The Journal of Social Media in Society
Spring 2018, Vol. 7, No. 1, Pages 295-312
thejsms.org

Framework-based Mapping and Filtering for Social Media

Tengku Adil Tengku Izhar
[1]Faculty of Information Management, Universiti Teknologi MARA, UiTM Selangor, Malaysia
*Corresponding Author: tengkuadil4540@salam.uitm.edu.my

The aim of this paper is to develop a framework to extract value from social media data based on mapping and filtering, using the data processing approach based on MapReduce. In order to test the application of this framework, researchers used NodeXL to obtain and analyze data from Twitter. The results show the application of framework to capture relevant data for efficient decision-making. The outcome of this paper contributes to a significant achievement that provides an important innovation in research methods in big data era to trace how data flows across the social media and how to analyze this data.

Keywords: Big data, filtering, mapping, MapReduce, NodeXl, social media

---

Society is becoming increasingly more instrumented and as a result, organizations are producing and storing vast amounts of data. Managing and gaining insights from the produced data is a challenge and key to competitive advantage (Assuncao, Calheiros, Bianchi, & Netto, 2015). Big data is a new way of thinking about enterprise data and how it can drive business value. The amount of data that is available to businesses is increasing, with social media and machine-to-machine as just two of the leading sources. Storage is getting cheaper and processing power is getting faster. The central role of business services in today's enterprises, and the more complex architecture through which they are delivered, make it essential to manage big data solutions from a business perspective. Business perspective focuses on business objectives and benefit, and prioritizes resources and activities according to the needs of the business. In this way, structuring the big data can ensure optimal relevance of data for more effective decision-making (Izhar, Torabi, Bhatti, & Liu, 2013).

With the development of smart devices and cloud computing, more and more data can be collected from various sources and can be analyzed in an unprecedented way. The huge social and academic impact of such developments caused a worldwide buzz for big data (Huang et al., 2015). For example, social media have been adopted by many businesses. More and more companies are using social media tools such as Facebook and Twitter to provide various services and interact with customers. As a result, a large amount of user-generated content is freely available on social media sites (He, Zha, & Li, 2013).

Social media have been a popular topic among scholars spanning several disciplines including communication, psychology, sociology and business. The bulk of existing academic literature on social media has been published in just the last few years and has focused on the social processes of social media and their effects in areas such as marketing, politics, health communication, and education (McIntyre, 2014). Social media platform Twitter has stormed onto the social media scene not only as an individual communication device but also as an information dissemination platform (Krause & Smith, 2014).

Currently, the majority of social media studies focus on individual companies or organizations. There are few studies performing social media competitive analysis on the leading companies in an industry in a systemic way (He et al., 2013). Social media can help decision makers to ensure efficient solutions to the problems raised (da Silva et al., 2014). However, the trustworthiness of this social data is often questionable due to the huge amount of data created in social media.

In recent years, valuable knowledge can be retrieved from petabyte scale datasets that can led to the development of solutions to process information based on parallel and distributed computing (Polato, Ré, Goldman, & Kon, 2014). This paper aims to develop a framework to extract value from social data based on mapping and filtering. In order to achieve this framework, the data processing approach based on MapReduce was adopted. The data processing strategy employed by MapReduce consists of two primitive functions: Map and Reduce. Behind this simple abstraction is a single fixed data flow. A MapReduce job is divided into Map and Reduce tasks, and assigned to idle slots of workers according to these two stages (Polato et al., 2014).

LITERATURE REVIEW

In recent years, the rapid development of Internet, Internet of Things, and Cloud Computing have led to the explosive growth of data in almost every industry and business area. Big data has rapidly developed into a hot topic that attracts extensive attention from academia, industry, and governments around the world. There are many challenges in harnessing the potential of big data today, ranging from the design of processing systems at the lower layer to analysis means at the higher layer, as well as a series of open problems in scientific research. Big data processing systems suitable for handling a diversity of data types and applications are the key to supporting scientific research of big data (Jin, Wah, Cheng, & Wang, 2015).

Social media are transforming the way information travels within and between networks of individuals (Gangadharbatla, Bright, & Logan, 2014). Although the research on social networks dates back to early 1920s, nevertheless, social media analytics is a nascent field that has emerged after the advent of Web 2.0 in the early 2000s. The key characteristic of the modern social media analytics is its data-centric nature. Social media analytics refer to the analysis of structured and unstructured data from social media channels. “Social media” is a broad term encompassing a variety of online platforms that allow users to create and exchange content. User-generated content (e.g., sentiments, images, videos, and bookmarks) and the relationships and interactions between the network entities (e.g., people, organizations, and products) are the two sources of information in social media (Gandomi & Haider, 2015).

Social media have profoundly changed our lives and how we interact with one another and the world around us (Qualman, 2009; Safko & Brake, 2009). Recent research indicates that more and more people are using social media applications such as Facebook and Twitter for reasons such as making new friends, socializing with old friends, receiving information, and entertaining themselves (Kaplan & Haelein, 2010; Keckley, 2010; Park, Kee, & Valenzuela, 2009; Raacke & Bonds-Raacke, 2008). Social media analysis will extract value from vast amount of social media data to detect and discover new knowledge to understand how industry is changing, and use the findings and improved understanding to achieve competitive advantage against their competitors (Governatori & Iannella, 2011; He et al., 2013). Social media competitive analysis allows a business to gain possible business advantage by analyzing the publicly available social media data of a business and its competitors (He et al., 2013). As social media have become a topic of interest for many industries, it is important to understand how social media data can be harvested for decision-making (He et al., 2013).

With the development of smart devices and cloud computing, more and more public data can be collected from various sources and can be analyzed in an unprecedented way. The huge social and academic impact of such developments caused a worldwide buzz for big data (Huang et al., 2015). Data flow is an ordered sequence which is consecutive, high-speed, infinite and time varying. It is also of great importance in Internet management, Internet security and Internet experiment. However, with the rapid development of Internet technology, the number of Internet applications and users keeps rising, and the Internet data are growing exponentially (Zhi et al., 2011). As a result, there is a stricter requirement about the efficiency, expandability and stability of the data flow in social media.

