2014 IEEE 18th International Enterprise Distributed Object Computing Conference

Fuzzy-Set Based Sentiment Analysis of Big Social Data

Raghava Rao Mukkamala
IT University of Copenhagen
Rued Langgaardsvej 7,
2300 Copenhagen, Denmark
rao@@itu.dk

Abid Hussain[1] and Ravi Vatrapu[1,2]
[1]Copenhagen Business School
Howitzvej 60, 2000 Frederiksberg, Denmark
[2]Norwegian School of Information Technology (NITH), Norway
{ah.itm, vatrapu}@cbs.dk

***Abstract*—Computational approaches to social media analytics are largely limited to graph theoretical approaches such as social network analysis (SNA) informed by the social philosophical approach of relational sociology. There are no other unified modelling approaches to social data that integrate the conceptual, formal, software, analytical and empirical realms. In this paper, we first present and discuss a theory and conceptual model of social data. Second, we outline a formal model based on fuzzy set theory and describe the operational semantics of the formal model with a real-world social data example from Facebook. Third, we briefly present and discuss the Social Data Analytics Tool (SODATO) that realizes the conceptual model in software and provisions social data analysis based on the conceptual and formal models. Fourth, we use SODATO to fetch social data from the facebook wall of a global brand, H&M and conduct a sentiment classification of the posts and comments. Fifth, we analyse the sentiment classifications by constructing crisp as well as the fuzzy sets of the artefacts (posts, comments, likes, and shares). We document and discuss the longitudinal sentiment profiles of artefacts and actors on the facebook page. Sixth and last, we discuss the analytical method and conclude with a discussion of the benefits of set theoretical approaches based on the social philosophical approach of associational sociology.**

***Keywords*—Formal Methods, Social Data Analytics, Computational Social Science, Data Science, Big Social Data.**

I. Introduction

Big data typically consists of large volumes of data in a variety of data formats that come into being at varying velocities in the form of historical archives to real-time streaming with differing degrees of data provenance. Currently, one such source for big data is user interactions on social media platforms and mobile applications. The participatory turn of the internet coupled with technological advancements in and consumer adoption of ubiquitous, pervasive and wearable technologies have resulted in big social data. Social media analytics is a term we use here to refer to the collection, storage, analysis, and reporting of these new data [1].

For example, recent studies have shown that social data on Facebook can be analysed for investigating political discourse on online public spheres for the United States Election [2], [3] and social data from twitter has been used for predicting Hollywood movies' box-office revenues [4] and quarterly sales of iphones [5].

Seminal work in computational social science (CSS) research has outlined four fundamental methodological principles [6]. First, an ensemble of models needs to be employed instead of using one model for social science investigations. Second, a set of invariant policies has to be followed that provide robust outcomes. Third, uncertainty can be dealt by adapting certain methods, for example the use of forecasting to adopt statistical models. Fourth and last, best results for social science investigations can be achieved through collaboration between humans and machines. Further, Conte and colleagues [7] also point that CSS is a model based science that analyses electronic trace data, builds predictive models and intends to provide instruments for enabling social science to inform decision makers for societal and organisational challenges. Adherering to the CSS methodological guidelines listed above and the answering the call for model based data science research, in this paper we propose and demonstrate a new approach to sentiment analysis based on a unified framework for big social data analytics based on a theory of socio-technical interactions [8], descriptive and formal models of social data [9], and a software schematic for the social data analytics tool [10]. Our objective is to turn big data sets into big data assests that generate competitive advantages for the companies. The application domain for this paper is brand sentiment analysis in the industry sector of fast fashion [9], [11].

A. Sentiment Analysis

At the enterprise level, as Li and Leckenby [12] 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 [13], [14]. Taken together this led to the emergence of organizations that strategically utilize the online channels including social media platforms for business purposes [1]. 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

1541-7719/14 $31.00 © 2014 IEEE
DOI 10.1109/EDOC.2014.19

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 [15] : 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) [16], [17] 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) [12]. 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 [18].

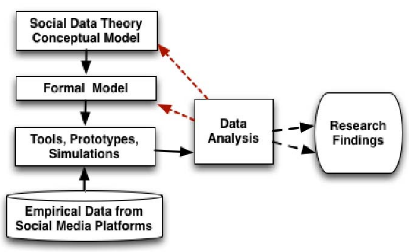


Figure 1. Overall Methodology

B. Formal Models

Formal modeling is a process of writing and analyzing formal descriptions of models and systems that represent real-world processes. It is a technique to model complex phenomena as mathematical entities so that rigorous analysis techniques can be applied on the models to understand the reality of the complex phenomenon. Formal specifications are abstract, precise and to some extent complete in nature [19], [20]. The abstraction of a formal specification allows to comprehend a complex phenomenon, whereas the precise semantics eliminates ambiguity in the model. The completeness ensures the study of all aspects of the behavior in the model [20].

