Big Social Data Analytics: Opportunities, Challenges and Implications on Society

Devakunchari R, Anna University (MIT campus), India

Valliyammai C, Anna University (MIT campus), India

Abstract

Big social data analysis is an interdisciplinary domain that uses advanced analytics to identify the key influence in a communication network that shrinks and grows rapidly in a near real-time fashion. It brings together various disciplines such as social media analysis, information retrieval techniques, reasoning, natural language processing (opinion mining), graph mining and analysis, linguistics, machine learning, multimedia management and big data processing. The role of big social data analysis in the growth of a business is larger than what meets the eye, but many enterprises are unaware of how this data influence their expansion. The analysis of information gathered from consumer transactions, communication devices, online activities and streaming devices are used to uncover hidden patterns. The right tools and a foolproof strategy are highly important to utilize information from social data to the fullest. The power of social media along with the emerging big data technology offers abundant opportunities for achieving business growth and if channeled properly, it could be leveraged for the betterment of society. The main goal of this paper is to emphasize the opportunities and challenges provided by dynamic social network data with its implications on the society

Keywords: Big data, Communication, Online social networks, Society

Introduction

In this iconic age of social media, people are transforming into data agents who share even the tiniest detail of their daily activity on social media. Online Social Networks (OSNs) churn out data in large volumes with every passing second. Globally, people are responsible for multiplying the data into twice as much for every 1.2 years. Statistics reveal that 5.1 billion people own mobile phones or cellular devices which are used to share billions of texts, stream numerous videos and search exhaustively on the internet every single day. To analyze the volumes of data shared on the Internet and to channel them in the right direction is the most intriguing topic of research in the near future. Borko Furht (2010) presented a collection of research articles on the importance of analytics on social networks.

What Is Big Social Data Analytics?

To simply put, big social data is the data generated by social networks. There is about 2.5 Exabyte of data being shared every second, loaded with endless stream of information. Big social data analytics is an important tool which can be used to understand the insights behind the information revealed by a person through social media. It is an approach which combines various statistical methods, machine learning, sentiment analysis and data mining techniques in forecasting, predicting people and analyzing trends. The Fig. 1 is a word cloud of the terms highly cohesive with social networks and its analytics using big data.

Big Social Data Characteristics

Online social networks (OSN) are one of the most important sources of big data as described by Min Chen, Shiwen Mao and Yunhao Liu (2014).

Volume: OSNs pour out huge volumes of information continuously and the information gathered instantly becomes the source for better business growth. Big social data does not just imply terabytes, petabytes or exabytes of data, but the context in which the gathered data are utilized.

Velocity: OSNs also has a high rate of data generation due to the exponential increase in streaming of social data in online networks based on the capacity of systems.

Variety: The data produced online is vast and the greatest challenge lies in categorizing the large quantity of data and processing them to retrieve potentially useful information. The data generated is of different types such as *Structured data* (data stored in a pre-defined schema like relational databases), *Unstructured data* (data which do not fall into any format or pattern like photos, images, audio and videos) and *Semi-Structured data* (does not have any prescribed data format, but arranged in some common pattern like emails, log files and word documents). Social Media is important for companies and institutions. The value of social media data lies in,

- What do most people do? -> Indicating behavior.
- How do the people feel? -> Indicating their sentiment.

Big social data contains three practices, namely monitor, predict and share. *Monitoring* includes listening, filter, sort, search, analyzes, action and refinement of social data. *Predicting* includes past, present and future findings. *Sharing* includes network visualization with simulation of information flow.