Today people have access to more data in single day than most people that have access to data in the previous decade. All this data captures in different formats and makes it almost impossible to understand the existing relationship between different data. For example, government agencies and large, medium and small private enterprises in many domains, such as engineering, education, manufacturing, are drowning in an ever-increasing deluge of data. Companies like Google, eBay, LinkedIn, and Facebook were built around big data from the beginning (Davenport & Dyche, 2013). Big data may be as important to business because more data can lead to more accurate analyses. More accurate analyses may lead to more confident decision-making and better decision can mean greater operational efficiencies in the organization (Davenport & Dyche, 2013).

Even though professionals such as data scientists are trained to analyze this data, the huge capacity of data created every day make it hard to identify which data is relevant from social media. As a result, it poses an issue on how effective these data are to support decision-making process (Izhar et al., 2013). Data scientists must somehow get along and work jointly with mere quantitative analysts (Davenport & Dyche, 2013). Thus, having an ability to analyze the data in a timely fashion can ensure domain experts have a competitive edge to improve productivity in their decision-making.

Recently, Google revealed that it has replaced the 10-year old MapReduce with its new systems such as DataFlow. It provides better performances and supports more sophisticated applications. Simultaneously, other new systems, such as Spark, Impala and epiC, are also being developed to handle new requirements for big data processing (Chen, Wu, & Wang, 2015). Digital technologies have changed the fabric of organizations, triggering novel organizational forms. Innovation and production are not confined to established organizations with clear cut boundaries. Capturing fine-grained and high quality interaction and collaboration data are also another important challenge that needs to be tackled (Brunswicker et al., 2015).

With the advance of mobile application, the way we access the Internet has changed. The number of electronic devices that can be connected to network services has increased dramatically. The Internet connectivity is also possible beyond these devices. Every object around us is connected, collecting, processing and sharing information. For example, running shoes can upload the time, speed and distance of your run to your devices through the Internet. As a result, it grows a memory capacity that has fueled the development of in-memory big data management and processing (Zhang et al., 2015). With the development of smart devices and cloud computing, more data can be collected from various sources and can be analyzed in an unprecedented way (Huang et al., 2015).

Nothing in technology stands still especially in the world of data and analytics. Systems built just a few years ago are starting to buckle under the explosion of data and the changing query needs of business users. Social media have now become important mediums of communication and interaction tools for social networks (Ghani & Kamal, 2015). Social media are also important for business platforms that can influence the corporative environment (Damianos, Cuomo, & Drozdetski, 2011; Damianos et al., 2007; Holtzblatt et al., 2013). For example, social networks involve agents in creating and processing information for the knowledge network (Roth & Cointet, 2010). At the same time, the role that causality can play in social network analysis is unclear (Doreian, 2001). Therefore, it is important to examine the flow of data in social media and to retrieve relevant data from large amount of it

Taking advantage of big data opportunities is challenging for the organizations (Berber, Graupner, & Maedche, 2014). Firms and other organizations have been using large databases and analytics for the last couple of decades. Transactions are stored in data warehouses and analyzed with data-mining algorithms to extract insights (Galbraith, 2014). In order to ensure the effectiveness of the data, organizations need to be able to store data reliably across a number of databases. Once data need to be distributed, organizations need a way to get them out again and they need to identify which data are needed, assemble them and analyze them. The challenge is how to capture data to be considered relevant for the specific organization's activities because determining relevant data are a key to delivering value from massive amounts of data as shown in Fig. 1. The real issue is not how the organizations acquire large amount of data but what they do with the data that counts (Davenport & Dyche, 2013). The technologies and concepts behind big data can allow organizations to achieve a variety of objectives.

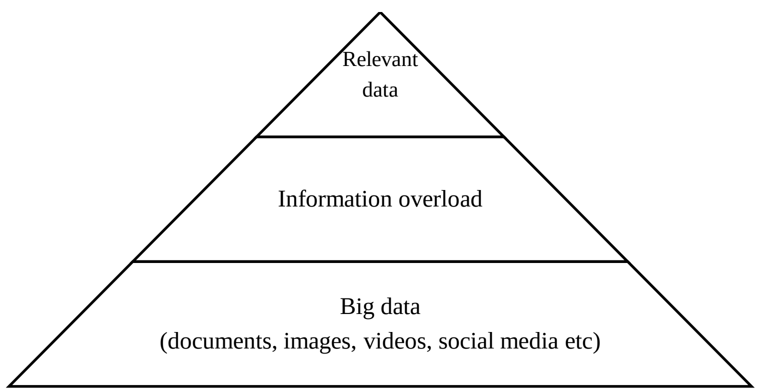


Figure 1. *Relevant data from large amounts of data.*

RESEARCH FRAMEWORK

In this paper, we adapt a framework by Ghani & Kamal (2015). The authors proposed a sentiment-based filtration and data analysis framework to identify relevant information from data generated by users in social media. In accordance to this, this paper proposes a framework to retrieve relevant information from large volume of social data based on mapping. The paper discusses an approach that explains the process in mapping, filtering and analytic in extracting relevant information, as shown in Fig. 2. In contrast to Ghani & Kamal (2015), the framework represents five stages and tasks:

1. **Data input**—Data are collected from databases. Social media can be defined as large continuous of data storage. Social media users create high volume of data and make it difficult to analyze this data for certain query. For example,

250 million tweets per day create high amounts of data and the amounts are increasing every day. A reliable analysis approach for this data is a big challenge. The framework will extract data from social media to be mapped in order to select data that can be considered relevant.

2. **Mapping**— This stage includes the process to map data that already selected in stage 1. The process is to identify specific query from databases. The query is mapped to identify the flow of data from the databases.
3. **Filtering**— This stage includes the processes involved in refining data that are already mapped during the previous stage. The process is to filter a selected query from databases in order to consider data related to the query that is relevant. This process identifies relevant data to be analyzed.
4. **Analytic**— In this stage, data analysis will be exercised.
5. **Data output**—In the final stage, the matched results will be presented as the output resulted. The results will be the most relevant results that can be used for decision-making.