Having said that, formal methods, models and tools for social data are largely limited to graph theoretical approaches informed by conceptual developments in relational sociology and methodological developments in social network analysis [21]–[23]. To the best of our knowledge, there are no other integrated modeling approaches to social data across the conceptual, formal and software realms. In a recent work [9], [11], we have used set theory to formalise the concepts of social data and developed a method based on sentiment analysis for profiling of artifacts and actors.

Sentiment analysis of social data is not an exact science and despite advances in computational linguistics methods and tools [24], there remains a fundamental vagueness in determining the sentiment of a text. Part of the problem is in determining expressed sentiment (as the original actor of that artefact intended) vs. impressed sentiment (as the other actor felt in the actual-space and real-time of interacting with the artefact posted by the original actor). Current practice in sentiment analysis of social data predominantly uses the classification of individual artefacts such as either positive or negative or neutral (e.g., [25]), and not the probabilities returned by the sentiment analysis method and/or tool.

Fuzzy set theory [26], [27] is a mathematical abstraction for the systemic treatment of vagueness and uncertainty both qualitatively and quantitatively. Fuzzy sets are well suited for the study of social systems owing to their ability to deal with vagueness, ambiguity and uncertainty of qualitative ideas and judgements [28]–[30]. In this paper, we propose the use of Fuzzy sets to model *raw* sentiment with classification probabilities of artifacts and develop a analysis method based on α-cut of Fuzzy sets [27], [30] to determine whether any given artifact expresses and impresses positive, negative, and/or neutral sentiment. Moreover, we also advocate an integrated modeling approach as shown in Figure 1, involving a conceptual model for social data, a formal model of the conceptual data based on Set Theory and Fuzzy Set Theory, a schematic model of a software application informed by the conceptual and formal models.

The remainder of the paper is organized as follows. We first present related work in respect of sentiment analysis, social network analysis and explain how our approach differs form the existing approaches. Then we present a brief discussion on social data and later we present a conceptual model of social data. In Sec. V, we outline our formal model based on Set Theory and Fuzzy Set theory [26]. In the next section, we present a methodology for sentiment analysis that is derived from the formal model. In Sec. VII, we briefly present and discuss the Social Data Analytics Tool (SODATO) that realizes the conceptual and formal models in software. Later, we present an empirical case study on sentiment analysis on social data of H&M from their Facebook page. Finally, we discuss our approach and in the end we present a brief conclusion.

II. Related Work

A. Social Network Analysis

The use of Social Network Analysis can be traced back to 1979, where Tichy et.al. [31] used it as a method of examining the relationships and social structures for the analysis of organisations. Later in 1987, David Krackhardt [32] proposed cognitive social structures as a solution for social network related problems.

Due to the advent of internet and the online social media in the last decade, the field of social computing attracted many researchers. It is not possible to refer to an extensive list of research articles in this emerging area, however we refer some of the important works here. First of all, Justin Zhan and Xing Fang in [33] provided an detailed overview about state of art in social networking analysis, social and human behavioural modeling and security on social networks. A framework for calculating reputations in multi-agent systems using social network analysis has been proposed in [34], where as social network analysis based on measuring social relations using multiple data sets has been explored in [35]. An algorithm to find overlapping communities via social network analysis was explored in [36]. Moreover, analysis of sub-graphs in the social network based on the characteristic features: leadership, bonding, and diversity was studied by the authors in [37]. All these works are primarily focussed on using social network analysis and other graph related formalisms, where as our work primarily focussed on using set theory and fuzzy theory for social graph analysis combined with social text analysis in general and sentiment analysis in particular. That is, 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*

Pang and Lee [24] provide a comprehensive state-of-the-art review of computational linguistics approaches to analysing natural language texts and identify three different technical terms: opinions, sentiments, and subjectivity. In this paper, we adopt Pang and Lee's [24] 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 [38]. Other methods and techniques for sentiment analysis are presented and discussed in [24], [38]–[42]. 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 [4], correlate with contemporaneous and subsequent stock returns [43], exploring cultural and linguistic differences in ratings and reviews [25], sentiment evolution in political deliberation on social media channels [44], assess sentiment towards a new vaccine [45], and explore semantic-level precedence relationships between participants in a blog network [46]. To briefly expand, [46] proposed a methodology for the detection of bursts of activity at the semantic level using linguistic tagging, term filtering and term merging. They used a probabilistic approach to estimate temporal relationships between the blogs. Asur and Huberman [4] showed that analysis of sentiment content on urls, retweets and their hourly rates of Twitter can predict box-office movies revenues.

