in Def. 5

Definition 4: The Interactions of *Social Data* are defined as a tuple $\mathsf{I} = (\mathsf{U}, \mathsf{R}, \text{Ac}, \mathsf{r}_{\text{type}}, \rhd, \rightarrow_{post}, \rightarrow_{share}, \rightarrow_{like}, \rightarrow_{tag}, \rightarrow_{\text{act}})$ where

i) U is a (finite) set of actors (or users) ranged over by u,
ii) R is a (finite) set of artifacts (or resources) ranged over by r,
iii) Ac is the activities set which is also finite,
iv) $\mathsf{r}_{\text{type}} : \mathsf{R} \to \mathbb{R}$ is typing function for artifacts that maps each artifact to an artifact type defined in 1,
v) $\rhd : \mathsf{R} \rightharpoonup \mathsf{R}$ is a partial function mapping artifacts to their parent artifact,
vi) $\rightarrow_{post} : \mathsf{U} \rightharpoonup \mathcal{P}_{disj}(\mathsf{R})$ is a partial function mapping actors to mutually disjoint subsets of artifacts created by them
vii) $\rightarrow_{share} \subseteq \mathsf{U} \times \mathsf{R}$ is a relation mapping between users to their artifacts (shared by them),
viii) $\rightarrow_{like} \subseteq \mathsf{U} \times \mathsf{R}$ is a relation mapping users to the artifacts liked by them,
ix) $\rightarrow_{tag} \subseteq \mathsf{U} \times \mathsf{R} \times (\mathcal{P}(\mathsf{U} \cup \text{Ke}))$ is a tagging relation mapping artifacts to power sets of actors and keywords indicating tagging of actors and keywords in the artifacts, where Ke is set of keywords defined in Def. 5,
x) $\rightarrow_{\text{act}} \subseteq \mathsf{R} \times \text{Ac}$ is a relation from artifacts to activities.

Formal definition of Interactions is provided in Def. 4, where the first three items (i, ii, x of Def. 4) contain a set of actors (U), a set of artifacts/resources (R) and a set of activities (Ac). Each artifact is mapped to an artifact type (such as status, photo etc) by artifact type function (Def. 4-iv). Furthermore, some of the artifacts are mapped to their parent artifact (if exists) by parent artifact function $\rhd$ (Def. 4-v). For example, a comment is an artifact which is made on a post, then it is mapped to its parent (which is the post), on the other hand, if the artifact is a status message or a new post, then there will not be any mapping for that artifact, as it has no parent.

Furthermore, each artifact is posted (created) by a single actor. As shown in Def. 4-vi, the $\rightarrow_{post}$ is a partial function mapping actors to mutually disjoint sub sets of artifacts, each set containing artifacts created or posted by an actor. On contrary, the $\rightarrow_{share}$ indicates a many-to-many relationship, indicating that an artifact can be shared by many actors and similarly each actor can share many artifacts (Def. 4-vii). Even though *share* and *post* actions seems to be similar, the $\rightarrow_{post}$ signifies the creator relationship of an artifact, where as $\rightarrow_{share}$ indicates share relationship between an artifact and an actor which can be many-to-many.

Similar to the *share* relation, the *like* relation ($\rightarrow_{like}$) maps between the artifacts and actors, indicating the artifacts liked by the actors. The *tagging* relation ($\rightarrow_{tag}$) is a bit different,

which is a mapping between actors, artifacts and power set of actors and keywords (Def. 4-ix). The basic intuition behind the tag relation is that, it allows an actor to tag other actors or keywords in an artifact. Finally, the $\rightarrow_{\mathsf{act}}$ relation indicates a mapping between artifacts to activities (Def. 4-x).

Definition 5: In Social Data $\mathsf{D} = (\mathsf{I}, \mathsf{C})$, we define Conversations as $\mathsf{C}= (\mathsf{To}, \mathsf{Ke}, \mathsf{Pr}, \mathsf{Se}, \rightarrow_{\mathsf{topic}}, \rightarrow_{\mathsf{key}}, \rightarrow_{\mathsf{pro}}, \rightarrow_{\mathsf{sen}})$ where

(i) To, Ke, Pr, Se are finite sets of topics, keywords, pronouns and sentiments respectively,
(ii) $\rightarrow_{\mathsf{topic}} \subseteq \mathsf{R} \times \mathsf{To}$ is a relation defining mapping between artifacts and topics,
(iii) $\rightarrow_{\mathsf{key}} \subseteq \mathsf{R} \times \mathsf{Ke}$ is a relation mapping artifacts to keywords,
(iv) $\rightarrow_{\mathsf{pro}} \subseteq \mathsf{R} \times \mathsf{Pr}$ is a relation mapping artifacts to pronouns,
(v) $\rightarrow_{\mathsf{sen}} \subseteq \mathsf{R} \times \mathsf{Se}$ is a realtion mapping artifacts to sentiments.

The *Conversations* of Social Data is formally defined in Def. 5 and it mainly contains sets of *topics* (To), *keywords* (Ke), *pronouns* (Pr), and *sentiments* (Se) as defined in Def. 5. The $\rightarrow_{\mathsf{topic}}$, $\rightarrow_{\mathsf{key}}$, $\rightarrow_{\mathsf{pro}}$ and $\rightarrow_{\mathsf{sen}}$ relations map the artifacts to the *topics* (To), *keywords* (Ke), *pronouns* (Pr), and *sentiments* (Se) respectively. One may note that all these relations allow many-to-many mappings, for example an artifact can be mapped to more than one sentiment and similarly a sentiment can contain mappings to many artifacts.

Finally, we define a time function to record the timestamp of actions performed on social data as follows.

Definition 6: In Social Data, let $\mathbf{T} : (u, r, ac) \mapsto \mathbb{N}$ be time function that keeps tracks of timestamp ($t \in \mathbb{N}$) of an action ($ac \in \mathbb{A}_{\mathbb{CT}}$) performed by an actor ($u \in \mathsf{U}$) on an artifact ($r \in \mathsf{R}$).

A. OPERATIONAL SEMANTICS

Operational semantics of Social Data model are defined in this section. More precisely, we define how actors perform actions on artifacts. As formally defined in Def. 7, the first action is *post*, which accepts a pair containing an actor and a new artifact (u, r). First, the actor will be added to the set of actors (i) and then the new artifact will be added to the set of artifacts (ii). Finally the post relation ($\rightarrow_{post}$) will be updated for the new mapping (iii).

Definition 7: In Social Data $\mathsf{D} = (\mathsf{I}, \mathsf{C})$ with Interactions $\mathsf{I} = (\mathsf{U}, \mathsf{R}, \mathsf{Ac}, \mathsf{r_{type}}, \triangleright, \rightarrow_{post}, \rightarrow_{share}, \rightarrow_{like}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$, we define a **post** operation of posting a new artifact r ($r \notin \mathsf{R}$) by an user u as $\mathsf{D} \bigoplus_p (u, r) = (\mathsf{I}', \mathsf{C})$ where $\mathsf{I}' = (\mathsf{U}', \mathsf{R}', \mathsf{Ac}, \mathsf{r_{type}}, \triangleright, \rightarrow_{post}{}', \rightarrow_{share}, \rightarrow_{like}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$,

i) $\mathsf{U}' = \mathsf{U} \cup \{u\}$
ii) $\mathsf{R}' = \mathsf{R} \cup \{r\}$
iii) $\rightarrow_{post}{}' = \begin{cases} \rightarrow_{post}(u) \cup \{r\} & \text{if } \rightarrow_{post}(u) \text{ defined} \\ \rightarrow_{post} \cup \{u, \{r\}\} & \text{otherwise} \end{cases}$

The *comment* action (e.g. on a post) accepts a tuple containing an actor, the parent artifact (on which the comment is made) and the comment content itself as shown in the Def. 8. As it creates a new artifact, it will first apply a *post* action to create the comment as a new artifact with the actor (i) and then followed by an update to the parent artifact function ($\triangleright$) by adding the respective mapping to its parent (ii).

Definition 8: In Social Data $\mathsf{D} = (\mathsf{I}, \mathsf{C})$ with Interactions $\mathsf{I} = (\mathsf{U}, \mathsf{R}, \mathsf{Ac}, \mathsf{r_{type}}, \triangleright, \rightarrow_{post}, \rightarrow_{share}, \rightarrow_{like}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$, the **comment** operation on an artifact r_p ($r_p \in \mathsf{R}$) by an user u for a new artifact r is formally defined as $\mathsf{D} \bigoplus_c (u, r, r_p) = (\mathsf{I}', \mathsf{C})$ where $\mathsf{I}' = (\mathsf{U}', \mathsf{R}', \mathsf{Ac}, \mathsf{r_{type}}, \triangleright', \rightarrow_{post}{}', \rightarrow_{share}, \rightarrow_{like}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$,

i) $\mathsf{D} \bigoplus_p (u, r) = (\mathsf{I}'', \mathsf{C})$ where $\mathsf{I}'' = (\mathsf{U}', \mathsf{R}', \mathsf{Ac}, \mathsf{r_{type}}, \triangleright, \rightarrow_{post}{}', \rightarrow_{share}, \rightarrow_{like}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$,
ii) $\triangleright' = \triangleright \cup \{r, r_p\}$

As mentioned before, the *share* operation does not create any new artifact, but it will updates the actors set and then makes an update to the share relation ($\rightarrow_{share}$) as formally defined in Def. 9.

Definition 9: Let Social Data be $\mathsf{D} = (\mathsf{I}, \mathsf{C})$ with Interactions $\mathsf{I} = (\mathsf{U}, \mathsf{R}, \mathsf{Ac}, \mathsf{r_{type}}, \triangleright, \rightarrow_{post}, \rightarrow_{share}, \rightarrow_{like}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$, then we define the **share** operation on an artifact r by an user u as $\mathsf{D} \bigoplus_s (u, r) = (\mathsf{I}', \mathsf{C})$ where $\mathsf{I}' = (\mathsf{U} \cup \{u\}, \mathsf{R}, \mathsf{Ac}, \mathsf{r_{type}}, \triangleright, \rightarrow_{post}, \rightarrow_{share} \cup \{(u, r)\}, \rightarrow_{like}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$.

The following definition (Def. 10) contains formal definitions of *like* and *unlike* operations as an update to the like relation ($\rightarrow_{like}$). A *like* action on an artifact will add a mapping to like relation ($\rightarrow_{like}$) (in addition to adding the actor to the actors set), where as an *unlike* action will simply remove the existing mapping.

Definition 10: In Social Data $\mathsf{D} = (\mathsf{I}, \mathsf{C})$ with Interactions $\mathsf{I} = (\mathsf{U}, \mathsf{R}, \mathsf{Ac}, \mathsf{r_{type}}, \triangleright, \rightarrow_{post}, \rightarrow_{share}, \rightarrow_{like}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$, we define the **like** operation by an user u on an artifact r as $\mathsf{D} \bigoplus_l (u, r) = (\mathsf{I}', \mathsf{C})$ where $\mathsf{I}' = (\mathsf{U} \cup \{u\}, \mathsf{R}, \mathsf{Ac}, \mathsf{r_{type}}, \triangleright, \rightarrow_{post}, \rightarrow_{share}, \rightarrow_{like} \cup \{(u, r)\}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$.

Similarly, we define the **unlike** operation on $\mathsf{D} = (\mathsf{I}, \mathsf{C})$ with Interactions $\mathsf{I} = (\mathsf{U}, \mathsf{R}, \mathsf{Ac}, \mathsf{r_{type}}, \triangleright, \rightarrow_{post}, \rightarrow_{share}, \rightarrow_{like}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$, as $\mathsf{D} \bigominus_l (u, r) = (\mathsf{I}', \mathsf{C})$ where $\mathsf{I}' = (\mathsf{U}, \mathsf{R}, \mathsf{Ac}, \mathsf{r_{type}}, \triangleright, \rightarrow_{post}, \rightarrow_{share}, \rightarrow_{like} \setminus \{(u, r)\}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$.

Finally, *tagging* action accepts a tuple $((u, r, t))$ containing an actor, an artifact and a set of hash words (i.e. keywords and actors) and an update to tagging relation ($\rightarrow_{tag}$) will be applied as shown in the Def. 11.

Definition 11: In a Social Data $\mathsf{D} = (\mathsf{I}, \mathsf{C})$ with Interactions $\mathsf{I} = (\mathsf{U}, \mathsf{R}, \mathsf{Ac}, \mathsf{r_{type}}, \triangleright, \rightarrow_{post}, \rightarrow_{share}, \rightarrow_{like}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$, we define the **tagging** operation by an user u on an artifact r with a set of hash words $t \in \mathcal{P}(\mathsf{U} \cup \mathsf{Ke})$ as $\mathsf{D} \bigoplus_t (u, r, t) = (\mathsf{I}', \mathsf{C})$ where $\mathsf{I}' = (\mathsf{U} \cup \{u\}, \mathsf{R}, \mathsf{Ac}, \mathsf{r_{type}}, \triangleright, \rightarrow_{post}, \rightarrow_{share}, \rightarrow_{like}, \rightarrow_{tag} \cup \{(u, r, t)\}, \rightarrow_{\mathsf{act}})$.

B. ILLUSTRATIVE EXAMPLE

In this section, we exemplify the formal model by taking an example post from the Facebook page of McDonald's

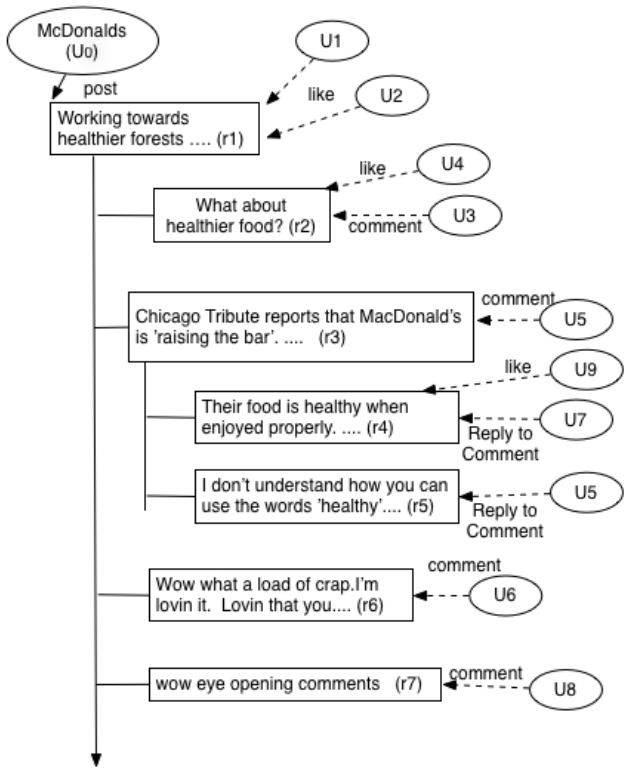

FIGURE 3. Facebook post example in formal model.

Food/Beverages as shown in the figure 3. In order to enhance the readability of the example, the artifacts (e.g. texts) have been annotated as $r1, r2$ etc and the annotated values will be used in encoding the example using the formal model.

Example 1: The following are some of the texts extracted from a sample post [60] from Facebook page of McDonald's Food/Beverages.

r_1 = Working towards healthier forests through more sustainable packaging. Learn more about how McDonald's is addressing climate change: http://McD.to/6188BrQzM

r_2 = What about healthier food?

r_3 = Chicago Tribute reports that MacDonald's is 'raising the bar'. You mean bars with nails in them to beat live chickens with? MacDonald's is one big lie. Don't believe them. Next they'll tell you their food is healthy.

r_4 = Their food is healthy when enjoyed properly. Their beef is amazing and that's what they move a lot of. The fattier menu items, if you have any modicum of a pallet, you'll notice are sides and not to be enjoyed in such an amount as whole meals themselves, but hey, I know some people who think raw sugar is a treat.

r_5 = *I* don't understand how you can use the words 'healthy' and MacDonald's in the same sentence. They manufacture (and I use that word deliberately) to have a perfect balance of salt, sugar and fat to hook children with their 'Happy Meals'. Sorry Keith, but healthy does not contain GMO's, Factory Farmed Animals, Chicken beaks, feathers etc, wood cellulose, fat, sugar and salt.

r_6 = Wow what a load of crap. I'm lovin it. Lovin that you are losing business and closing stores. Serving gmo.poison and promoting health. I want to puke

r_7 = wow eye opening comments

The example shown in Fig. 3 can be encoded as follows,

The social Data $\mathsf{D} = (\mathsf{I}, \mathsf{C})$ contains two components:

$\mathsf{I} = (\mathsf{U}, \mathsf{R}, \mathrm{Ac}, \mathsf{r}_{\mathsf{type}}, \rhd, \rightarrow_{post}, \rightarrow_{share}, \rightarrow_{like}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$ is the Interactions and

$\mathsf{C} = (\mathrm{To}, \mathrm{Ke}, \mathrm{Pr}, \mathrm{Se}, \rightarrow_{\mathsf{topic}}, \rightarrow_{\mathsf{key}}, \rightarrow_{\mathsf{pro}}, \rightarrow_{\mathsf{sen}})$ is the Conversations.

Initailly, let us assume that the sets of activities, topics, keywords, pronouns and sentiments will have the following values.