Fig.1.Terms highly correlated with big social data analytics

Analytics Tools

Apache Hadoop [1], the leading technology used for big data analytics is an open source, powerful and reliable tool which comprises of two components, namely, Hadoop Distributed File System and Map Reduce framework [2]. Apache spark [3] and Apache storm [4] are used for real-time distributed computation. Facebook uses presto, which is an open source SQL query engine that helps in querying data at their sources using Hive, Cassandra, relational databases or any other software. Twitter acquired Gnip, a social data company which collects the tweets, filters and resells the data which is valuable for advertisers and marketing campaigns of companies. IBM collects the data distributed on Twitter and Facebook regarding human thoughts, complaints and comments over a specific topic to derive real-time insights. This will help people to improve their marketing campaigns, promotions, discounts and pricing offers. Agreste, De Meo, Ferrara, Piccolo, and Provetti [5] created aNobii, a heterogeneous structure, analyzing the user interests on tag based, group based and wish-list based profiles. The Fig.2 Infographic depicts the various tools and their solutions given for analyzing big social data with different social networks.
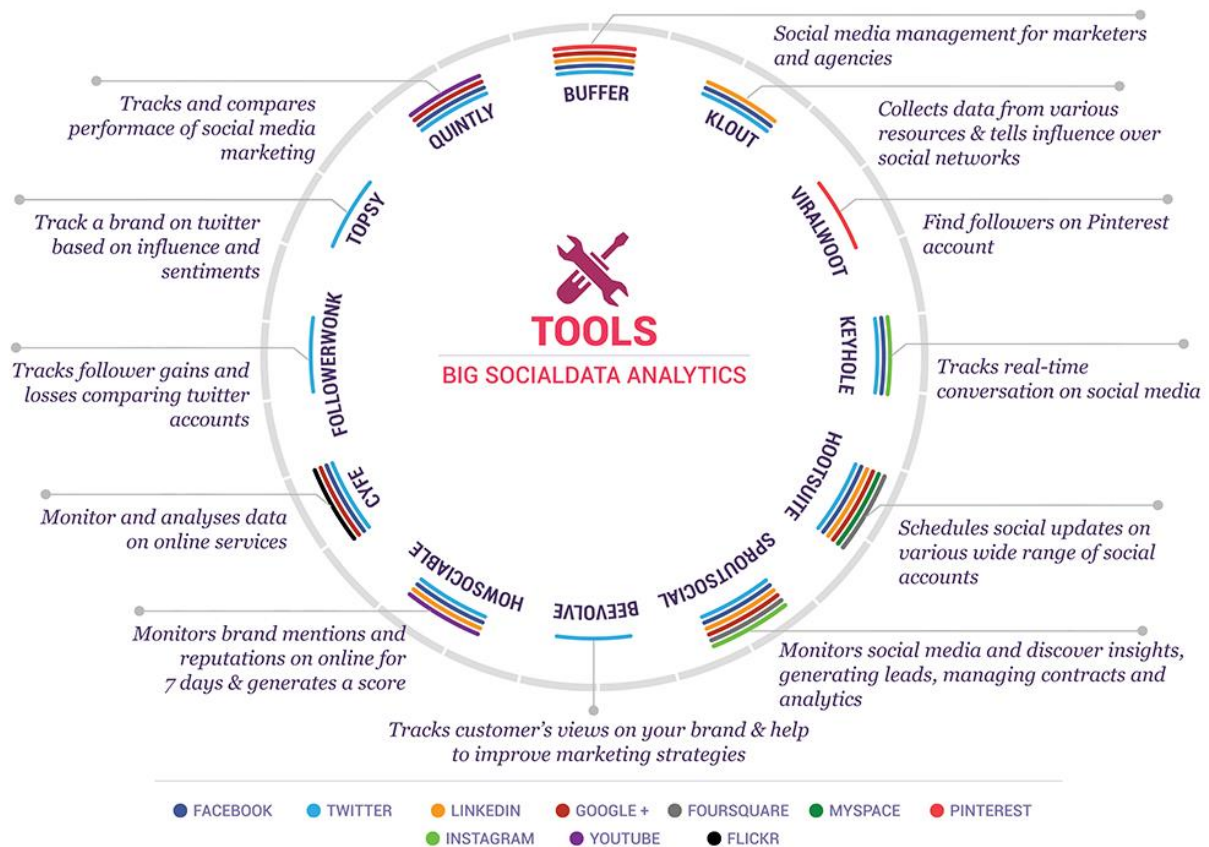
Fig.2. Big social data analytics tools of OSNs

Opportunities

Big data analytics techniques can be used by companies to understand their customers, find hidden opportunities, mitigate frauds and inspire millions of startups for analytics.

