

# REVIEW:STUDY ABOUT VISUALIZING BIG DATA IN SOCIAL NETWORK WITH VARIOUS SOCIAL MEDIA ANALYSIS

Preeti Sharma | Dr. Subhajit Ghosh

<sup>1</sup>(Computer Science & Engg, Galgotias University, Greater Noida, India, eng.preeti@gmail.com)

<sup>2</sup>(ComputerScience&Engg,GalgotiasUniversity,GreaterNoida,India,subhajit.ghosh@galgotiasuniversity.edu.in)

**Abstract**—This paper presents a conversations on social media (e.g., wordpress, Twitter, and Facebook) discusses learning experiences— opinions, feelings, and concerns about the learning process. A social network is a set of people (or organizations or other social entities) connected by a set of social relationships, such as friendship, co-working or information exchange. Social network analysis focuses on the analysis of patterns of relationships among people, organizations, states and such social entities. Social network analysis provides both a visual and a mathematical analysis of human relationships. In this paper, we analyzed a workflow to integrate both qualitative analysis and large-scale data mining techniques. We focused on engineering students' Twitter posts to understand issues and problems in their educational experiences.

**Keywords**—visualization big data, social networking, social media, twitter, Facebook etc.

## 1. INTRODUCTION

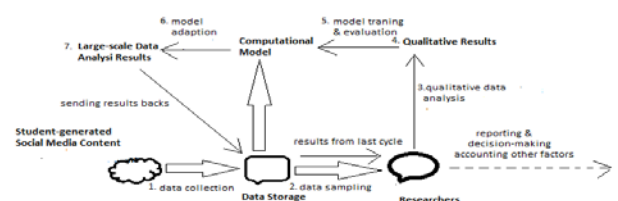
### 1.1 Background

Since their introduction, social network sites (SNSs) such as MySpace, Facebook, Cyworld, and Bebo have attracted millions of users, many of whom have integrated these sites into their daily practices. As of this writing, there are hundreds of SNSs, with various technological affordances (affordances to describe the relationship between human perception and usability.) By identifying social affordances impacting adoption, I begin to build a generalizable description of social factors in adoption. My goal is to describe, in a principled manner, the notion that social affordances impact adoption. Hypothesis and Plan My approach to understanding social and technical factors in adoption is inspired by Norman's work in object perception (Norman, 1988). I appropriate the term 'object affordance,' which is a theory human perception vis-à-vis object interaction, and apply it to social interaction.), supporting a wide range of interests and practices. While their key technological features are fairly consistent, the cultures that emerge around SNSs are varied. Most sites support the maintenance of preexisting social networks, but others help strangers connect based on shared interests, political views, or activities. Some sites cater to diverse audiences, while others attract people based on common language or shared racial, sexual, religious, or nationality based identities. Sites also vary in the extent to which they incorporate new information and communication tools, such as mobile connectivity, blogging, and photo/video-sharing

Social network sites (SNSs) are increasingly attracting the attention of academic and industry researchers intrigued by their affordances and reach.

This special theme section of the Journal of Computer-Mediated Communication brings together scholarship on these emergent phenomena. In this introductory article, we describe features of SNSs and propose a comprehensive definition. We then present one perspective on the history of such sites, discussing key changes and developments. After briefly summarizing existing scholarship concerning SNSs, we discuss the articles in this special section and conclude with considerations for future research. We chose to focus on engineering students' posts on Twitter about problems in their educational experiences mainly because:

1. Engineering schools and departments have long been struggling with student recruitment and retention issues [8]. Engineering graduates constitute a significant part of the nation's future workforce and have a direct impact on the nation's economic growth and global competency [9].
2. Based on understanding of issues and problems in students' life, policymakers and educators can make more informed decisions on proper interventions and services that can help students overcome barriers in learning.