$\mathrm{Ac} = \{\text{promotion}\}$,
$\mathrm{To} = \{\text{healthy food, sustainable packaging}\}$,
$\mathrm{Ke} = \{\text{healthy, sustainable, beef, chicken, . .}\}$
$\mathrm{Pr} = \{We, I\}$, $\mathrm{Se} = \{+, 0, -\}$,
$\mathsf{U} = \{u_0, u_1, ...\}$
$\mathsf{R} = \{r_1\}$
$\rightarrow_{\mathsf{act}} = \{(r_1, promotion)\}$

post action by u_0

$\mathsf{D} \bigoplus_p (u_0, r_1) = \mathsf{D}_1 = (\mathsf{I}_1, \mathsf{C})$ where
$\mathsf{I}_1 = (\mathsf{U}_1, \mathsf{R}_1, \mathrm{Ac}, \mathsf{r}_{\mathsf{type}}, \rhd, \rightarrow_{post\ 1}, \rightarrow_{share}, \rightarrow_{like}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$ with the following values
$\mathsf{U}_1 = \mathsf{U} \cup \{u_0\}$, $\mathsf{R} = \mathsf{R} \cup \{r_1\}$ and
$\rightarrow_{post\ 1} = \rightarrow_{post} \cup \{(u_0, \{r_1\})\}$

like action by u_2 and u_1

Let's imagine that the post was liked by user u_2 first and then liked by user u_1.

$\mathsf{D}_1 \bigoplus_l (u_2, r_1) \bigoplus_l (u_1, r_1) = \mathsf{D}_2 = (\mathsf{I}_2, \mathsf{C})$ where
$\mathsf{I}_2 = (\mathsf{U}_2, \mathsf{R}, \mathrm{Ac}, \mathsf{r}_{\mathsf{type}}, \rhd, \rightarrow_{post\ 1}, \rightarrow_{share}, \rightarrow_{like\ 1}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$ with the following values
$\mathsf{U}_2 = \mathsf{U}_1 \cup \{u_2\} \cup \{u_1\}$, and
$\rightarrow_{like\ 1} = \rightarrow_{like} \cup \{(u_2, r_1), (u_1, r_1)\}$

comment action by u_5 on the post r_1

Let's imagine that the user u_5 posted a comment (r_3) on the Facebook post and let D_3 be the social data before the comment action.

$\mathsf{D}_3 \bigoplus_c (u_5, r_3, r_1) = \mathsf{D}_4 = (\mathsf{I}_4, \mathsf{C})$ where
$\mathsf{I}_4 = (\mathsf{U}_4, \mathsf{R}_3, \rhd_1, \mathsf{r}_{\mathsf{type}}, \mathrm{Ac}, \rightarrow_{post\ 3}, \rightarrow_{share}, \rightarrow_{like\ 1}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$ with the following values
$\mathsf{U}_3 = \mathsf{U}_3 \cup \{u_5\}$, $\mathsf{R}_3 = \mathsf{R}_2 \cup \{r_3\}$, $\rightarrow_{post\ 3} = \rightarrow_{post\ 2} \cup \{(u_5, \{r_3\})\}$ and $\rhd_1 = \rhd \cup \{(r_3, r_1)\}$.

Reply to comment by u_7 on the comment r_3

Let's imagine that the user u_7 posted a reply (r_4) on the comment (r_4).

$\mathsf{D}_4 \bigoplus_c (u_7, r_4, r_3) = \mathsf{D}_5 = (\mathsf{I}_5, \mathsf{C})$ where
$\mathsf{I}_5 = (\mathsf{U}_5, \mathsf{R}_4, \rhd_2, \mathsf{r}_{\mathsf{type}}, \mathrm{Ac}, \rightarrow_{post\ 4}, \rightarrow_{share}, \rightarrow_{like\ 1}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$ with the following values
$\mathsf{U}_5 = \mathsf{U}_4 \cup \{u_7\}$, $\mathsf{R}_4 = \mathsf{R}_3 \cup \{r_4\}$, $\rightarrow_{post\ 4} = \rightarrow_{post\ 3} \cup \{(u_7, \{r_4\})\}$ and $\rhd_2 = \rhd_1 \cup \{(r_4, r_3)\}$.

The rest of the operations shown in Fig. 3 can expressed similarly in the formal model.

VI. CASE STUDY 1: FUZZY-SET BASED SENTIMENT ANALYSIS

At the enterprise level, as Li and Leckenby [61] observed, technological advances such as the Internet have resulted in the vertical integration of business channel capacities such as production, distribution, transaction (e.g., Amazon and other e-commerce websites) and a horizontal integration of marketing functions such as advertising, promotions, public relations (e.g., Facebook and other social media platforms). At the agentic level of consumers, Internet and social media platforms resulted in changes not only to consumers' attitudes, perceptions and behaviours but also to the decision-making process itself in terms of the consideration set, search criteria, heuristics, and time [2], [62]. Taken together this led to the emergence of organizations that strategically utilize the online channels including social media platforms for business purposes [4]. This results in vast amounts of social data related to an enterprise's products, services, policies and processes. As such, one key application domain for sentiment analysis in enterprises is to monitor brand image, loyalty, and reputation.

Sentiment analysis can help in the understanding the user motivations for social media engagement, the different phases of consumer decision-making process and the potential business value and organizational impact of positive, negative and neutral sentiments. To illustrate this point, let us consider the following instance of socially shared consumption [63]: a positive mention about a product resulting from an automated status update of digital consumption on social media platform such as Facebook. In terms of consumer decision-making, this Facebook post can play a role in all three different orderings of the Hierarch of Effects (HoE) [64], [65] in terms of learning about the product, evaluating one's own experience of it with those of others, and engaging with the product as a brand loyalist by following that particular product related Facebook pages and posts. Similarly, the interactional dynamics of users sentiments on social media platforms might help companies better understand the sales funnel models such as AIDA (Attention, Interest, Desire and Action) [61]. Sentiments of users' posts might provide value in terms of social capital and/or signaling by turning the private individual act of consumption into a public social event and thereby signaling the user's characteristics such as taste, class, conscientiousness, and/or wealth. In other words, sociological dynamics and marketing implications similar to the conspicuous consumption [66].

A. FORMAL MODEL OF FUZZY SOCIAL DATA

In this section, we will extend the formal semantics of social data presented in Sec. V with the semantics of Fuzzy sets. Regarding notations for the formal model of Fuzzy Social Data, we will follow the same notations mentioned in Sec. V.

1) FUZZY SETS

First, we will recall necessary basic definitions of Fuzzy sets [67].

Definition 12: If X is a set of elements denoted by x, then a fuzzy set A over X is defined as a set of ordered pairs $A = \{(x, \mu_A(x)) \mid x \in X)\}$ where $\mu_A : X \rightarrow [0, 1]$ is the membership function.

Each member or element of a fuzzy set A is mapped to real number between 0 and 1 ($[0, 1]$), which represents the degree of membership of an element in the fuzzy set. A membership value of 1 indicates full membership, while a value of 0 indicates no membership.

Definition 13: For a (finite) fuzzy set A, the *cardinality* is defined as $|A| = \sum_{x \in X} \mu_A(x)$, which is the summation of all membership values of a fuzzy set. The *relative cardinality* $\|A\|$ is defined as $\|A\| = \frac{|A|}{|X|}$, where $|X|$ is the number of elements in set X.

Definition 14: The support of a fuzzy set A is a crisp set of all $x \in X$ such that $\mu_A(x) > 0$. The crisp set of elements that belongs to fuzzy set A at least to a degree α is called α-level or α-cut is defined as $A_\alpha = \{x \mid x \in X \wedge \mu_A(x) \geq \alpha\}$.

Definition 15: The *Union* operation on two fuzzy sets $A = \{(x, \mu_A(x)) \mid x \in X)\}$ and $B = \{(x, \mu_B(x)) \mid x \in X)\}$ with membership functions μ_A and μ_B respectively is defined as a fuzzy set $\{(x, \mu_{A \cup B}(x)) \mid \mu_{A \cup B}(x) = Max(\mu_A(x), \mu_B(x))\}$.

Definition 16: A fuzzy relation R from a set A to B with its membership function $\mu_R : A \times B \rightarrow [0, 1]$ is defined as $R = \{((a, b), \mu_R(a, b)) \mid (a, b) \in A \times B\}$.

Similar to a fuzzy set, the membership function of a fuzzy relation indicates strength of its relationship. Moreover a fuzzy relation is nothing but a fuzzy set where the elements are ordered pairs of the relation.

2) FUZZY SOCIAL DATA

Following the definitions of Artifact Type 1, Actions 2 and Social Data 3 from Sec. V, we redefine fuzzy *Interactions* by redefining the activity relation ($\rightarrow_{\mathsf{act}}$) as a fuzzy relation as follows.

Definition 17: In Social Data $\mathsf{D} = (\mathsf{I}, \mathsf{C})$, fuzzy Interactions is defined as a tuple $\mathsf{I} = (\mathsf{U}, \mathsf{R}, \mathrm{Ac}, \mathsf{r}_{\mathsf{type}}, \triangleright, \rightarrow_{post}, \rightarrow_{share}, \rightarrow_{like}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$ where

i) $\mathsf{U}, \mathsf{R}, \mathrm{Ac}, \mathsf{r}_{\mathsf{type}}, \triangleright, \rightarrow_{post}, \rightarrow_{share}, \rightarrow_{like}, \rightarrow_{tag}$ are same as defined in 4,

ii) $\rightarrow_{\mathsf{act}} = \{((r, a), \mu_{\rightarrow_{\mathsf{act}}}(r, a)) \mid r \in \mathsf{R}, a \in \mathrm{Ac}\}$ is a fuzzy relation mapping artifacts to activities with membership function $\mu_{\rightarrow_{\mathsf{act}}} : \mathsf{R} \times \mathrm{Ac} \rightarrow [0, 1]$

As shown in Def. 17-i, except $\rightarrow_{\mathsf{act}}$ relation, semantics of the rest of the items in fuzzy Interactions remain same as defined in the Interactions of core Social Data formal model. The $\rightarrow_{\mathsf{act}}$ is a fuzzy relation indicates a mapping between artifacts to activities (Def. 17-ii) with a membership function ($\mu_{\rightarrow_{\mathsf{act}}}$) indicating the strength of relationship, varies between 0 to 1. A membership value of 0 indicates complete non-existence of relationship between an artifact to an activity, where as value of 1 indicates full existence of such relationship. A value in between 0 to 1 indicates partial existence of the relationship.

Similarly, we define fuzzy Conversations of Social data as follows by redefining all relations as fuzzy relations.

Definition 18: In Social Data $\mathsf{D} = (\mathsf{I}, \mathsf{C})$ we define fuzzy Conversations as $\mathsf{C} = (\mathrm{To}, \mathrm{Ke}, \mathrm{Pr}, \mathrm{Se}, \rightarrow_{\mathsf{topic}}, \rightarrow_{\mathsf{key}}, \rightarrow_{\mathsf{pro}}, \rightarrow_{\mathsf{sen}})$ where

(i) To, Ke, Pr, Se are the sets of topics, keywords, pronouns and sentiments respectively as defined in 5

(ii) $\rightarrow_{\mathsf{topic}} = \{\big((r, to), \mu_{\rightarrow_{\mathsf{topic}}}(r, to)\big) \mid r \in \mathsf{R}, to \in \mathrm{To}\}$ is a Fuzzy relation mapping artifacts to topics with membership function $\mu_{\rightarrow_{\mathsf{topic}}} : \mathsf{R} \times \mathrm{To} \rightarrow [0, 1]$,

(iii) $\rightarrow_{\mathsf{key}} = \{\big((r, ke), \mu_{\rightarrow_{\mathsf{key}}}(r, ke)\big) \mid r \in \mathsf{R}, ke \in \mathrm{Ke}\}$ is a Fuzzy relation mapping artifacts to keywords with membership function $\mu_{\rightarrow_{\mathsf{key}}} : \mathsf{R} \times \mathrm{Ke} \rightarrow [0, 1]$,

(iv) $\rightarrow_{\mathsf{pro}} = \{\big((r, pr), \mu_{\rightarrow_{\mathsf{pro}}}(r, pr)\big) \mid r \in \mathsf{R}, pr \in \mathrm{Pr}\}$ is a Fuzzy relation mapping artifacts to pronouns with membership function $\mu_{\rightarrow_{\mathsf{pro}}} : \mathsf{R} \times \mathrm{Pr} \rightarrow [0, 1]$,

(v) $\rightarrow_{\mathsf{sen}} = \{\big((r, se), \mu_{\rightarrow_{\mathsf{sen}}}(r, se)\big) \mid r \in \mathsf{R}, se \in \mathrm{Se}\}$ is a Fuzzy relation mapping artifacts to sentiments with membership function $\mu_{\rightarrow_{\mathsf{sen}}} : \mathsf{R} \times \mathrm{Se} \rightarrow [0, 1]$.

The semantics of sets of *topics* (To), *keywords* (Ke), *pronouns* (Pr), and *sentiments* (Se) in fuzzy Conversations remain the same as in the case of Conversations (Def. 5) of Core formal model of Social data. Furthermore, one may note that all the relations in fuzzy Conversations ($\rightarrow_{\mathsf{topic}}$, $\rightarrow_{\mathsf{key}}$, $\rightarrow_{\mathsf{pro}}$ and $\rightarrow_{\mathsf{sen}}$) are defined as fuzzy relations with membership function varies from [0, 1], indicating the strength of relationships, where as the relations in Conversations of core formal model of social data are crisp relations.

B. METHODOLOGY

In this section, we will outline a method for calculating the sentiments of artifacts and actors based on formal model presented in previous section.

1) SENTIMENT ANALYSIS

In contrast to the analytical focus on relationships in traditional social network analysis (SNA) methods, our analytical focus is on associations of actors and artefacts as sets and fuzzy sets based on certain criteria for actions, activities, sentiments, topics etc. In our associational approach, we model set and fuzzy set memberships of *Actors* performing *Actions* in *Activities* on *Artifacts*. Artifacts carry direct sentiment as they can be analysed by a sentiment engine and assigned a sentiment score and label by the sentiment engine. Individually, an action does not carry any sentiment, but it is the artifacts on which these actions are carried over, that contain sentiments. Similarly, even though actors does not carry sentiment directly, but they express their sentiments by performing actions on the artifacts, which contain the direct sentiment. Therefore, the sentiment attributed to an actor can be inferred or derived from the artifacts on which the actions are performed. Let us assume that the set of sentiments in the Conversations contain some predefined labels: *positive* (+), *neutral* (0) and *negative* (−) as indicated in $\mathrm{Se} = \{+, 0, -\}$.

2) SENTIMENT ANALYSIS OF ARTIFACTS

In this sentiment analysis of artifacts, let us assume that we are confined to textual types of artifacts, i.e. $\mathsf{r}_{\mathsf{type}}(r) = (\mathrm{post} \vee \mathrm{comment})$. Using an automatic method (for example using a natural language processing engine) for categorising sentiment of artifacts, an artifact can be mapped to different sentiment labels with a score indicating probability of relevance between the artifact and sentiment label. Normally, these scores are expressed as either percentages or real numbers (between 0 to 1), and the sum of such scores of an artifact for multiple sentiment labels will be equal to 1.

Therefore, in this sentiment analysis, we consider the sentiment score of an artifact as it's membership value of relationship between an artifact and a sentiment label ($\rightarrow_{\mathsf{sen}}$). For example, if the sentiment of an artifact r_1 is categorised among three sentiment labels as 0.43 *positive*, 0.26 *neutral* and 0.31 *negative*, then it is encoded in the sentiment fuzzy relation ($\rightarrow_{\mathsf{sen}}$) as $\rightarrow_{\mathsf{sen}} = \{., \big((r_1, +), 0.43\big), \big((r_1, 0), 0.26\big), \big((r_1, -), 0.31\big), .\}$.

Furthermore, we can perform an $\alpha-cut$ operation (Def. 14) on a Fuzzy set, to convert it to a crisp set containing set members, whose membership value is at least to the degree of of $\alpha \in [0, 1]$.

$$\mathsf{R}^{se}_{\alpha} = \{r \mid (\mu_{\rightarrow_{\mathsf{sen}}}(r, se) \geq \alpha)\}$$

Finally the crisp set R^{se}_{α} contains all the desired artifacts whose sentiment is more than certain minimum value (α). Based on the context and requirements, one could apply different $\alpha-cuts$ to the fuzzy set to $\rightarrow_{\mathsf{sen}}$, to get the crisp sets containing artifacts meeting to certain minimum sentiment score as criteria (α).

Especially, the method of application $\alpha - cuts$ is quite useful when we want to explore a phenomena which is very feebly represented in the data corpus. For example, in order to explore a weak negative sentiment in response to an event in the data corpus, one could go for a very low value of $\alpha - cut$ (e.g. $\alpha = 0.2$ or even less), to further analyse the data in a magnified view to get fine grained data visualisations. On there other hand, if some one wants to get a more abstract view on a dominantly represented sentiment values, adopting higher values of $\alpha - cut$ (e.g. $\alpha > 0.6$ or even more) will results in a view with a course grained data visualisations where only strong sentiments are represented.

ACTORS ASSOCIATED WITH ARTIFACTS

Several actors are associated with an artifact. For example actors can perform *post*, *comment share* and *like* actions on an artifact. Of course, actors can also perform *tag* action on an artifact, but we will ignore tagging operation for sentiment analysis in this paper. The set of actors that are associated with the given set of artifacts (e.g. R^{se}_{α}), can be computed as follows,

$$\begin{aligned} &\forall r \in \mathsf{R}^{se}_{\alpha}. \\ &\mathsf{U}_{\mathsf{R}^{se}_{\alpha}} = \{u \mid r \in \rightarrow_{post} (u)\} \\ &\qquad \cup \{u \mid r' \in \mathsf{R} \wedge r' \in \rightarrow_{post} (u) \wedge \triangleright(r') = r\} \\ &\qquad \cup \{u \mid (u, r) \in \rightarrow_{share}\} \\ &\qquad \cup \{u \mid (u, r) \in \rightarrow_{like}\}. \end{aligned}$$

As formally expressed above, the set of actors ($\mathsf{U}_{\mathsf{R}_\alpha^{se}}$) associated with given set of artifacts (R_α^{se}) contains sets of users who posted the artifacts, who commented on the artifacts, who shared the artifacts and who liked the artifacts. One could notice that both the set of actors ($\mathsf{U}_{\mathsf{R}_\alpha^{se}}$) and set of artifacts ($\mathsf{R}_\alpha^{se}$) are crisp sets and taking the cardinality of these sets will provide us the number of members in them. One of the ways to analyse the sentiment over a time scale could be to compute these sets (R_α and $\mathsf{U}_{\mathsf{R}_\alpha^{se}}$) for each sentiment label ($\forall .se \in \{+, 0, -\}$) for given time span intervals to plot them across the time horizon.