We find that the existing models are primarily focused on using social network analysis and other graph theory related formalisms. In contrast, 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.

III. THEORY OF SOCIAL DATA

Our theory of social data is drawn from the theory of socio-technical interactions by Vatrapu [8]. Social media platforms such as Facebook and Twitter, at the highest level of abstraction, involve individuals interacting with (a) technologies and (b) other individuals. These interactions are termed socio-technical interactions. There are two types of socio-technical interactions: 1) interacting with the technology per se (for example, using the Facebook app on the user's smartphone) and 2) interacting with social others using the technology (for example, liking a picture of a friend in the Facebook app of the user's smartphone). These socio-technical interactions are theoretically conceived as (a) perception and appropriation of socio-technical affordances, and (b) structures and functions of technological intersubjectivity. Briefly, socio-technical affordances are action-taking possibilities and meaning-making opportunities in an actor-environment system bounded by the cultural-cognitive competencies of the actor and the technical capabilities of the environment. Technological intersubjectivity (TI) refers to a technology supported interactional social relationship between two or more actors. A more detailed explication of the theoretical framework in terms of its ontological and epistemological assumptions and principles is beyond the scope of this paper but for details, please confer Vatrapu [8].

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

IV. DESCRIPTIVE MODEL

Social data consists of two types: *Social Graph* and *Social Text*. Social Graph maps on to the first aspect of socio-technical interactions that involve perception and appropriation of affordances (which users/actors act up on which technological features to interact with what other social actors in the systems). Social Text maps on to the second aspect of socio-technical interactions that constitute the structures and functions and technological intersubjectivity (what the users/actors are trying to communicate to each other and how they are trying to influence each other through language).

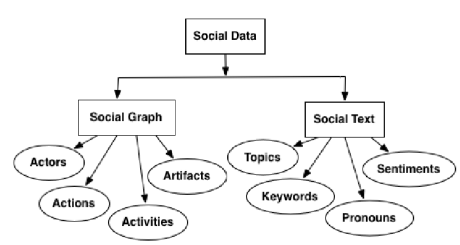


Figure 2. Social Data Model

There is no distinction between a user and an actor in the model. With respect to action/activity, an action (post, comment, like etc.) is atomic event done by an actor on an artifact, where an activity (e.g. promotion, campaign etc.) can spread across many actions, artifacts and actors.

Social graph consists of the structure of the relationships emerging from the apprproiation of social media affordances such as posting, linking, tagging, sharing, liking etc. It focuses on identifying the **actors** involved, the **actions** they take, the **activities** they undertake, and the **artifacts** they create and interact with. Social text consists of the communicative and linguistic aspects of the social media interaction such as the **topics** discussed, **keywords** mentioned, **pronouns** used and **sentiments** expressed. We now turn our attention to formalizing the conceptual model as we believe that formal models are essential for the application of computational techniques and tools given not only the large volumes of data involved but also their ambiguity and unstructured nature.

V. SOCIAL DATA FORMAL MODEL

In this section, we will extend the formal semantics of social data that as originally proposed in [9], [11] with Fuzzy sets.

Notation: The cardinality (number of elements) of set A is represented as $|A|$. 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. 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 recall necessary basic definitions of Fuzzy sets [27].

Definition 1: 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 $A = \{\big(x, \mu_A(x)\big) \mid x \in X)\}$ where $\mu_A : X \to [0, 1]$ 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 2: 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 3: 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 4: The *Union* operation on two fuzzy sets $A = \{\big(x, \mu_A(x)\big) \mid x \in X)\}$ and $B = \{\big(x, \mu_B(x)\big) \mid x \in X)\}$ 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 5: A fuzzy relation R from a set A to B with its membership function $\mu_R : A \times B \to [0, 1]$ is defined as $R = \{\big((a, b), \mu_R(a, b)\big) \mid (a, b) \in A \times B\}$.

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.

A. Formal Model

First, we define type of artifacts in a socio-technical system as $\mathsf{R_T} = \{\text{post, comment, link, photo, video}\}$. The social data model contains Social Graph and Social Text which are formally defined in Def. 6 as follows,

Definition 6: Formally, the Social Data Model is defined as a tuple $\mathsf{S} = (\mathsf{G}, \mathsf{T})$ where

(i) G is the social graph as defined in Def. 7

(ii) T is the social text as further defined in Def. 8

Definition 7: The Social Graph is defined as a tuple $\mathsf{G} = (\mathsf{U}, \mathsf{R}, \mathrm{Ac}, \mathsf{r_{type}}, \triangleright, \rightarrow_{post}, \rightarrow_{share}, \rightarrow_{like}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$ where

(i) U is a finite set of actors/ users ranged over by u,

(ii) R is the finite set of artifacts ranged over by r,

(iii) Ac is a set of activities,

(iv) $\mathsf{r_{type}} : \mathsf{R} \to \mathsf{R_T}$ is a function mapping each artifact to its type.

