

Mapping Twitter Topic Network for Information Sharing using NodeXL

¹Tengku Adil Tengku Izhar, ²Trieu Minh Nhut Le

¹Universiti Teknologi MARA, UiTM Selangor, Malaysia, ²Saigon University
Ho Chi Minh City, Vietnam

Abstract

Social media has a potential to be used as a professional and social networking to share interest. Users share similar interests on the Internet use social media such as Twitter. Despite the various existing methodologies on social media, there are many opportunities that need to be explored. The aim of this paper is to map the information shared on social media in order to identify relevant information. In order to achieve this aim, we examine user trends on sharing information on social media. We use NodeXL to obtain and analyse information from Twitter. 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 information is shared across the social media and how to retrieve this information to assist decision-making process.

1. Introduction

Today, there are tons of data collected from devices all over the world [8,10,17,18]. Devices such as mobile phones, laptops, tablets and even video game such as Play Stations connected to the Internet. These devices transmitted millions of data everywhere in the world and it highlights the development of big data. Billions of devices will be connected to the Internet and soon hundreds of billions of devices. As these devices connected to each other, these connected devices become an intelligent system of systems. Intelligence systems of systems can share data and analyze it over cloud and they can transform our businesses, our lives and our world in countless ways. For example, improving the medical outcomes, building better product faster with lower development cost, making shopping more enjoyable and optimizing energy generated and consumption.

Companies scramble to adapt to this new and unpredictable data reality, they are also beginning to see increasing challenges in meeting demands from the business as they relate to not only being able to cope with the volume of the data but also being able

to turn the data into information. Most of the data we think about from an enterprise perspective is sort of modular, it's transactional, it's a call center call or it's a web transaction or it's a sale or it's a quote or it's a piece of data about a product. Nothing in technology stands still and this is especially true for 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 has 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 its effects in areas such as marketing, politics, health communication, and education [1]. Social media platform such as Twitter has stormed onto the social media scene not only as an individual communication device but also as an information dissemination platform [2]. People on social media express opinions on different topics [3].

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 [4]. Social media can help decision makers to ensure efficient solutions to the problems raised [3]. However, the trustworthiness of this social data is often questionable due to the huge amount of data created in social media.

Social media has now become an important medium of communication and interaction tools for social networks [5]. Social media is also important for business platform that can influence the corporate environment [6], [7], [8]. For example, social networks involve agents in creating and processing information for knowledge network [9]. At the same time, the role that causality can play in social network analysis is unclear [10]. Therefore, it is important to examine the flow of information share on social media and to retrieve relevant information from large amount of it. While most people take such technologies for granted, our understanding about social media is very limited [11]. Most existing

studies on social media remain descriptive, focusing on what people do with social media [11], [12]. Despite the abundant research on IT adoption in general, our understanding about the effectiveness of sharing information through social media is still at the early stage. Moreover, the uniqueness of social media from the other IT applications may require further theoretical extension.

This paper is to address these issues related to social media usage by developing a theoretical framework that extends the existing IT adoption theories, e.g., the Theory of Acceptance and Use of Technology (UTAUT) [13], Technology Acceptance Model (TAM) [14], [15] to accounts for the uniqueness of sharing information through social media. The aim of this paper is to identify the information shared on social media by mapping the topic discusses on social media. The approach is important to identify the flow of the topic on social media. As a result, this paper assists domain experts to identify the relevant information from the discussed topic.

This research will lead to some experimental results that can lead to some new methods and evaluation in sharing information from different perspectives. The outcome will offer an enormous opportunity to advance the science of data analysis so future researchers have new understanding of what is needed to improve their data analysis process. The presence of this research will ensure the projects agility in responding to unfolding events, and substantially enhance its ability to engage in and impact on social media and societies.

The remainder of this paper is organized as follows. Section 2 is background. Section 3 discusses information sharing on social media. Case study is in Section 4. Discussion is in Section 5. The final section contains some concluding remarks.