Business Opportunities

Marketing through social media creates lots of opportunities by attracting customers with engaging content and creating communities to leverage their products. It attracts more customers with campaigns and promotions within the communities, thereby increasing the product exposure. The strategies employed through the social media explosion results in gaining of customers at a highly exponential rate. It is vital to compare the campaigns made by an organization against their competitors in social media to take necessary steps to promote their products better. Analytics over the social data make clear discussion on the current trends followed by the people.

Social media makes money out of its data. The post or tweets shared about a product will provide a Return On Investment (ROI) by visiting the product's site and purchasing it. Social data are vital in analyzing the possibilities for customer retention. It is challenging to retain a customer amidst of growing competitors in the market. Keen analytics on social data helps in tackling this challenge. Personalized recommendations for users are made possible based on the analytics over their social data. People feel contented with their preferred products being recommended and they would in turn suggest the product to somebody else, step by step increasing the sale of a product. Providing customer support through social media is highly rewarding for the business and the option seems attractive for a customer as well. But, as in Fig.3, revealing the Personally Identifiable Information (PII) on social networks makes them

Fig.3.Business opportunities associated with social networks employing big social data analytics

vulnerable to many attackers which leads to the misuse of identity in all kinds of malpractices. Krombholz, Hobel, Huber and Weippl (2015) described about the kinds of social engineering attacks possible in online social networks.

Recent Activities

- *IBM* provides a wide variety of services such as,
  - Social media analytics as a service - consuming social media data and understanding the trends, opinions and attitude of the user over online
  - Predicting the likeliness of user and recommend their satisfactory products
  - Creation of campaigns and promotions of preferred products in order to attract and satisfy customers
  - Identifying most dominant influences on social networks
  - IBM also partnered with Twitter to do big data analytics to derive insights from their user's information. This will largely improve decision making. Different domains of various industries can refer across boundaries for large set of social purposes
- *Twitter* has a streaming visualization tool which reveals the firehouse of information from Twitter

- On *Facebook,* based on semantic analysis of data, regressions and predictions of trends about the products are done
- Retailer databases are used for logistics, financial and health data analysis

Research Opportunities

Due to increase in usage, spread, and adoption of cloud computing and Internet of Things, along with the integration of social network platforms, there is an overwhelming data. There is a scarcity of tools used for structuring the social data and inferring knowledge out of it. Researchers face challenges in understanding the conceptual tools available for big data. The biggest opportunity in social media revolution is that the availability of free large open data sets for all kinds of research projects including text classification and graphical analysis. Data Science has reduced the works of agencies in social media. Each organization has their own analytics research team to analyze social data for their own benefits. Chen, Chiang and Poor (2013) described about the opportunity for new research directions towards technological networking through social networks.

Many researchers around the world are claiming that they are doing big data analytics. But, many of them lack enough subject knowledge of. The major aspects considered for social data analytics: how to get the data, what to get out of those data and how to put the information to use. Some people specialize in collecting data from various resources. Some are skilled at analytical processing over the data obtained. Some are good at identifying the right channel for the information retrieved from the data. Companies struggle to have all the three aspects of big data to make good sense out of it. Social media companies try to acquire data from those sites and offer it for sale to other analytical companies. Dean and Ghemawat (2008) described about the programming model with mapper and reducer to ease the work on processing large scale social networks. Facebook has started collecting data from users on acceptance of new terms, including personal data, photos, locations, smart phone travel locations and like interests on Facebook server for the purpose of providing personal interests over a specific thing to the companies.