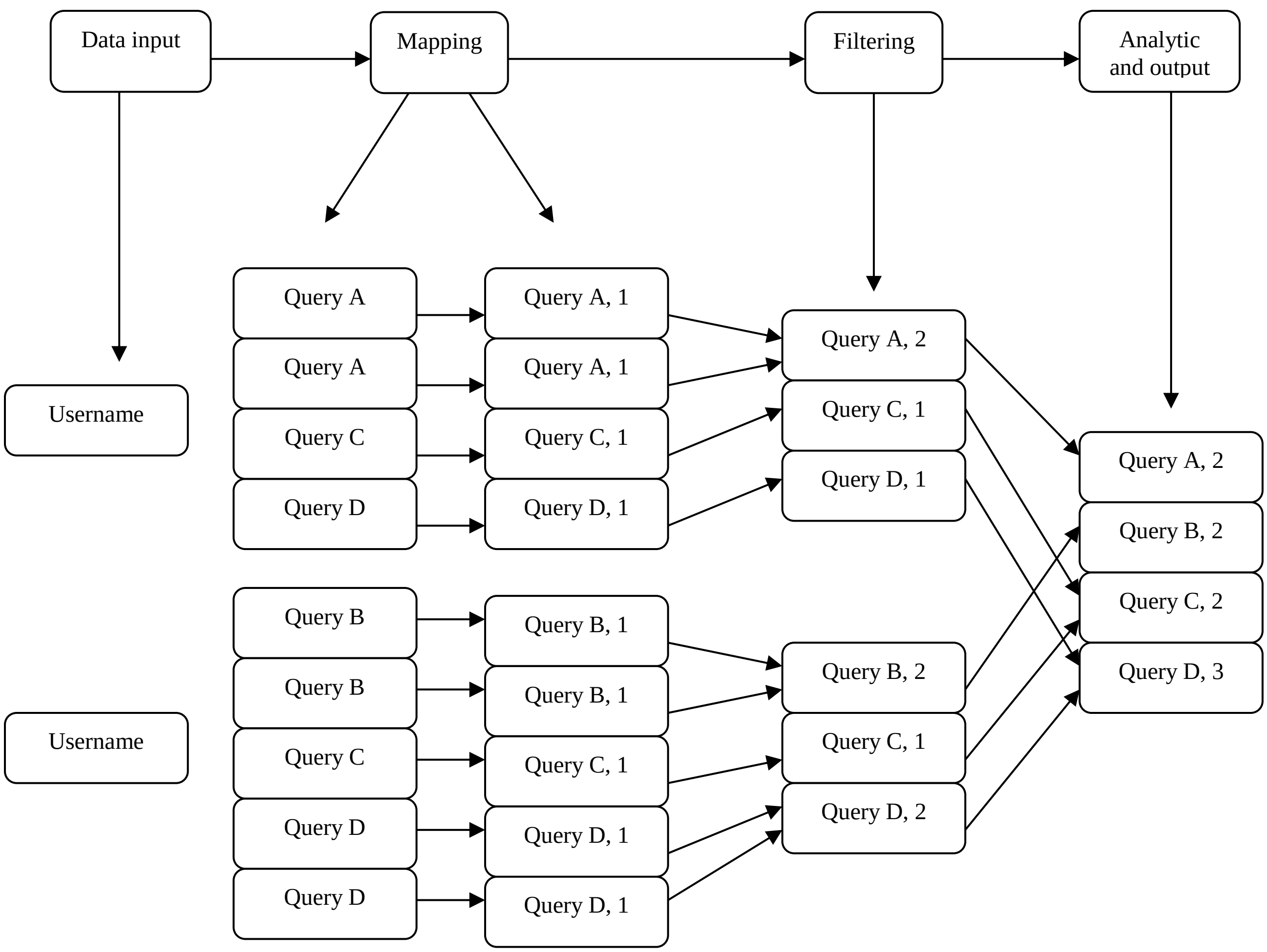


Figure 2. *Framework based mapping, filtering and analytic for social media (adapted from MapReduce).*

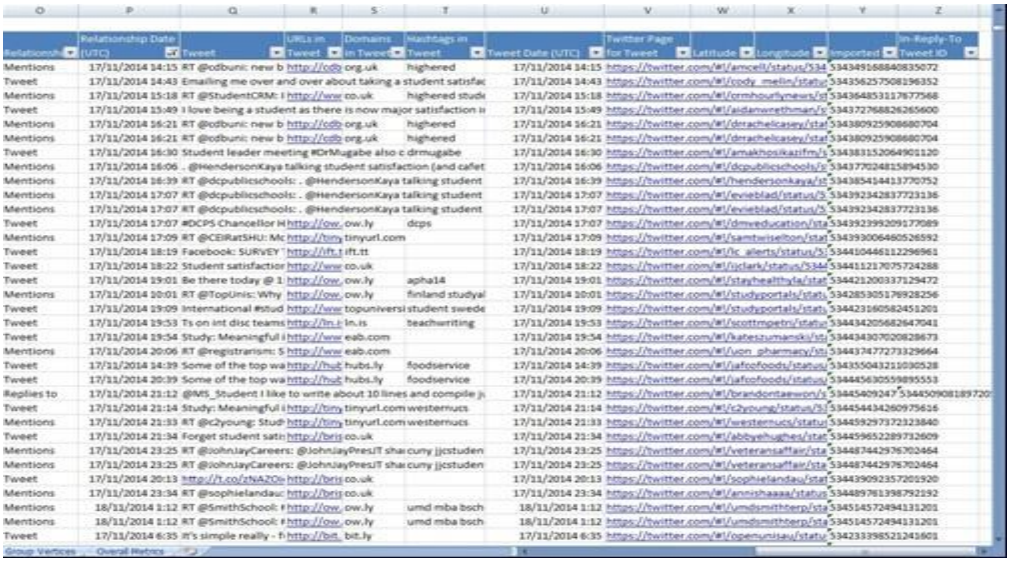
CASE STUDY

NodeXL is an extendible toolkit for network overview, discovery and exploration implemented as an add-in to the Microsoft Excel 2007 spreadsheet software (Smith et al., 2010). NodeXL is applied to retrieve data from social media and import this data into the spreadsheet as shown in Fig. 5. NodeXL demonstrate data analysis and visualization features with a social media data sample drawn from an enterprise intranet social network. NodeXL is applied to visualize the ontology for people who tweet about service popularity and service satisfaction.