3) SENTIMENT ANALYSIS OF ACTORS

As explained in the previous section, the sentiment attributed to an actor can be derived from the artifacts on which actions are performed by the actor. An actor can perform different actions: *post*, *comment*, *share*, *like* and *tag* on different artifacts. However *tag* action is not considered for the sentiment analysis as mentioned previously. From the formal model, for any given actor, we can compute the sets of artifacts over which the actor performed actions as mentioned previously. Building on that, for any given artifact we can also compute the sentiment scores associated with different sentiment labels from the sentiment relation ($\rightarrow_{\mathsf{sen}}$).

Therefore, the sentiment associated with an actor (u^{se}) can be defined as a tuple containing the following fuzzy sets,

$$(\rightarrow_p^{se}, \rightarrow_c^{se}, \rightarrow_s^{se}, \rightarrow_l^{se})$$

1) $\rightarrow_p^{se} = \{\big((r, se), \mu_p(r, se)\big) \mid r \in \rightarrow_{post} (u) \wedge \triangleright(r) \text{ is not defined}\}$ is a fuzzy set containing all the artifacts that are posted by the user with $\mu_p(r, se) = \mu_{\rightarrow_{\mathsf{sen}}}(r, se)$ as membership function,
2) $\rightarrow_c^{se} = \{\big((r, se), \mu_c(r, se)\big) \mid r \in \rightarrow_{post} (u) \wedge \exists r' \in \mathsf{R}. \triangleright(r) = r'\}$ is a fuzzy set containing all the comment artifacts that are posted by the user, with $\mu_c(r, se) = \mu_{\rightarrow_{\mathsf{sen}}}(r, se)$ as membership function,
3) $\rightarrow_s^{se} = \{\big((r, se), \mu_s(r, se)\big) \mid (u, r) \in \rightarrow_{share}\}$ is a fuzzy set containing all the artifacts that are shared by the user, with $\mu_s(r, se) = \mu_{\rightarrow_{\mathsf{sen}}}(r, se)$ as membership function,
4) $\rightarrow_l^{se} = \{\big((r, se), \mu_l(r, se)\big) \mid (u, r) \in \rightarrow_{like}\}$ is a fuzzy set containing all the artifacts that are liked by the user, with $\mu_l(r, se) = \mu_{\rightarrow_{\mathsf{sen}}}(r, se)$ as membership function,

where $r \in \mathsf{R}$, $se \in \mathsf{Se}$, $\mu_{\rightarrow_{\mathsf{sen}}}(r, se)$ is the membership function of the sentiment fuzzy relation ($\rightarrow_{\mathsf{sen}}$).

The the sentiment associated with an actor (u^{se}) can calculated by application of union operation (Def. 15) on the above fuzzy sets ($\rightarrow_p^{se} \cup \rightarrow_c^{se} \cup \rightarrow_s^{se} \cup \rightarrow_l^{se} \cup \rightarrow_t^{se}$). Therefore, sentiment associated with an actor (u^{se}) can be computed as follows

$$u^{se} = \{\big((r, se), \mu_u(r, se)\big) \mid r \in \mathsf{R}_u\}, \text{ where}$$

1) R_u is set of artifacts for an actor (u) over which the actions are performed
$\mathsf{R}_u = \rightarrow_{post} (u) \cup \rightarrow_{share} (u) \cup \rightarrow_{like} (u)$.
Notice that, the set $\rightarrow_{post} (u)$ contains all artifacts that are posted and commented by the user.
2) the membership function is defined as
$$\mu_u(r, se) = \text{Max}(\mu_p(r, se), \mu_c(r, se), \mu_s(r, se), \mu_l(r, se))$$

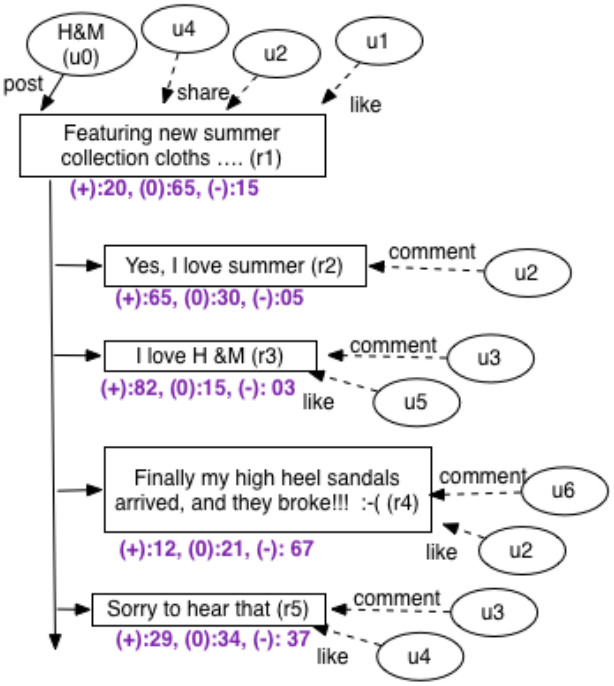

FIGURE 4. Example in formal model.

One could observe that the associated sentiment of an actor is a fuzzy set with artifacts and sentiment labels with membership values as the sentiment scores. Therefore, one could apply the $\alpha - cuts$ on the fuzzy set to extract a crisp set (u_α^{se}) meeting up the criteria for each sentiment label ($\forall .se \in \{+, 0, -\}$).

Furthermore, the same method can applied to get such sets for different time span intervals with in a time period. One of the ways to analyse the associated actor sentiment over a time scale could be to compute these sets (u_α^{se}) for each sentiment label ($\forall .se \in \{+, 0, -\}$) for given time span intervals and plot their cardinalities (e.g. number of artifacts in the set for + sentiment) across the time horizon. In this way, we could profile the associated sentiment of an actor over a period of time by computing how the cardinalities of the sets of the associated sentiment labels of an actor varies over timeline.

C. ILLUSTRATIVE EXAMPLE

In this section, we will exemplify the formal model with fuzzy sets by taking an example post from the Facebook page of H&M cloth stores as shown in the figure 4. In order to enhance the readability of the example, the artifacts (e.g. texts) have been annotated as $r1, r2$ etc and the annotated values will be used in encoding the example using the formal model.

Moreover, as our focus is to mainly to demonstrate sentiment analysis, we will abstract away from the details of the sets (e.g. Topics, Keywords etc) which are not directly involved in the sentiment analysis. As shown in Figure 4, the sentiments of the artifacts (e.g. (+):20, (0):65, (−):15) are represented in the boxes below the artifacts.

Example 2: The example shown in Fig. 4 will be encoded as follows,

$\mathsf{D} = (\mathsf{I}, \mathsf{C})$ where $\mathsf{I} = (\mathsf{U}, \mathsf{R}, \mathrm{Ac}, \mathsf{r}_{\mathsf{type}}, \rhd, \rightarrow_{post}, \rightarrow_{share}, \rightarrow_{like}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$ is the Interactions and C= (To, Ke, Pr, Se, $\rightarrow_{\mathsf{topic}}, \rightarrow_{\mathsf{key}}, \rightarrow_{\mathsf{pro}}, \rightarrow_{\mathsf{sen}}$) is the Conversations.

Initially, the sets of actors, artifacts and other relations have the following values.

$\mathsf{U} = \{u_0, u_1, u_2, u_3, u_4, u_5, u_6, ...\}$
$\mathsf{R} = \{r_1, r_2, r_3, r_4, r_5, ...\}$
$\rhd = \{(r_2, r_1), (r_3, r_1), (r_4, r_1), (r_5, r_1), ..\}$
$\rightarrow_{post} = \{(u_0, \{r_1, ..\}), (u_2, \{r_2\}), (u_3, \{r_3, r_5\}), (u_6, \{r_4\}), ..\}$
$\rightarrow_{share} = \{(u_4, r_1), (u_2, r_1)..\}$
$\rightarrow_{like} = \{(u_1, r_1), (u_5, r_3), (u_2, r_4), (u_4, r_5), ..\}$
$\mathrm{Se} = \{+, 0, -\}$

After the artifacts are analysed for the sentiments, the sentiment relation becomes a fuzzy set contain the pairs of artifacts and sentiment labels with the sentiment score as membership value as shown below,

$$\begin{aligned}\rightarrow_{\mathsf{sen}} = \{&((r_1, +), 0.20), ((r_1, 0), 0.65), ((r_1, -), 0.15),\\ &((r_2, +), 0.65), ((r_2, 0), 0.30), ((r_2, -), 0.05),\\ &((r_3, +), 0.82), ((r_3, 0), 0.15), ((r_3, -), 0.03),\\ &((r_4, +), 0.12), ((r_4, 0), 0.21), ((r_4, -), 0.67),\\ &((r_5, +), 0.29), ((r_5, 0), 0.34), ((r_5, -), 0.37)\}\end{aligned}$$

Regarding temporal dimension ($\mathbf{T}$), let us assume that the post (in Figure 4) and all its conversation happened in same time frame ($t_1 - t_2$), then sentiment relation for time period ($t_1 - t_2$) is same as $\rightarrow_{\mathsf{sen}}$.

From the sentiment fuzzy set, one can extract different crisp sets (R^{se}_{α}) for artifacts based different values of $\alpha - cuts$. For example for a value of $\alpha = 0.4$, the artifact sets for + and − will be

$\mathsf{R}^{+}_{\alpha=0.40} = \{r_2, r_3\}$ and $|\mathsf{R}^{+}_{\alpha=0.40}| = 2$
$\mathsf{R}^{-}_{\alpha=0.40} = \{r_4\}$ and $|\mathsf{R}^{-}_{\alpha=0.40}| = 1$

On the other hand, if some one wants a fine grained analysis of the data, they could use a lower value for $\alpha - cut$, which will include more elements into the analysis.

$\mathsf{R}^{+}_{\alpha=0.20} = \{r_1, r_2, r_3, r_5\}$ and $|\mathsf{R}^{+}_{\alpha=0.20}| = 4$
$\mathsf{R}^{-}_{\alpha=0.20} = \{r_4, r_5\}$ and $|\mathsf{R}^{-}_{\alpha=0.20}| = 2.$

Similarly, we can also compute the actor sets ($\mathsf{U}_{\mathsf{R}^{se}_{\alpha}}$) that are associated with the artifact sets as follows.

$$\begin{aligned}\mathsf{U}_{\mathsf{R}^{+}_{\alpha=0.40}} &= \{u_2\} \cup \emptyset \cup \emptyset \cup \emptyset \cup \{u_3\} \cup \emptyset \cup \emptyset \cup \{u_5\}\\ &= \{u_2, u_3, u_5\}\\ \mathsf{U}_{\mathsf{R}^{-}_{\alpha=0.20}} &= \{u_6\} \cup \emptyset \cup \emptyset \cup \{u_2\} \cup \{u_3\} \cup \emptyset \cup \emptyset \cup \{u_4\}\\ &= \{u_6, u_2, u_3, u_4\}\end{aligned}$$

Notice that, here we have an advantage due to fuzzy set modelling that an can an actor can be present in more than one set (e.g. $\mathsf{U}_{\mathsf{R}^{+}_{\alpha=0.2}}$ and $\mathsf{U}_{\mathsf{R}^{-}_{\alpha=0.2}}$), as an actor can express more than one sentiment by performing the actions on artifacts in reality. When once crisp sets for artifacts (R^{se}_{α}) and actors ($\mathsf{U}_{\mathsf{R}^{se}_{\alpha}}$) are computed on a time scale for given time spans, one can plot their cardinalities against the time scale.

1) INFERRED SENTIMENT AND ACTOR PROFILING:

As explained in the previous section, the inferred sentiment for actors can be calculated in the similar line as above. In this example, we will show how one can compute inferred sentiment for the actor u_2, where we take union of fuzzy sets containing artifacts with sentiment labels for the artifacts posted, shared and liked by actor u_2 as follows.

$$\begin{aligned}u_2^{+} &= \{((r_2, +), 0.65)\} \cup \{((r_1, +), 0.20)\} \cup \{((r_4, +), 0.12)\}\\ &= \{\big((r_2, +), 0.65\big), \big((r_1, +), 0.20\big), \big((r_4, +), 0.12\big)\}\\ u_2^{-} &= \{((r_2, -), 0.05)\} \cup \{((r_1, -), 0.15)\} \cup \{((r_4, -), 0.67)\}\\ &= \{\big((r_2, -), 0.05\big), \big((r_1, -), 0.15\big), \big((r_4, -), 0.67\big)\}\end{aligned}$$

After computing the fuzzy sets as above, one could apply $\alpha - cut$ with the required granularity to get crisp sets similar to the sentiment analysis of the artifacts. After that many such sets can be computed for a given time intervals and can be plotted on a time scale to analyse how the sentiment of an actor varies in the time frame.

D. CASE STUDY AND FINDINGS

In this section, we present a case study where big social data of the fast fashion company, H&M is collected from its Facebook page. We empirically analyse the sentiment of artifacts on social data collected by Social Data Analytics Tool (SODATO) [38] from the Facebook page of H&M and analysis using the methodology presented in the previous section that is based on formal modelling of social data.

1) CONVERSATION ANALYSIS

Google Prediction API [68] was utilized in order to calculate sentiments for the posts and comments on the wall. Google Prediction API provides RESTful API access to the service. Configuration for computation of sentiment began with the setting up a model which was trained with the manually labelled data subset from the H&M data corpus fetched by SODATO. This training dataset consisted of 11,384 individual posts and comments randomly selected from H&M data corpus and their corresponding sentiment labels as coded by five different student analysts. Training data was labelled *Positive*, *Negative* or *Neutral* and the file was uploaded on the Google Cloud Storage using the console explorer interface provided by the Google.

After successful training of the model, Sentiment module provided by SODATO was utilized to calculate sentiment for posts and comments for the entire conversations corpus of H&M. The sentiment results for each individual post/comment returned by the Google Prediction API were

TABLE 3. Parent artifact (posts) sentiment distribution.

	α-cuts				
sentiment	$\geq$ **0.1**	$\geq$ **0.3**	$\geq$ **0.5**	$\geq$ **0.7**	$\geq$ **0.9**
$+$	17,752	25,949	30,343	25,869	19,974
$-$	9,166	14,503	16,577	13,494	10,397
0	12,566	21,607	26,826	24,312	21,830
$+ \cap -$	5,661	5,184	2,067	1,489	913
$+ \cap 0$	16,674	14,401	8,550	7,069	4,673
$- \cap 0$	10,017	9,984	6,541	5,381	3,892
$+ \cap - \cap 0$	39,001	19,209	6,512	4,567	2,739
Total artifacts	110,837	110,837	97,416	82,181	64,418

TABLE 4. Total artifact (posts + comments + likes) sentiment distribution.

	α-cuts				
sentiment	$\geq$ **0.1**	$\geq$ **0.3**	$\geq$ **0.5**	$\geq$ **0.7**	$\geq$ **0.9**
$+$	36,114	57,653	77,551	80,540	83,378
$-$	19,433	32,472	42,145	35,388	28,310
0	37,511	62,037	85,334	80,404	79,981
$+ \cap -$	28,788	33,929	49,315	109,785	237,156
$+ \cap 0$	94,094	99,339	119,822	141,158	297,516
$- \cap 0$	54,756	56,527	50,176	44,660	35,520
$+ \cap - \cap 0$	16,537,774	13,810,588	11,477,670	10,189,815	7,742,858
Total artifacts	16,808,470	14,152,545	11,902,013	10,681,750	8,504,719

TABLE 5. Actors sentiment with different α − *cuts*.

	α-cuts				
sentiment	$\geq$ **0.1**	$\geq$ **0.3**	$\geq$ **0.5**	$\geq$ **0.7**	$\geq$ **0.9**
$+$	331,891	441,290	549,159	563,600	555,964
$-$	211,783	311,861	382,082	317,912	199,815
0	1,074,602	1,335,933	1,469,989	1,413,921	1,168,176
$+ \cap -$	67,496	92,901	111,438	76,491	51,868
$+ \cap 0$	647,315	712,828	523,046	511,667	508,537
$- \cap 0$	411,821	248,707	149,532	122,645	66,889
$+ \cap - \cap 0$	979,718	581,106	400,186	338,565	231,158
Total actors	3,724,626	3,724,626	3,585,432	3,344,801	2,782,407

saved back to the relational database. In order to calculate quarterly aggregation of the sentiment classified conversations, further segmentation and grouping was performed using SQL queries and relational database entities were used to store data and it was made available for Analytical calculations.

2) DATA ANALYSIS

The H&M Facebook wall was fetched for a time period from 12-March-2007 to 31-December-2013 using SODATO tool. The total data corpus for that period contains 12.58 million data elements including posts, comments, likes on posts and comments and shares. The sentiment scores for the 12.58 million data elements were analysed using Google Prediction API [68].

3) FINDINGS

Compared to existing sentiment analysis methods and tools in academia and industry, the set theory and fuzzy set theory approach that we demonstrated in the tables (3, 4 and 5) and figures (5, 6 and 7) above reveal the longitudinal sentiment profiles of actors and artefacts for the entire corpus.

