Towards a Formal Model of Social Data

Raghava Rao Mukkamala[1]
Abid Hussain[2]
Ravi Vatrapu[2,3]

[1] **IT University of Copenhagen, Denmark**
[2] **Copenhagen Business School, Denmark**
[3] **Norwegian School of Information Technology (NITH), Norway**

IT University Technical Report Series **TR-2013-169**

ISSN 1600–6100 **November 2013**

Raghava Rao Mukkamala[1]
Abid Hussain[2]
Ravi Vatrapu[2,3]

[1] IT University of Copenhagen, Denmark
[2] Copenhagen Business School, Denmark
[3] Norwegian School of Information Technology (NITH), Norway

ISSN 1600–6100

ISBN 978-87-7949.307-0

Copies may be obtained by contacting:

IT University of Copenhagen
Rued Langgaards Vej 7
DK-2300 Copenhagen S
Denmark

Telephone: +45 72 18 50 00
Telefax: +45 72 18 50 01
Web www.itu.dk

Towards a Formal Model of Social Data

Raghava Rao Mukkamala[1], Abid Hussain[2], and Ravi Vatrapu[2,3]

[1] IT University of Copenhagen
Rued Langgaardsvej 7, 2300 Copenhagen, Denmark
rao@itu.dk, http://www.itu.dk
[2] Copenhagen Business School
Howitzvej 60, Frederiksberg, 2000, Denmark
{ah.itm, rv.itm}@cbs.dk, http://cssl.cbs.dk
[3] Norwegian School of Information Technology (NITH), Norway

Abstract. Computational social science (CSS) is an emerging field of research that seeks to apply computational methods and tools to important and interesting social science questions and problems. Situated within CSS, Social data analytics as a research stream aims to collect, archive, retrieve, process, transform, analyse, and report social data from social media platforms such as Facebook and twitter. Formal methods, models and tools for social data are largely limited to graph theoretical approaches informing conceptual developments in relational sociology and methodological developments in social network analysis. As far as we know, there are no integrated modelling approaches to social data across the conceptual, formal and software realms. Social media analytics can be undertaken in two main ways - ”Social Graph Analytics” and ”Social Text Analytics” (Vatrapu, in press/2013). Social graph analytics is concerned with the structure of the relationships emerging from social media use. It focuses on identifying the actors involved, the activities they undertake, and the artifacts they create and interact with. Social text analytics is more concerned with the substantive nature of the interactions, it focuses on the topics discussed and how they are discussed. What keywords appear? What pronouns are used? How far are negative or positive sentiments expressed? In this report, we first present and discuss a conceptual model of social data followed by a formal model based on set theory. Second, we exemplify the semantics of the formal model with real-world social data examples. Third, we briefly present and discuss the Social Data Analytics Tool (SODATO) that realizes the conceptual model in software and provisions social data for computational social science analysis based on the formal model. Finally, we exemplify our approach with help of a case study on big social data of the fast fashion company, H&M. from its Facebook page.

Keywords: Formal Methods, Social Data Analytics, Computational Social Science, Data Science

1 Introduction

There are ever growing pools of data everywhere around us. While surfing internet, doing day to day transactions with IT tools, talking on mobile phones, travelling on plane, walking and interacting with digital gadgets, sending a tweet or a posting a comment on Facebook, all these actions are accumulating data.

These very large sets of data are commonly termed as *Big Data*. Big data sets demand exploration and evaluation of new analytical methods, techniques, and tools as existing data analysis techniques are increasingly becoming inadequate [29] and require multi-disciplinary skills.

For example, Chen, Chiang, and Storey [3] draw attention towards the importance of gaining knowledge in the area of databases, machine learning, statistics and visualization that should enable analysts to analyze data for better sense making. Such challenges with big data have led to the emergence of the new research field of data science. In this report, we are concerned with outlining a formal methods approach to data science involving big data sets from social media platforms such as Facebook and twitter. But before we get to the actual problematics of formal methods for social data analytics, we need to situate our work in the fields of data science and computational social science first.

1.1 Data Science

The coining of the term *data science* is generally attributed to Cleveland [6] but the use of term can also be traced in literature in 2000 when Ohsumi [20] highlighted the need of moving from data analysis towards data science, arguing that merely statistical methods and data mining do not cover the challenges of and opportunities with making sense of data.

Data science is a multi-disciplinary field that brings together cross-disciplinary investigations of the subject under consideration utilizing statistical models, methods, theories and power of computing. One of the aims of the field of data science is to train data analysts and data scientists [6, 16]. Data scientists should be able to use computational methods, database management systems, data analysis tools and data mining techniques that work together towards a solution that is driven by data analysis [6]. Data science can also be defined as a process of dealing with very large sets of data [8].

Data science demands diverse sets of techniques and knowledge that are exceptionally hard to find in one person. Hence data science is a team process of analysing data where data scientists involve in the process of collecting data, analysing data and eventually finding and reporting interesting results. It is this deep involvement of data scientist in the whole process of analysis differentiates data science from traditional field of statistics [15]. Data science is about turning data into data products and services. A data product or service could be an interesting factual finding, finding hidden patterns from large sets of data, sense making for unstructured textual data or a visualization of findings from big data.

1.2 Computational Social Science

As a sub-field of data science, computational social science (CSS) is an emerging field of study that involves the application of computational methods, techniques, and tools for the study of social science issues, questions, and problems. It is an inter-disciplinary research field that brings together the fields of mathematics, social and behavioural science, agent modeling and computer science in order to find answers to complex challenges of the society today [14]. The early days of computational social science can be traced back to 1960's when scientists started using computers for statistical methods [5]. Conte and colleagues [7] define Computational Social Science as "the integrated, inter-disciplinary investigation of social systems as information-processing organizations and through the medium of advanced computational system stems.

In the past, social science findings were based on individually reported relationships and/or analysis of aggregate data collected by state agencies or researchers. Advancements in information technology and the availability of data through diverse mediums have enabled researchers to perform analysis in a way that was never possible before. Different sets of data can provide answers to questions of different domains of social science. For example, digital video data of a child's early life could provide insight about development of language and interactions.

Data collection from work related inter-organisational email communication could potentially provide directions towards how different types of communications affect performance of workers and how it has an effect of workers efficiency, if at all it does. The most important revolution has been that the internet data is a potential source of revealing behaviour and predicting concerns [14]. Usage of automated information extraction could be useful for data mining to detect anomalies or to analyse real time streams of data under a setup that is supported by visualization [5].

1.3 Social Data Analytics

The growth of social media use in society is generating large quantities of new digital information about individuals, organizations and institutions that is now commonly labeled *Big Data*. Social media analytics is a term we use here to refer to the collection, storage, analysis, and reporting of these new data [25]. These social data sets carry valuable information and if analysed utilizing proper methods, techniques, and tools of computational social science in particular and data science in general. They can provide meaningful facts and actionable insights that go beyond traditional social science research methods.