2. Background

2.1. Big Data

Big data provides significant opportunities for enterprises to impact a wide range of business processes in organizations [16]. Organizations create large number of data every day. However, data was created and captured in many different formats. Therefore, it is difficult to understand the existing relationships between different formats of data. As a result, it is hard to identify which data is relevant to the organization's goals. For example, big and small agencies are lost in an ever-increasing overflow of data. Companies like Google, eBay, LinkedIn, and Facebook were built around Big Data from the beginning [17].

Large number of data makes it harder for data scientists to identify which data is relevant to the organization. It cause an issue on how to effectively

use this data for better decision-making [16]. Data created from the smartphone also poses an issue on how to identify relevant information that matter to the organization.

Big data will be important to organizations that can lead to better decision-making. Organizations should take the advantage of big data opportunities [18]. In order to ensure the effectiveness of the data, organizations need to store the data reliably across a number of databases. Once the data is distributed, and when the needs arise, the organization must find a way to extract the data again, identify which data is needed, assemble it and analyze it. The challenge is to capture relevant data and to harness the value of this data [17].

2.2. Managing Big Data

Volume of data has grown rapidly every year since 2000 with pervasive digitalization content [19]. A recent survey by the Independent Oracle Users Group (IOUG) found that approximately 48% of enterprises expect a moderate increase of data in every five years [20]. In recent year, we have seen a significant increase in the use of big data analytics. Big data provides significant opportunities for enterprises to impact a wide range of business processes in the organizations [21].

High structure and high quality of data is important. It is the foundation for organizations to analyse relevant data into information in formal reports and dashboards. This information has to be used to make actions and decisions based on actual facts. Organization should benefit from information spectrum as a foundation for organization to analyse relevant data into useful information.

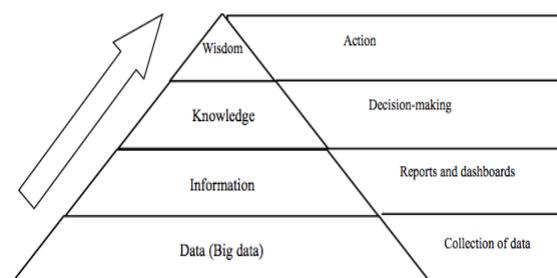


Figure 3. Big Data and Information Management

The implementation of big data highlights the development of big data analytics. Big data analytics could be used to examine large amounts of data to discover useful information. Such information can provide a competitive advantage for the organization. The goal of big data analytics is to assist organization with better decision-making. Organization can enable data analysts to analyze huge volumes of transaction

data, which remains an unresolved challenge for many organizations

2.3. Social Media

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 re-search of big data [22]. Social networking tools (e.g., Twitter, LinkedIn, Facebook, MySpace) have been proliferating on the Internet.

Social media are transforming the way information moves within and between networks of individuals [23]. Social media analytics is an important field that has emerged after the advent of Web 2.0. The key element 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. Social media is a broad term encompassing a variety of online platforms that allow users to create and exchange content. There are two sources of information in social media. User-generated such as sentiments, images, videos, and bookmarks and the relationships and interactions between the network entities such as people, organizations, and products [24].

Social media have profoundly changed our lives and how we interact with one another and the world around us [25, 26]. Recent research indicates that more and more people are using social media applications such as Facebook and Twitters for various reasons such as making new friends, socializing with old friends, receiving information, and entertaining themselves [27], [28], [29], [30]. 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 [4, 31]. 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 [4]. 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 [4].

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 [32]. Data flow is an ordered sequence, which is consecutive, high-speed, infinite and time varying. It's 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 is growing exponentially [33]. There are stricter requirements about the efficiency, expandability and stability of the data flow in social media.

3. Sharing Information on Twitter

As of September 2012, Twitter reported more than 140 million active users (Kolampokis, 2013). Twitter is a popular microblogging service in which users post status messages, called "tweets", with no more than 140 characters. In most cases, its users enter their messages with much fewer characters than the limit established. Twitter represents one of the largest and most dynamic datasets of user generated content, approximately 200 million users post 400 million tweets per day [3], [34]. Twitter users post messages on a variety of topics, unlike blogs, news, and other sites, which are tailored to specific topics.