(v) $\triangleright : \mathsf{R} \rightharpoonup \mathsf{R}$ is a partial function mapping artifacts to their parent artifact (if defined),

(vi) $\rightarrow_{post} : \mathsf{U} \rightharpoonup disj(\mathcal{P}(\mathsf{R}))$ is a partial function mapping actors to mutually disjoint sets of artifacts,

(vii) $\rightarrow_{share} \subseteq \mathsf{U} \times \mathsf{R}$ is a relation mapping users to artifacts shared by them

(viii) $\rightarrow_{like} \subseteq \mathsf{U} \times \mathsf{R}$ is a relation mapping users to the artifacts indicating the artifacts liked by the users,

(ix) $\rightarrow_{tag} \subseteq \mathsf{U} \times \mathsf{R} \times (\mathcal{P}(\mathsf{U} \cup \mathrm{Ke}))$ 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. 8

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

As shown in the first two items (i, ii, x) of Def. 7, the social graph primarily contains a set of actors or users (U), a set of artifacts or 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. 7-iv). In addition to that, some of the artifacts are mapped to their parent artifact (if exists) by parent artifact function $\triangleright$ (Def. 7-v). For example, if the artifact is a comment on a post, then it is mapped to its parent (which is the post), on the other hand, it won't have any parent if it is a new post or a status message.

Furthermore, each artifact is mapped to a unique actor, who is the creator of that artifact. As shown in Def. 7-vi, the $\rightarrow_{post}$ is a partial function mapping actors to mutually disjoint 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. 7-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 ($\rightarrow_{like}$) models mapping 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. 7-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 $\rightarrow_{\mathsf{act}}$ is a fuzzy relation indicates a mapping between artifacts to activities (Def. 7-x) 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.

Definition 8: In Social Data Model S = (G, T) we define Social Text as T= (To, Ke, Pr, Se, $\rightarrow_{\mathsf{topic}}, \rightarrow_{\mathsf{key}}, \rightarrow_{\mathsf{pro}}, \rightarrow_{\mathsf{sen}}$) where

(i) $\mathrm{To}, \mathrm{Ke}, \mathrm{Pr}, \mathrm{Se}$ are the sets of topics, keywords, pronouns and sentiments respectively,

(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 $\mu_{\rightarrow_{\mathsf{sen}}} : \mathsf{R} \times \mathrm{Se} \rightarrow [0, 1]$.

As explained in the conceptual model, the Social Text mainly contains sets of *topics* (To), *keywords* (Ke), *pronouns* (Pr), and *sentiments* (Se) as defined in Def. 8.

Further, one may note that all the relations in Social Text ($\rightarrow_{\mathsf{topic}}$, $\rightarrow_{\mathsf{key}}$, $\rightarrow_{\mathsf{pro}}$ and $\rightarrow_{\mathsf{sen}}$) are defined as fuzzy relations with membership function varies from 0 to 1, indicating the strength of relationships.

VI. 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.

A. 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 Social Text contains some predefined labels: *positive* (+), *neutral* (0) and *negative* (−) as indicated in $\mathrm{Se} = \{+, 0, -\}$.

B. Sentiment Analysis of Artifacts

In this sentiment analysis of artifacts, let us assume that we are confined to textual types of artifacts, i.e. $\mathsf{r}_{\mathsf{type}}(r) = (\text{post} \vee \text{comment})$. 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. 3) 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 $\alpha \in [0, 1]$.

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

Finally the crisp set R_{α}^{se} 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.

1) 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,

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

As formally expressed above, the set of actors ($\mathsf{U}_{\mathsf{R}^{se}_{\alpha}}$) associated with given set of artifacts (R^{se}_{α}) 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 ($\mathsf{U}_{\mathsf{R}^{se}_{\alpha}}$) and set of artifacts ($\mathsf{R}^{se}_{\alpha}$) 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 $\mathsf{U}_{\mathsf{R}^{se}_{\alpha}}$) for each sentiment label ($\forall.se \in \{+, 0, -\}$) for given time span intervals to plot them across the time horizon.