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 [21, 22] and social data from twitter has been used for predicting Hollywood movies' box-office revenues [1]. Bankes et al. [2] defined early guidelines for computational social science (CSS) research providing four fundamental methodological principles for CSS research.

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. Conte and colleagues [7] also point that Computational Social Science 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.

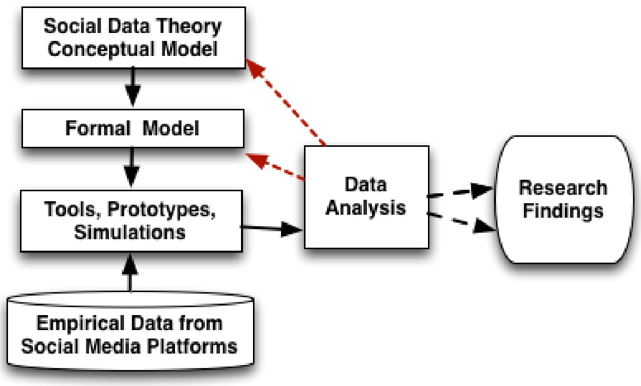

Fig. 1. Overall Methodology

1.4 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 [10, 19]. The abstraction of a formal specification allows to comprehend a complex phenomenon, where as the precise semantics eliminates ambiguity in the model. The completeness ensures the study of all aspects of the behavior in the model [10].

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. As far as we know, there are no integrated modeling approaches to social data across the conceptual, formal and software realms. This report seeks to address this problem by proposing an integrated modeling approach involving a conceptual model for social data, a formal model of the conceptual data based on set theory, a schematic model of a software application informed by the conceptual and formal models as shown in Fig. 1 and an outline of research questions that will be empirically answered using the integrated modeling approach.

The remainder of the report is organized as follows. We first present and discuss theory of social data (Sec. 3) and a conceptual model of social data (Sec. 4). Second, we outline a formal model based on set theory and discuss the semantics of the formal model with a real-world social data example from facebook in Sec. 5. 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 (Sec. 6). In Sec. 7, we present a case study on social data of H&M, where we empirically analyse the relationship of H&M's quarterly revenues and big social data from Facebook. Finally, we present a brief discussion on our methodology (Sec. 8), outline of the research questions that are currently being addressed using the formal and computational models presented in this report (Sec. 9) and in the end a brief conclusion (Sec. 10).

2 Related Work

The use of Social network analysis can be traced back to 1979, where Tichy et.al. [24] used it as a method of examining the relationships and social structures for the analysis of organisations. Later in 1987, David Krackhardt [13] 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 refer 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 [30] provided an extensive 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 [23], where as social network analysis based on measuring social relations using multiple data sets has been explored in [11]. An algorithm to find overlapping communities in an social network analysis explored in [9]. Moreover, analysis of sub-graphs in the social network based on the characteristic features: leadership, bonding, and diversity was studied by the authors in [17]. All these works are primarily focussed on using social network analysis and other graph related formalisms, where as our work primarily focussed on using social graph analysis combined with social Text analysis.

Semantic-level precedence relationships between participants in a blog network are studied in [18], where the authors 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. However in an another interesting work, Sitaram Asur and Bernardo A. Huberman [1] showed that social media feed can be used as effective indicators of the real-world performance. In their work, they used analysis of sentiment content on urls, retweets and their hourly rates of Twitter to estimate to forecast the box-office movies revenue.

3 Theory of Social Data

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 Face book 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 report but for details, please confer [27, 28].

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 report.

Based on the theory of social data described above, we present a conceptual model of social data in the next section.

4 Conceptual 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).

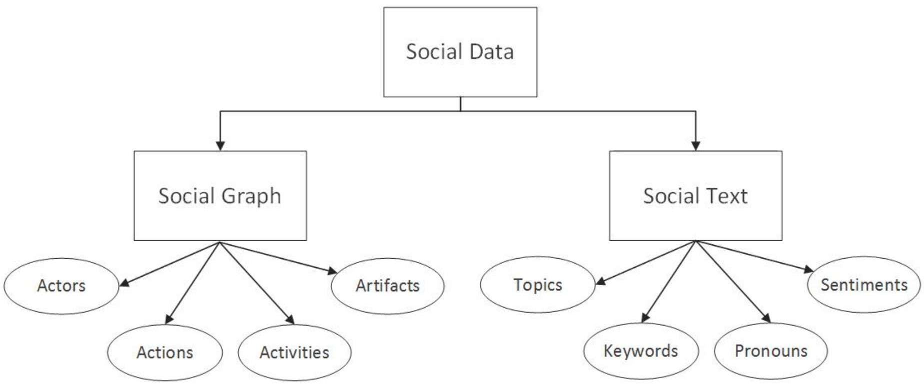

Fig. 2. Social Data Model [26]

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.

5 Formal Model

In this section, we will provide the formal semantics for social data mode introduced in the previous section 4.

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. 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 $f : A \rightarrow B$, 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 $\mathbb{R}$ = { status, comment, link, photo, video }.*

As explained in the conceptual model, the social data model contains Social Graph and Social Text which is formally defined in Def. 2 as follows,

Definition 2. *Formally, the Social Data Model is defined as a tuple $\mathsf{S} = (\mathsf{G}, \mathsf{T})$ where*

(i) G *is the social graph representing the structural aspects of social data as defined further in Def. 3*

(ii) T *is the social text representing the content of social data and is further defined in Def. 4*

As shown in the first two items (i, ii, x) of Def. 3, 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. 3-iv). In addition to that, some of the artifacts are mapped to their parent artifact (if exists) by parent artifact function $\triangleright$ (Def. 3-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, if the artifact is a status message or a new post, then it will not have any parent.

Furthermore, each artifact is mapped to a unique actor, who is the creator of that artifact. As shown in Def. 3-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. 3-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. 3-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. 3-x).

Definition 3. *The Social Graph is defined as a tuple $\mathsf{G} = (\mathsf{U}, \mathsf{R}, \mathrm{Ac}, \mathsf{r}_{\mathsf{type}}, \triangleright, \rightarrow_{post}, \rightarrow_{share}, \rightarrow_{like}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$ where*

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

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

(iii) Ac *is a set of activities,*

(iv) $\mathsf{r}_{\mathsf{type}} : \mathsf{R} \rightarrow \mathbb{R}$ *is the artifact type function mapping each artifact to a artifact type defined in 1,*

(v) $\triangleright : \mathsf{R} \rightharpoonup \mathsf{R}$ *is parent artifact function, which 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,*

(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. 4,*

(x) $\rightarrow_{\mathsf{act}} \subseteq \mathsf{R} \times \mathrm{Ac}$ *is a relation mapping artifacts to activities.*

As explained in the conceptual model, the Social Text mainly contains set of *topics* (To), *keywords* (Ke), *pronouns* (Pr), and *sentiments* (Se) as defined in Def. 4. The $\rightarrow_{\mathsf{topic}}$, $\rightarrow_{\mathsf{key}}$ and $\rightarrow_{\mathsf{pro}}$ relations map the artifacts to the topics, keywords and pronouns respectively. One may note that all these relations allow many-to-many mappings, for example an artifact can be mapped to more than one topics and similarly a topic can contain mappings to many artifacts. On contrary, the $\rightarrow_{\mathsf{sen}}$ is a partial function allowing an artifact to be mapped to at most only one sentiment (Def. 4-vi).