Sample and Data Collection

To identify participants, we used the sampling pool of Twitter that appears in the NodeXL, a software tool that imports data from outside data providers. Users create data on Twitter every second and data are collected based on the date of the tweet is created in order to avoid huge volume of data. The tool imported data from Twitter username LaTrobe into a spreadsheet. In this case study, we evaluated data for two weeks, from 17 November 2014 to 30 November 2014. Therefore, data are filtered and captured within this period of time using NodeXL. In the LaTrobe network, we capture data that matched to the query of people who were discussing student satisfaction. We expanded this query by looking at people who discussed service popularity and service satisfaction. These data are based on people who tweet, mention and reply.

Service popularity is based on the number of time students use the services. The more services student use, the more popular the services are. Student satisfaction is based on the quality of the services. It means that if the student uses the services many times, we define that the services fulfill the student satisfaction. In this paper, the unit of measurement is based on the level of student satisfaction and service popularity. The number of the frequency is identified based on the data capture from Twitter using NodeXL from 17 November 2014 to 30 November 2014.



Relationship	Relationship Date (UTC)	Tweet	URLs in Tweet	Domains in Tweet	Hashtags in Tweet	Tweet Date (UTC)	Twitter Page for Tweet	Latitude	Longitude	Imported ID	In-Reply-To Tweet ID
Mentions	17/11/2014 14:15	RT @cdbunc: new b	http://cdb	org.uk	highered	17/11/2014 14:15	https://twitter.com/#!/amcell/status/534			534349168840835072	
Tweet	17/11/2014 14:43	Emailing me over and over about taking a student satisfac				17/11/2014 14:43	https://twitter.com/#!/cody_mellin/statu			534356257508196352	
Mentions	17/11/2014 15:18	RT @StudentCRM: I	http://ww	co.uk	highered stude	17/11/2014 15:18	https://twitter.com/#!/crmhourlynews/st			534364853117677568	
Tweet	17/11/2014 15:49	I love being a student as there is now major satisfaction in				17/11/2014 15:49	https://twitter.com/#!/aidanwrethman/s			534372768826265600	
Mentions	17/11/2014 16:21	RT @cdbunc: new b	http://cdb	org.uk	highered	17/11/2014 16:21	https://twitter.com/#!/drrachelcasey/stat			534380925908680704	
Mentions	17/11/2014 16:21	RT @cdbunc: new b	http://cdb	org.uk	highered	17/11/2014 16:21	https://twitter.com/#!/drrachelcasey/stat			534380925908680704	
Tweet	17/11/2014 16:30	Student leader meeting #DrMugabe also c			drmugabe	17/11/2014 16:30	https://twitter.com/#!/amakhosikazifm/s			534383152064901120	
Mentions	17/11/2014 16:06	. @HendersonKaya talking student satisfaction (and cafet				17/11/2014 16:06	https://twitter.com/#!/dcpublicschools/s			534377024815894530	
Mentions	17/11/2014 16:39	RT @dcpublicschools: . @HendersonKaya talking student				17/11/2014 16:39	https://twitter.com/#!/hendersonkaya/st			534385414413770752	
Mentions	17/11/2014 17:07	RT @dcpublicschools: . @HendersonKaya talking student				17/11/2014 17:07	https://twitter.com/#!/evieblad/status/5			534392342837723136	
Mentions	17/11/2014 17:07	RT @dcpublicschools: . @HendersonKaya talking student				17/11/2014 17:07	https://twitter.com/#!/evieblad/status/5			534392342837723136	
Tweet	17/11/2014 17:07	#DCPS Chancellor H	http://ow.	ow.ly	dcps	17/11/2014 17:07	https://twitter.com/#!/dmveducation/sta			534392399209177089	
Mentions	17/11/2014 17:09	RT @CEIRatSHU: Mc	http://tiny	tinyurl.com		17/11/2014 17:09	https://twitter.com/#!/samtwiselton/stat			534393006460526592	
Tweet	17/11/2014 18:19	Facebook: SURVEY	http://ift.t	ift.tt		17/11/2014 18:19	https://twitter.com/#!/lc_alerts/status/5			534410446112296961	
Tweet	17/11/2014 18:22	Student satisfactior	http://ww	co.uk		17/11/2014 18:22	https://twitter.com/#!/jclark/status/534			534411217075724288	
Tweet	17/11/2014 19:01	Be there today @ 1	http://ow.	ow.ly	apha14	17/11/2014 19:01	https://twitter.com/#!/stayhealthyla/stat			534421200337129472	
Mentions	17/11/2014 10:01	RT @TopUnis: Why	http://ow.	ow.ly	finland studyal	17/11/2014 10:01	https://twitter.com/#!/studyportals/stat			534285305176928256	
Tweet	17/11/2014 19:09	International #stud	http://ww	topuniversi	student swede	17/11/2014 19:09	https://twitter.com/#!/studyportals/stat			534423160582451201	
Tweet	17/11/2014 19:53	Ts on int disc teams	http://ln.i	ln.is	teachwriting	17/11/2014 19:53	https://twitter.com/#!/scottmpetri/statu			534434205682647041	
Tweet	17/11/2014 19:54	Study: Meaningful i	http://ww	eab.com		17/11/2014 19:54	https://twitter.com/#!/kateszumanski/st			534434307020828673	
Mentions	17/11/2014 20:06	RT @registrarism: S	http://ww	eab.com		17/11/2014 20:06	https://twitter.com/#!/uon_pharmacy/st			534437477273329664	
Tweet	17/11/2014 14:39	Some of the top wa	http://hut	hubs.ly	foodservice	17/11/2014 14:39	https://twitter.com/#!/iafcofoods/status			534355043211030528	
Tweet	17/11/2014 20:39	Some of the top wa	http://hut	hubs.ly	foodservice	17/11/2014 20:39	https://twitter.com/#!/iafcofoods/status			534445630559895553	
Replies to	17/11/2014 21:12	@MS_Student I like to write about 10 lines and compile ju				17/11/2014 21:12	https://twitter.com/#!/brandontaewon/s			5344540924?	5344509081897203
Tweet	17/11/2014 21:14	Study: Meaningful i	http://tiny	tinyurl.com	westernucs	17/11/2014 21:14	https://twitter.com/#!/c2young/status/5			534454434260975616	
Mentions	17/11/2014 21:33	RT @c2young: Stud	http://tiny	tinyurl.com	westernucs	17/11/2014 21:33	https://twitter.com/#!/westernucs/statu			534459297372323840	
Tweet	17/11/2014 21:34	Forget student satis	http://bris	co.uk		17/11/2014 21:34	https://twitter.com/#!/abbyehughes/stat			534459652289712609	
Mentions	17/11/2014 23:25	RT @JohnJayCareers: @JohnJayPresJT sha			cuny jjcstuden	17/11/2014 23:25	https://twitter.com/#!/veteransaffair/sta			534487442976702464	
Mentions	17/11/2014 23:25	RT @JohnJayCareers: @JohnJayPresJT sha			cuny jjcstuden	17/11/2014 23:25	https://twitter.com/#!/veteransaffair/sta			534487442976702464	
Tweet	17/11/2014 20:13	http://t.co/zNA2Oi	http://bris	co.uk		17/11/2014 20:13	https://twitter.com/#!/sophielandau/stat			534439092357201920	
Mentions	17/11/2014 23:34	RT @sophielandau:	http://bris	co.uk		17/11/2014 23:34	https://twitter.com/#!/annishaaaa/status			534489761398792192	
Mentions	18/11/2014 1:12	RT @SmithSchool: #	http://ow.	ow.ly	umd mba bsch	18/11/2014 1:12	https://twitter.com/#!/umdsmithterp/sta			534514572494131201	
Mentions	18/11/2014 1:12	RT @SmithSchool: #	http://ow.	ow.ly	umd mba bsch	18/11/2014 1:12	https://twitter.com/#!/umdsmithterp/sta			534514572494131201	
Tweet	17/11/2014 6:35	It's simple really - f	http://bit.	bit.ly		17/11/2014 6:35	https://twitter.com/#!/openunisau/statu			534233398521241601	