C. 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,

$$(\rightarrow^{se}_{p}, \rightarrow^{se}_{c}, \rightarrow^{se}_{s}, \rightarrow^{se}_{l})$$

1) $\rightarrow^{se}_{p} = \{\big((r, se), \mu_p(r, se)\big) \mid r \in \rightarrow_{post} (u) \wedge \rhd(r)$ 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) $\rightarrow^{se}_{c} = \{\big((r, se), \mu_c(r, se)\big) \mid r \in \rightarrow_{post} (u) \wedge \exists r' \in \mathsf{R}. \rhd (r) = r'\}$ 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) $\rightarrow^{se}_{s} = \{\big((r, se), \mu_s(r, se)\big) \mid (u, r) \in \rightarrow_{share}\}$ 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) $\rightarrow^{se}_{l} = \{\big((r, se), \mu_l(r, se)\big) \mid (u, r) \in \rightarrow_{like}\}$ 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 $r \in \mathsf{R}$, $se \in \mathrm{Se}$, $\mu_{\rightarrow_{\mathsf{sen}}}(r, se)$ 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. 4) on the above fuzzy sets ($\rightarrow^{se}_{p} \cup \rightarrow^{se}_{c} \cup \rightarrow^{se}_{s} \cup \rightarrow^{se}_{l} \cup \rightarrow^{se}_{t}$). 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
$\mu_u(r, se) = \mathrm{Max}(\mu_p(r, se), \mu_c(r, se), \mu_s(r, se), \mu_l(r, se))$

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^{se}_{α}) 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^{se}_{α}) 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.

D. 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 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.

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 3, the sentiments of the artifacts (e.g. (+):20, (0):65, (-):15) are represented in the boxes below the artifacts.

Example 1: The example shown in Fig. 3 will be encoded as follows,
$\mathsf{S} = (\mathsf{G}, \mathsf{T})$ where $\mathsf{G} = (\mathsf{U}, \mathsf{R}, \mathrm{Ac}, \mathsf{r}_{\mathsf{type}}, \rhd, \rightarrow_{post}, \rightarrow_{share}, \rightarrow_{like}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$ is the social graph and T= (To, Ke, Pr, Se, $\rightarrow_{\mathsf{topic}}, \rightarrow_{\mathsf{key}}, \rightarrow_{\mathsf{pro}}, \rightarrow_{\mathsf{sen}}$) is the Social Text.

Initially, the sets of actors, artifacts and other relations have the following values.
$\mathsf{U} = \{u_0, u_1, u_2, u_3, u_4, u_5, u_6, ...\}$
$\mathsf{R} = \{r_1, r_2, r_3, r_4, r_5, ...\}$
$\rhd = \{(r_2, r_1), (r_3, r_1), (r_4, r_1), (r_5, r_1), ..\}$
$\rightarrow_{post} = \{(u_0, \{r_1, ..\}), (u_2, \{r_2\}), (u_3, \{r_3, r_5\}), (u_6, \{r_4\}), ..\}$
$\rightarrow_{share} = \{(u_4, r_1), (u_2, r_1)..\}$

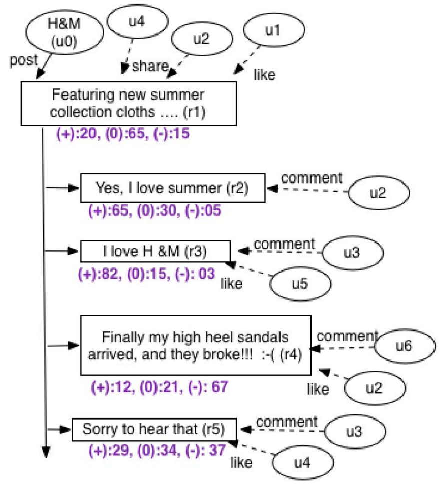


Figure 3. Example in formal model

$\rightarrow_{like} = \{(u_1, r_1), (u_5, r_3), (u_2, r_4), (u_4, r_5), ..\}$
$\mathsf{Se} = \{+, 0, -\}$
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,

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

Regarding temporal dimension ($\mathsf{T_{ime}}$), let us assume that the post (in Figure 3) 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_α^{se}) 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 ($\mathsf{U}_{\mathsf{R}_\alpha^{se}}$) 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 modeling 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_α^{se}) and actors ($\mathsf{U}_{\mathsf{R}_\alpha^{se}}$) 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.

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

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.

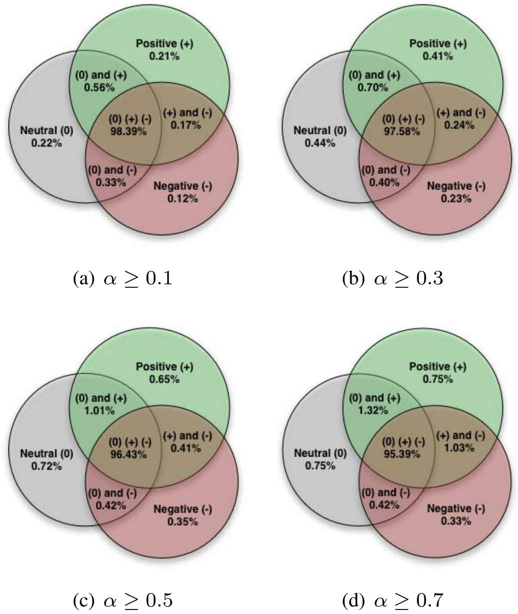
VII. Demonstrative Case Study

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) [10] from the Facebook page of H&M using the methodology presented in the previous section that is based on formal modeling of social data. SODATO [10] is an IT artefact that is a custom built software solution that features collection and archival of Big Social data from online social network platforms, the collected data is then preprocessed and aggregated to make it available on demand for Analytics engine and at the end to the visualization module.