Definition 4. *In Social Data Model* $\mathsf{S} = (\mathsf{G}, \mathsf{T})$*, we define Social Text as* $\mathsf{T} = (\mathrm{To}, \mathrm{Ke}, \mathrm{Pr}, \mathrm{Se}, \rightarrow_{\mathsf{topic}}, \rightarrow_{\mathsf{key}}, \rightarrow_{\mathsf{pro}}, \rightarrow_{\mathsf{sen}})$ *where*

(i) To *is the set of topics and* Ke *is the set of keywords,*

(ii) Pr, Se *are the sets of pronouns and sentiments respectively,*

(iii) $\rightarrow_{\mathsf{topic}} \subseteq \mathsf{R} \times \mathrm{To}$ *is a relation defining mapping between artifacts and topics,*

(iv) $\rightarrow_{\mathsf{key}} \subseteq \mathsf{R} \times \mathrm{Ke}$ *is a relation mapping artifacts to keywords,*

(v) $\rightarrow_{\mathsf{pro}} \subseteq \mathsf{R} \times \mathrm{Pr}$ *is a relation mapping artifacts to pronouns,*

(vi) $\rightarrow_{\mathsf{sen}} : \mathsf{R} \rightharpoonup \mathrm{Se}$ *is a partial function mapping artifacts to sentiments.*

5.1 Operational Semantics

In this section, we will define the operational semantics of the model. As part of the operational semantics of the model, we have primarily identified five different actions: *post*, *comment*, *share*, *like* and *tagging*, which are further defined as follows.

As formally defined in Def. 5, the first action is *post*, which accepts a tuple containing an actor and an artifact (u, r) and creates a new artifact. First, the actor will be added to the set of actors (i) and then the new artifact will be will be added to the set of artifacts (ii). Finally the post relation ($\rightarrow_{post}$) will be updated for the new mapping (iii).

Definition 5. *In Social Data Model* $\mathsf{S} = (\mathsf{G}, \mathsf{T})$ *with* $\mathsf{G} = (\mathsf{U}, \mathsf{R}, \triangleright, \mathsf{r}_{\mathsf{type}}, \mathrm{Ac}, \rightarrow_{post}, \rightarrow_{share}, \rightarrow_{like}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$*, we define a* ***post*** *operation of posting an artifact* r *by an user* u *as* $\mathsf{S} \bigoplus_p (u, r) = (\mathsf{G}', \mathsf{T})$ *where* $\mathsf{G}' = (\mathsf{U}', \mathsf{R}', \triangleright, \mathsf{r}_{\mathsf{type}}, \mathrm{Ac}, \rightarrow_{post}', \rightarrow_{share}, \rightarrow_{like}, \rightarrow_{tag}, \rightarrow_{\mathsf{act}})$*,*

(i) $\mathsf{U}' = \mathsf{U} \cup \{u\}$

(ii) $\mathsf{R}' = \mathsf{R}' \cup \{r\}$

(iii) $\rightarrow_{post}' = \rightarrow_{post} \cup \{(u, r)\}$

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. 6. 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 for comment with its parent (ii).

Definition 6. *Let Social Data Model be* $\mathsf{S} = (\mathsf{G}, \mathsf{T})$ *with* $\mathsf{G} = (\mathsf{U}, \mathsf{R}, \rhd, \mathsf{r}_{\mathsf{type}}, \mathrm{Ac}, \rightharpoonup_{post}, \rightharpoonup_{share}, \rightharpoonup_{like}, \rightharpoonup_{tag}, \rightharpoonup_{\mathsf{act}})$*, the* ***comment*** *operation on an artifact* r *by an user* u *is formally defined as* $\mathsf{S} \bigoplus_c (u, r, r') = (\mathsf{G}', \mathsf{T})$ *where* $\mathsf{G}' = (\mathsf{U}', \mathsf{R}', \rhd', \mathsf{r}_{\mathsf{type}}, \mathrm{Ac}, \rightharpoonup_{post}', \rightharpoonup_{share}, \rightharpoonup_{like}, \rightharpoonup_{tag}, \rightharpoonup_{\mathsf{act}})$*,*

(i) $\mathsf{S} \bigoplus_p (u, r') = (\mathsf{G}'', \mathsf{T})$ *where* $\mathsf{G}'' = (\mathsf{U}', \mathsf{R}', \rhd, \mathsf{r}_{\mathsf{type}}, \mathrm{Ac}, \rightharpoonup_{post}', \rightharpoonup_{share}, \rightharpoonup_{like}, \rightharpoonup_{tag}, \rightharpoonup_{\mathsf{act}})$*,*

(ii) $\rhd' = \rhd \cup \{r, r'\}$

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 ($\rightharpoonup_{share}$) as formally defined in Def. 7.

Definition 7. *Let Social Data Model be* $\mathsf{S} = (\mathsf{G}, \mathsf{T})$ *with* $\mathsf{G} = (\mathsf{U}, \mathsf{R}, \rhd, \mathsf{r}_{\mathsf{type}}, \mathrm{Ac}, \rightharpoonup_{post}, \rightharpoonup_{share}, \rightharpoonup_{like}, \rightharpoonup_{tag}, \rightharpoonup_{\mathsf{act}})$*, then we define the* ***share*** *operation consisting of sharing an artifact* r *by an user* u *as* $\mathsf{S} \bigoplus_s (u, r) = (\mathsf{G}', \mathsf{T})$ *where* $\mathsf{G}' = (\mathsf{U} \cup \{u\}, \mathsf{R}, \rhd, \mathsf{r}_{\mathsf{type}}, \mathrm{Ac}, \rightharpoonup_{post}, \rightharpoonup_{share} \cup \{(u, r)\}, \rightharpoonup_{like}, \rightharpoonup_{tag}, \rightharpoonup_{\mathsf{act}})$.

In the Def. 8, we formally define the *like* and *unlike* operations as an update to the like relation ($\rightharpoonup_{like}$). A *like* action on an artifact will add a mapping to like relation ($\rightharpoonup_{like}$) (in addition to adding the actor to the actors set), where as an *unlike* action will simply remove the existing mapping.