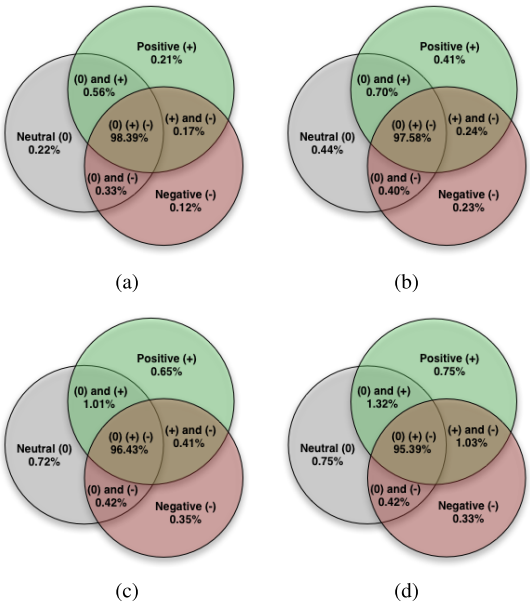

FIGURE 5. Artifact sentiments. **(a)** $\alpha \geq 0.1$. **(b)** $\alpha \geq 0.3$. **(c)** $\alpha \geq 0.5$. **(d)** $\alpha \geq 0.7$.

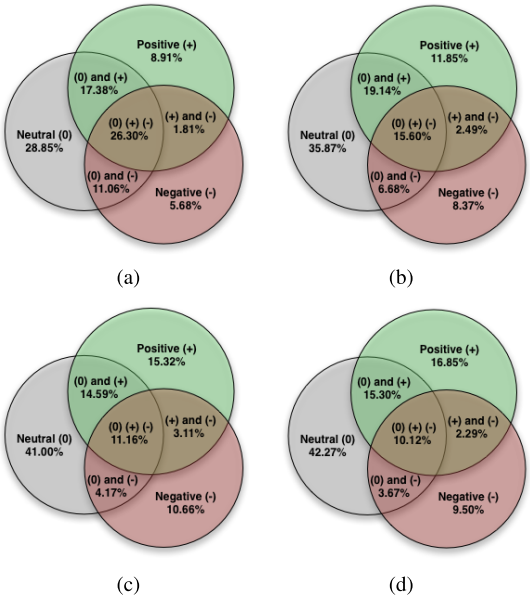

FIGURE 6. Actor sentiments. **(a)** $\alpha \geq 0.1$. **(b)** $\alpha \geq 0.3$. **(c)** $\alpha \geq 0.5$. **(d)** $\alpha \geq 0.7$.

The α-cut approach to sentiment analysis allows analysts (marketing professionals and/or academic researchers) to specify their own probability level for sentiment categories of positive, negative and neutral. Further, it allows the individual analyst to identify the intersections of positive, negative, and neutral sentiment for any given α-cut. This allows the analyst to identify strong-weak expressions of positive, negative, and neutral sentiment.

For example, let us consider the α-cut of 0.9 for actors in Table 5 and Figure 7(b). The graph shows that 7.18%

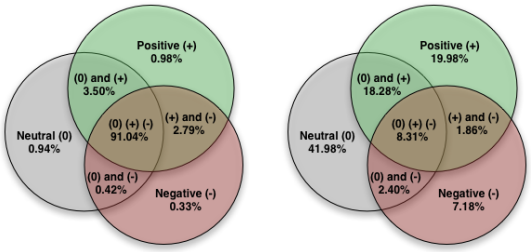

FIGURE 7. Artifact and Actor sentiments for $\alpha \geq 0.9$. (a) Artifact sentiments $\alpha \geq 0.9$. (b) Actor sentiments $\alpha \geq 0.9$.

of the entire Facebook user group for the company are always expressing negative sentiments whereas 19.9% of the user group is always expressing positive sentiments. With the caveat that not all of those positive and negative sentiments could be about the company itself (they could be directed towards other brands and/or other social actors on the Facebook page of the company), the results can help identify the strong brand loyalists (always positive) and strong brand critics (always negative). Similar analysis for the α-cut of 0.1 for actors will yield weak brand loyalists and critics.

With respect to the sentiment analysis of artifacts, at the α-cut of 0.9 (Table 3 and Figure 6(a)), we find that 0.33% of all conversations on the Facebook page were entirely negative. A quick test for the social media marketing effectiveness can be constructed by extracting the number of completely negative conversations started by the company itself. That is, it is marketing problem if the company's posts are being categorized as negative sentiment and the all ensuing interactions by its Facebook users are also negative. This might have implications for the brand reputation and image even discounting attempts at humor by self-depreciation and/or irony. Applying the crisp set and fuzzy set modelling of sentiments of actors and artefacts over critical time periods can reveal the temporal dynamics of how different users express their sentiments for different products, campaigns, and events. Having said that, our primary objective in this paper is not to provide a detailed interpretation of the results but to propose and demonstrate a new approach to sentiment analysis.

VII. CASE STUDY 2: SOCIAL SET ANALYSIS OF CORPORATE SOCIAL MEDIA CRISES ON FACEBOOK

Corporate crises are by nature unpredictable but post hoc, crises appear not unexpected. Corporate crisis can trigger negative reactions from stakeholders and thereby affect the overall performance of the company. Therefore, it is important for the companies to respond to the crises in order to limit the damage [69], [70]. This paper addresses the topic of corporate crises on social media channels. Social media crises pose significant challenges for organizations in terms of their rapid propagation and deterioration of brand parameters that can have sustained negative business impacts. This paper addresses the following research questions.

Research Questions

1) What were the characteristics of big social data before, during, and after the corporate crisis?
2) What strategies and tactics does a companies employ, if at all they do, in order to manage the social media crisis?
3) How can a company in general best manage a social media crisis?

A. SELECTED CORPORATE SOCIAL MEDIA CRISES

In order to address the above research questions, we selected four recent social media crises. The objective was to uncover temporal dyanamics and interactional patterns of big social data and to investigate the strategies and tactics adopted by the companies that have experienced social media crisis in order to manage them. We purposefully limited the selection of social media crises to Denmark and the social media platform to Facebook to hold invariant the technological,linguistic and socio-cultural aspects of interacting with social media [32], [37] invariant : Copenhagen Zoo, Telenor, Jensen's Bøfhus (translation: Jensens Steak House), and Imerco. Next, we briefly describe each corporate social media crisis.

Copenhagen Zoo experienced a social media crisis, which started on February 8th 2014, due to an impending euthanizing of a young giraffe they had chosen to call Marius and lasted until February 13th 2014. Also, major international media has also participated in the case of Marius. British BBC and The Guardian newspaper has also referred to the killing, CNN followed the case on both network and TV, and The New York Times has also written about Marius' death [71].

Telenor experienced a social media crisis on Facebook, which started on August 3rd 2014 and lasted until August 8, 2014, due to a farewell salute from an unsatisfied customer who wrote in the evening on August 2nd 2014 at Telenor's Facebook page that he had ended his mobile subscription with the telecom company. In his post, the dissatisfied customer described that Telenor could not manage to collect money by Direct Debit and that the company had repeatedly sent reminders before he had received the normal expense. This post brought Telenor into a social media crisis on Facebook[1] and more than 30,000 "liked it".[2]

Jensen's Bøfhus experienced a social media crisis on Facebook, which started on September 19, 2014 and lasted until September 27, 2014, due to a dispute between Jensen's Bøfhus, and a fish restaurant named *Jensens Fiskerestaurant* (ed. Jensen's Seafood Restaurant). The case involved a conviction in the Supreme Court that caused great debate in Denmark, since Jensen's Bøfhus were successful at that the name, Jensen Fiskerestaurant, is too similar to the steakhouse chain restaurant. This meant that the owner of Jensen's Fiskerestaurant, Jacob Jensen, had to change the name of his restaurant. According to Jensen's Bøfhus they were trying to protect their trademark in the catering industry as Jensens

[1] Telenor on tv2.dk.
[2] Telenor on politiken.dk.

TABLE 6. Actions of admin/non-admin actors on post artifact.

artifact (post) by	actions	by actor
admin actor	comment or like	admin actor
non-admin actor		non-admin actor

Fiskerestaurant were planning to expand with new restaurants in other cities.[3] According to the judgment, the small restaurateur, Jacob Jensen, had to pay 200,000 Danish kroner to Jensen's Bøfhus, 150,000 Danish kroner to the costs that Jensen's Bøfhus have had his own lawyer and other expenses.[4]

Imerco experienced a social media crisis, which started on August 25th, 2014 and lasted until August 26th 2014, due to a fast sold out anniversary vase from the brand Kähler. 16,000 customers wanted to buy a special anniversary vase from the company Kähler on offer at Imerco's website. However, this tumbled the website, after which angry customers vented their displeasure on Imerco's Facebook page.[5]

B. METHODOLOGY

In this section, we will outline the methodology adopted to conduct big social data analytics on the Facebook walls of the companies based on the formal definitions of social data as defined in Sec. V. In the analysis, we also distinguish between admin-actor (denoted by u_a), who manages the Facebook wall of an enterprise from non-admin actors (denoted by $u \in \mathsf{U} \backslash u_a$), who are the social media users. To simply the matters, we have excluded *share* action from our analysis as we did not noticed any share actions in the datasets. Moreover, the terms *user* and *actor* are used interchangeably throughout the paper without any difference in semantics.

1) ARTIFACT ANALYSIS (CRISIS DETECTION)

Social media crises are characterized by marked increase in interaction levels on the social media channels. Further, based on traditional crisis communication and management theories and frameworks discussed earlier, we conducted temporal analysis of interactions in terms of two kinds of actions (like and comment) with respect to two kinds of artifacts (posts and comments) made by two different kinds of actors (admins/companies and non-admins) over temporal dimension of daily, weekly and yearly as further explained below.

a: POST ARTIFACT ANALYSIS

As shown in Table 6, kind of actions that can be performed on a post artifact are comment and like. As one of the possible interactions in Table 6, comment and like actions made by non-admin actors over the post artifact created by the admin-actor can be defined as,

[3]Jensen's Bøfhus on tv2.dk.
[4]Jensen's Bøfhus on politiken.dk.
[5]Imerco on politiken.dk.

1) Comments by non-admin actors on admin-actor posts: $\mathsf{R}_c^{u|u_a} = \{r_c \mid (u_a, r_p) \wedge (u, r_c) \in \rightarrow_{post}\}$
2) Likes by non-admin actors on admin-actor posts: $\mathsf{L}^{u|u_a} = \{(u, r_p) \mid (u_a, r_p) \in \rightarrow_{post} \wedge (u, r_p) \in \rightarrow_{like}\}$.

The set $(\mathsf{R}_c^{u|u_a})$ contains comment artifacts (r_c) made by non-admin actors (u) on the post artifact (r_p) created by admin-actor (u_a). Similarly, the set $\mathsf{L}^{u|u_a}$ contains pairs of non-admin actors (u) with their liked post artifacts (r_p), that were created by the admin-actor (u_a). Finally, total number of actions made by the non-admin actors on admin-actor posts can be calculated by taking sum of set cardinalities $(|\mathsf{R}_c^{u|u_a}| + |\mathsf{L}^{u|u_a}|)$. Using this method, we have calculated weekly distribution of actions made by non-admin actors over the admin posts for the case study companies. As an example, such a distribution for Copenhagen Zoo crisis is plotted as shown in Figure 8(a). The other interactions from Table 6 can be defined similarly.

b: COMMENT ARTIFACT ANALYSIS

Like is the only type of interaction that can be performed on a comment artifact. Therefore, we have conducted temporal analysis of like action (by admin vs non-admin actors) on comments made (by admin vs non-admin actors) over the posts (made by admin vs non-admin actors) on a temporal dimension of daily, weekly and yearly as are shown in Table 7.

TABLE 7. Likes on comments by admin/non-admin actors over posts by admin/non-admin actors.

artifact (post) by	artifact (comment) by	action
admin actor	admin	like by admin/non-admin actor
	non-admin	
non-admin actor	admin	
	non-admin	

As one of the possible interactions from Table 7, we define likes by non-admin actors on comments made by non-admin actors over posts by admin-actor as follows.

Let $u_1, u_2 \in \mathsf{U} \setminus u_a$ be the non-admin actors, $r_p, r_c \in \mathsf{R}$ be the post and comment artifacts such that the comment is made on post $(r_p \rhd r_c)$, then
$\mathsf{L}^{u|u|u_a} = \{(u_2, r_c) \in \rightarrow_{like} \mid (u_a, r_p), (u_1, r_c) \in \rightarrow_{post}\}$.

The set $\mathsf{L}^{u|u|u_a}$ indicates likes by non-admin actors (u) on the comments (r_c) made by non-admin actors (u) over the posts (r_p) made by admin actor (u_a).

Similarly, the likes by non-admin actors on the comments made by the admin-actor over the admin posts can be defined as,
$\mathsf{L}^{u|u_a|u_a} = \{(u_1, r_c) \in \rightarrow_{like} \mid (u_a, r_p), (u_a, r_c) \in \rightarrow_{post}\}$.

Using the above methodology, comparison of likes on comments made by admin actor verses non-admin actors over the admin posts for the Jensen Bøfhus company is computed and plotted as shown in Figure 8(c).

2) ACTOR ANALYSIS (SOCIAL SET ANALYSIS)

As part of social set analysis, sets containing unique actors who performed interactions *during* (U_d), *before* (U_b) and *after* (U_a) the crisis period are computed. Let $\mathbf{ts}_d$, $\mathbf{ts}_b$ and $\mathbf{ts}_a$

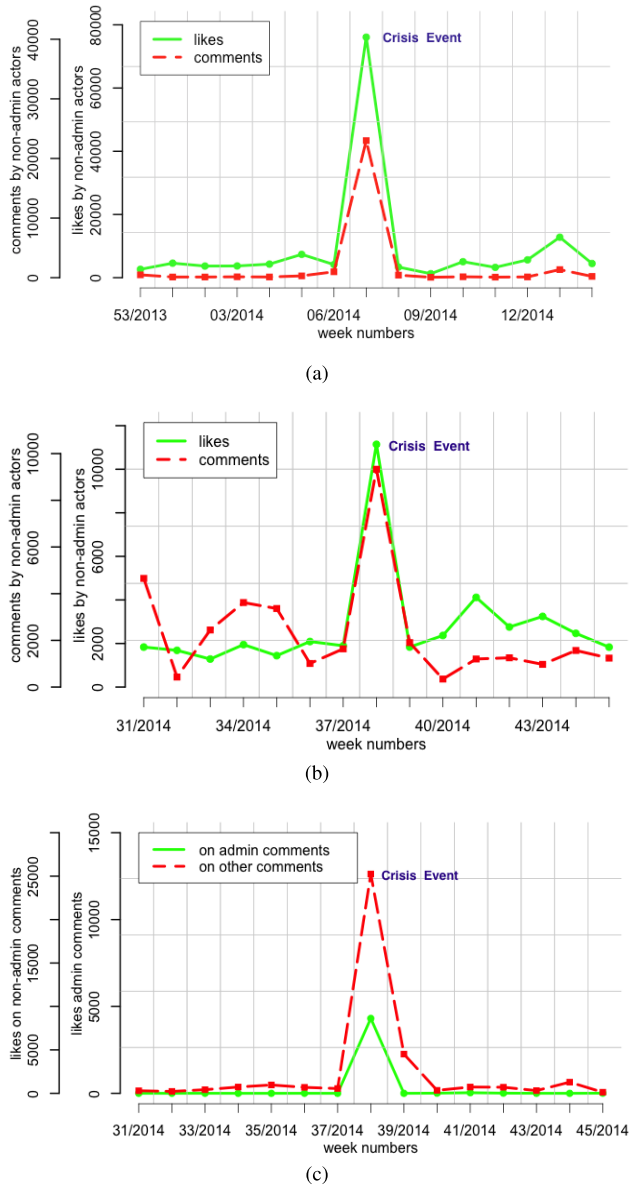

FIGURE 8. Artifact Analysis of Copenhagen Zoo and Jensen Bøfhus Crises [72]. (a) Zoo - comments, likes by non-admin actors on admin posts. (b) Jensen Bøfhus - comments, likes by non-admin actors on admin posts. (c) Jensen Bøfhus - comparison of likes on comments made by admin vs non-admin actors.

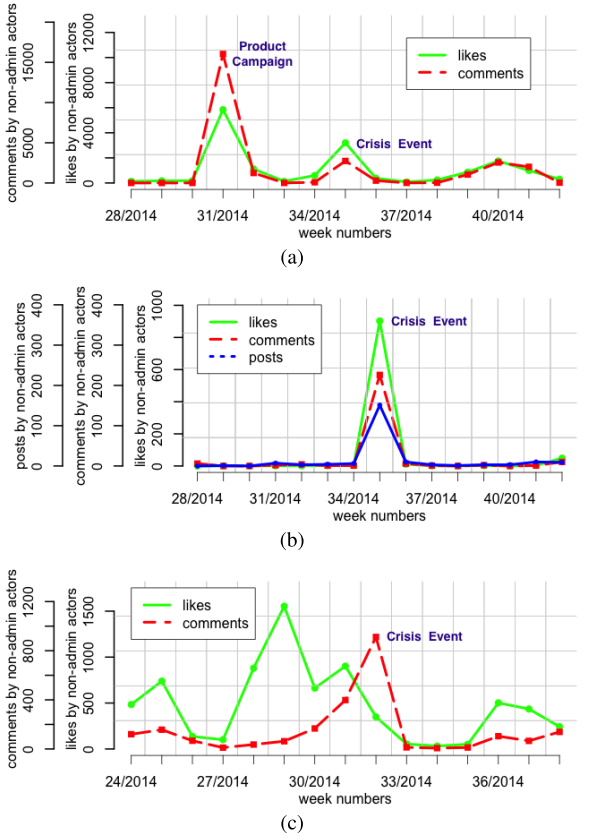

FIGURE 9. Artifact Analysis of Imerco and telenor Crises [72]. (a) Imerco - comments, likes by non-admin actors on admin posts. (b) Imerco - posts, comments, likes by non-admin actors on non-admin posts. (c) Telenor - comments, likes made by non-admin actors on admin posts.

be time spans for *during*, *before* and *after* the crisis respectively containing respective sets of time stamps for those periods. In the social set analysis conducted on the four companies presented in this paper, we observed that the crisis period spans around two weeks on the social media platforms, therefore timespan $\mathbf{ts}_d$ contains time stamps belonging two weeks of the crisis period, where as $\mathbf{ts}_b$, $\mathbf{ts}_a$ contains timestamps belonging to two weeks before the start of the crisis and two weeks after the end of the crisis respectively.

a: ACTORS ANALYSIS FOR CRISIS PERIOD

The *during* (U_d) actors set contains the actors who have either posted or commented or liked an artifact during the crisis period ($\mathbf{ts}_d$), as defined below. Let $ac \in \{post, comment\}$, then

$$\mathsf{U}_d = \{u \mid \exists r \in \mathsf{R}.(u, r) \in \rightarrow_{post} \ \wedge \mathbf{T}(u, r, ac) \in \mathbf{ts}_d\} \cup \{u \mid \exists r \in \mathsf{R}.(u, r) \in \rightarrow_{like} \ \wedge \mathbf{T}(u, r, like) \in \mathbf{ts}_d\}$$

where $\mathbf{T}(u, r, ac)$ and $\mathbf{T}(u, r, like)$ are timestamps of the respective actions. As indicated above the set U_d contains all the unique actors that have performed either a *post*, or a *comment* or a *like* on an artifact during the crisis period. Similarly, the unique actor sets U_b and U_a can computed where the time stamp of the actions belongs to time spans: before ($\mathbf{ts}_b$) and after ($\mathbf{ts}_a$) the crisis period respectively. Finally intersections between actor sets ($\mathsf{U}_d, \mathsf{U}_b, \mathsf{U}_a$) have been computed to represent actor Venn diagrams as shown in Fig. 10. As an example, the set of unique actors who have performed actions only during crisis (neither before nor after) can be computed using the principle of Venn diagram as: $\mathsf{U}_d \cup (\mathsf{U}_d \cap \mathsf{U}_b \cap \mathsf{U}_a) \setminus \big((\mathsf{U}_d \cap \mathsf{U}_b) \cup (\mathsf{U}_d \cap \mathsf{U}_a)\big)$.