**Figure 1. The workflow for making sense of social media data integrates qualitative analysis and data mining algorithms [9].**

The width of gray arrows represents data volumes wider indicates more data volume. Black arrows represent data analysis, computation, and results flow. The dashed arrows represent the parts that do not concern the central work of this paper. This workflow can be an iterative cycle. Twitter is a popular social media site. Its content is mostly public and very concise (no more than 140 characters per tweet). Twitter provides free APIs that can be used to stream data. Therefore, we chose to start from analyzing students' posts on Twitter.

### 1.2 Toward Visualization based Big Data Discovery

“By visualizing information, we turn it into a landscape that can explore with our eyes, a sort of information map. And when you're lost in information, an information map is kind of useful.” Big data is creating unprecedented opportunities for businesses to achieve deeper, faster insights that can strengthen decision making, improve the customer experience, and accelerate the pace of innovation. But today, most big data yields neither meaning nor value. Businesses are so overwhelmed by the amount and variety of data cascading into and through their operations that they struggle just to store the data—much less analyze, interpret, and present it in meaningful ways.

#### 1.2 Real Value of Big Data

While Apache\* Hadoop\* and other technologies are emerging to support back-end concerns such as storage and processing, visualization-based data discovery tools focus on the front end of big data—on helping businesses explore the data more easily and understand it more fully. Visualization-based data discovery tools allow business users to mash up disparate data sources to create custom analytical views with flexibility and ease of use that simply didn't exist before. Advanced analytics are integrated in the tools to support creation of interactive, animated graphics on desktops, as well as on powerful mobile devices such as the Ultrabook™ and laptops powered by Intel® Core™ processors. End users can view the graphics on the same devices, or on even smaller mobile devices such as tablets or, in limited cases, smartphones.

#### 1.4 Key Features of Visualization-based Data Discovery Tools

- Enable real-time data analysis
- Support real-time creation of dynamic, interactive presentations and reports
- Allow end users to interact with data, often on mobile devices
- Hold data in-memory, where it is accessible to multiple users

- Allow users to share and collaborate securely

#### 1.4.1 Additional Features to Look For

- Ability to visualize and explore data in-database as well as in-memory
- Governance dashboard that displays user activity and data lineage
- In-memory data compression to enable handling of large datasets without driving up hardware costs
- Touch optimization for use with touch-enabled mobile devices such as the Ultrabook™

## 2. LITERATURE REVIEW

Before In this section we will review numerous traditional social network analyses with big data. In addition to the themes identified above, a growing body of study learning addresses other aspects of SNSs, their users, and the practices they enable. For example, Visualizing Big Data in Social Network Analysis, hybrid approach of emotional-based and word-based for automatic sentimental analysis of twitter messages, big data content, its scope, methods, samples, advantages and challenges of Data and so on:

**Lieberman, M. et al.[1]** The author showed Visualizing Big Data in Social Network Analysis. Mainly according to author Social researchers became entrepreneurs, and the phenomenal success of social media platforms such as Twitter, Facebook, Flickr, YouTube, and Wikipedia perceived their output as networks. That is, interconnected actors relate to each other. Sociologists introduced the language of social communities, and in academic research laboratories researchers studied motivation, influence, and social roles on the internet.

**Lima, Ana C.E.S, et al.[2]** presents hybrid approach of emotional-based and word-based for automatic sentimental analysis of twitter messages (i.e Tweets) and they also use basic text mining techniques and naive-Bayes classification algorithm which provide good efficiency. Generally sentimental word dictionaries will be used for labeling of Small piece of data called “crunches”. These kinds of dictionaries contain certain threshold value for sentiment word and the defined value is used to decide sentiment of word is positive or negative for subjective sentences. SentiWordNet V3.0 or WordNet are the online available sentiment word dictionaries.