Definition 8. *In a Social Data Model* $\mathsf{S} = (\mathsf{G}, \mathsf{T})$ *with Graph* $\mathsf{G} = (\mathsf{U}, \mathsf{R}, \rhd, \mathsf{r}_{\mathsf{type}}, \mathrm{Ac}, \rightharpoonup_{post}, \rightharpoonup_{share}, \rightharpoonup_{like}, \rightharpoonup_{tag}, \rightharpoonup_{\mathsf{act}})$*, we define the* ***like*** *operation by an user* u *on an artifact* r *as* $\mathsf{S} \bigoplus_l (u, r) = (\mathsf{G}', \mathsf{T})$ *where* $\mathsf{G}' = (\mathsf{U} \cup \{u\}, \mathsf{R}, \rhd, \mathsf{r}_{\mathsf{type}}, \mathrm{Ac}, \rightharpoonup_{post}, \rightharpoonup_{share}, \rightharpoonup_{like} \cup \{(u, r)\}, \rightharpoonup_{tag}, \rightharpoonup_{\mathsf{act}})$.

Similarly, we also define the ***unlike*** *operation on* $\mathsf{S} = (\mathsf{G}, \mathsf{T})$ *with Graph* $\mathsf{G} = (\mathsf{U}, \mathsf{R}, \rhd, \mathsf{r}_{\mathsf{type}}, \mathrm{Ac}, \rightharpoonup_{post}, \rightharpoonup_{share}, \rightharpoonup_{like}, \rightharpoonup_{tag}, \rightharpoonup_{\mathsf{act}})$ *as* $\mathsf{S} \bigominus_l (u, r) = (\mathsf{G}', \mathsf{T})$ *where* $\mathsf{G}' = (\mathsf{U}, \mathsf{R}, \rhd, \mathsf{r}_{\mathsf{type}}, \mathrm{Ac}, \rightharpoonup_{post}, \rightharpoonup_{share}, \rightharpoonup_{like} \setminus \{(u, r)\}, \rightharpoonup_{tag}, \rightharpoonup_{\mathsf{act}})$.

Finally, the *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 ($\rightharpoonup_{tag}$) will be applied as shown in the Def. 9.

Definition 9. *In a Social Data Model* $\mathsf{S} = (\mathsf{G}, \mathsf{T})$ *with Graph* $\mathsf{G} = (\mathsf{U}, \mathsf{R}, \rhd, \mathsf{r}_{\mathsf{type}}, \mathrm{Ac}, \rightharpoonup_{post}, \rightharpoonup_{share}, \rightharpoonup_{like}, \rightharpoonup_{tag}, \rightharpoonup_{\mathsf{act}})$*, we define the* ***tagging*** *operation by an user* u *on an artifact* r *with a set of hash words* $t \in \mathcal{P}(\mathsf{U} \cup \mathrm{Ke})$ *as* $\mathsf{S} \bigoplus_t (u, r, t) = (\mathsf{G}', \mathsf{T})$ *where* $\mathsf{G}' = (\mathsf{U} \cup \{u\}, \mathsf{R}, \rhd, \mathsf{r}_{\mathsf{type}}, \mathrm{Ac}, \rightharpoonup_{post}, \rightharpoonup_{share}, \rightharpoonup_{like}, \rightharpoonup_{tag} \cup \{(u, r, t)\}, \rightharpoonup_{\mathsf{act}})$.

5.2 Example

In this section, we will exemplify the formal model by taking an example 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.

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

Initailly, the sets of activities, topics, keywords, pronouns and sentiments have the following values.
$\mathrm{Ac} = \{promotion\}, \mathrm{To} = \{\text{summer collection, new store request}\}$,
$\mathrm{Ke} = \{H\&M, Dallas, Singapore\}$
$\mathrm{Pr} = \{We, I\}, \mathrm{Se} = \{positive, negative\}$,

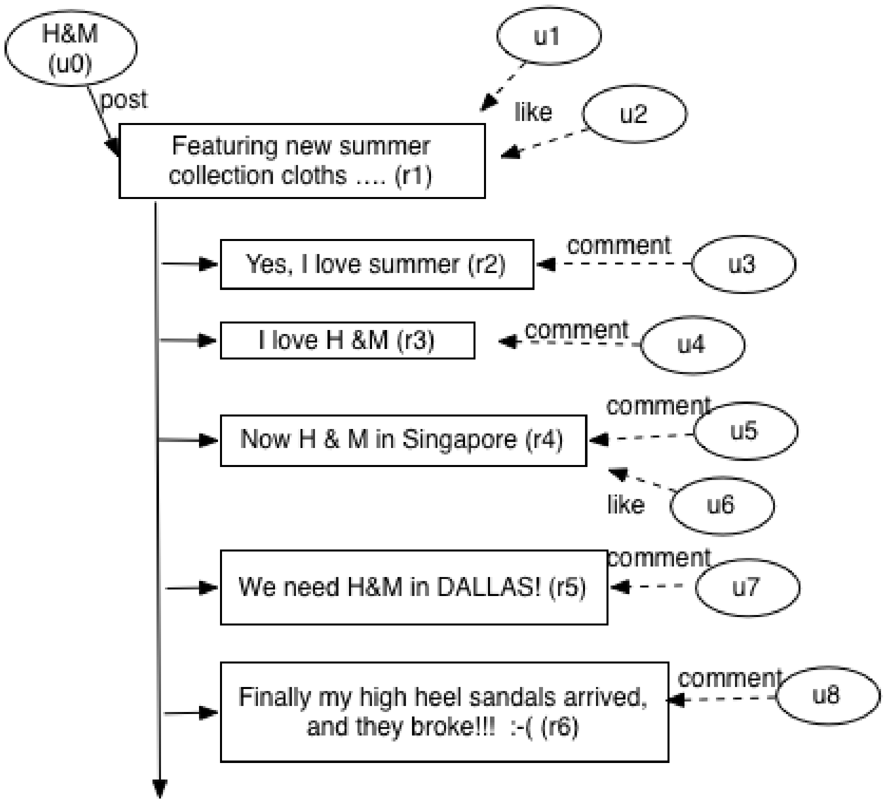

Fig. 3. Example in formal model

$\mathsf{U} = \{u_0, u_1, u_3, \}$
$\rightarrow_{\mathsf{act}} = \{(r_1, promotion)\}$