Figure 3. *Example of imported Twitter data into spreadsheet using NodeXL from 17/11/2014 to 30/11/2014.*

Mapping Diagram for Twitter

In social media, users create data every second and minute. As a result, number of social data keep changing all the time and make it hard to evaluate. For example, today data might be important but tomorrow this data might not be important anymore. Data mapping for social media is illustrated as shown in Fig. 4.

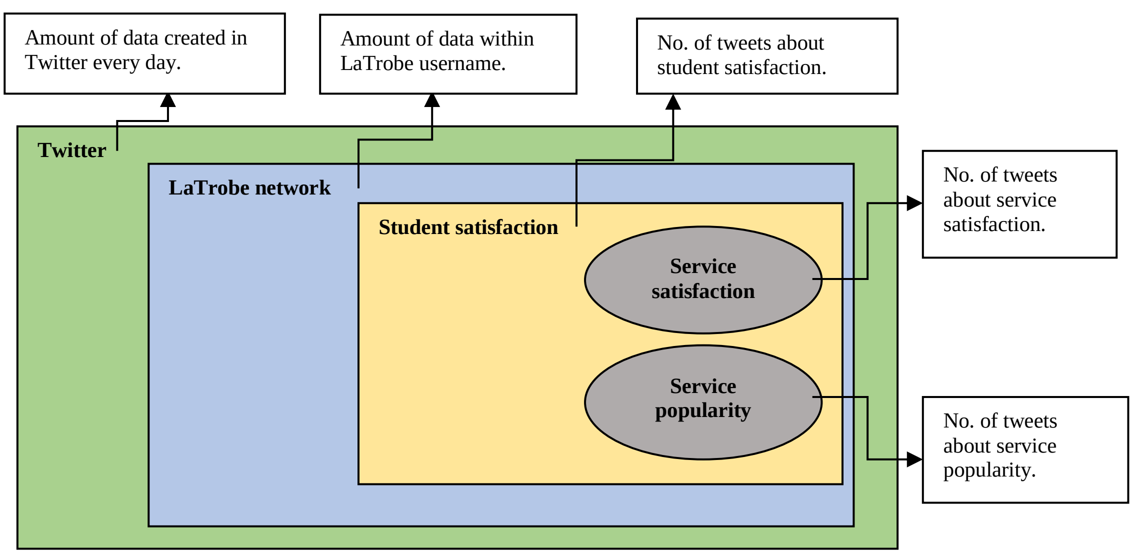


Figure 4. *Mapping diagram for Twitter to capture data that relate to service popularity and service satisfaction.*

Data filtering. This section provides steps to demonstrate how we filter data from large amounts of data from social media to allow us to evaluate the specific query. These preferences are used to configure the steps in filtering the data from Twitter using NodeXL. The step-by-step guideline explains how we selected data from huge volumes of social data that related to the query. Using NodeXL as a tool to capture this data, this guideline provides systematic steps for domain experts to capture which data they want to use in their decision-making in relation to the query. However, in other cases, domain experts might use different tools to capture this data.

1. Import from Twitter users network, as shown in Fig. 5. It optionally clears the NodeXL workbook, and then gets the network of specified Twitter users.
2. Specify the Twitter users with specific username. We were interested in username @LaTrobe.
3. Import basic network plus followers and following who replies, mentions and tweet. Limit it to 100 recent tweets per user.
4. Import from Twitter search network, as shown in Fig. 6. It optionally clears the NodeXL workbook, and then gets the network of people who tweet certain specified word(s).
5. Search for the tweets that match to the specific query. We searched for student satisfaction.
6. Import basic networks to specifically show who replied or mentioned in the tweets. Limit to 100 tweets.
7. Filter by relationships (tweet, mentions and replies).
8. Filter by specific date (day, week, month).
9. We applied the steps for tweet(s) that match to service popularity and service satisfaction in the user network.