**Sagiroglu, S. et al.[3]** (20-24 May 2013), “Big Data: A Review” describe the big data content, its scope, methods, samples, advantages and challenges of Data. The critical issue about the big data is the privacy and security. Big data samples describe the review about the atmosphere, biological science and research. Life sciences etc. By this

paper, we can conclude that any organization in any industry having big data can take the benefit from its careful analysis for the problem solving purpose. Using Knowledge Discovery from the Big data easy to get the information from the complicated data sets. The overall Evaluation describe that the data is increasing and becoming complex. The challenge is not only to collect and manage the data also how to extract the useful information from that collected data. According to the Intel IT Center, there are many challenges related to Big Data which are data growth, data infrastructure, data variety, data visualization, data velocity.

**Garlasu, D. et al.[4]** ( 17-19 Jan. 2013),”A Big Data implementation based on Grid Computing”, Grid Computing offered the advantage about the storage capabilities and the processing power and the Hadoop technology is used for the implementation purpose. Grid Computing provides the concept of distributed computing. The benefit of Grid computing center is the high storage capability and the high processing power. Grid Computing makes the big contributions among the scientific research, help the scientists to analyze and store the large and complex data.

**Mukherjee, A. et al.[5]** (18-22 Dec. 2012) “Shared disk big data analytics with Apache Hadoop” Big data analytics define the analysis of large amount of data to get the useful information and uncover the hidden patterns. Big data analytics refers to the Map reduce Framework which is developed by the Google. Apache Hadoop is the open source platform which is used for the purpose of implementation of Google’s Map reduces Model.

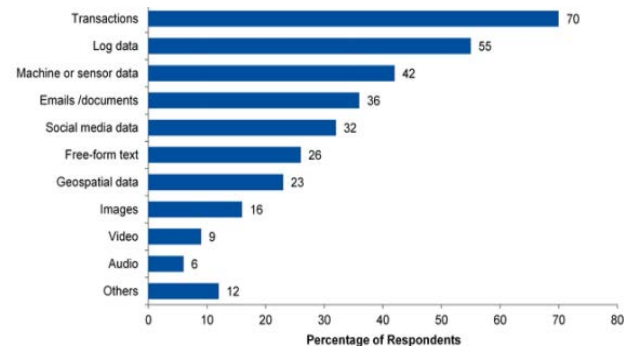
**Patel, Aditya B. et al.[6]** (6-8 Dec. 2012) “Addressing Big Data Problem Using Hadoop and Map Reduce” reports the experimental work on the Big data problems. In this the performance of SF-CFS is compared with the HDFS using the SWIM by the Facebook job traces .SWIM contains the workloads of thousands of jobs with complex data arrival and computation patterns. It describe the optimal solutions using Hadoop cluster, Hadoop Distributed File System (HDFS) for storage and Map Reduce programming framework for parallel processing to process large data sets

**Vaswani, Gaurav et al.[7]** Real Time Literature Review about the big data According to 2013, facebook has 1.11 billion people active accounts from which 751 million using facebook from a mobile. Another example is flicker having feature of Unlimited photo uploads (50MB per photo), Unlimited video uploads (90 seconds max, 500MB per video), the ability to show HD Video, Unlimited storage, Unlimited bandwidth. Flickr had a total of 87 million registered members and more than 3.5 million new images uploaded daily.

**Welser,Howard T et al.[8]** show that distinct connecting patterns among users are related to a variety of social roles that, in turn, form complex ecosystems in social media spaces. Social network diagrams were used to illustrate key social roles found in discussion spaces and wiki documents, including ‘answer people’, ‘discussion people’,

‘discussion starters’, and people who specialized in improving the formatting of wiki pages.

**Adamic,L.A. et al.[9]** illustrate the value of social network analysis for understanding the social connections within question and answer discussions in the Yahoo! Answers system. Their visualizations of different collections of messages, grouped by a common tag or category, illustrate a range of social practices and patterns, from question and answer exchanges to long debates and arguments.