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

like action by u_2
$\mathsf{S}_1 \bigoplus_l (u_2, r_1) = \mathsf{S}_2 = (\mathsf{G}_2, \mathsf{T})$ where $\mathsf{G}_2 = (\mathsf{U}_2, \mathsf{R}_1, \rhd, \mathsf{r}_{\mathsf{type}}, \mathrm{Ac}, \rightarrow_{post\ 1}, \rightarrow_{share}\ , \rightarrow_{like\ 1}, \rightarrow_{tag}\ , \rightarrow_{\mathsf{act}})$ with the following values
$\mathsf{U}_2 = \mathsf{U}_1 \cup \{u_2\}$, and $\rightarrow_{like\ 1} = \rightarrow_{like}\ \cup \{(u_0, r_1)\}$
comment action by u_3
$\mathsf{S}_2 \bigoplus_c (u_3, r_1, r_2) = \mathsf{S}_3 = (\mathsf{G}_3, \mathsf{T})$ where $\mathsf{G}_3 = (\mathsf{U}_3, \mathsf{R}_2, \rhd_1, \mathsf{r}_{\mathsf{type}}, \mathrm{Ac}, \rightarrow_{post\ 2}, \rightarrow_{share}\ , \rightarrow_{like\ 1}, \rightarrow_{tag}$ $, \rightarrow_{\mathsf{act}})$ with the following values
$\mathsf{U}_3 = \mathsf{U}_2 \cup \{u_3\}$, $\mathsf{R}_2 = \mathsf{R}_1 \cup \{r_2\}$, $\rightarrow_{post\ 2} = \rightarrow_{post\ 1} \cup \{(u_3, r_2)\}$ and
$\rhd_1 = \rhd \cup \{(r_1, r_2)\}$.

After applying data analysis techniques on the social data, the Social Text will be transformed as follows, T= (To, Ke, Pr, Se, $\rightarrow_{\mathsf{topic}}, \rightarrow_{\mathsf{key}}, \rightarrow_{\mathsf{pro}}, \rightarrow_{\mathsf{sen}}$) evolves as follows,
$\rightarrow_{\mathsf{topic}} = \{(r1, \text{summer collection}), (r_2, \text{summer collection}), (r_5, \text{new store request})\}$,
$\rightarrow_{\mathsf{key}} = \{(r_3, H\&M), (r_4, H\&M), (r_5, H\&M), (r_4, Singapore), (r_4, Dallas)\}$
$\rightarrow_{\mathsf{pro}} = \{(r_2, I), (r_3, I), (r_5, We)\}$,
$\rightarrow_{\mathsf{sen}} = \{(r_2, positive), (r_3, positive), (r_6, negative)\}$.

6 Software Architecture: SODATO

Social Data Analytics Tool (SODATO) 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. Conceptual model discussed earlier in the article is employed in order to group different Analysis units so that one module comprehensively provides analysis for sense making of each element of the model for big social data. For example if we are interested in studying actors belonging to certain social data set then we are referring to Social Graph area of our conceptual model.

The software provides a logical module which consists of a web page that is used to demand Actor Analysis, fetching engine functions that facilitate data fetch for actors from requested online Social channel followed by execution of set of methods for data formatting and storage (that again are part of Actor Analysis package) and finally data is made available for Analytics & Visualization engine that present the results either as visual dashboard or possibility of taking the actual data as export.

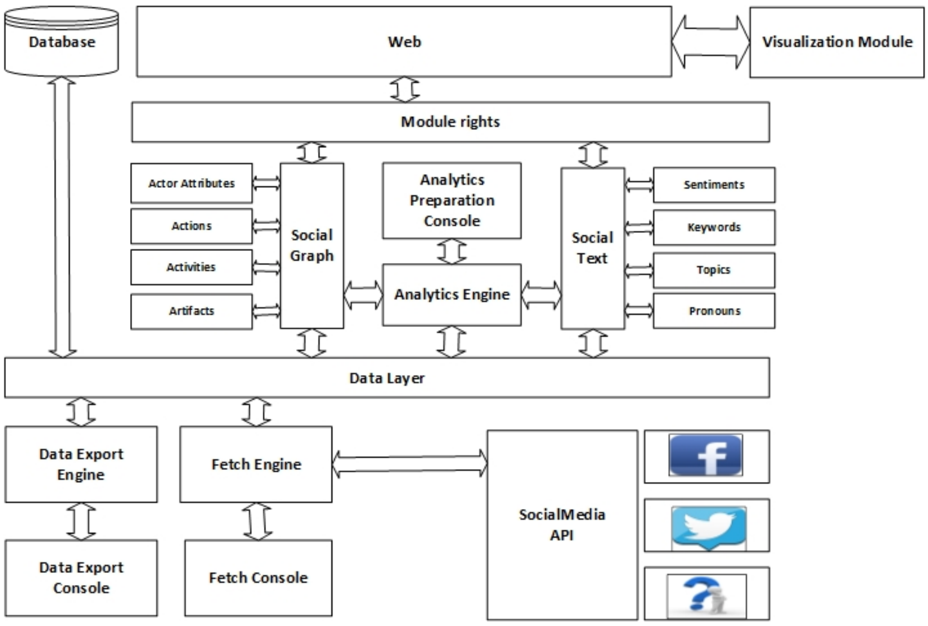

Fig. 4. Social Data Analytics Tool Architecture

Similarly different modules are built that are informed by the model. It can be observed from the diagram that it is not only the end user interaction or the output of Analysis that employs the model but actual underlying software architecture as well is informed by the model. Independent modules are built that seamlessly integrate together to provide a platform for sense making of big social data that covers all the elements from our conceptual social data model.

It is important to highlight that not all elements of the model have all similar components in the software. For example, Actor Analysis, Sentiment Analysis and Keyword Analysis utilise all the components including fetch, preprocess, aggregate, processed by Analytics Engine and Visualization however on the other hand Activity and Actions are inferred from existing data and from other elements like Keywords Analysis and do not need their own fetch capabilities and hence would consist of elements belonging to Analytics engine and Visualization.

Technically SODATO utilizes the APIs provided by the social network vendors for example Facebook open source API named as Graph API. SODATO is a combination of web as well as windows based console applications that run in batches to fetch data and prepare data for analysis. The Web part of the tool is developed using

HTML, JavaScript, Microsoft ASP.NET and C#. Console applications are developed using C#. Microsoft SQL Server is used for data storage and data pre-processing for social graph analytics and social text analytics. The schematic of architecture can be further understood from the Fig 4.

7 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 relationship of H&M's quarterly revenues and big social data from Facebook. Results show a positive correlation between H&M's facebook measures and its quarterly revenues. We discuss the results, present substantive interpretations of the findings, implications of big data analytics for IS and conclude with directions for future work.

7.1 Methodology

We obtained the quarterly revenues for H&M from their publicly accessible financial reports[4]. We copied the numbers from the H&M quarter reports (consolidated income statements[5]) and double-checked them for potential data-entry errors.

7.1.1 H&M's Social Data As mentioned earlier, H&M's Facebook wall data was fetched using SODATO. The fetch request was generated from 01-January-2009 to 31-July-2013 using the web interface of SODATO. This triggered the console fetch job that retrieved the data using Facebook's Graph API and stored into the local MS SQL Server 2008 R2 database. Based on our netnographic observations of the actual Facebook wall of H&M, we specified a fetching window of 1 day. The fetched data was aggregated by the Metric Preparation Console (Figure 4).