A. Social Text Analysis

Google Prediction API [47] 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 social text corpus of H&M. The sentiment results for each individual post/comment returned by the Google Prediction API were saved back to the relational database. In order to calculate quarterly aggregation of the sentiment classified social texts, further segmentation and grouping was performed using SQL queries and relational



(a) $\alpha \geq 0.1$ (b) $\alpha \geq 0.3$

(c) $\alpha \geq 0.5$ (d) $\alpha \geq 0.7$

Figure 4. Artifact sentiments

database entities were used to store data and it was made available for Analytical calculations.

B. Data Analysis

sentiment	α-cuts				
	$\geq$ **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 I. PARENT ARTIFACT (POSTS) SENTIMENT DISTRIBUTION

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 [47].

C. 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 (I, II and III) and figures (4, 5 and 6) above reveal the longitudinal sentiment profiles of actors and artefacts for the entire corpus. 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

sentiment	α-cuts				
	$\geq$ **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

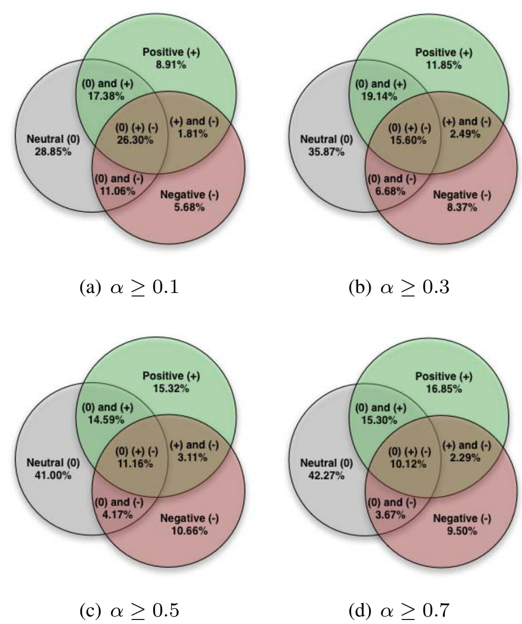
Table II. TOTAL ARTIFACT (POSTS + COMMENTS + LIKES) SENTIMENT DISTRIBUTION

sentiment	α-cuts				
	$\geq$ **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

Table III. ACTORS SENTIMENT WITH DIFFERENT $\alpha - cuts$

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.

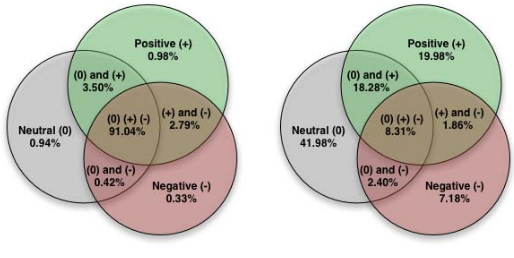
For example, let us consider the α-cut of 0.9 for actors in Table III and Figure 6(b). The graph shows that 7.18% of the entire Facebook user group for the company are always expressing negative sentiments whereas 19.9% of the user group is always expressing positive sentiments. With the caveat that not all of those positive and negative sentiments could be about the company itself (they could be directed towards



(a) $\alpha \geq 0.1$ (b) $\alpha \geq 0.3$

(c) $\alpha \geq 0.5$ (d) $\alpha \geq 0.7$

Figure 5. Actor sentiments



(a) Artifact sentiments $\alpha \geq 0.9$ (b) Actor sentiments $\alpha \geq 0.9$

Figure 6. Artifact and Actor sentiments for $\alpha \geq 0.9$

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 I 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 modeling 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.

VIII. Discussion

In this paper, we have presented an integrated modeling approach for analysis of social data with the sentiment analysis based on the Fuzzy set theory. We articulated a theory of social data that is drawn from the theory of socio-technical interactions for better understanding perception and action on the screen with regard to social media platforms such as Facebook. We then presented a conceptual model of social data emanating from socio-technical interactions on social media platforms such as Facebook. This was followed by a formal model based on Fuzzy Set Theory and a schematic model of a software application developed based on the conceptual and formal models. The formalization of the conceptual model allows the necessary abstraction to comprehend the complex scenarios of social media interactions. On top of that, the formal model also served as a bridge between the conceptual model and schematic model of the software application and helped in concretizing the abstract ideas from the conceptual model to schematic model in the process of developing the Social Data Analytics Tool.