**Figure 2: Big data Sources[7]**

**Christakis, Nicholas A.2009 et al.[10]** which examines how social networks impact on our everyday lives. What differentiates Analyzing Social Media Networks with NodeXL: Insights from a Connected World from this canon of work, is that the book focuses on examining how NodeXL can be used to create visualizations of social networks and assist in their analysis. With NodeXL, researchers can input network data into a table format and, via the click of a button, produce a customisable visualisation of the network. These visualisations assist in the quantitative dissection of social networks, as they visually depict key players in a network, highlight those participants who rarely communicate, display participants in a network who regularly communicate, and illuminate participants in a network who link together other people within the network.

### 3. CHALLENGES

Distributed programming framework utilize parallelism in computations and storage to process massive amounts of the data .A popular example is map reduce framework, which splits an input file into multiple chunks in the first phase of map reduce, a mapper for each chunk reads the data, perform some computation, and outputs a list of key/value pairs. In the next phase, a reducer combines the values belonging to each distinct key and outputs the result. There are two major attack prevention measures: securing the manners and securing the data in the presence of an untrusted manner.

Non-relational data stores popularized by NoSQL databases are still evolving with respect to security infrastructure. For instance, robust solutions to NoSQL injection are still not mature each NoSQL DBs were built

to tackle different challenges posed by the analytics world and hence security was never part of the model at any point of its design stage. Developers using NoSQL databases usually embed security in the middleware. NoSQL databases do not provide any Support for Enforcing it explicitly in the database. However, clustering aspect of NoSQL databases poses additional challenges to the robustness of such security practices.

Data and transaction logs are stored in multi-tiered storage media manually moving data between tiers gives the it manager direct control over exactly what data is moved and when. However as the size of data set has been and continues to be, growing exponentially, scalability and availability necessitated auto tiering for big data storage management. Auto tiering solutions do not keep track of where the data is stored, which poses new challenges to secure data storage. New mechanisms are imperative to thwart unauthorised access and maintain 24/7 availability.

Many big data use cases in Enterprise settings require data collection from many sources, such as end point devices for example, a security information and event management system (SIEM) may collect event logs from millions of hardware devices and software application in an enterprise network. A key challenge in the data collection process is input validation: how can we trust the data? How can we validate that a source of input data is not malicious and how can we filter malicious input from our collection? Input validation and filtering is a daunting challenge posed by untrusted input sources, especially with the bring your own device (BYOD) model.

Real time security monitoring has always been a challenge, given the number of alerts generated by (security) devices. These alerts (correlated or not) lead to many false positive, which are mostly ignored or simply "clicked away", as humans cannot cope with the sheer amount. This problem might even increase with the big data given the volume and velocity of data streams however, big data technologies might also provide an opportunity, in the sense that these technologies do allow for fast processing and analytics of different types of data. Which in its turn can be used to provide, for instance, real time anomaly detection based on scalable security analytics.

#### 4. IMPORTANCE

**"Big Data offers endless streams of information. How your corporation locates actionable trends is key to leveraging these vast data stores."** Big Data is the buzzword of the business world these days and its benefits are being touted by every business professional. However, to some it seems like all this talk about Big Data is just speculation. People don't think that Big Data can be more useful than other manageable forms of data that organizations are already using these days.

As technology advances and we are able to gather increasingly complex information, Big Data may soon become the norm. Also, if used correctly, Big Data can

have huge benefits for any business. Big Data analytics can give insights that can be very profound and highly profitable when used correctly. This is why Big Data is now being used extensively by companies when it comes to marketing and it has proven to be extremely useful when it comes to social media marketing.

Examining the role of Big Data in social media we looked at how it has the potential to transfer customer insight. The importance that corporations can adopt in order to sift through the data and filter out vital content that will deliver commercial opportunities. When the world is producing around 2.5 exabytes of data a day, pulling out elements of value suddenly make finding a needle in a haystack look strangely simple. However, Big Data is big money as evidenced by investments made by the big social networks such as Twitter who recently acquired Gnip and Facebook's investing in Presto to gain the insights these businesses need to improve their services. Social data provides a wealth of information to retailers, but are they taking advantage of the opportunities that social media metrics offer? Many are not. For any retail business with a digital or physical storefront, there's no reason to stay away from social media. A recent study from Yes Lifecycle Marketing, reported by GlobeNewswire, found that 91 percent of retailers are present on two or more social channels.