7.1.1.1 Social Graph (Facebook posts/likes/comments by H&M and Non-H&M actors)

For this report, with regard to the social graph, for artefacts, we selected posts, likes, and comments and for actors, we divided all Facebook users posting/liking/commenting on H&M's Facebook wall into two types of actors: H&M and Non-H&M. In order to compute the relevant social graph measures, we built and executed custom SQL scripts and stored results aggregated by the quarter into the database.

7.1.1.2 Social Text (sentiment of posts/comments by H&M and Non-H&M actors)

Google Prediction API[6] was utilized in order to calculate sentiments for the posts and comments on the wall. An account was created with access to Google Prediction API. Google Prediction API provides a web interface as well as RESTful API access to the service. Configuration for computation of sentiment began with the setting up a model which was trained with the human-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's 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. Console explorer was then used to execute Prediction API command training data command.

[4] http://about.hm.com/en/About/Investor-Relations/Financial-Reports/Financial-Reports.html
[5] http://about.hm.com/content/dam/hm/about/documents/en/cision/1634480_en.pdf
[6] https://developers.google.com/prediction/

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

7.2 Results

We organize the results into two sub-sections. First we present descriptive statistics and then present the correlational analysis statistics results on the relationship between measures of H&M's quarterly revenues and measures of social graph and social text.

	N	Min	Max	Sum	Mean	Std. Deviation
Revenues (millions SEK)	18	27282	37930	585589	32532.72	3176.388
Social Graph Measures						
Total Posts	18	509	12104	98461	5470.06	2950.985
Posts by H&M	18	7	175	1420	78.89	38.423
Posts by Non-H&M	18	502	11999	97041	5391.17	2932.096
Total Likes	18	6	1895557	7029851	390547.28	614728.618
Total Likes on H&M Posts	18	6	1459642	5291025	293945.83	445010.526
Total Likes on Non-H&M Posts	18	0	584724	1738826	96601.44	176637.248
Total Comments	18	44	37242	253900	14105.56	10397.186
Total Comments on H&M Posts	18	44	18293	147954	8219.67	5011.945
Total Comments on Non-H&M Posts	18	0	19033	105946	5885.89	5860.389
Social Text Measures						
Positive Sentiment-Total Posts	18	309	4584	40638	2257.67	1069.975
Negative Sentiment-Total Posts	18	77	5626	24326	1351.44	1242.817
Neutral Sentiment-Total Posts	18	123	3657	33328	1851.56	1051.699
Positive Sentiment-Posts by H&M	18	1	27	197	10.94	5.876
Negative Sentiment-Posts by H&M	18	1	29	258	14.33	7.685
Neutral Sentiment- Posts by H&M	18	5	123	965	53.61	28.755
Positive Sentiment- Posts by Non-H&M	18	308	4573	40441	2246.72	1069.283
Negative Sentiment- Posts by Non-H&M	18	76	5597	24068	1337.11	1238.996
Neutral Sentiment- Posts by Non-H&M	18	118	3592	32363	1797.94	1037.627
Positive Sentiment- Comments on Posts by H&M	18	26	7357	64509	3583.83	2073.563
Negative Sentiment- Comments on Posts by H&M	18	6	7037	38136	2118.67	1535.653
Neutral Sentiment- Comments on Posts by H&M	18	12	5267	41592	2310.67	1430.012
Positive Sentiment- Comments on Posts by Non-H&M	18	0	6863	35471	1970.61	2143.297
Negative Sentiment- Comments on Posts by Non-H&M	18	0	4937	25515	1417.5	1503.582
Neutral Sentiment- Comments on Posts by Non-H&M	18	0	7672	44720	2484.44	2265.156

Table 1. Descriptive Statistics of Quarterly Revenues and Social Data Measures for H&M

7.2.1 Descriptive Statistics

Overall, the revenue data consists of quarterly revenues for H&M for 2009, 2010, 2011, 2012 and the 2 quarters of 2013 (total of 18 quarters). Facebook social data fetched by SODATO was from 01-Jan-2009 to 31-July-2013.The Facebook data corpus consists of 100,465 posts, 262,588 comments on posts, and 7, 779,411 likes on posts and comments. A total of 3,134,249 unique Facebook ids/users were present in the corpus (that is, each at least performed an action of post, comment, or, like). We assume that each Facebook id is a unique actor in the real world and categorize them into H&M and Non-H&M actors.

For sentiment analysis, we investigated the distribution of positive, negative and neutral sentiment on posts and comments for H&M as well as Non-H&M actors. Table 1 presents the cumulative values for the revenues and social graph and social text measures for H&M's facebook wall across the 18 quarters. Table 1 presents the descriptive statistics for the raw data values of the different social graph and social text measures from the Facebook wall of H&M and its reported revenues for 18 quarters.

	Spearman's rho	**Zscore: Sales (millions) SEK**
Zscore: Total Posts	Correlation Coefficient	0.748**
	Sig. (2-tailed)	0.000
Zscore: Posts by H&M	Correlation Coefficient	0.474*
	Sig. (2-tailed)	0.047
Zscore: Posts by Non-H&M	Correlation Coefficient	0.748**
	Sig. (2-tailed)	0.000
Zscore: Total Likes	Correlation Coefficient	0.835**
	Sig. (2-tailed)	0.000
Zscore: Total Likes on H&M Posts	Correlation Coefficient	0.839**
	Sig. (2-tailed)	0.000
Zscore: Total Likes on Non-H&M Posts	Correlation Coefficient	0.800**
	Sig. (2-tailed)	0.000
Zscore: Total Comments	Correlation Coefficient	0.723**
	Sig. (2-tailed)	0.001
Zscore: Total Comments on H&M Posts	Correlation Coefficient	0.544*
	Sig. (2-tailed)	0.020
Zscore: Total Comments on Non-H&M Posts	Correlation Coefficient	0.822**
	Sig. (2-tailed)	0.000

' Correlation is significant at the 0.01 level (2-tailed)

* Correlation is significant at the 0.05 level (2-tailed)

Table 2. Correlational Analysis of Quarterly Revenues and Social Graph Measures for H&M

7.2.2 Correlational Analysis To statistically assess the relationship between social data measures of H&M and its real-world business outcomes, we calculated Spearman's rank correlation coefficient (Spearman's rho) between quarterly revenues and selected social graph and social text measures. We selected the Spearman?s rank correlation over Pearson product-moment correlation coefficient as we neither assume the normal distribution of the variables of analytical interest nor posit a linear relationship between them.

Standardized scores were used instead of the nominal values of quarterly sales in millions of Swedish Kroner, raw counts of social graph measures such as posts, comments and likes, and raw counts of social text measures of sentiment classification for posts and comments.