In order to unleash the positive outcomes of social big data, researchers need to uplift their skills, including the learning of statistical analysis, machine learning techniques like association rules, classification, clustering, time series analysis and regression analysis to

identify the social and behavioural changes. Al-Jarrah, Yoo, Muhaidat, Karagiannidis and Taha (2015) reported the overview of machine learning techniques with respect to experimental aspects, large scale data intensity, energy efficiency and computational requirements. Researchers need to make sense of the data by utilizing the complex relationships between nodes in a social network, their platform, third party systems and databases. It is also needed to face the challenges arising due to geo-spatial and spatio-temporal data from OSNs as they are represented in different formats.   The location of the user in the earth can be inferred from the geo-spatial data and (Global Positioning System) GPS data can be obtained from GPS enabled ubiquitous devices. Embedded sensors connect physical world and the digital world with a timestamp and spatial location. Analyzing the spatio-temporal data provide a particular pattern of a user at a certain time interval. This will be helpful in identifying their future actions and anomaly behaviour. Vision recognition extracts meaningful information from image data sets.

Natural Language Processing (NLP)

Many researchers do natural language processing of social media data which will help to analyze the behaviour of the user with respect to folksonomy, tagging behaviour of users, analysis of current trends, detection of communities based on similar interests, analysis of privacy schemes, structuring of blog contents, trust computation and analysis etc. The result of the processing may greatly help to better understand the nature of each user in social networks. Making opinions on products and recommendations for personalized schemes can be suggested using natural language processing.

There are kinds of NLP like individual element analysis, relationship analysis and event analysis, which consumes lots of social data. User's opinions about a particular topic either a social issue, a product, an organization, a trend or a lifestyle through social data around the world can be utilized for sentiment analysis.

Society

In the case of any natural calamity, the updates on social media can help in identifying potential victims of the affected area. The information shared on social media is highly useful for critical data discovery, awareness about the situation, and analytics for gaining the inference. Social data are used for creating awareness during emergency situations. Cameron

and his researchers developed a prototype capable of making use of tweets on Twitter to create an emergency awareness during a crisis by analyzing the tweets using text classifier and clustered similar tweets together. There are watch officers allotted for aggregating and summarizing the realization of tweets. This will help to address the emergency situation by informing the respective officials to take precautionary steps. The Police department can be benefited from the social data to find a missing person, to identify the suspects, gaining a reputation for the department and to spread awareness about any critical issues.

Crowdsourcing

Social media have also stepped forward to provide relief for the victims of Chennai floods in India, through all means. The kind of relief includes crowd sourced information regarding the flooded streets in the city, sharing emergency numbers, contact details for food and shelter, etc., which can be shown by a Facebook post as in Fig.4 (a). Pretz (2013) explored the usage of social networks to provide awareness during the emergencies. Arun and team (2015) from Map box have generated a map-based tool for visualizing the flooded streets and also to mark the streets as flooded. They have used crowdsourcing technology to inform a person in Singapore about his parents in Chennai. Relatives and friends are able to know about the inhabitants of Chennai through social media. Facebook launched an application called *safety check* as in Fig.4 (b), for people to know whether the person is in crisis or non-crisis region. Due to the problems in telecommunication, friends of a person can also mark that person safe because of the network problem. Other friends in social media came to know about the current situation of the person in Chennai. Similarly, Facebook used this safety check application not only for natural, but also for man-made disasters after the Paris attack.
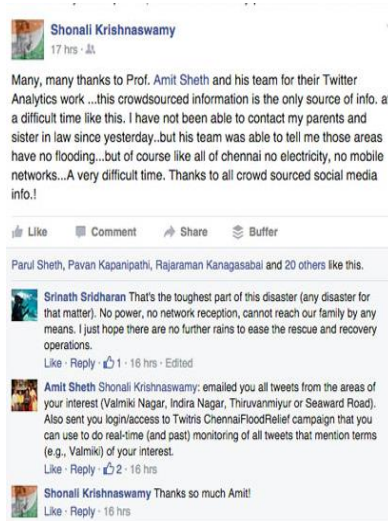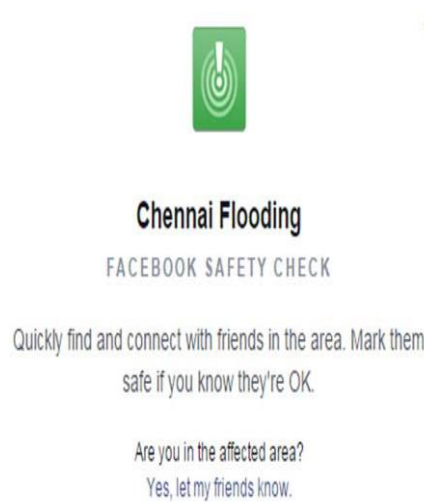
(a)   (b)