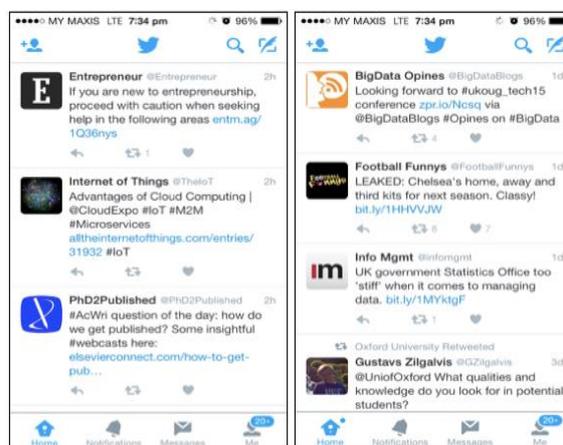


Figure 4. Information shared on Twitter

4. Case Study

NodeXL, a software tool that import data from outside data providers. It is an extendible toolkit for network overview, discovery and exploration implemented as an add-in to the Microsoft Excel 2007 spreadsheet software. NodeXL is applied to retrieve data from social media and import this data. NodeXL demonstrate data analysis with a social media data sample drawn from an enterprise intranet social network. NodeXL is designed to support students who are learning social network analysis and professionals

interested in applying network analysis to business problems. It builds on the familiar spreadsheet paradigm to provide an easy to use tool for non-programmers. Available for free download at: www.codeplex.com/nodexl.

We use Twitter as information source [35]. Twitter is a micro blogging site in which users can post updates (tweets) to friends (followers and following) [3]. We apply data from Twitter to test our ontology. However, Twitter users create high volume of data and make it difficult to analyse 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. Therefore, it is important to structure this data by setting up a goal.

We select one username (@Write4Research) as an example. Write4Research is a Twitter profile that discuss on research activity and it is popular among researchers and PhD students. We set up a goal to examine the trend of people (followers and followings) who mention the word “PhD” in their tweets. In this example, there are three main variables in Twitter. There are other attributes that relate to these variables such as “favorites” and “message” but at this stage we only apply these variables to analyse the goal. The variables are:

1. Username (@Write4Research)
2. Followers
3. Following
4. Tweets



Figure 5. @Write4Research Twitter profile

Figure 5 shows the main profile of @Write4Research. The information shares on this profile help researchers and scientists to be informed on any tips, news and suggestions regarding to their research. The discussions from both “followers” and

“following” are very wide in any discipline so it creates large volume of information. The issue is to evaluate this information into insight. It is important to identify the trend of users in sharing the information on Twitter. For example, the users who share (retweet) the information from this profile might use the #hashtag in order to specify the discussion. Hashtag is a metadata on social media, which makes it easier for users to find message with a specific topic or theme. Some of the popular #hashtag are:

#TheForceAwakens – tweets about the new Star Wars movie.

#worldcup2014 – tweets around the world about the 2014 FIFA World Cup.

#illridewithyou – created to tag message of support for Australian Muslims after the 2014 Sydney hostage crisis.