3) ACTOR ANALYSIS FOR LIKES ON ADMIN POSTS

The *like* action on a post is an indication of definitive support over the opinion expressed by the post. The sets of unique actors who performed like actions on the posts made by admin actor (u_a) during the crisis period (U_d^l) is computed

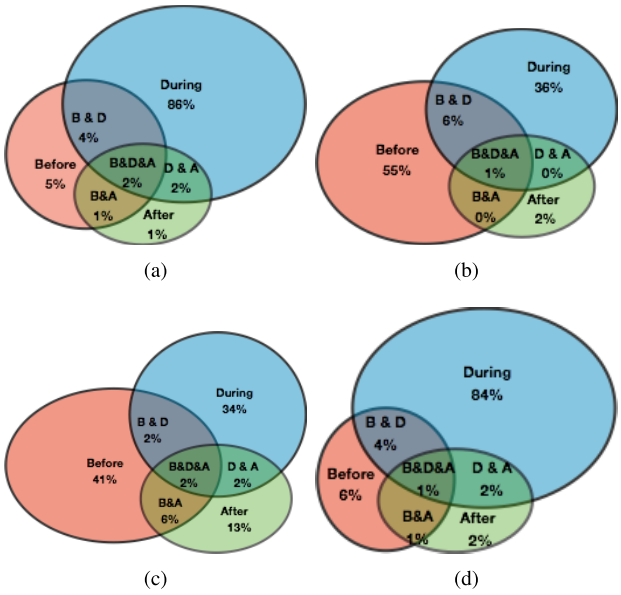

FIGURE 10. Social Set Analysis of actors during crisis. (a) Copenhagen Zoo. (b) Telenor. (c) Jensen's Bøfhus. (d) Imerco.

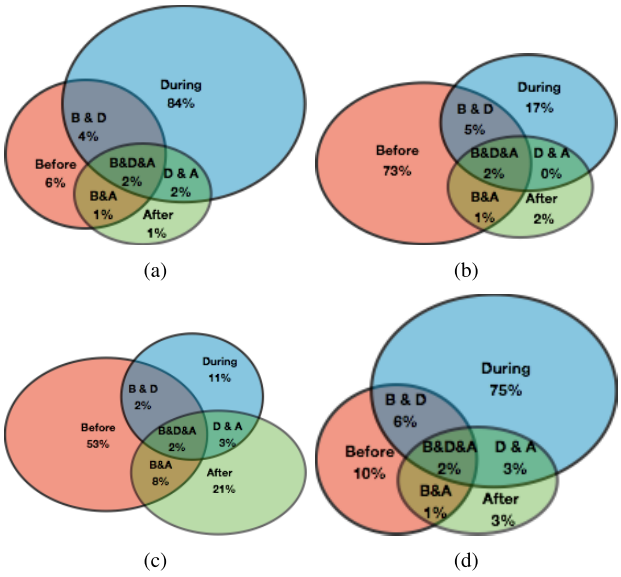

FIGURE 11. Set Analysis of actors who liked admin posts during crisis. (a) Copenhagen Zoo. (b) Telenor. (c) Jensen's Bøfhus. (d) Imerco.

as follows.

$$\mathsf{U}_d^l = \{u \mid \exists r \in \mathsf{R}.(u_a, r) \in \rightarrow_{post} \ \wedge (u, r) \in \rightarrow_{like} \\ and\ \mathbf{T}(u, r, like), \mathbf{T}(u_a, r, post) \in \mathbf{ts}_d\}$$

As defined above the set U_d^l contains the unique actors who have performed like action on the posts made by the admin actor on the Facebook wall of the enterprise. In the similar lines, the set of unique actors who liked the admin posts before (U_b^l) and after (U_a^l) the crisis period can be computed by considering the timestamps belonging to $\mathbf{ts}_b$ and $\mathbf{ts}_a$ time periods respectively. The Venn diagrams representing the sets of unique actors who liked admin posts are computed for four companies and shown in Fig. 11.

4) ACTOR ANALYSIS FOR COMMENTS ON ADMIN POSTS

Unlike the *like*, the *comment* action is not a definitive action in support of the opinion expressed by a post. Therefore, we have computed the sets of unique actors who commented on the posts made by admin actor (u_a) during the crisis period (U_d^c) as follows.

$$\mathsf{U}_d^c = \{u \mid \exists r_p, r_c \in \mathsf{R}.r_p \rhd r_c \wedge (u_a, r_p), (u, r_c) \in \rightarrow_{post} \\ \wedge\ \mathbf{T}(u_a, r_p, post), \mathbf{T}(u, r_c, comment) \in \mathbf{ts}_d\}$$

As shown above, the set U_d^c contains all the unique actors who have commented on the posts made by the admin actor (u_a) during the crisis period. The other two sets: U_b^c, U_a^c containing the unique actors who have commented on the posts made by admin actor (u_a) before and after the crisis can be computed in the similar lines. The Venn diagrams containing the intersections of the actors who have commented on admin posts before, during and after the crisis period can be computed as shown in Fig. 12.

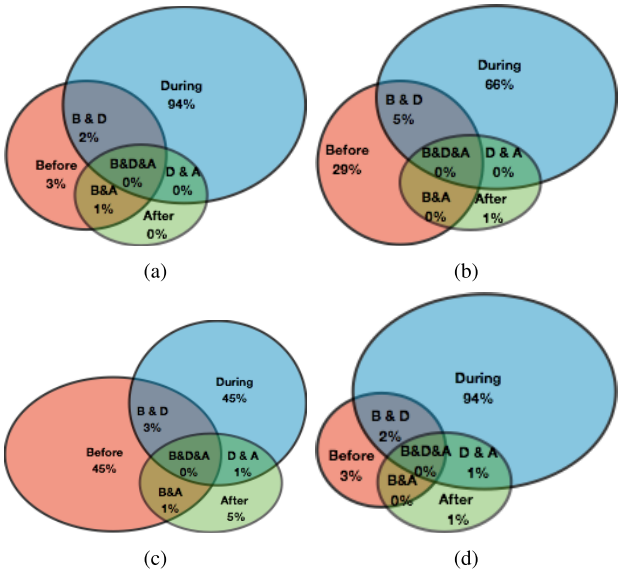

FIGURE 12. Analysis of actors who commented on admin posts during crisis. (a) Copenhagen Zoo. (b) Telenor. (c) Jensen's Bøfhus. (d) Imerco.

C. FINDINGS

In this section, we first present the interactional patterns revealed by the Social Set Analysis and deeper substantive analysis of the social data using netnograpic analysis, manual sentiment analysis and topic discovery.

1) CRISIS DETECTION

Figures 8 and 9 present the results from the temporal analysis of interactions. Figure 8(a) reveals the interactional spikes by non-admin actors on the Copenhagen Zoo's posts as well as an preliminary indication of the nature of the crises. To be specific, the spike of likes on the admin's posts and comments

is an indicator of positive endorsement of the Copenhagen Zoo's activities during the crises. As can be in seen from figure 8(c), in case of Jensen's Bøfhus, the admin comments received far less number of likes when compared to likes on comments made by non-admin users during the interactional peak, which is an indicator of negative endorsement of Jensen's Bøfhus activities.

Thus, we can not only detect the interactional peaks (in this case, known social media crises) but also obtain preliminary indicators of the nature of net user sentiments towards the companies duing the crises. We supplement this with a deeper analysis of users' actions before, during and after the crises, a netnographic analysis of the Facebook walls, and sentiment and topic analysis of the posts and comments during the crises as presented and discussed next.

2) SOCIAL SET ANALYSIS

The analytical objective for conducting Social Set Analysis (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 10), (b) distribution of likes by facebook users on the artefacts (posts and comments) created by the company (Figure 11), and (c) distribution of comments by facebook users on posts created by the company (Figure 12).

As can be seen in Figure 10, a disproportionately high proportion of facebook users only interacted with the facebook walls of Copenhagen Zoo (86%) and Imerco (84%) during the crises period. Even for Telenor (36%) and Jensen's Bøfhus (34%), 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 many more actors interacting during the crises but they stop interacting after the crises has 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 11 shows the temporal distribution of facebook users' likes to the artefacts (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 artefact (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 11, surprisingly high proportion of total likes were received during the crises for the Copenhagen Zoo (84%) and Imerco (75%). 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 12 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 are comparable in for Jensen's Bøfhus (45% and 46%) and Telenor (29% and 66%) whereas Copenhagen Zoo (3% and 94%) and Imerco (3% and 94%) have highly skewed distribution of comments during the before and during periods of the social media crises. Since facebook doesn't have a ''dislike'' button, comments are the only artefact for users to express negative associations, sentiments and expressions (also positive sentiments and expressions). Given the distribution of likes for Copenhagen Zoo and Imerco's posts and comments, the SSA of comments reveals an interesting pattern of higher likes for the company's artefacts as well as higher number of comments.

Taken together, SSA results suggest that the crisis type 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) nethnographic 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 analysis help shed further analytical light on the nature of the crises and the crises communication and management strategies, if any, employed by the companies.

VIII. CASE STUDY 3: SOCIAL SET VISUALIZER: A SET THEORETICAL APPROACH TO BIG SOCIAL DATA ANALYTICS OF REAL-WORLD EVENTS

Event studies is a finance methodology to assess an impact on corporate wealth (e.g. stock prices) caused by events such as restructuring of companies, leadership change, mergers & acquisitions [73]–[75]. It has been a powerful tool since the late 1960s to assess financial impact of changes in corporate policies and used exclusively in the area of investments and accounting to examine stock price performance and the dissemination of new information [76].

While there is no unique structure for event study methodology, at a higher level of abstraction, it contains identifying three important time periods or windows. First, defining an event of interest and identifying the period over which it is active (event window), the second involves identifying the estimation period for the event (pre-event or estimation window) and the final one being identifying the post-event window [75]. In social set analysis of a real-world event, we have applied event study methodology to identify the three important time periods of user interactions on social media platforms: *before* (pre-event window), *during* (event window) and *after* (post-event window).

A. METHODOLOGY

Building on the formal definitions of social data from Sec. V, we further define the notion of a Facebook wall as follows,

Definition 19: With Social Data D, let $\mathbb{W}$ be a set of Facebook walls such that each wall $\mathsf{w} \in \mathbb{W}.\mathsf{w} \in \mathsf{R} \wedge \mathsf{r}_{\mathsf{type}}(w) = \mathit{wall}$.

Definition 20: With Social Data D, we define $\mathbf{match} \subseteq \mathsf{U} \times \mathbb{W}$ as a relation associating actors to walls as follows,

$$\mathbf{match}(u, \mathsf{w}) = \begin{cases} \top & \text{if } (u, \mathsf{w}) \in \rightarrow_{post} \quad (1) \\ \top & \text{if } (-, \mathsf{w}) \in \rightarrow_{post} \wedge (u, \mathsf{w}) \in (\rightarrow_{like} \vee \rightarrow_{share}) \quad (2) \\ \top & \text{if } \exists r.(u, r) \in \rightarrow_{post} \wedge r \rhd \mathsf{w} \quad (3) \\ \top & \text{if } \exists r.(-, r) \in \rightarrow_{post} \wedge r \rhd \mathsf{w} \wedge (u, r) \in (\rightarrow_{like} \vee \rightarrow_{share}) \quad (4) \\ \top & \text{if } \exists r, r'.(-, r), (u, r') \in \rightarrow_{post} \wedge (r \rhd \mathsf{w}) \wedge (r \rhd r') \quad (5) \\ \top & \text{if } \exists r, r'.(-, r), (-, r') \in \rightarrow_{post} \wedge (r \rhd \mathsf{w}) \wedge (r \rhd r') \wedge (u, r') \in \rightarrow_{like} \quad (6) \\ \bot & \text{otherwise} \quad (7) \end{cases}$$

In the def. 20, we define a boolean function **match** that keeps track whether an actor (u) interacted with a Facebook wall (w) or not. It returns true ($\top$), if the actor is the creator of the wall (1), or if he likes the wall (2), or if he posts messages on the wall (3). Similarly making comments on posts (5) or liking or sharing (4) of posts pertaining to wall or even liking a comment will also makes the actor to belongs to a wall as formally explained in Def. 20.

In the analysis, the terms *user* and *actor* are used interchangeably throughout the paper without any difference in semantics.

1) MOBILITY OF ACTORS ACROSS TIME

As part of social set analysis, we have considered three different time frames for an event: before, during and after, which corresponds to pre-event, event and post-event timelines of the event methodology. For an event, sets containing unique actors who performed interactions *during* (U_d), *before* (U_b) and *after* (U_a) are computed. Let $\mathbf{ts}_d$, $\mathbf{ts}_b$ and $\mathbf{ts}_a$ be the sets of time spans for *during*, *before* and *after* periods respectively.

The *during* (U_d) actors set contains the actors who have either posted or commented or liked an artifact in the pre-event time period ($\mathbf{ts}_d$), can be computed as below. Let $ac \in \{post, comment\}$, then

$$\begin{aligned} \mathsf{U}_d = &\{u \mid \exists r \in \mathsf{R}.(u, r) \in \rightarrow_{post} \wedge \mathbf{T}(u, r, ac) \in \mathbf{ts}_d\} \\ &\cup \{u \mid \exists r \in \mathsf{R}.(u, r) \in \rightarrow_{like} \wedge \mathbf{T}(u, r, like) \in \mathbf{ts}_d\} \\ &\cup \{u \mid \exists r \in \mathsf{R}.(u, r) \in \rightarrow_{share} \wedge \mathbf{T}(u, r, share) \in \mathbf{ts}_d\} \end{aligned}$$

where $\mathbf{T}(u, r, ac)$ and $\mathbf{T}(u, r, like)$ are timestamps of the respective actions. As indicated above the set U_d contains all the unique actors that have performed at least either a *post*, or a *comment* or a *like* or a *share* on an artifact during the event period. Similarly, the unique actor sets U_b and U_a can computed easily by replacing the $\mathbf{ts}_d$ with $\mathbf{ts}_b$ and $\mathbf{ts}_a$ in the above equation, where the time stamp of the actions belongs to time spans: before ($\mathbf{ts}_b$) and after ($\mathbf{ts}_a$) the event period respectively. Finally intersections between actor sets ($\mathsf{U}_d, \mathsf{U}_b, \mathsf{U}_a$) are computed using standard set operations. As an example, the set of unique actors who have performed actions only during the event period (neither before nor after) can be computed using the principle of Venn diagram as: $\mathsf{U}_d \setminus \big((\mathsf{U}_d \cap \mathsf{U}_b) \cup (\mathsf{U}_d \cap \mathsf{U}_a)\big)$.

2) MOBILITY OF ACTORS ACROSS SPACE

In social set analysis, mobility across space corresponds to a notion of actors interacting with different Facebook walls. Given a set of Facebook walls (W), actors mobility across space can be computed as follows.

$$\mathsf{U}^{\mathsf{W}} = \{u \mid \forall w \in \mathsf{W}.\mathbf{match}(u, w) = \top\}$$

where U^{W} is the set of actors who have interacted with all the walls in a given set of Facebook walls ($w \in \mathsf{W}$). Mobility across space is useful for analytical purposes in domains ranging from brand loyalty (actors who have visited only one wall) to social activism (actors who might be visiting many walls to express their protest over the companies).

3) MOBILITY OF ACTORS ACROSS TIME AND SPACE

By combining mobility across time and space, we can compute the set of actors that have interacted within a specific time period (e.g. during event), who also have interacted with given set of walls (W) by taking intersection of two sets: $\mathsf{U}_d^{\mathsf{W}} = \mathsf{U}_d \cap \mathsf{U}^{\mathsf{W}}$.

B. TOOL AND CASE STUDY

The garment industry in Bangladesh is the second-largest exporter of clothing after China, and employs more than 3 million - mainly female - workers. This is emphasized by [77] in reference to a large factory fire in Bangladesh at the 25th of November 2012 which killed 112 workers.