Fig.4 Social media and crowdsourcing as a disaster management tool

Sociomine is to study the wisdom of the crowds which involves two steps. One is to model how information spreads across a social network, how people learn or herd. Another is to correlate between crowd behavior and exogenous phenomena.

Challenges

One cannot judge the interests of people based on likes for the status update, clicks on a link, or retweets for a message. More deep research is required to understand what the social media reveals and the inference collected legitimately based on the analysis of social data. Each social network operates in a different way with respect to delivering the information through public shared data. The huge amount of public data shared does not reveal more insight. There are lots of possibilities for misunderstanding which creates difficulty in producing accurate results. Aggregation of large datasets does not show out all the human activities. These aggregated data are partial and so it has to be filtered, distorted and hence complex and reflective. The data from the social networks have the greatest impact on the design of technological networks with respect to e-commerce activities, trust, cooperation, friendship, etc.

The challenges in handling big data from social networks are:

- Mastering three aspects of social data, including collection, analytics and visualization

- Deciphering the right data to make the best sense out of it

- Non-traditional approaches have to be utilized for processing unstructured data

- Collating and segregating the data for better understanding of the context hidden in the social data

- Synchronization of data from various sources

- Lack of skilled professionals capable of working with new technologies for finding business insights

- Working with social data and making use of it for other organizations requires new creative ways for collaborating with social networks

- Security concerns over data hinder the availability of specific data available in social media

Implications on Society

Social media have influenced the society in numerous ways. It has changed in the way of communication and interaction on day to day life. Social networks are duly created for serving the purpose of communication even though there is a downside getting involved with social networking.

Addiction

- Hours spent on social media every day lead to distraction of an individual resulting in a decrease in productivity

- Teenagers are influenced by the attention gained from social media for their updates of daily happenings in their life and their popularity on social media serves as an ego-boost

- Constantly connected to the internet through smart phones, laptops and other smart gadgets

Cyber-bullying

- Children, teenagers and women are more exploited to the practice of cyber bullying which leads to fear of ramification, feeling alone and frightened

- Social networks are becoming the breeding grounds for bullying because of its open platform and lack of privacy by making open posts and sending private messages

- According to the Washington post, it was reported that 9 out of 10 teenagers in Facebook were bullied in the social networking sites

- Bullying through social media data constitutes of streams of abusive comments, derogatory posts on the user's wall, posting harassing negative comments on the user's wall and making fun of others using their pictures or videos

Business- Productivity reduction

- Many organizations have blocked social networks to reduce the addiction and distraction of the employees in the office, which in turn would affect the productivity of the organization

- Many organizations try to attract clients by contacting them with social media through which employees may be highly distracted showing more interests on their friends, updates and likes etc., than actual work, thereby reducing the productivity in business

Reduction in academic performance

- The present generation is highly attracted to online networking and spend too much of their time in social media

- Distractions due to social networks result in lower academic performances

- The highly private content shared on social media due to vulnerability may result in negative consequences.