7.2.2.1 Social Graph Analytics

As can be seen from Table 2 below, statistically significant positive correlations were observed between quarterly revenues and each of the social graph measures. In particular, strong correlations were observed between quarterly sales and total likes, total likes on the company's posts as well as users' posts and total comments on users' posts.

7.2.2.2 Social Text Analytics

Table 3 presents the results from the correlation analysis for selected social text measures. As can be seen from 3, correlational analysis results for sentiment expressions of posts and comments by H&M and its users were a mixed bag.

Statistically significant strong positive correlations were observed for positive sentiment expression only for Comments on Posts by Non-H&M users on the facebook wall. On the other hand, strong correlations were observed, surprisingly, for the negative sentiment expressions on Total Posts, Posts by Non-H&M and Comments on Posts by Non-H&M facebook users.

	Spearman's rho	**Zscore: Sales (millions) SEK**
Zscore: Positive Sentiment-Total Posts	Correlation Coefficient	0.459
	Sig. (2-tailed)	0.055
Zscore: Negative Sentiment-Total Posts	Correlation Coefficient	0.812**
	Sig. (2-tailed)	0.000
Zscore: Neutral Sentiment-Total Posts	Correlation Coefficient	0.851**
	Sig. (2-tailed)	0.000
Zscore: Positive Sentiment-Posts by H&M	Correlation Coefficient	0.756**
	Sig. (2-tailed)	0.000
Zscore: Negative Sentiment-Posts by H&M	Correlation Coefficient	0.501*
	Sig. (2-tailed)	0.034
Zscore: Neutral Sentiment- Posts by H&M	Correlation Coefficient	0.404
	Sig. (2-tailed)	0.097
Zscore: Positive Sentiment- Posts by Non-H&M	Correlation Coefficient	0.447
	Sig. (2-tailed)	0.063
Zscore: Negative Sentiment- Posts by Non-H&M	Correlation Coefficient	0.818**
	Sig. (2-tailed)	0.000
Zscore: Neutral Sentiment- Posts by Non-H&M	Correlation Coefficient	0.845**
	Sig. (2-tailed)	0.000
Zscore: Positive Sentiment- Comments on Posts by H&M	Correlation Coefficient	0.571*
	Sig. (2-tailed)	0.013
Zscore: Negative Sentiment- Comments on Posts by H&M	Correlation Coefficient	0.467
	Sig. (2-tailed)	0.050
Zscore: Neutral Sentiment- Comments on Posts by H&M	Correlation Coefficient	0.525*
	Sig. (2-tailed)	0.025
Zscore: Positive Sentiment- Comments on Posts by Non-H&M	Correlation Coefficient	0.827**
	Sig. (2-tailed)	0.000
Zscore: Negative Sentiment- Comments on Posts by Non-H&M	Correlation Coefficient	0.822**
	Sig. (2-tailed)	0.000
Zscore: Neutral Sentiment- Comments on Posts by Non-H&M	Correlation Coefficient	0.794**
	Sig. (2-tailed)	0.000

** Correlation is significant at the 0.01 level (2-tailed)

* Correlation is significant at the 0.05 level (2-tailed)

Table 3. Correlational Analysis of Quarterly Revenues and Social Graph Measures for H&M

Next, we discuss the results in terms of the research question and offer substantive interpretations.

8 Discussion

Based on the reported results, we provide a preliminary but promising answer the overarching research question:

To what extent are social data measures of H&M correlated with its quarterly revenues?

With regard to social graph measures, we found statistically significant positive correlations between H&M's quarterly revenues and the social graph measures of posts, likes, and comments for both H&M and Non-H&M users. Global brands such as H&M are increasingly socially constructed in discursive practices in real-world and online contexts such as facebook and twitter.

As the cliché goes, the brands are increasing what consumers are telling each other instead of what the marketers are advertising in terms of offline and online advertisements. In line with Asur and Huberman (2010), we believe that social graph actions such as liking, commenting, sharing, tweeting, re-tweeting, mentioning, rating, reviewing are proxies for the individual user's attention toward the company/brand.

With regard to social text measures, our report surprising findings of strong positive correlations of quarterly revenues with negative sentiments on total posts, posts by Non-H&M users and comments on posts by Non-H&M facebook users. An interpretive qualitative analysis is needed to understand and explain this. We speculate that negative sentiment in itself is not detrimental to the brand identity and business value if it not directed towards the company itself.

In other words, our social text analytics results indicate that sentiment polarity is necessary but not sufficient for predicting business outcomes and methodological advances must be made to also determine sentiment directions. That is, if the sentiment is directed towards the company/brand/ product/service or not. In terms of Decision-Making and Big Data, taken together, social graph and social text measures can be good indicators of social influence [4, 12] in the online communities. For fast moving consumer goods as H&Mwell as luxury brands, social influence plays a critical role in shaping consumers' perception and behaviours.

9 Future Work

The future work can be subdivided into sub sections into two sections as follows,

9.1 Formal Modeling

First of all, there is need for extending the 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 formal model to formalise the whole socio-technical interactions to develop it as a abstract formal model for modeling most of the social media platforms.

Furthermore, modeling social concepts in general involves fuzziness. For example, in modeling of Facebook posts in Sec. 5.2, even though comments (artifacts) $r2$ (*"Yes, I love summer"*) and $r3$ (*"I love H & M"*) are both categorised as positive sentiments in favour of *H & M*, but there is difference in levels of expressiveness of the sentiments. In this example, the comment $r3$ is unambiguous and can be categorised as a strong sentiment, where as the comment $r2$ is not a direct indication of positive sentiment in favour of *H & M*, and therefore it needs to be categorised as a weak sentiment. As part of future work, we would like to use Fuzzy set theory to model this type of fuzzy behaviour in the social data.

9.2 Empirical Analysis

Based on the unified conceptual, formal, and technological model for social data presented and discussed, ongoing research in our lab is empirically addressing the following topics in different research fields

1. Discourse analysis of social data from online political interactions (Political Science)
2. Knowledge diagnostics of social data from classrooms (Learning Sciences)
3. Cohort analysis of social data from interactions of fans with sports clubs and their sponsors (Marketing)
4. Correlational analysis of social data from facebook wall interactions and financial performance for companies (Finance)
5. Predictive analytics of social data for sustainability related topics and sentiments (Management)

10 Conclusion

In this report, we have presented an integrated modeling approach for analysis of social data using a conceptual model on social data, a formal model modeling the key concepts of the conceptual model 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 data. 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 concretising the abstract ideas from the conceptual model to schematic model in the process of developing the Social Data Analytics Tool.