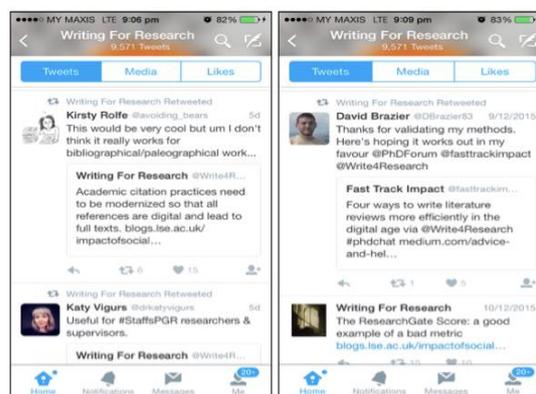


Figure 6. Information share on @Write4Research profile

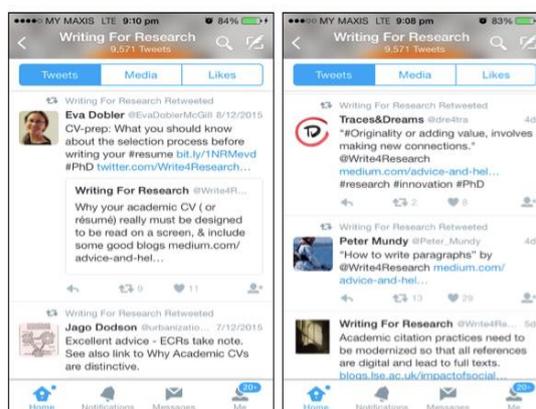


Figure 7. Information with hashtag #PhD on @Write4Research profile

In this network, somewhere there might be someone who mentions the word “PhD” based on three relationships. The relationships are:

1. "tweet_the" relationship.
2. "replies_to" relationship.
3. "mentions_the" relationship.

The relationships shows when the person in this network who tweet, mentions or replies to one another tweets about PhD. We analyse these relationships that represents three different queries. We develop the relationship for the queries as follows:

Query 1: People who tweet about PhD

tweet_the (Followers, PhD), as followers tweet PhD.
 tweet_the (Following, PhD), as following tweet PhD.

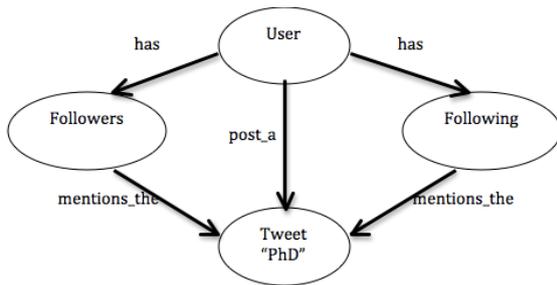


Figure 8. Relationship for query 1: People who follow the tweet about PhD

Query 2: People who replies to the tweets about PhD.
 replies_to (Followers, PhD), as followers replies to PhD.
 replies_to (Following, PhD), as following replies to PhD.

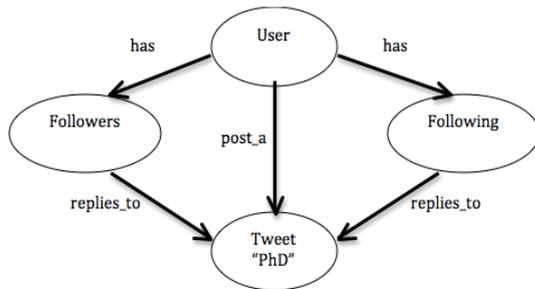


Figure 9. Relationship for query 2: People who reply to the tweet about PhD

Query 3: People who mention the tweets about PhD.
 mentions_the(Followers, PhD) as followers mentions the PhD.
 mentions_the(Following, PhD) as following mentions the PhD.

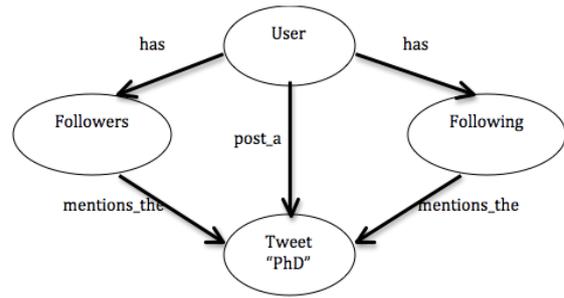


Figure 10. Relationship for query 3: People who mention the tweet about PhD

4.1. Data collection and analysis

In order to identify participants, we used the sampling pool of Twitter that appears in NodeXL, a software tool that import data from outside data providers. It is an extendible toolkit for network overview, discovery and exploration implemented as an add-in to the Microsoft Excel 2007 spreadsheet software. NodeXL is applied to retrieve data from social media and import this data. NodeXL demonstrate data analysis with a social media data sample drawn from an enterprise intranet social network. We apply NodeXL to visualize the ontology for three different queries. When it comes to collecting, computing, analyzing, and acting on social data, technical challenges are quite different because number of social data always increase and make it difficult to be evaluated. For example, today data might be important but tomorrow this data might not be important anymore. Therefore, we filter this data based on certain date and for each relationship. The preferences are used to configure the steps in filtering the data from Twitter using NodeXL.