- Reading practices of students have fallen drastically caused by an unhealthy practice of skimmed reading due to the influence of short texts on social media

Health

- Individuals try to compare their lifestyle with others and feel sorry for themselves due to their inability to share   interesting happenings in their life leading to poor mental health and low self esteem

- People try to relax by using social media, ironically, it makes them busy and stressful

- Children who use social media daily have the high impact on using tobacco, alcohol, etc.

- Frequent users of social media are found unhappy due to lack of satisfaction with what they are on the social network

Privacy Leakage

- Companies do a background check on a person's social media activity before recruiting them

- Intimate details of an individual are being lost by exposing more about their personal lives in social media

- Social media is used by the people to expose their personal information like name, birth, age family members, hobbies and interests, their work histories, likes and dislikes through third party APIs, tools and frameworks. As discussed by Dwyer (2011), Google and Facebook extract the online behavior of their users to increase their revenue.

- The dark side of big social data is the shortage of privacy and security policies in social networks. There are many threats caused by the frauds due to the exposed data in social networks as discussed by Fire, Goldschmidt and Elovici (2014). Zhang, Sun, Zhu and Fang (2010) reported about the privacy and security concerns rising due to the usability and sociability existing in social networks. With some knowledge about the users in social networks, there are possible privacy breaches and attacks on social data.

- The potential victims may be targeted or unknown by an imposter who utilizes the loopholes within the same OSN or from different OSN. Some of the threats, their potential victims and their possible solutions are given in Fig. 5.
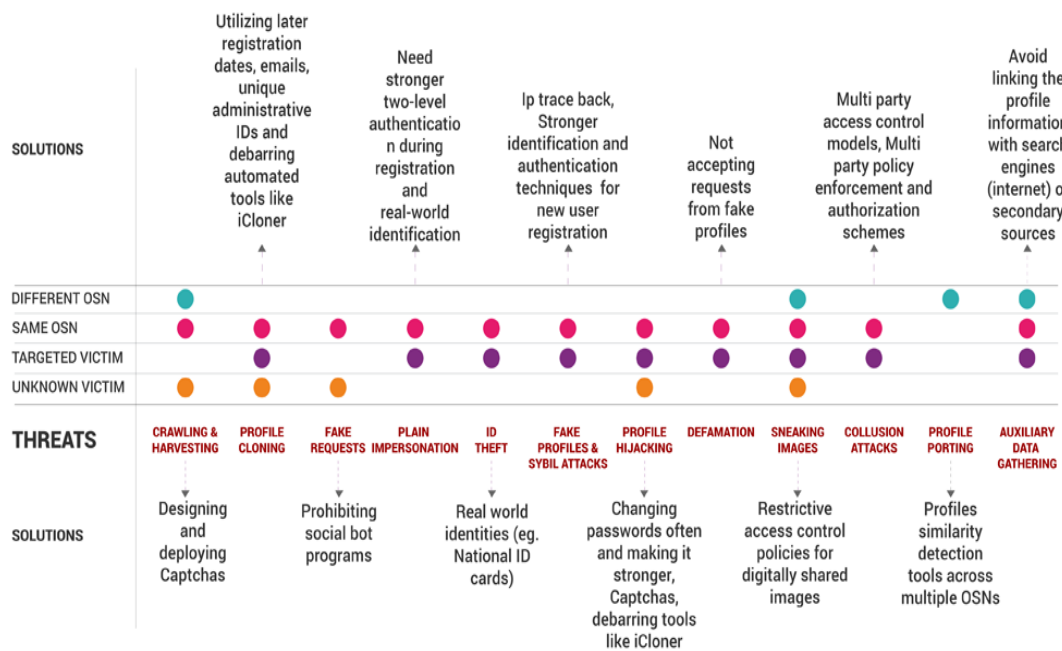

Fig.5. Threats caused due to social data leakage in OSN

Relationships

- People try to reconnect with old friends through social networks

- Due to text messaging, photos and posts, social data cause occasional jealousy and diffidence between partners
- Public display of affection through social forums by romantic partners might make a few uncomfortable
- People become less interactive with their family by spending excessive time saying stories to school friends, resulting in breaking up of relationships