#### 5. DISCUSSION

In this discussion part various Author had studied the social network analysis and big data using Twitter, facebook etc. Big data security system in order to verify together the uniqueness of the person sentiment, combining statistical classification analysis of approaches based on data mining. To meet current security requirement, one of the analysis method, i.e. social media systems have found to be more accurate, compact, and efficient with satisfactory.

As we know today's world is becoming a narrower. We get reaction of people for particular products, events, issues very fast on web. Automatic sentiment analysis is very useful to identify and predict current and future trends, product reviews, people opinion for social issues, effect of some specific event on people; ROI and Business Intelligence applications use the sentiment analysis at big organizations like SAP, SAS and TCS.

Here, we had discussed importance that can be launched with social network system. We have also discussed various challenges and importance of Big data techniques that can be used for a secure social media system.

#### 6. CONCLUSION

In this paper we surveyed the some author analysis on the social media with visualization big data security. The percentage of unstructured data production on social media

makes it difficult to analyze using traditional methods that rely on human analysts. Social media analytics is a new field of study that is developing automated or semi-automated methods for analyzing data. One prominent technique is called sentiment analysis. This can be a useful tool to assess public reaction to a particular event, such as a Twitter, facebook Google, smartphones—there is so much marketing noise that even big TV networks and news giants no longer have the ability to guarantee overnight product success. However, outside of specific contexts the insights that can be drawn from this technique are currently limited. Research is underway to improve the technology and apply it to wider settings. One example is the WeGov project, funded by the European Commission, which is building tools to analyse responses to government policies on social media.

## REFERENCES

- [1]. Michael Lieberman” Visualizing Big Data: Social Network Analysis” digital research conference online+ Mobile +big data San Antonio, texas .march 11-12 2014.
- [2]. Ana C.E.S Lima and Leandro N.de Castro, “Automatic Sentiment Analysis of Twitter Messages”, IEEE Fourth International Conference on Computational Aspect .of Social Networks (CASoN), p.52-57, 2012.
- [3]. Sagioglu, S.; Sinanc, D. Big Data: A Review” Collaboration Technologies and Systems (CTS), 2013 International Conference on 20-24 May 2013 Print ISBN: 978-1-4673-6403-4
- [4]. Garlasu, D.; Sandulescu, V. ; Halcu, I. ; Neculoiu, G. ;( 17-19 Jan. 2013),”A Big Data implementation based on Grid Computing”, Grid Computing.
- [5]. Mukherjee, A.; Datta, J.; Jorapur, R.; Singhvi, R.; Haloi, S.; Akram, W., (18-22 Dec.,2012) , “Shared disk big data analytics with Apache Hadoop”
- [6]. Aditya B. Patel, Manashvi Birla, Ushma Nair “Addressing Big Data Problem Using Hadoop and Map Reduce” ,(6-8 Dec. 2012)
- [7]. Gaurav Vaswani, Anuradha Bhatia “A Real Time Approach with BIG Data – A review” Volume 3, Issue 9, September 2013 ISSN: 2277 128X International Journal of Advanced Research in Computer Science and Software Engineering.
- [8]. Welser, Howard T., Eric Gleave, Danyel Fisher, and Marc Smith (2007). Visualizing the signatures of social roles in online discussion groups, *The Journal of Social Structure*.8(2). <http://www.cmu.edu/joss/content/articles/volume8/Welser/>
- [9]. Adamic, L.A., Zhang, J., Bakshy, E., and Ackerman, M. (2008). Knowledge sharing and Yahoo Answers: Everyone knows something, *Proc. World Wide Web Conference, WWW2008.org*.
- [10]. Christakis, Nicholas A and James H. Fowler, *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*, London: Little, Brown and Company, 2009.