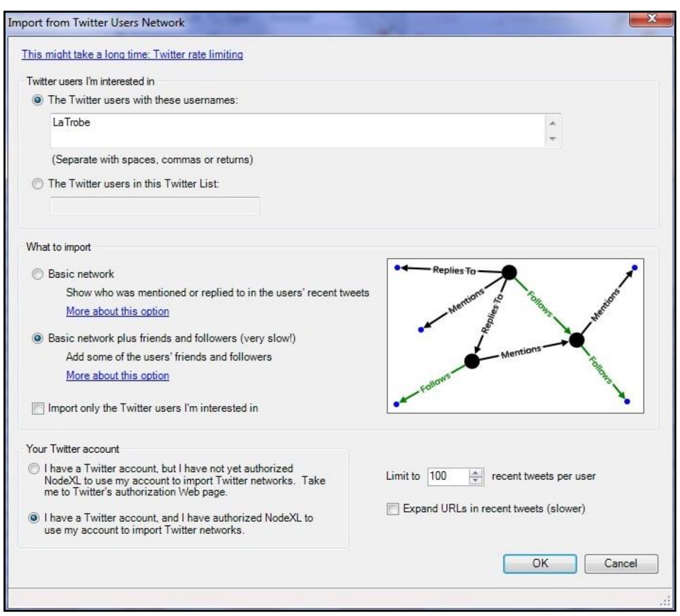


Figure 5. *Import Twitter username.*

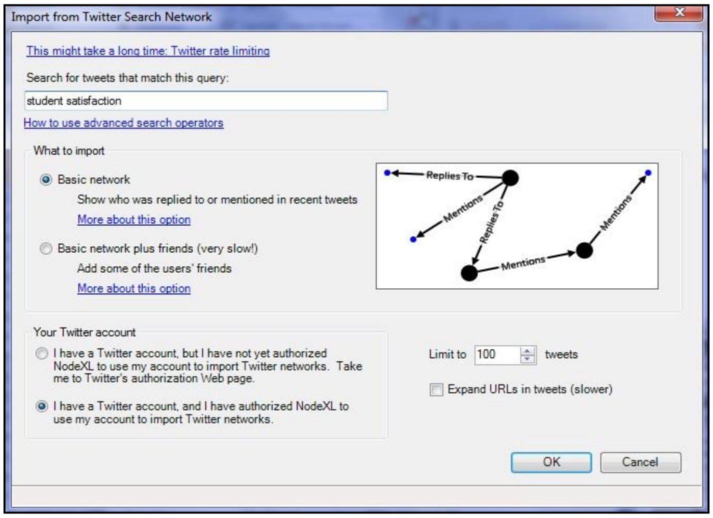


Figure 6. *Import tweets from Twitter network.*

DATA ANALYSIS AND RESULTS

Data are evaluated to identify the level of student satisfaction from the LaTrobe username. The results show the evaluation of service satisfaction and service popularity among La Trobe university students. After we capture the data from Twitter, we summarize the number of tweets for two weeks as shown in Tables 1 and 2. Results are summarized (Fig. 7, 8, 9) to assist the decision-making process to evaluate the level of student satisfaction.

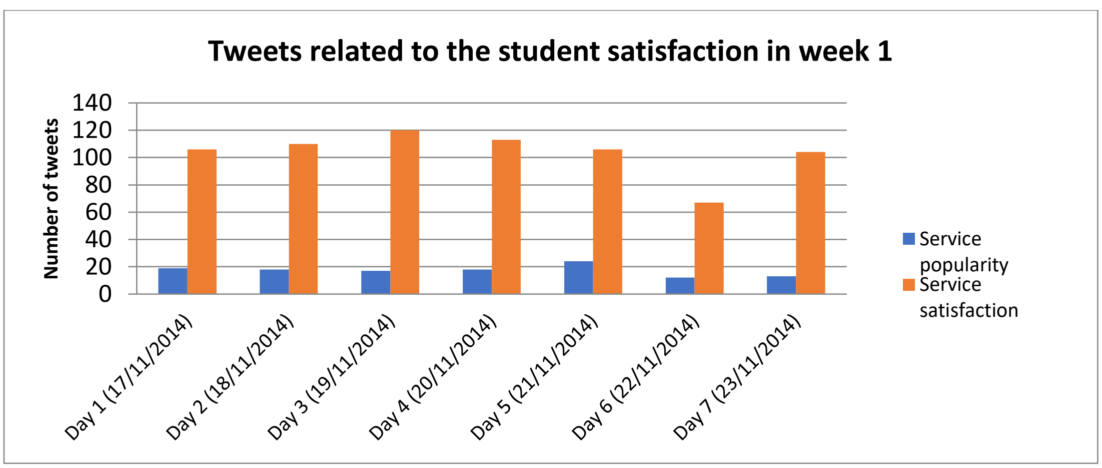


Figure 7. *Comparison of daily tweets about service popularity and service satisfaction in week 1.*

Table 1
Summary of Tweets on Student Satisfaction in the LaTrobe Network in Week 1

Summary metrics/ Week 1	Day 1 17/11/ 2014	Day 2 18/11/ 2014	Day 3 19/11/ 2014	Day 4 20/11/ 2014	Day 5 21/11/ 2014	Day 6 22/11/ 2014	Day 7 23/11/ 2014
Vertices	49	40	41	33	43	8	6
Unique Edges	43	32	34	35	44	6	5
Edges With Duplicates	2	2	0	0	0	0	0
Total Edges (Total tweet)	45	34	34	35	44	6	5
Self-Loops	21	10	10	11	13	1	3
Reciprocated Vertex Pair Ratio	0.043478261	0	0	0.090909091	0.068965517	0	0
Reciprocated Edge Ratio	0.083333333	0	0	0.166666667	0.129032258	0	0
Connected Components	29	21	19	16	19	3	4
Single-Vertex Connected Components	14	8	7	9	10	0	2
Maximum Vertices in a Connected Component	4	5	7	8	11	3	2
Maximum Edges in a Connected Component	4	8	8	14	11	3	2
Maximum Geodesic Distance (Diameter)	2	2	3	2	3	2	1
Average Geodesic Distance	0.574257	0.6875	1.054264	1.059829	1.363184	0.818182	0.4
Graph Density	0.010204082	0.01474359	0.014634146	0.022727273	0.017165006	0.089285714	0.066666667

Note: Modularity was n/a for all days and NodeXL Version 1.0.1.334 was used for each day.