The garment industry in Bangladesh has rapidly grown during the past 20 years while approving of lax safety regulations and frequent accidents [78]. "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" [79].

At April 24th, 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 [80], in which more than 1100 garment workers died from the factory's collapse and subsequent fires [81].

TABLE 8. Overview of Facebook dataset of Retail clothing companies.

Facebook Wall	Posts	Comments	Likes	Total
1) Benetton	2,411	51,156	3,760,914	3,814,481
2) Calvin Klein	12,390	44,224	3,196,564	3,253,178
3) Carrefour	3,711	18,651	79,855	102,217
4) E.C. Ingles	21,211	121,684	3,168,950	3,311,845
5) H&M	100,461	262,588	7,779,411	8,142,460
6) JC Penny	24,744	154,620	3,064,581	3,243,945
7) Mango	3,498	204,695	18,661,291	18,869,484
8) Primark	1,343	73,229	1,333,181	1,407,753
9) PVH	66	80	1,801	1,947
10) Walmart	284,523	2,147,994	44,812,653	47,245,170
11) Zara	3,136	12,437	246,294	261,867
Total:	**457,494**	**3,091,358**	**86,105,495**	**89,654,347**

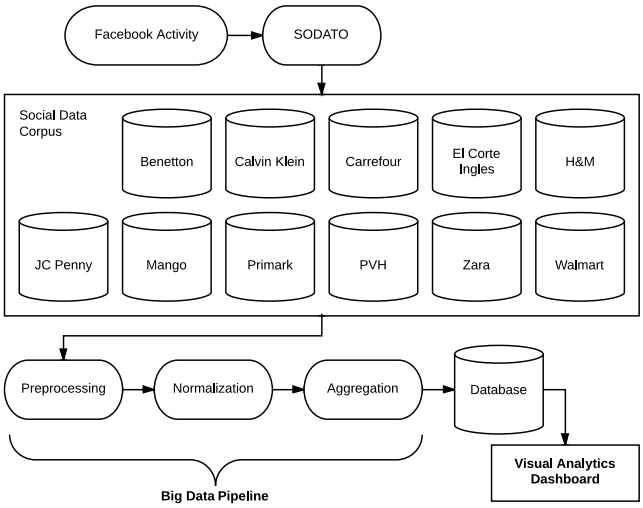

FIGURE 13. Big Data Acquisition Pipeline of the Social Data from Facebook used later on in the Visual Analytics Tool.

This event has been reported by media outlets all over the world and deeply shocked many end consumers of clothing products originating from Bangladesh.

In various research publications, safety and struggles of workers in the Bangladesh garment industry have been widely discussed [79], also with special regard to ongoing protests [82], globalization-related problems [83] and ethical aspects of the factory disasters [84].

Nevertheless, the lack of publicly shown empathy by many major textile industry companies created a public outcry against perceived unethical behavior 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 consumer-facing textile industry brands like *H&M* and *Walmart* to join campaigns supporting textile workers' rights in Bangladesh. A more sustainable, but lagging impact is felt by the introduction of better methods of supply chain management such as social contracts in supply chains [85].

1) METHODOLOGY

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 (as shown in Table. 8) 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 SODATO [8]. 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 using the dashboard. Seventh and last, we deployed the dashboard internally to support ongoing research by CSR researchers and practitioners.

2) DATA COLLECTION & PROCESSING

The event timeline of Bangladesh factory accidents and media reports was collected through desk research including systematic searches in web and media databases. Facebook data was collected through the Social Data Analytics Tool (SODATO) [37], [38], [86]. SODATO-provided Facebook activity datasets are generated as independent files for each company's Facebook wall, and were combined into one for using them as a whole data set that can be filtered or expanded on demand. Figure 13 shows SoSeVi's system schematic for the data acquisition, processing and visualization. The general concept follows the stages of the "Big Data Value Chain" introduced by Miller and Mork [87], with steps of preparation, organization and integration of the data prior to visualization and analysis. Data preparation tasks are performed in a pre-processing step which converts all CSV files to from their character encoding UTF-16 to the more commonly used UTF-8 and handles edge cases in which the generated SODATO output lacks proper data type encapsulation. Subsequently, a data normalization phase performs sanity checks on the input data and identifies malformed data or unneeded information. Lastly, all distinct data sets are aggregated while conserving information regarding their original source in an additional variable. The aggregated data is then imported into a database management system (DBMS), from which it can be accessed for visual analytics purposes.

3) DESIGN

In this section, the design process of the visual analytics dashboard of SoSeVi is outlined.

4) DESIGN GOALS & OBJECTIVES

The visual analytics dashboard has the following design goals.

a: MULTIDIMENSIONALITY

A visual analytics dashboard consists of a mash-up of multiple visualizations which can be utilized by the user in combination to maximize efficiency. The type and size of each visualization need to be carefully evaluated.

b: ACCESSIBILITY

The dashboard should be accessible as easily as possible for users. It should therefore have as few hard dependencies in terms of installed software, operating system or device type as possible.

c: RESPONSIVENESS

The dashboard needs to be responsive to different device types and screen sizes. It should be able to display both on a 4K display used in a conference room and a normal tablet.

d: PERFORMANCE

A key objective for the visual analytics dashboard displays the performance in terms of both server and client side software components. As the dashboard needs to deal with large-scale data sets it should be able to process the data efficiently. In order to achieve higher performance sharing of data processing between server and client software components needs to be established. Thereby, workload may be shifted as needed and user interface waiting times are reduced.

e: EASE OF USE

For end users, ease of use depicts an important non-functional requirement. The visual analytics dashboard should be designed in a way that enables users to work with the dashboard without any prior briefing or training on how to use it.

f: EXTENSIBILITY

Lastly, during realization of the visual analytics dashboard, an extensible framework should be used so that future changes can be implemented with only moderate effort and without unnecessary technical hindrances.

5) DESIGN PRINCIPLES

The design of a visual analytics dashboard such as SoSeVi needs to follow a set of core principles, through which the above stated goals can be achieved. The following design principles are adopted:

a: DETAIL ON DEMAND

The detail on demand principle strives to first present an easily graspable overview to the user, as that it can be processed visually and intellectually in short time. Only subsequently, when the user decides to, the level of detail shown in the visual analytics tool can be increased.

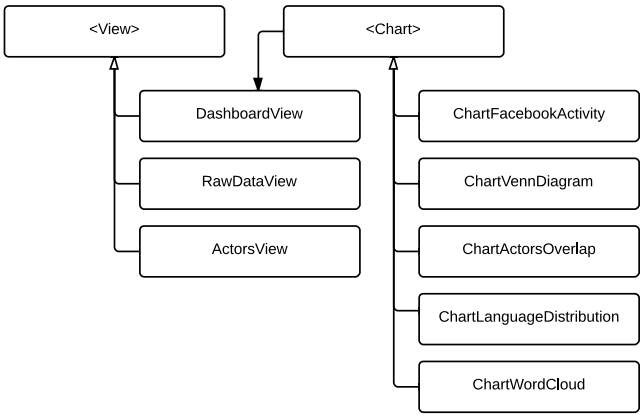

FIGURE 14. Software Architecture.

b: READY-MADE VISUALIZATIONS

The SoSeVi Visual Analytics dashboard is based on social media data from Facebook. The dashboard may consist of a combination of multiple visualizations and each visualization needs to highlight unique features of the underlying social interactions between actors and artifacts. This allows the dashboard as a whole to be kept clean and organized, preventing it from becoming too complex.

c: USER-CENTRIC DESIGN (UCD)

Reference [88] emphasizes that in user-centric design, "the role of the designer is to facilitate the task for the user and to make sure that the user is able to make use of the product as intended and with a minimum effort to learn how to use it". When designing the interface, a focus is put on optimization of the user experience.

6) SoSeVi: VISUALIZATION FRAMEWORK

The technology choice for realizing the dashboard visualizations is the D3.js Javascript-based visualization framework which uses dynamic SVG images for data visualization. D3.js constitutes a lightweight and very extendable Javascript visualization framework which can display visualizations for a multitude of browser-based clients. The flexibility provided by D3.js enables the creation of new kinds of interactive visualizations which are able to run on any device with decent processing resources including Windows, MacOS and Linux based systems with screen sizes up to 4K devices.

Figure 14 presents the software architecture of SoSeVi. DashboardView is the main view of the web application which contains the SoSeVi and is initially shown to the user. RawdataView presents a detailed search interface for the Facebook activity data. Many visualizations in DashboardView refer to RawdataView in order to provide the user with further information. ActorsView presents a dedicated interface for analysis tasks related to Actor Mobility across time and space of companies' facebook walls. The visualizations of actor mobility in DashboardView refer to ActorsView in order to provide the user with further details when requested.

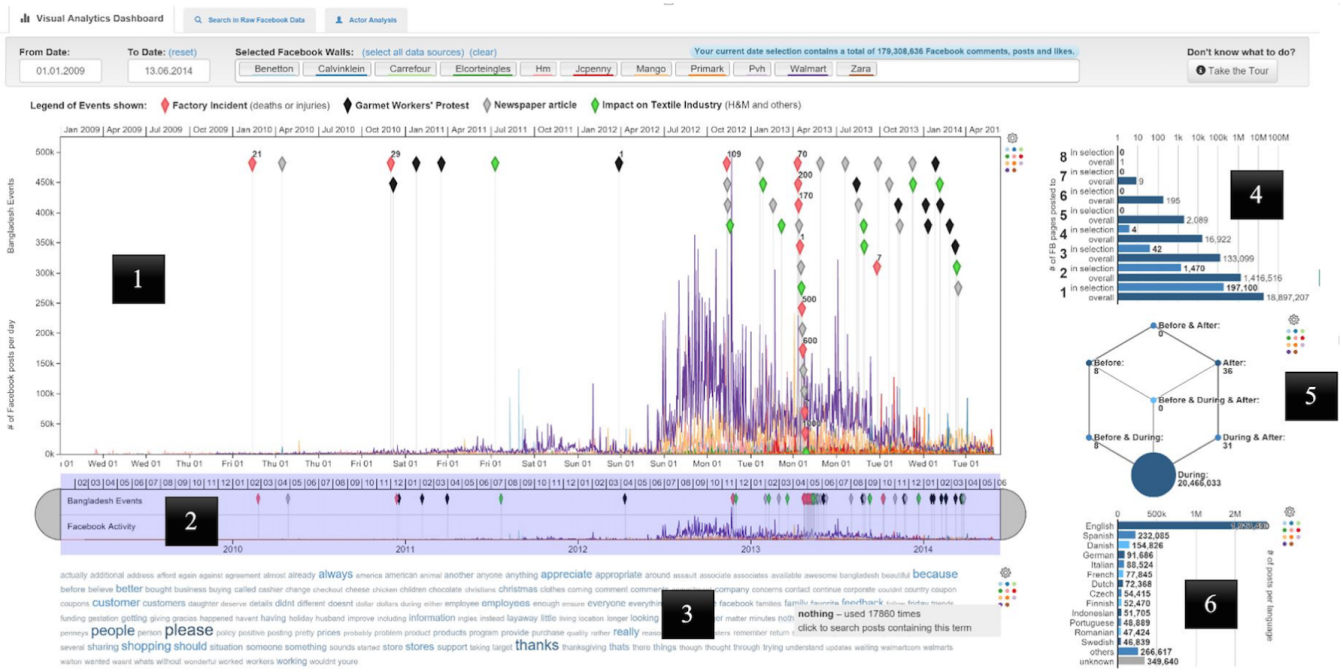

FIGURE 15. Social Set Visualizer: For the selected time period (see date range fields in top-left) and selected facebook walls (see colour coded selection bubble chart next to visualizations) [1] Facebook Activity Chart; [2] Timeline of Bangladesh Factory Accidents & Facebook Actions; [3] Word Cloud 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.

Furthermore, ActorsView presents a handy set of tools for analysis of actor mobility and cross-postings between different time frames and Facebook walls.

7) DEVELOPMENT OF SoSeVi: DASHBOARD INTERFACE

Figure 15 presents the SoSeVi dashboard and its constituent visualizations for the full dataset.

The *Facebook activity visualization* displays the social media activity on Facebook over the whole time period. It consists of a large main chart and a smaller mini chart underneath. Both charts use a line plot to display activity. The mini-chart can be used as a brush to change the time period of the data shown in the main chart. The *Actor Mobility across Space visualization* at the top right of the dashboard displays the number of different Facebook walls on which Actors have posted. For this visualization, a bar chart is used. The chart depicts the number of Actors based on the number of Facebook walls they have posted to. The *Actor Mobility across Time visualization* at the center right of the dashboard displays the number of Actors within each time period and their respective overlaps. For this visualization, an exploded Venn diagram is used which is aligned hexagonally. The *Language Distribution visualization* at the bottom right of the dashboard displays the number of social media Artifacts based on their language. For this visualization, a bar chart is used. It presents each language and the respective number of social media activities during the selected timeframe. The *Word Cloud visualization* located right beneath the Facebook Activity chart displays the results of the word frequency analysis based on conversation artifacts in the available social data. The font size of each word is determined by its overall frequency within all conversations that happened during the selected time period.

A Legend for the event timeline is placed at the very top of the dashboard between the user-driven filtering interface and the Facebook Activity visualization. It conveys information about different types of events which are part of the event timeline. In the case at hand, the event timeline is based on the Bangladesh factory disaster events, which means that the event types classified are encoded in the legend.

The user-driven filtering interface contains two components. On the left hand side, the user may input start and end dates of the timeframe to be visualized in the dashboard. Mouse or touch interactions with the input fields will reveal a hidden date picker component. This date picker enables the user to either input dates using a keyboard or specifying the day, month and year using their mouse or even a touch screen. Secondly, on the right hand side, the user may select the companies whose Facebook walls are shown in the dashboard. User interaction with the available input field can be performed in various ways. The user can directly type Facebook walls into the field, which are then displayed in the visualization. An alternative method is that the user selects an item from a drop down menu that ap-pears when the input field is focused.

To summarize, the SoSeVi big data visual analysis dashboard empowers users to use it in different ways. The dashboard adheres to the user's preferred interaction

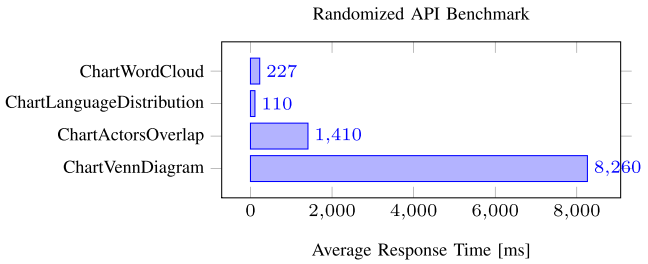

FIGURE 16. Performance Benchmark of four API Endpoints.

method without making any assumptions. This means tablet users may also type in their selection of the Facebook walls, or desktop users may use the Datepicker to manually select a date. The dashboard may be accessed at http://5.9.74.245:3000/, access credentials will be provided to the research community upon request.

8) EVALUATION

Benchmarking: Figure 16 displays benchmarking results of the dashboard's underlying API. The results underline the varying complexity in calculating data needed for the visualizations of different event windows. According to the presented benchmark, visualizations of conversation content (*ChartLanguageDistribution* and *ChartWordCloud*) are much faster calculated and presented to the dashboard user than visualizations of actor mobility (ChartVennDiagram, ChartActorsOverlap). This can be explained by the fact that visualizations of Actor Mobility need to take each single actor into account, whereas visualizations of conversation content have access to much better speed improvements through precalculated datasets which derive from the main dataset. Due to the bad benchmark results of *ChartVennDiagram*, and a general discrepancy in performance, further optimizations are performed to the database as described further.

Query Optimization: When using a RDBMS such as PostgreSQL in big data analytics, many opportunities for increased performance can be realized through query optimization. The systematic optimization of slow database queries is demonstrated on the visualization of Language Distribution. All optimizations are benchmarked against the initial query in order to assess their effectiveness. The benchmarking process follows a strict methodology, in which each query will be executed $n = 10$ times and query execution time is logged. Then, the average execution time is used to decide on the feasibility of the optimization at hand. If the average execution time is reduced, the optimization step will be applied to the query. The optimization process may be repeated until sufficient reduction of the average query execution time is reached. Out of all queries, the language distribution query was identified as a very slow query and therefore we have performed optimizations on it. The initial query is displayed in Listing 1. It returns 24 rows after an execution time of approximately 10 seconds, which is way slower than the users' anticipated loading time of a visual analytics dashboard. Based on the precalculation of as much data as possible and separating this data into its own database table, we optimized the performance of the query as shown in Listing 2.

```
SELECT lang, COUNT(*) as count
FROM fbdata WHERE eventname != 'LIKE' AND
  "date" BETWEEN '2009-01-01' AND '2014-06-12'
  AND source in ('carrefour', 'walmart')
GROUP BY lang ORDER BY count DESC;
```

Listing 1. Initial Query for Language Distribution .

```
CREATE TABLE fbdata_language_distribution AS
  SELECT date, source, lang, count(*) as count
  FROM fbdata GROUP BY date, source, lang
  ORDER BY date, source, lang ASC;
SELECT lang, sum(count) as count
FROM fbdata_language_distribution
WHERE "date" BETWEEN '2009-01-01' AND '2014-06-12'
  AND source in ('carrefour', 'walmart')
```

Listing 2. First Optimization of Language Distribution Query .

This performance improvement of the database query shown in listing 2 is based on the fact that the new query does not need to access the much larger *fbdata* table, but only uses a small subset which is available in the derived table. In the second round, further optimization are performed on the query in listing 2 by creating indexes on suitable columns such as datetime and others. After creation of the indexes on the derived table, the performance of the query is increased marginally as shown in Fig. 17. A performance improvement of 300 times was realized in the first optimization, whereas the second optimization step yielded only a 1.77 times improvement.