Regarding formal modelling of temporal dimensions of social media interactions (such as Twitter interactions), we are currently developing a hybrid approach by constructing crisp sets as well as fuzzy sets. For example, given an event 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.

Philosophically speaking, the current paradigm in computational social science that is dominated by a theoretical focus on relationships of actors, actors and artifacts and the mathematical modeling of those relationships as social networks based on graph theory . Our argument is not that the triad of relational sociology, graph theory, and social network analysis are invalid or ineffective. Instead, our argument is that other sociological approaches, mathematical theories, and analysis techniques need to be explored towards the description, explanation, prescription, and prediction of social media interactions. To make a bold claim, as far as we know, our paper is the first to propose an alternate holistic framework to the dominant triad of relational sociology, graph theory and social network analysis with regard to sentiment analysis of big social data. Moreover, our approach is one of the very first few models for the analysis of Facebook data.

We acknowledge that many works exist in fuzzy sentiment analysis and social networks and we cited relevant papers in the related work section. But as also 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 more 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 Bruno Latour's ([48]) 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 [11]. 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.

As part of future work, we would like to extend the Fuzzy Set Theoretical formal model to encompass modeling 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.

References

[1] R. Vatrapu, "Understanding social business." in *Emerging Dimensions of Technology Management*, K. B. Akhilesh, Ed. Springer, 2013, pp. 147–158.

[2] S. P. Robertson, R. K. Vatrapu, and R. Medina, "Off the wall political discourse: Facebook use in the 2008 u.s. presidential election," *Info. Pol.*, vol. 15, no. 1,2, pp. 11–31, Apr. 2010.

[3] S. Robertson, R. K. Vatrapu, and R. Medina, "Online video friends social networking: Overlapping online public spheres in the 2008 u.s. presidential election," *Journal of Information Technology & Politics*, vol. 7, no. 2-3, pp. 182–201, 2010.

[4] S. Asur and B. Huberman, "Predicting the future with social media," in *International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010*, vol. 1, 2010, pp. 492–499.

[5] N. Lassen, R. Madsen, and R. Vatrapu, "Predicting iphone sales from iphone tweets," in *Proceedings of the 18th IEEE Enterprise Computing Conference (EDOC 2014)*, 2014.

[6] S. Bankes, R. Lempert, and S. Popper, "Making computational social science effective: Epistemology, methodology, and technology," *Soc. Sci. Comput. Rev.*, vol. 20, no. 4, pp. 377–388, Oct. 2002. [Online]. Available: http://dx.doi.org/10.1177/089443902237317

[7] R. Conte, N. Gilbert, G. Bonelli, C. Cioffi-Revilla, G. Deffuant, J. Kertesz, V. Loreto, S. Moat, J.-P. Nadal, A. Sanchez, A. Nowak, A. Flache, M. San Miguel, and D. Helbing, "Manifesto of computational social science," *The European Physical Journal Special Topics*, vol. 214, no. 1, 2012.

[8] R. K. Vatrapu, "Explaining culture: An outline of a theory of socio-technical interactions," in *Proceedings of the 3rd International Conference on Intercultural Collaboration*, ser. ICIC '10. New York, NY, USA: ACM, 2010, pp. 111–120.

[9] R. R. Mukkamala, A. Hussain, and R. Vatrapu, "Towards a formal model of social data," IT University of Copenhagen, Denmark, IT University Technical Report Series TR-2013-169, November 2013.

[10] A. Hussain and R. Vatrapu, "Social data analytics tool (sodato)," in *DESRIST-2014 Conference (in press)*, ser. Lecture Notes in Computer Science (LNCS). Springer, 2014.

[11] R. R. Mukkamala, A. Hussain, and R. Vatrapu, "Towards a set theoretical approach to big data analytics," in *3rd International Congress on Big Data (IEEE BigData 2014)*, June 2014.

[12] H. Li and J. D. Leckenby, *Internet Advertising: Theory and Research*. Psychology Press; 2 edition, 2007, ch. Internet Advertising Formats and Effectiveness.

[13] C. P. Haugtvedt, P. M. Herr, and F. R. Kardes, *Handbook of consumer psychology*. Psychology Press, 2012.

[14] C. Vollmer and G. Precourt, *Always On: Advertising, Marketing, and Media in an Era of Consumer Contro*. McGraw Hill Professional, 2008.

[15] K. Kunst and R. Vatrapu, "Towards a theory of socially shared consumption: Literature review, taxonomy and research agenda," in *Proceedings of the European Conference on Information Systems (ECIS), Tel Aviv, Israel*, 2014.