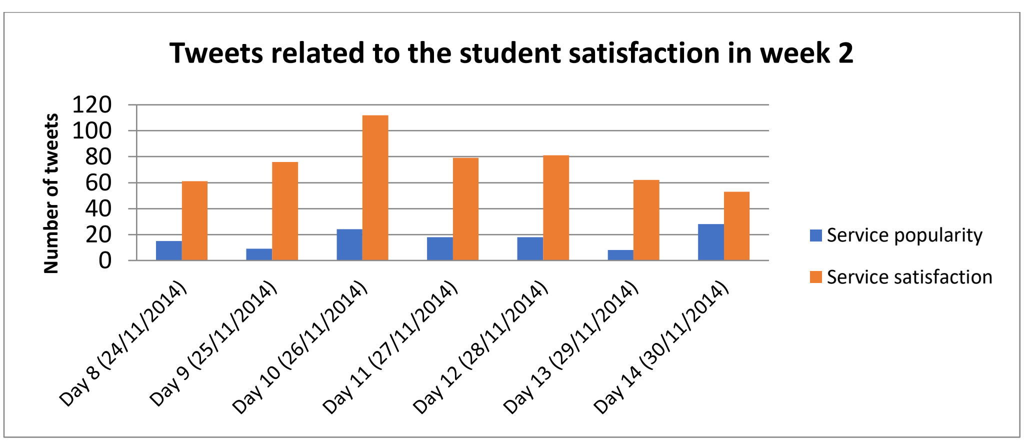


Figure 8. *Comparison of daily tweets about service popularity and service satisfaction in week 2.*

Table 2
Summary of Tweets on Student Satisfaction in the LaTrobe Network in Week 2

Summary metrics/ Week 2	Day 8 24/11/ 2014	Day 9 25/11/ 2014	Day 10 26/11/ 2014	Day 11 27/11/ 2014	Day 12 28/11/ 2014	Day 13 29/11/ 2014	Day 14 30/11/ 2014
Vertices	17	27	40	24	43	10	4
Unique Edges	14	24	37	21	40	11	3
Edges With Duplicates	2	7	4	2	0	2	0
Total Edges (Total tweet)	16	31	41	23	40	13	3
Self-Loops	6	13	16	9	11	4	2
Reciprocated Vertex Pair Ratio	0	0.0625	0	0.076923077	0	0	0
Reciprocated Edge Ratio	0	0.117647059	0	0.142857143	0	0	0
Connected Components	8	15	17	12	19	5	3
Single-Vertex Connected Components	4	8	10	6	8	3	2
Maximum Vertices in a Connected Component	7	6	9	5	5	5	2
Maximum Edges in a Connected Component	6	11	9	6	5	8	1
Maximum Geodesic Distance (Diameter)	2	3	3	2	3	2	1
Average Geodesic Distance	1.2	0.849315	1.284091	0.878788	0.931298	0.8125	0.333333
Graph Density	0.033088235	0.024216524	0.015384615	0.025362319	0.016057586	0.1	0.083333333

Note: Modularity was n/a for all days and NodeXL Version 1.0.1.334 was used for each day.

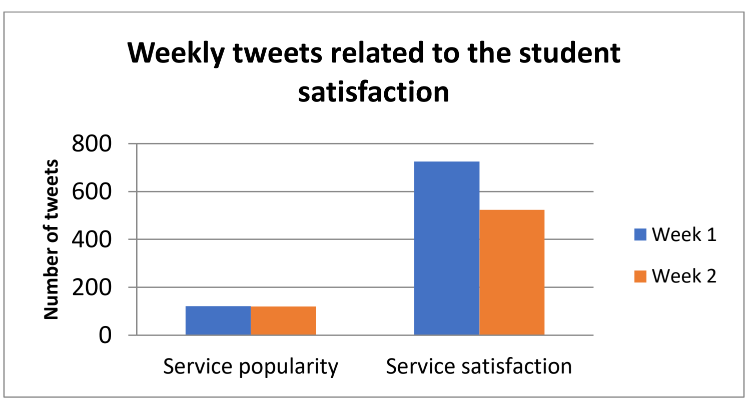


Figure 9. *Weekly score of tweets related to the student satisfaction.*

Table 3 shows that service satisfaction is a popular topic discussed by student during week 1 in day 3 with 120 tweets and service popularity is a popular topic discussed during week 2 in day 14 with 28. Service satisfaction is a popular topic discussed by student in week 1 with 726 tweets and service popularity is a popular topic discussed in week 1 with 121 tweets.

Table 3
Daily query score sheet

Summary of tweets	
Username network	LaTrobe
Current week	2
Tweet query	Student satisfaction
Current day	14
Total tweets	370
Higher tweet	Day 1
Week	Week 1
Number of tweet	45
Tweet query	Service popularity
Current day	14
Total tweets	241
Higher tweet	Day 14
Week	Week 2
Number of tweets	28
Tweet query	Service satisfaction
Current day	14
Total tweets	1250
Higher tweet	Day 3
Week	Week 1
Number of tweets	120

Data were collected for two weeks, as shown in Table 4. Based on these datasets, we evaluate the percentage of tweets in relation to the student satisfaction in the LaTrobe network. Therefore, we can make a decision of what people think about La Trobe University Student Support Services.

Table 4
Total Weekly Tweets

Week/Days	Service popularity	Service satisfaction
Week 1		
Day 1 (17/11/2014)	19	106
Day 2 (18/11/2014)	18	110
Day 3 (19/11/2014)	17	120
Day 4 (20/11/2014)	18	113
Day 5 (21/11/2014)	24	106
Day 6 (22/11/2014)	12	67
Day 7 (23/11/2014)	13	104
Total	121	726
Week 2		
Day 8 (24/11/2014)	15	61
Day 9 (25/11/2014)	9	76
Day 10 (26/11/2014)	24	112
Day 11 (27/11/2014)	18	79
Day 12 (28/11/2014)	18	81
Day 13 (29/11/2014)	8	62
Day 14 (30/11/2014)	28	53
Total	120	524
2-week Total	241	1,250

Table 5 shows that service satisfaction has the highest percentage with 70% and service popularity with 67%. The total number of tweets for week 1 and week 2 also shows that service satisfaction has the highest number of tweets. People from the network believe that service satisfaction is important to them when they use any services in the university.

Table 5
Score of Tweets Related to Student Satisfaction

Weeks	Service popularity	Rank	Service satisfaction	Rank
Week 1	121	1	726	1
Week 2	120	2	524	2
Total/Total rank	241	361	1250	1774
Percentage	67		70	

Limitations

Twitter users create high volume of data, which makes it difficult to analyze this data for certain query. More than 250 million tweets per day create high amounts of data and the amounts are increasing every day. A reliable analysis approach for this data set is a big challenge. When it comes to collecting, computing, analyzing, and acting on social data, technical challenges are quite different because the number of social data always increase and make it difficult to evaluate.

Terms that have similar meaning also create difficulties. Users from different backgrounds or nationalities might tweet different terms that have same meaning. For example, users from Malaysia will use “kamu”,”engkau” or “awak”, which mean “you” in their tweets.