Military

- Young soldiers should be responsible in the manner in which they portray themselves in social networks
- Professional tactics of being soldiers, both offline and online are required to maintain the dignity of the army in the society
- Soldiers should be careful to not disclose any confidential information on social networks pertaining to their job

Conclusion

Social networks try to bring the world closer. Data crawling through social networks is highly important. The information observed from the study of social data can help us in a great way for the recommendation, personalized marketing, customized promotions, etc. People will enjoy their activities on the internet when they could retrieve all the relevant and personally liked things displayed on their social networks and other search engines. Highly relevant technologies and expertise are required for enhancing the smartness in the technology through analytics.

Acknowledgments

References

Agreste, S, De Meo, P, Ferrara, E, Piccolo, S and Provetti, A. (2015). Analysis of Heterogeneous Social Network of   Humans   and   Cultural   Objects. *IEEE Transactions on Systems, Man and Cybernetics: Systems, 45* (4), 559 -   570. http://dx.doi.org /10.1109/TSMC.2014.2378215

Al-Jarrah, O.Y, Yoo, P.D, Muhaidat, S, Karagiannidis, G.K, Taha, K (2015). Efficient Machine Learning for Big   Data: A Review. *Science Direct-Elsevier, Big Data, Analytics, and High-Performance Computing, 2* (3),   87–93. http://dx.doi.org/10.1016/j.bdr.2015.04.001

Arun Ganesh. (2015, Dec.2). *Crowdsourcing flood data for Chennai* [Blog Post]. Retrieved from:https://www.mapbox.com/blog/chennai-flood-map/

Borko Furht. (2010), *Handbook of Social Network Technologies and Applications*, New York, Springer.   http://dx.doi.org /10.1007/978-1-4419-7142-5

Chen K.C, Chiang, M, Poor, H.V. (2013). From Technological Networks to Social Networks. *IEEE Journal on Selected   Areas   in   Communications,   31* (9),   548-572. http://dx.doi.org/10.1109/JSAC.2013.1308.

Dan Worth (2015, March), *IBM teams with Twitter for social media big data analysis* [Blog Post], Retrieved from: http://www.v3.co.uk/v3-uk/news/2400032/ibm-teams-with-twitter-for-social-media-big-data-analysis.

Dean, J and Ghemawat, S. (2008). Map Reduce: Simplified data processing on large clusters. *Communications of*

the ACM, 51 (1), 107-113. http://dx.doi.org /10.1145/1327452.1327492

Dwyer, C. (2011). Privacy in the Age of Google and Facebook. *IEEE Technology and Society Magazine, 30* (3), 58 - 63. http://dx.doi.org/10.1109/MTS.2011.942309

Fire, M, Goldschmidt, R, Elovici, Y. (2014). *Online Social Networks: Threats and Solutions.* IEEE Communication Surveys and Tutorials, 16 (4), 2019 - 2036. http://dx.doi.org/10.1109/COMST.2014.2321628

Krombholz, K, Hobel, H, Huber, M, Weippl, E. (2015). Advanced social engineering attacks. *Elsevier - Journal of Information Security and Applications, 22* (C), 113-122. http://dx.doi.org/10.1016/j.jisa.2014.09.005

Min Chen, Shiwen Mao and Yunhao Liu. (2014). Big Data: A Survey. *Journal on Mobile Networks and Applications, Springer, 19* (2), 171-209. http://dx.doi.org/10.1007/s11036-013-0489-0

Pretz, K. (2013, May), Leveraging *Social Media to Help during Emergencies*, http://theinstitute.ieee.org/technology-focus/technology-topic/leveraging-social-media-to-help-during-emergencies.

Zhang, C, Sun, J, Zhu, X and Fang, Y. (2010). Privacy and security for online social networks: challenges and opportunities*, IEEE Network, 24* (4), 13 - 18. http://dx.doi.org/10.1109/MNET.2010.5510913