[16] T. E. Barry, "The development of the hierarchy of effects: An historical perspective," *Current issues and Research in Advertising*, vol. 10, no. 1-2, pp. 251–295, 1987.

[17] R. J. Lavidge and G. A. Steiner, "A model for predictive measurements of advertising effectiveness." *Journal of marketing*, vol. 25, 1961.

[18] T. Veblen, *The theory of the leisure class; an economic study of institutions*. Oxford University Press, 1899(2009).

[19] A. Nowak, A. Rychwalska, and W. Borkowski, "Why simulate? to develop a mental model," *Journal of Artificial Societies and Social Simulation*, vol. 16, no. 3, 2013.

[20] A. Hall, "Realising the benefits of formal methods," in *Formal Methods and Software Engineering*, ser. LNCS. Springer Berlin Heidelberg, 2005, vol. 3785.

[21] M. Emirbayer, "Manifesto for a relational sociology," *The American Journal of Sociology*, vol. 103(2), pp. 281–317, 1997.

[22] S. P. Borgatti, A. Mehra, D. J. Brass, and G. Labianca, "Network analysis in the social sciences," *Science*, vol. 323(5916), pp. 892–895, 2009.

[23] S. Wasserman and K. Faust, *Social network analysis: Methods and applications (Vol. 8)*. Cambridge university press, 1994.

[24] B. Pang and L. Lee, "Opinion mining and sentiment analysis," *Foundations and trends in information retrieval*, vol. 2, no. 1-2, pp. 1–135, 2008.

[25] D. Hardt and J. Wulff, "What is the meaning of 5 *'s? An investigation of the expression and rating of sentiment," in *Proceedings of KONVENS 2012*, J. Jancsary, Ed. OGAI, September 2012, pp. 319–326, pATHOS 2012 workshop.

[26] L. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, no. 3, pp. 338 – 353, 1965.

[27] H.-J. Zimmermann, "Fuzzy set theory," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 2, no. 3, pp. 317–332, 2010. [Online]. Available: http://dx.doi.org/10.1002/wics.82

[28] V. Dimitrov, "Social choice and self-organization under fuzzy management," *Kybernetes*, vol. 6, no. 3, pp. 153–156, 1977.

[29] ——, "Use of fuzzy logic when dealing with social complexity," *Compexity International*, vol. 04, January 1997.

[30] C. C. Ragin, *Fuzzy-set social science*. University of Chicago Press, 2000.

[31] N. M. Tichy, M. L. Tushman, and C. Fombrun, "Social network analysis for organizations," *The Academy of Management Review*, vol. 4, no. 4, October 1979.

[32] D. Krackhardt, "Cognitive social structures," *Social Networks*, vol. 9, no. 2, pp. 109–134, Jun. 1987.

[33] J. Zhan and X. Fang, "Social computing: the state of the art," *International Journal of Social Computing and Cyber-Physical Systems*, vol. 1, no. 1, pp. 1–12, 01 2011.

[34] J. Sabater and C. Sierra, "Reputation and social network analysis in multi-agent systems," in *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems: Part 1*, ser. AAMAS '02. New York, NY, USA: ACM, 2002, pp. 475–482.

[35] J. Karikoski and M. Nelimarkka, "Measuring social relations with multiple datasets," *IJSCCPS*, vol. 1, no. 1, pp. 98–113, 2011.

[36] M. Goldberg, S. Kelley, M. Magdon-Ismail, K. Mertsalov, and A. Wallace, "Finding overlapping communities in social networks," in *IEEE Conference on Social Computing (SocialCom)*, 2010, pp. 104–113.

[37] O. Macindoe and W. Richards, "Comparing networks using their fine structure," *International Journal of Social Computing and Cyber-Physical Systems*, vol. 1, no. 1, 2011.

[38] 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 Technical Digest*, vol. 30, no. 1, pp. 22–30, 2011.

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

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

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

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

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

[44] S. P. Robertson, "Changes in referents and emotions over time in election-related social networking dialog," in *System Sciences (HICSS), 2011 44th Hawaii International Conference on*, 2011, pp. 1–9.

[45] M. Salathé and S. Khandelwal, "Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control," *PLoS computational biology*, vol. 7, no. 10, 2011.

[46] T. Menezes, C. Roth, and J.-P. Cointet, "Finding the semantic-level precursors on a blog network." *IJSCCPS*, vol. 1, pp. 115–134, 2011.

[47] Google-Inc, "Google prediction api," September 2012, https://developers.google.com/prediction/.

[48] B. Latour, *Reassembling the Social: An Introduction to Actor-Network-Theory*. Oxford University Press, USA, 2005.