DISCUSSION

Social media such as Facebook and Twitter have become an indispensable part of our lives. This article described the main features of a proposed framework when researchers developed the relationship for social data. In addition, we have proposed an alternative way to capture social data using NodeXL. We extend the application of the framework for social media based on mapping and filtering approach. This paper enables researchers to classify and evaluate existing research, to design scientific research, and to identify the gaps and weaknesses prior to future research directions.

References

Assuncao, M. D., Calheiros, R. N., Bianchi, S., & Netto, M. A. S. (2015). Big data computing and clouds: trends and future directions. *Journal of Parallel and Distributed Computing, 79*, 3-15.

Berber, M., Graupner, E., & Maedche, A. (2014). The information panopticon in the big data era. *Journal of Organization Design, 3*(1), 14-19.

Brunswicker, S., Bertino, E., & Matei, S. (2015). Big data for open digital innovation- A research roadmap. *Big Data Research, 2*, 53-58.

Chen, G., Wu, S., & Wang, Y. (2015). The evolvement of big data systems: from the perspective of an information security application. *Big Data Research, 2*, 65-73.

Damianos, L. E., Cuomo, D. L., & Drozdetski, S. (2011). *Handshake: A case study for exploring business networking for enterprise, inside and out.* Paper presented at the International Conference on Human Computer Interaction.

Damianos, L. E., Cuomo, D., Griffith, J., Hirst, D. M., & Smallwood, J. (2007). *Exploring the adoption, utility and social influences of social bookmarking in a corporate environment*. Paper presented at the International Conference on System Sciences, Hawaii.

Davenport, T. H., & Dyche, J. (2013). Big data in big companies: International Institute for Analytics.

Doreian, P. (2001). Causality in social network analysis. *Sociological Methods & Research, 30*(1), 81-114.

Galbraith, J. R. (2014). Organization design challengers resulting from big data. *Journal of Organization Design, 3*(1), 2-13.

Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods and anlytics. *International Journal of Information Management, 35*, 137-144.

Gangadharbatla, H., Bright, L. F., & Logan, K. (2014). Social media and news gathering: tapping into the millennial mindset. *The Journal of Social Media in Society, 3*(1), 45-63.

Ghani, N. A., & Kamal, S. S. M. (2015). A Sentiment-based filteration and data analysis framework for social media *5th International Conference on Computing and Informatics ICOCI* (pp. 632-637). Istanbul, Turkey.

Governatori, G., & Iannella, R. (2011). A modeling and reasoning framework for social networks policies. *Enterprise Information Systems, 5*(1), 145–167.

He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in pizza industry. *International Journal of Information Management, 33*, 464-472.

Holtzblatt, L., Drury, J. L., Weiss, D., Damianos, L. E., & Cuomo, D. (2013). Evaluating the uses and benefits of an enterprise social media platform. *Journal of Social Media for Organizations, 1*(1), 1-21.

Huang, T., Lan, L., Fang, X., An, P., Min, J., & Wang, F. (2015). Promises and challenges of big data computing in health sciences. *Big Data Research, 2*(1), 2-11.

Izhar, T. A. T., Torabi, T., Bhatti, M. I., & Liu, F. (2013). Recent developments in the organization goals conformance using ontology. *Expert Systems with Applications, 40*(10), 4252-4267.

Jin, X., Wah, B. W., Cheng, X., & Wang, Y. (2015). Significance and challenges of big data research. *Big Data Research, 2*, 59-64.

Kaplan, A. M., & Haelein, M. (2010). Users of the worl, united! The challenges and opportunities of Social Media. *Business Horizons, 53*(1), 59-68.

Keckley, P. H. (2010). Social networks in health care: Communication, collaboration and insights. from http://www.deloitte.com/ assets/Dcom-UnitedStates/Local%20Assets/Documents/US CHS 2010Social Networks 070710.pdf

Krause, D., & Smith, M. (2014). Twitter as mythmaker in storytelling: The emergence of hero status by the Boston police department in the aftermath of the 2013 marathon bombing. *The Journal of Social Media in Society, 3*(1), 8-27.

McIntyre, K. (2014). The evolution of social media from 1969 to 2013: A change in competition and a trend toward complementary. *The Journal of Social Media in Society, 3*(2), 6-25.

Park, N., Kee, K. F., & Valenzuela, S. (2009). Being immersed in social networking environment: Facebook groups, uses and gratifications, and social outcomes. *CyberPsychology & Behavior, 12*(6), 729–733.

Polato, I., Ré, R., Goldman, A., & Kon, F. (2014). A comprehensive view of Hadoop research- A systematic literature review. *Journal of Networking and Computer Application, 46*, 1-25.

Qualman, E. (2009). *Socialnomics how social media transforms the way we live and do business*. Hoboken: Wiley John & Sons, Inc.

Raacke, J., & Bonds-Raacke, J. (2008). MySpace and Facebook: Applying the uses and gratifications theory to exploring friend-networking sites. *CyberPsychology & Behavior, 11*(2), 169–174.

Roth, C., & Cointet, J. P. (2010). Social and semantic coevolution in knowledge networks. *Social Networks, 32*(1), 16-29.

Safko, L., & Brake, D. K. (2009). *The social media bible: Tactics, tools, and strategies for business success*. Hoboken: Wiley John & Sons, Inc.

da Silva, N. F. F., Hruschka, E. R., & Hruschka, Jr., E. R. (2014). Tweet sentiment analysis with classifier ensembles. *Decision Support Systems, 66*, 170-179.

Smith, M., Ceni A., Milic-Frayling, N., Shneiderman, B., Mendes Rodrigues, E., Leskovec, J., Dunne, C., (2010). NodeXL: a free and open network overview, discovery and exploration add-in for Excel 2007/2010/2013/2016, from the Social Media Research Foundation: https://www.smrfoundation.org

Zhang, H., Chen, G., Ooi, B. C., Tan, K.-L., & Zhang, M. (2015). In-memory big data management and processing: A survey. *IEEE Transaction on Knowledge and Data Engineering, 27*(7), 1920-1948.

Zhi, Q., Zhao-Wen, L., & Yan, M. (2011). Research of Hadoop-based data flow management system. *The Journal of China Universities of Post and Telecommunications, 18*, 164-168.