9) SELECTED EMPIRICAL FINDINGS FROM SoSeVi

Due to space restrictions, we present only a subset of the empirical findings resulting from the use of the Social Set Visualizer (SoSeVi) tool by researchers and practitioners in the field of Corporate Social Responsibility (CSR). These empirical findings demonstrate the analytical utility of our proposed set theoretical approach to big social data and our social set analysis approach to visual analytics dashboards. The following points outline some of the key issues that were investigated using the SoSeVi:

1) The global supply chain concerns with regard to Bangladesh garment factories have been expressed by Facebook users from as far back as 2009
2) With regard to Conversation analysis of big social data, the distribution of the keyword ''bangladesh'' across time and space of 11 different Facebook walls is proportional to the severity of fatalities in Bangladesh garment factories and peaks for the Rana Plaza disaster that killed more than 1100 factory workers.
3) With regard to Interaction analysis of big social data, in terms of actors, SoSeVi helped identify the most influential negative critics as well as positive advocates for each of the 11 companies before, during, and after the maximum accident density time period

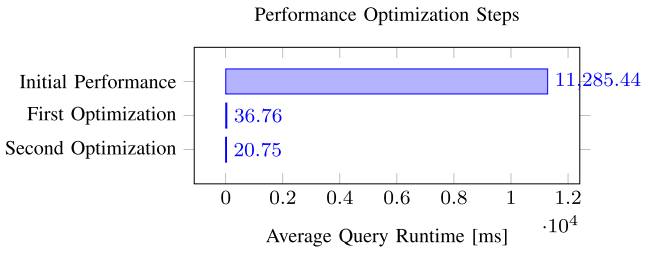

Figure 17. Three-Step Performance Optimization of the Language Distribution Visualization.

4) There are many instances of authentic displays of support and expressions of empathy from Facebook users as well as robotic incidents of slacktivism
5) Surpisingly, majority usage the keyword *please* was with respect to opening of new stores in the case of H&M
6) 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)
7) Companies followed not only different CSR strategies but also different social media strategies before, during and after the Bangladesh garment factory accidents. Further, companies adopted different crises communication and management strategies yielding different outcomes
8) For almost all of the accidents, a majority of the users are posting during the news cycle, e.g. the coverage of the event through traditional media outlets, and they do not return to the Facebook walls again. This emphasizes that social media engagement during factory accidents is episodic and burst-y with little overlap to the *business-as-usual* period before or after the accident.

10) REFLECTIONS ON THE IT-ARTIFACT

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, the data at hand presents numerous opportunities for attaining deep insights. In this context, visual analytics present the means of reaching those insights to many users with different backgrounds, both experts and novices alike. The novel implementation of the present Social Set Visualizer (SoSeVi) dashboard showcases that the creation of visual analytics software, which meets the high technical, analytical and user experience requirements of present-day computing, is viable (and can be achieved by an academic research group with limited resources). Furthermore, the developed IT artifact leverages open-source visual analytics frameworks to maximum extent in order to achieve a pure implementation of important concepts in visual analytics.

IX. DISCUSSION

In this paper we have presented an new approach for analysis of big social data using a conceptual model of social data, a set-theoretical formalisation of the conceptual model and an analytical framework called ''Social Set Analysis''. The set-theoretical formalization of the conceptual model provides the necessary abstraction to comprehend the complex and complicated interactional scenarios and conversational contexts of big social data. Further, the formal model informed the schematic model of the software application and helped realise the abstract ideas from the conceptual model into the analytical framework for Social Set Analysis. We now briefly discuss the methods of and findings from the three illustrative case studies.

A. CASE STUDY 1: FUZZY-SET BASED SENTIMENT ANALYSIS

In the first case study (Sec. VI), we have presented an integrated modelling approach for analysis of big social data with the sentiment analysis based on the Fuzzy set theory. We have presented a method for profiling of artifacts and actors and applied this technique to the analysis of big social data collected from Facebook page of the fast fashion company, H&M. Regarding formal modelling of temporal dimensions of social media interactions, we are currently developing a hybrid approach by constructing crisp sets as well as fuzzy sets. For example, given an event of of analytical interest such as a marketing campaign, we construct crisp sets of sentiment categories (positive, negative, & neutral) for actors and artefacts and fuzzy sets of the interactional time-periods (before_event, during_event, and after_event). This allows us to model and analyze the different user characteristics, behaviours, and dynamics within the intersections and unions of the temporal categories of before-the-event, during-the-event, and after-the-event at analyst determined fuzzy set membership criteria for sentiment categories.

We acknowledge that many works exist in fuzzy sentiment analysis and social networks and we have cited relevant papers in the related work section. But as stated before, our approach primarily differs from the current approaches of social network analysis based on relational sociology. Our approach is based on associational sociology, where we focussed on finding ''association-ship'' among actors and artifacts, based on set theoretical approach, rather than only focussing on the relationship between the actors. Our approach of ''associational sociology'' is drawn from Latour's ([25]) term ''sociology of associations''. We postulate that Set Theory in general and Fuzzy Set Theory is well-suited from sociological and mathematical standpoints to model human associations [27]. Beyond the immediate social network and particularly on large scale social media platforms such as Facebook, twitter and Tencent QQ, we believe that this fundamental change in the foundational mathematical logic of the formal model of social data from graphs to sets will generate new insights. This paper is a first

attempt to articulate such an alternate integrated approach across the theoretical, conceptual, formal and computational realms.

B. CASE STUDY 2: SOCIAL SET ANALYSIS OF CORPORATE SOCIAL MEDIA CRISES ON FACEBOOK

In the second case study (Sec. VII), we first proposed a set-theoretical formal method for social set analysis drawn from the event-study framework to investiage corporate social media cries on Facebook. The proposed method was then applied to big social data for four different social media crises. Event studies is a finance methodology to assess an impact on corporate wealth (e.g. stock prices) due to events such as restructuring of companies, leadership change, mergers & acquisitions [73]–[75]. It has been a powerful tool since late 1960s to assess financial impact of changes and continues to be used extensively to examine stock price performance and the dissemination of new information [76]. While there is no unique structure for event study methodology, at an higher level of abstraction, it contains identifying three important time periods or windows. First, defining an event of interest and identify the period over which it is active (event window), the second involves identifying the estimation period for the event (pre-event or estimation window) and the final one being identifying the post-event window [75]. In social set analysis of social media crises, we have adopted the event study methodology to identify the three important time periods of user interactions on social media platforms: *before* (pre-event window), *during* (event window) and *after* (post-event window). SSA results showed the voluminous but also transient nature of interactions during the social media crises and a diversity of aggregate user behavioural patterns. SSA combined with netnography and content analysis in terms of sentiment analysis and topic discovery revealed the different strategies employed by the organizations to manage the crises and their outcomes.

C. CASE STUDY 3: SOCIAL SET VISUALIZER: A SET THEORETICAL APPROACH TO BIG SOCIAL DATA ANALYTICS OF REAL-WORLD EVENTS

In the third case study (Sec. VIII), we designed, developed and evaluated a visual analytics tool, SoSeVi(Social Set Visualiser). SoSeVi incorporated new set theoertical visualizations of big social data in terms of actor mobility across space (wall crossings) and actor mobility across time (before, during and after). SoSeVi leverages open-source visual analytics frameworks to maximum extent in order to achieve a pure implementation of important concepts in visual analytics such as the detail on demand principle. A thorough evaluation showcased the effectiveness of the tool's approach on visual analytics. Both the client- and a server-side components of the Visual Analytics Dashboard present performance at par with commercial tools, and can seamlessly be used under many circumstances. Additionally, the results of the user study performed indicate that the presented Visual Analytics dashboard combines a high ease of use with the ability of performing many different interactive analyses on a large dataset. Moreover, the Visual Analytics tool put forward may be utilized through any modern browser on a multitude of different devices and screen sizes, with backend response times as low as in the hundreds of milliseconds. Complementing benchmarks of the database optimizations applied to the Visual Analytics dashboard in real-world deployments showcase a good performance and satisfactory handling of large amounts of social data.

D. REFLECTIONS ON THE SET-THEORETICAL APPROACH

We now briefly discuss the current adoption of and future prospects for the set-theortical approach with regard to Social Science, Computer Science, and Computational Social Science.

1) SET THEORY AND SOCIAL SCIENCE

Recent advancements in set theory and readily available software have enabled social science researchers to bridge the variable-centered quantitative and case-based qualitative methodological paradigms in order to analyze multi-dimensional associations beyond the linearity assumptions, aggregate effects, unicausal reduction, and case specificity. In the social sciences, application of set theory has seen a dramatic increase over the last decade which can be attributed to the method called Qualitative Comparative Analysis [89] developed by the political scientist Charles Ragin [90], [91]. Qualitative Comparative Analysis (QCA) models causal relations as subset or superset relations corresponding to necessity and sufficiency conditions. QCA aims to derive casually complex patterns in terms of equifinality, conjunctural causation and asymmetry [90], [92], [93]. Although developed initially by Ragin [91] for qualitative case study researchers (medium sample size of $n < 90$), the proponents and supporters of QCA have argued about its unique advantages over regression-based approaches [93], [94] and its application for analysis of large-N datasets. In the adoption of set theoretical methods in social sciences [89] three variants of QCA methodology have surfaced: crisp-set QCA(CsQCA), fuzzy-set QCA (fsQCA) [90] and multi-set QCA (MvQCA) [93] with a number of software tools supporting set-theoretical social science research (e.g. R packages like QCA and QCAPro, fs/QCA, Tosmana).

2) SET THEORY AND COMPUTER SCIENCE

In order to further systematic research on set-theoretical algorithms, data structures and programs in Computer Science, we envision *Computational Set Analysis* as a research program. In this regard, the SetVR workshop series[6] augurs well for the formalisation and computational implementation of set-theoretical reasoning and visualisations. In terms of visual analytics, recent advancements with regard to set intersections include the *Upset* project [95] on the visualizations of set intersections based on innovative approaches

[6]SetVR workshop: https://sites.google.com/site/setvr2kn/current-workshop.

to combination matrices and the *Euler Diagrams* project on creating area-proportional Euler diagrams using ellipses instead of the traditionally used circles [96].

3) SET THEORY AND COMPUTATIONAL SOCIAL SCIENCE

As discussed in the Conceptual Framework section, set-theoretical approaches big social data analytics hold several advantages in terms of modelling the implicit vagueness of many social science concepts and combining the strengths and addressing the weakness of variable vs. case based empirical approaches in social science research. For example, automated sentiment annotation of social data artifacts based on computational linguistics methods such as supervised machine learning produce both classifications of tokens into types (such as positive, negative and neutral) as well as probabilistic estimates. As we have demonstrated in Case Study 1, these classifications and probabilities can be modelled using Crisp and Fuzzy Set theories respectively and analyzed to reveal historical developmental patterns as well as overlapping categories. Practical implications from the analysis could help inform an organization to assess the size of the different actor sets (sub-communities) such as entirely positive, partially positive, entirely negative etc. Investigating the absolute and relative size of entirely negative conversations might enable the organization to identify the underlying customer service issues and/or content problems. Similarly, knowing the absolute and relative number of social media users that exclusively express positive sentiments towards the organization helps identify and nurture the advocacy group.

E. LIMITATIONS

One of this paper's limitations is that we do not present domain-specific social science findings in terms of visual analytics, crisis communication, crisis management, labor rights, industrial safety and/or corporate social responsibility. That said, first attempts at domain-specific empirical findings of the set-theoretical approach can be found in [72] and [97]. A second limitation is the lack of exposition of the full range of set theory beyond the classical crisp sets and fuzzy sets discussed in the paper (for example: Rough sets, Random sets, Bayesian sets). A third and final limitation is the limited space given to the computational aspects of the visuala analytics tool, SoSeVi.

F. FUTURE RESEARCH

Current and planned future work in our computational social sciences laboratory is addressing some of the theoretical limitations identified above. In particular, we are exploring novel set-theoretical visualisations of large number of set interestions and to indicate set migrations of actors across space and time with a focus on dynamic set composition and decomposition. In terms of formal models and analytical methods, we are extending Social Set Analysus to include Rough and Random sets. Furthermore, we plan to release a software library for ''Social Set Analysis'' that will allow researchers and practitioners to easily integrate set-theoretical analytics into their Big Data Analytics workbenches.

X. CONCLUSION

In conclusion, one of the contributions of this paper is to demonstrate the suitability and effectiveness of Social Set Analysis for conceptualizing, formalizing and analyzing big social data from content-driven social media platforms like Facebook for event studies such as unexpected crises and/or coordinated marketing campaigns. 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, user actions on various organisation's Facebook walls, Big Social Data presents numerous opportunities for attaining deep insights. As illustrated by the three case studies above, SSA covers the range of prescriptive, visual, and descriptive analytics. Taken together, the three demonstrative case studies illustrate the viability of Social Set Analysis as a holistic approach to Computational Social Science in general and Big Data Analytics in particular.

As part of future work, we would like to extend the Fuzzy Set Theoretical formal model to encompass modelling of networks of groups and friends of users in an online social media platform. We also have plans to extend the Fuzzy Set methods and techniques to other kinds of socio-technical interactions and further develop our abstract formal model. Modelling social concepts in general involves fuzziness and we would like to use Fuzzy set theory to model fuzzy behaviour in the social data.

ACKNOWLEDGMENTS

The authors thank the members of the Computational Social Science Laboratory (http://cssl.cbs.dk) for their valuable feedback. Thanks to the master theses and course project students that have helped in the design, development, use and evalaution of the methods and tools.

Any opinions, findings, interpretations, conclusions or recommendations expressed in this paper are those of its authors.

REFERENCES

[1] R. E. Montalvo, ''Social media management,'' *Int. J. Manage. Inf. Syst.*, vol. 15, no. 3, pp. 91–96, 2011
[2] C. Vollmer and G. Precourt, *Always On: Advertising, Marketing, and Media in an Era of Consumer Control (Strategy + Business)*. New York, NY, USA: McGraw-Hill, 2008.
[3] A. McAfee, *Enterprise 2.0: New Collaborative Tools for Your Organization's Toughest Challenges*. Boston, MA, USA: Harvard Business Press, 2009.
[4] R. Vatrapu, ''Understanding social business,'' in *Emerging Dimensions of Technology Management*. India: Springer, 2013, pp. 147–158.
[5] W. S. Cleveland, ''Data science: An action plan for expanding the technical areas of the field of statistics,'' *Int. Statist. Rev.*, vol. 69, no. 1, pp. 21–26, 2001. [Online]. Available: http://dx.doi.org/10.1111/j.1751-5823.2001.tb00477.x
[6] M. Loukides, *What Is Data Science?* Sebastopol, CA, USA: O'Reilly Media, 2012.

[7] N. Ohsumi, "From data analysis to data science," in *Data Analysis, Classification, and Related Methods*. Berlin, Germany: Springer, 2000, pp. 329–334. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-59789-3_52

[8] D. Lazer *et al.*, "Computational social science," *Science*, vol. 323, no. 5915, pp. 721–723, 2009.

[9] J. Sterne, *Social Media Metrics: How to Measure and Optimize Your Marketing Investment*. New York, NY, USA: Wiley, 2010.

[10] M. Sponder, *Social Media Analytics: Effective Tools for Building, Interpreting, and Using Metrics*. New York, NY, USA: McGraw-Hill, 2011.

[11] Z. Tufekci. (2014). "Big questions for social media big data: Representativeness, validity and other methodological pitfalls." [Online]. Available: http://arxiv.org/abs/1403.7400

[12] C. Cioffi-Revilla, *Introduction to Computational Social Science: Principles and Applications*. London, U.K.: Springer, 2013.

[13] R. Vatrapu, R. R. Mukkamala, and A. Hussain, "Towards a set theoretical approach to big social data analytics: Concepts, methods, tools, and empirical findings," in *Proc. Int. Conf. Social Media Soc. (SMSociety)*, 2014

[14] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, vol. 8. Cambridge, U.K.: Cambridge Univ. Press, 1994.

[15] M. Emirbayer, "Manifesto for a relational sociology," *Amer. J. Sociol.*, vol. 103, no. 2, pp. 281–317, 1997.

[16] R. Vatrapu, A. Hussain, N. B. Lassen, R. R. Mukkamala, B. Flesch, and R. Madsen, "Social set analysis: four demonstrative case studies," in *Proc. Int. Conf. Soc. Media Soc.*, 2015, p. 3.

[17] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, "Network analysis in the social sciences," *Science*, vol. 323, no. 5916, pp. 892–895, 2009.

[18] J. L. Gross and J. Yellen, *Graph Theory and Its Applications*. Boca Raton, FL, USA: CRC Press, 2005.

[19] M. S. Mizruchi, "Social network analysis: Recent achievements and current controversies," *Acta Sociol.*, vol. 37, no. 4, pp. 329–343, 1994.

[20] R. Vatrapu, R. R. Mukkamala, and A. Hussain, "Set theoretical approach to big social data analytics: Concepts, methods, tools, and findings," in *Proc. Comput. Social Sci. Workshop Eur. Conf. Complex Syst. (ECSS)*, 2014.

[21] F. Cusset, *French Theory: How Foucault, Derrida, Deleuze, & Co. Transformed the Intellectual Life of the United States*. Minneapolis, MN, USA: Univ. Minnesota Press, 2008.

[22] I. Hacking, *The Social Construction of What?* Cambridge, MA, USA: Harvard Univ. Press, 1999.

[23] D. Boyd and K. Crawford, "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon," *Inf., Commun. Soc.*, vol. 15, no. 5, pp. 662–679, 2012.

[24] J. Lovett, *Social Media Metrics Secrets*, vol. 159. New York, NY, USA: Wiley, 2011.

[25] B. Latour, *Reassembling the Social: An Introduction to Actor-Network-Theory*. New York, NY, USA: Oxford Univ. Press, 2005.

[26] C. C. Ragin, *Fuzzy-Set Social Science*. Chicago, IL, USA: Univ. Chicago Press, 2000.