Furthermore, we have also presented a case study where we empirically evaluate the relationship between real-world business outcomes and social media interactions of a Global brand, H&M. The report is intended as a demonstrative case study of an integrated approach of theory, model, tool, and analytics of big social data analytics and business outcomes. Towards this end, 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 for social media platforms such as Facebook. We then presented a descriptive model of social data emanating from socio-technical interactions on social media platforms such as Facebook.

The SODATO was used to collect analyse Facebook wall data of H&M from 01-Jan-2009 to 31-July-2013 and consisting of 100,465 posts, 262,588 comments on posts, 7, 779,411 likes on posts and comments across 3,134,249 unique Facebook ids/users We then empirically analysed the relationship between social data measures and revenues for a global brand, H&M and found statistically significant correlations for measures of social graph (posts, likes, comments) as well as social text (positive, negative or neutral sentiment expressions in posts and comments). In contrast to prior related work that uses either analytical methods for sentiment analysis of the content or the social network analysis techniques to study social relationships, our approach in this paper is novel in the sense that we use both social graph analysis combined with social text analysis to empirically evaluate correlations between the social data and financial of the companies. Furthermore, as far as we know, we are the first to use Facebook data in measuring the relationship to business outcomes instead of social data from twitter, internet search trends, website visits, blogs or discussion forums.

Bibliography

[1] S. Asur and B.A. Huberman. Predicting the future with social media. In *Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on*, volume 1, pages 492–499, 2010. 3, 4

[2] Steven Bankes, Robert Lempert, and Steven Popper. Making computational social science effective: Epistemology, methodology, and technology. *Soc. Sci. Comput. Rev.*, 20(4):377–388, October 2002. 3

[3] Hsinchun Chen, Roger H. L. Chiang, and Veda C. Storey. Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4):1165–1188, 2012. 1

[4] Robert B. Cialdini and Noah J. Goldstein. Social influence: Compliance and conformity. *Annual Review of Psychology*, 55(1):591–621, 2004. PMID: 14744228. 14

[5] Claudio Cioffi-Revilla. Computational social science. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(3):259–271, 2010. 2

[6] William S. Cleveland. Data science: an action plan for expanding the technical areas of the field of statistics. *International Statistical Review*, 69(1):21–26, 2001. 2

[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*, 214(1):325–346, 2012. 2, 3

[8] Alfredo Cuzzocrea and MohamedMedhat Gaber. Data science and distributed intelligence: Recent developments and future insights. In Giancarlo Fortino, Costin Badica, Michele Malgeri, and Rainer Unland, editors, *Intelligent Distributed Computing VI*, volume 446 of *Studies in Computational Intelligence*, pages 139–147. Springer Berlin Heidelberg, 2013. 2

[9] M. Goldberg, S. Kelley, M. Magdon-Ismail, K. Mertsalov, and A. Wallace. Finding overlapping communities in social networks. In *Social Computing (SocialCom), 2010 IEEE Second International Conference on*, pages 104–113, 2010. 4

[10] Anthony Hall. Realising the benefits of formal methods. In Kung-Kiu Lau and Richard Banach, editors, *Formal Methods and Software Engineering*, volume 3785 of *Lecture Notes in Computer Science*, pages 1–4. Springer Berlin Heidelberg, 2005. 3

[11] Juuso Karikoski and Matti Nelimarkka. Measuring social relations with multiple datasets. *IJSCCPS*, 1(1):98–113, 2011. 4

[12] H.C. Kelman. Compliance, identification, and internalization: Three processes of attitude change. *Journal of Conflict Resolution*, 2:51–60, 1958. 14

[13] David Krackhardt. Cognitive social structures. *Social Networks*, 9(2):109–134, June 1987. 4

[14] D. Lazer. Life in the network: the coming age of computational social science. *Science (New York, NY)*, 323(5915):721, 2009. 2

[15] Mike Loukides. *What is Data Science ? and people that turn data into products What is Data Science ?* O'Reilly (An OReilly Radar Report), 2010. 2

[16] Mike Loukides. *What Is Data Science?* O'Reilly Media, 2012. 2

[17] Owen Macindoe and Whitman Richards. Comparing networks using their fine structure. *International Journal of Social Computing and Cyber-Physical Systems*, 1(1):79–97, 2011. 4

[18] Telmo Menezes, Camille Roth, and Jean-Philippe Cointet. Finding the semantic-level precursors on a blog network. *IJSCCPS*, 1(2):115–134, 2011. 4

[19] Andrzej Nowak, Agnieszka Rychwalska, and Wojciech Borkowski. Why simulate? to develop a mental model. *Journal of Artificial Societies and Social Simulation*, 16(3):12, 2013. 3

[20] Noboru Ohsumi. From data analysis to data science. In *Data Analysis, Classification, and Related Methods*, pages 329–334. Springer Berlin Heidelberg, 2000. 2

[21] Scott P. Robertson, Ravi K. Vatrapu, and Richard Medina. Off the wall political discourse: Facebook use in the 2008 u.s. presidential election. *Info. Pol.*, 15(1,2):11–31, April 2010. 3

[22] Scott P. Robertson, Ravi K. Vatrapu, and Richard Medina. Online video â”friends”, social networking: Overlapping online public spheres in the 2008 u.s. presidential election. *Journal of Information Technology & Politics*, 7(2-3):182–201, 2010. 3

[23] Jordi Sabater and Carles 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*, AAMAS '02, pages 475–482, New York, NY, USA, 2002. ACM. 4

[24] Noel M Tichy, Michael L Tushman, and Charles Fombrun. Social network analysis for organizations. *The Academy of Management Review*, 4(4):507–519, October 1979. 4

[25] Ravi Vatrapu. Understanding social business. In *Emerging Dimensions of Technology Management*, pages 147–158. Springer, 2013. 3

[26] Ravi Vatrapu, Abid Hussain, Daniel Hardt, and Zeshan Jaffari. Social data analytics tool: A demonstrative case study of methodology and software. In R Gibson, editor, *Analysing Social Media Data and Web Networks*. Palgrave Macmillan, 2014. 5

[27] Ravi K. Vatrapu. *Technological Intersubjectivity and Appropriation of Affordances in Computer Supported Collaboration*. PhD thesis, University of Hawaii at Manoa, USA, 2007. AAI3302125. 5

[28] Ravi K. Vatrapu. Explaining culture: An outline of a theory of socio-technical interactions. In *Proceedings of the 3rd International Conference on Intercultural Collaboration*, ICIC '10, pages 111–120, New York, NY, USA, 2010. ACM. 5

[29] Daniel Zeng, Hsinchun Chen, R. Lusch, and Shu-Hsing Li. Social media analytics and intelligence. *Intelligent Systems, IEEE*, 25(6):13–16, 2010. 1

[30] Justin Zhan and Xing Fang. Social computing: the state of the art. *International Journal of Social Computing and Cyber-Physical Systems*, 1(1):1–12, 01 2011. 4