[27] R. R. Mukkamala, A. Hussain, and R. Vatrapu, "Towards a set theoretical approach to big data analytics," in *Proc. 3rd IEEE Int. Congr. Big Data*, Jun./Jul. 2014, pp. 629–636. [Online]. Available: http://www.itu.dk/people/rao/pubs_accepted/2014_IEEE-BigData-socialdata-set-theory.pdf

[28] M. J. Smithson and J. Verkuilen, *Fuzzy Set Theory: Applications in the Social Sciences* (Quantitative Applications in the Social Sciences). New York, NY, USA: SAGE Publications, Feb. 2006. [Online]. Available: http://www.worldcat.org/isbn/076192986X

[29] A. Kechris, *Classical Descriptive Set Theory*, vol. 156. New York, NY, USA: Springer-Verlag, 2012.

[30] R. K. Vatrapu, "Technological intersubjectivity and appropriation of affordances in computer supported collaboration," Ph.D. dissertation, Dept. Commun. Inf. Sci., Univ. Hawaii Manoa, Honolulu, HI, USA, 2007.

[31] R. K. Vatrapu, "Towards a theory of socio-technical interactions," in *Learning in the Synergy of Multiple Disciplines*. Berlin, Germany: Springer, 2009, pp. 694–699.

[32] R. K. Vatrapu, "Explaining culture: An outline of a theory of socio-technical interactions," in *Proc. 3rd Int. Conf. Int. Collaboration (ICIC)*, 2010, pp. 111–120.

[33] J. J. Gibson, *The Ecological Approach to Visual Perception*. Boston, MA, USA: Houghton Mifflin, 1979.

[34] A. Noë, *Action in Perception*. Cambridge, MA, USA: MIT Press, 2004.

[35] A. Schutz, *The Phenomenology of the Social World*. Evanston, IL, USA: Northwestern Univ. Press, 1967.

[36] H. Garfinkel, *Studies in Ethnomethodology*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1967.

[37] R. Vatrapu, A. Hussain, D. Hardt, and Z. Jaffari, "Social data analytics tool: A demonstrative case study of methodology and software," in *Analyzing Social Media Data and Web Networks*, M. Cantijoch, R. Gibson, S. Ward, Eds. Basingstoke, U.K.: Palgrave Macmillan, 2014.

[38] A. Hussain and R. Vatrapu, "Social data analytics tool (SODATO)," in *Advancing the Impact of Design Science: Moving from Theory to Practice* (Lecture Notes in Computer Science). Switzerland: Springer, 2014.

[39] M. Kryszkiewicz, "Rough set approach to incomplete information systems," *Inf. Sci.*, vol. 112, nos. 1–4, pp. 39–49, 1998.

[40] N. M. Tichy, M. L. Tushman, and C. Fombrun, "Social network analysis for organizations," *Acad. Manage. Rev.*, vol. 4, no. 4, pp. 507–519, Oct. 1979.

[41] D. Krackhardt, "Cognitive social structures," *Soc. Netw.*, vol. 9, no. 2, pp. 109–134, Jun. 1987.

[42] J. Zhan and X. Fang, "Social computing: The state of the art," *Int. J. Soc. Comput. Cyber-Phys. Syst.*, vol. 1, no. 1, pp. 1–12, 2011.

[43] J. Karikoski and M. Nelimarkka, "Measuring social relations with multiple datasets," *Int. J. Soc. Comput. Cyber-Phys. Syst.*, vol. 1, no. 1, pp. 98–113, 2011.

[44] J. Sabater and C. Sierra, "Reputation and social network analysis in multiagent systems," in *Proc. 1st Int. Joint Conf. Auton. Agents Multiagent Syst. (AAMAS)*, 2002, pp. 475–482.

[45] M. Goldberg, S. Kelley, M. Magdon-Ismail, K. Mertsalov, and A. Wallace, "Finding overlapping communities in social networks," in *Proc. IEEE 2nd Int. Conf. Soc. Comput. (SocialCom)*, Aug. 2010, pp. 104–113.

[46] O. Macindoe and W. Richards, "Comparing networks using their fine structure," *Int. J. Soc. Comput. Cyber-Phys. Syst.*, vol. 1, no. 1, pp. 79–97, 2011.

[47] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retr.*, vol. 2, nos. 1–2, pp. 1–135, 2008.

[48] C. R. Fink, D. S. Chou, J. J. Kopecky, and A. J. Llorens, "Coarse-and fine-grained sentiment analysis of social media text," *Johns Hopkins APL Tech. Dig.*, vol. 30, no. 1, pp. 22–30, 2011.

[49] M. Grassi, E. Cambria, A. Hussain, and F. Piazza, "Sentic Web: A new paradigm for managing social media affective information," *Cognit. Comput.*, vol. 3, no. 3, pp. 480–489, 2011.

[50] T. Nguyen, D. Phung, B. Adams, and S. Venkatesh, "Prediction of age, sentiment, and connectivity from social media text," in *Web Information System Engineering*. Berlin, Germany: Springer, 2011, pp. 227–240.

[51] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede, "Lexicon-based methods for sentiment analysis," *Comput. Linguistics*, vol. 37, no. 2, pp. 267–307, 2011.

[52] T. Wilson, J. Wiebe, and P. Hoffmann, "Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis," *Comput. Linguistics*, vol. 35, no. 3, pp. 399–433, 2009.

[53] S. Asur and B. A. Huberman, "Predicting the future with social media," in *Proc. IEEE/WIC/ACM Int. Conf. Web Intell. Intell. Agent Technol. (WI-IAT)*, vol. 1. Aug./Sep. 2010, pp. 492–499.

[54] H. Chen, P. De, Y. Hu, and B.-H. Hwang, "Sentiment revealed in social media and its effect on the stock market," in *Proc. IEEE Statist. Signal Process. Workshop (SSP)*, Jun. 2011, pp. 25–28.

[55] D. Hardt and J. Wulff, "What is the meaning of 5 */s? An investigation of the expression and rating of sentiment," in *Proc. KONVENS*, Sep. 2012, pp. 319–326.

[56] S. P. Robertson, "Changes in referents and emotions over time in election-related social networking dialog," in *Proc. 44th Hawaii Int. Conf. Syst. Sci. (HICSS)*, Jan. 2011, pp. 1–9.

[57] M. Salathé and S. Khandelwal, "Assessing vaccination sentiments with online social media: Implications for infectious disease dynamics and control," *PLoS Comput. Biol.*, vol. 7, no. 10, p. e1002199, 2011.

[58] T. Menezes, C. Roth, and J.-P. Cointet, "Finding the semantic-level precursors on a blog network," *Int. J. Soc. Comput. Cyber-Phys. Syst.*, vol. 1, no. 2, pp. 115–134, 2011.

[59] R. R. Mukkamala, A. Hussain, and R. Vatrapu, "Towards a formal model of social data," IT Univ. Copenhagen, Copenhagen, Denmark, Tech. Rep. TR-2013-169, Nov. 2013.

[60] McDonalds. *McDonalds Facebook Post*, accessed on May 18, 2016. [Online]. Available: https://www.facebook.com/McDonalds/photos/a.10150151878945584.414818.10150097174480584/10156340092750584/?type=3

[61] H. Li and J. D. Leckenby, "Internet advertising formats and effectiveness," in *Internet Advertising: Theory and Research*, 2nd ed. Hove, U.K.: Psychology Press, 2007.

[62] C. P. Haugtvedt, P. M. Herr, and F. R. Kardes, Eds., *Handbook of Consumer Psychology*. Hove, U.K.: Psychology Press, 2012.
[63] K. Kunst and R. Vatrapu, "Towards a theory of socially shared consumption: Literature review, taxonomy and research agenda," in *Proc. Eur. Conf. Inf. Syst. (ECIS)*, Tel Aviv, Israel, 2014, pp. 1–12.
[64] T. E. Barry, "The development of the hierarchy of effects: An historical perspective," *Current Issues Res. Advertising*, vol. 10, nos. 1–2, pp. 251–295, 1987.
[65] R. J. Lavidge and G. A. Steiner, "A model for predictive measurements of advertising effectiveness," *J. Marketing*, vol. 25, no. 6, pp. 59–62, 1961.
[66] T. Veblen, *The Theory of the Leisure Class: An Economic Study of Institutions*. Oxford, U.K.: Oxford Univ. Press, 2009.
[67] H.-J. Zimmermann, "Fuzzy set theory," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 2, no. 3, pp. 317–332, 2010. [Online]. Available: http://dx.doi.org/10.1002/wics.82
[68] Google Inc. (Sep. 2012). *Google Prediction API*. [Online]. Available: https://developers.google.com/prediction/
[69] W. T. Coombs, "Choosing the right words: The development of guidelines for the selection of the 'appropriate' crisis-response strategies," *Manage. Commun. Quart.*, vol. 8, no. 4, pp. 447–476, 1995.
[70] W. T. Coombs, *Ongoing Crisis Communication: Planning, Managing, and Responding*. New York, NY, USA: Sage Publications, 2014.
[71] C. Zimmerman, Y. Chen, D. Hardt, and R. Vatrapu, "Marius, the giraffe: A comparative informatics case study of linguistic features of the social media discourse," in *Proc. Conf. Collaboration Across Boundaries: Culture, Distance Technol.*, 2014, pp. 131–140.
[72] R. R. Mukkamala, J. I. Sørensen, A. Hussain, and R. Vatrapu, "Detecting corporate social media crises on facebook using social set analysis," in *Proc. IEEE Int. Congr. Big Data (BigData Congress)*, Jun./Jul. 2015, pp. 745–748.
[73] P. Bromiley, M. Govekar, and A. Marcus, "On using event-study methodology in strategic management research," *Technovation*, vol. 8, no. 1, pp. 25–42, 1988.
[74] A. McWilliams and D. Siegel, "Event studies in management research: Theoretical and empirical issues," *Acad. Manage. J.*, vol. 40, no. 3, pp. 626–657, 1997.
[75] A. C. MacKinlay, "Event studies in economics and finance," *J. Econ. Literature*, vol. 35, no. 1, pp. 13–39, 1997.
[76] J. Binder, "The event study methodology since 1969," *Rev. Quant. Finance Accounting*, vol. 11, no. 2, pp. 111–137, 1998.
[77] V. Bajaj. (Nov. 25, 2012). *Fatal Fire in Bangladesh Highlights the Dangers Facing Garment Workers*, New York Times, accessed on May 18, 2016. [Online]. Available: http://www.nytimes.com/2012/11/26/world/asia/bangladesh-fire-kills-more-than-100-and-injures-many.html?ref=world&_r=0
[78] H. Sato, "Cournot competition and reduction of corruption to prevent garment factory fires in bangladesh," *Adv. Manage. Appl. Econ.*, vol. 4, no. 4, pp. 17–20, Aug. 2014.
[79] P. Khanna, "Making labour voices heard during an industrial crisis: Workers' struggles in the Bangladesh garment industry," *Labour, Capital Soc.*, vol. 44, no. 2, pp. 106–129, 2011.
[80] J. A. Manik, J. Yardley, and B. Dhaka. (2014). Building collapse in bangladesh leaves scores dead. NY TIMES. [Online]. Available: http://www.nytimes.com/2013/04/25/world/asia/bangladesh-buildingcollapse.html
[81] J. Burke. (Jun. 6, 2013). *Bangladesh Factory Collapse Leaves Trail of Shattered Lives*, Guardian, accessed on May 18, 2016. [Online]. Available: http://www.theguardian.com/world/2013/jun/06/bangladesh-factory-building-collapse-community
[82] S. A. Himi and A. Rahman, "Workers unrest in garment industries in Bangladesh: An exploratory study," *J. Org. Human Behaviour*, vol. 2, no. 3, pp. 49–55, 2013.
[83] S. Rahman, *Broken Promises of Globalization: The Case of the Bangladesh Garment Industry*. Lexington, KY, USA: Lexington Books, 2013.
[84] K. L. Stewart, "An ethical analysis of the high cost of low-priced clothing," *J. Acad. Bus.Ethics*, vol. 8, pp. 1–9, Jul. 2014.
[85] M. Azizul, C. Deegan, and R. Gray. (Jul. 14, 2014). *Social Audits and Multinational Company Supply Chain: A Study of Rituals of Social Audits in the Bangladesh Garment Industry*. [Online]. Available: http://ssrn.com/abstract=2466129
[86] A. Hussain and R. Vatrapu, "Social data analytics tool: Design, development, and demonstrative case studies," in *Proc. IEEE 18th Enterprise Distrib. Object Comput. Conf. Workshops Demonstrations (EDOCW)*, Sep. 2014, pp. 414–417.
[87] H. G. Miller and P. Mork, "From data to decisions: A value chain for big data," *IT Prof.*, vol. 15, no. 1, pp. 57–59, 2013.
[88] C. Abras, D. Maloney-Krichmar, and J. Preece, "User-centered design," in *Encyclopedia of Human-Computer Interaction*, W. Bainbridge, Ed. Thousand Oaks, CA, USA: Sage Publications, 2004.
[89] A. Thiem and A. Dusa, *Qualitative Comparative Analysis with R A User's Guide*, vol. 5. New York, NY, USA: Springer-Verlag, 2012.
[90] C. C. Ragin, *Redesigning Social Inquiry: Fuzzy Sets and Beyond*, vol. 240. New York, NY, USA: Wiley, 2008.
[91] C. C. Ragin, *The Comparative Method: Moving Beyond Qualitative and Quantitative Strategies*. Berkeley, CA, USA: Univ. California Press, 1987.
[92] P. C. Fiss, "A set-theoretic approach to organizational configurations," *Acad. Manage. Rev.*, vol. 32, no. 4, pp. 1180–1198, 2007.
[93] C. Wagemann and C. Q. Schneider, "Qualitative comparative analysis (QCA) and fuzzy-sets: Agenda for a research approach and a data analysis technique," *Comparative Sociol.*, vol. 9, no. 3, pp. 376–396, 2010.
[94] P. Emmenegger, D. Schraff, and A. Walter, "QCA, the truth table analysis and large-n survey data: The benefits of calibration and the importance of robustness tests," presented at the 2nd Int. QCA Expert Workshop, Zurich, Switzerland, 2014.
[95] A. Lex, N. Gehlenborg, H. Strobelt, R. Vuillemot, and H. Pfister, "UpSet: Visualization of intersecting sets," *IEEE Trans. Vis. Comput. Graphics*, vol. 20, no. 12, pp. 1983–1992, Dec. 2014. [Online]. Available: http://vcg.github.io/upset
[96] L. Micallef and P. Rodgers, "*eulerAPE*: Drawing area-proportional 3-Venn diagrams using ellipses," *PloS One*, vol. 9, no. 7, p. e101717, 2014.
[97] R. R. Mukkamala, J. I. Sorensen, A. Hussain, and R. Vatrapu, "Social set analysis of corporate social media crises on Facebook," in *Proc. IEEE 19th Int. Enterprise Distrib. Object Comput. Conf. (EDOC)*, Sep. 2015, pp. 112–121.

RAVI VATRAPU received the B.Tech. degree in computer science and systems engineering from Andhra University, the M.Sc. degree in computer science and applications from Virginia Tech, and the Ph.D. degree in communication and information sciences from the University of Hawaii at Manoa. He is currently a Professor of Human–Computer Interaction with the Department of IT Management, Copenhagen Business School, a Professor of Applied Computing with the Westerdals Oslo School of Arts Communication and Technology, and the Director of the Computational Social Science Laboratory. His current research focus is on big social data analytics. Based on the enactive approach to the philosophy of mind and phenomenological approach to sociology and the mathematics of classical, fuzzy, and rough set theories, his current research program seeks to design, develop, and evaluate a new holistic approach to computational social science and social set analytics (SSA). SSA consists of novel formal models, predictive methods, and visual analytics tools for big social data.

RAGHAVA RAO MUKKAMALA received the B.Tech. degree from Jawaharlal Nehru Technological University, India, and the M.Sc. degree in information technology and the Ph.D. degree in computer science from the IT University of Copenhagen, Denmark. He is currently an Assistant Professor of Computational Social Science with the Department of IT Management, Copenhagen Business School; an External Lecturer of applied computing with the Westerdals Oslo School of Arts Communication and Technology; and the Co-Director of the Computational Social Science Laboratory. His current research focus is on interdisciplinary approach to big data analytics. Combining formal/mathematical modeling approaches with data/text mining techniques and machine learning methodologies, his current research program seeks to develop new algorithms and techniques for big data analytics such as social set analytics. Before moving to research, he has many of years of programming and IT development experience from the Danish IT industry.

ABID HUSSAIN received the M.Sc. degree in software development from the IT University of Copenhagen, Denmark, the Graduate Diploma degree in information systems management from Central Queensland University, Australia, and the Diploma degree in information technology and software development from the Holmesglen Institute, Australia. He is currently an Assistant Professor of Computational Social Science with the Department of IT Management, Copenhagen Business School, and an Associate Researcher with the Computational Social Science Laboratory. His research focus is on the design, development, and evaluation of design principles and design patterns for the systematic collection, storage, retrieval, and processing of big social data. He is the Lead Researcher and Developer of the Social Data Analytics Tool, the first research-based big social data analytics tool for Facebook. He has more than ten years of software development experience in the IT industry, where he has served as the System Architect and Lead Developer on software teams ranging up to 40 members.

BENJAMIN FLESCH received the M.Sc. degree in business administration and information systems from the Copenhagen Business School, Denmark, and the M.Sc. degree in business administration from the University of Mannheim, Germany. He is a Ph.D. Fellow of Computational Social Science with the Department of IT Management, Copenhagen Business School, and a Research Fellow with the Computational Social Science Laboratory. His research aims to formulate and evaluate a new field of study, computational set analysis in general and computational set visualizations in particular. His Ph.D. project aims to design, develop, and evaluate social set visualizer based on set-theoretical approach to computational social science.