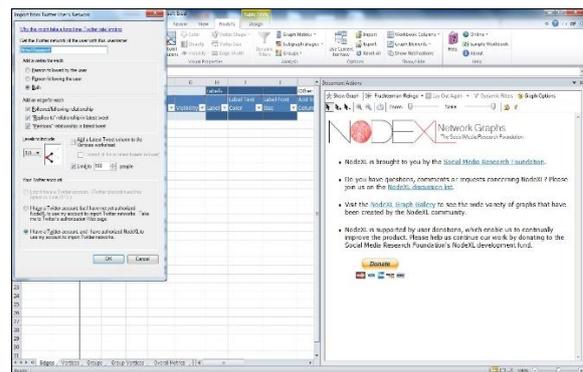


Figure 11. Twitter users network

1. Import from Twitter users network.
 - 1.1. It optionally clears the NodeXL workbook, then get the network of specified Twitter users.
2. Specify the Twitter users with specific username.

- 2.1. Search for @Write4Research.
3. Import basic network plus followers and following who replies, mentions and tweet.
- 3.1. Limit it to 100 recent tweets per user.

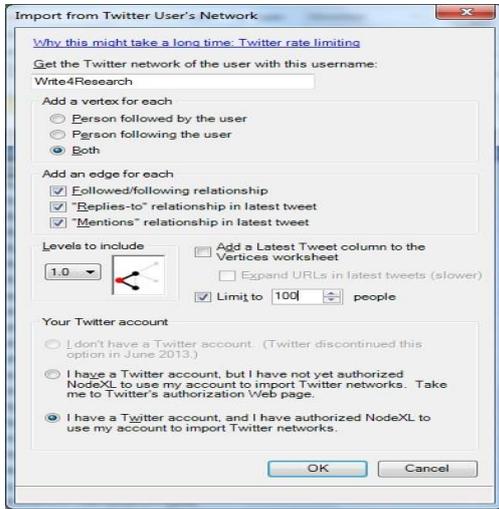


Figure 12. Import tweet username

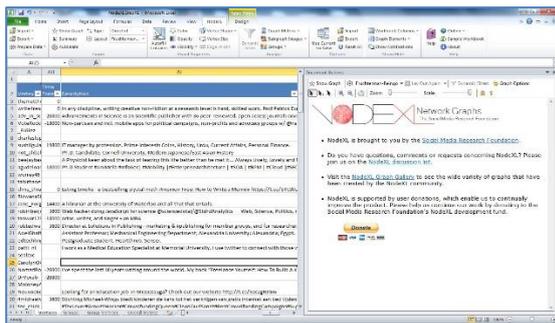


Figure 13. List of usernames in Write4Research network

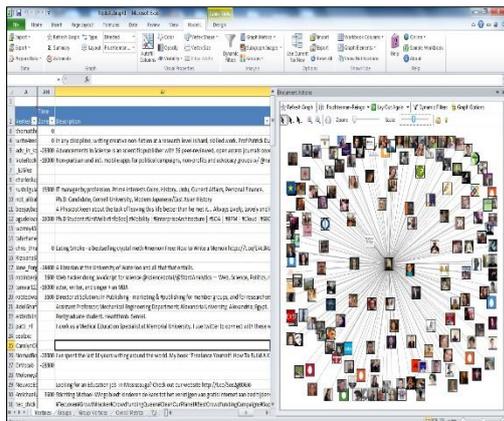


Figure 14. Graph for usernames

4. Import from Twitter search network.
- 4.1. It optionally clears the NodeXL workbook, then get the network of people who tweets certain specified word.
5. Search for the tweets that match to the specific query.
- 5.1. Search for PhD.
6. Import basic network to specifically show who replied or mentioned in the tweets.
- 6.1. Limit to 100 tweets.
7. Filter by relationships (tweet, mentions and replies).

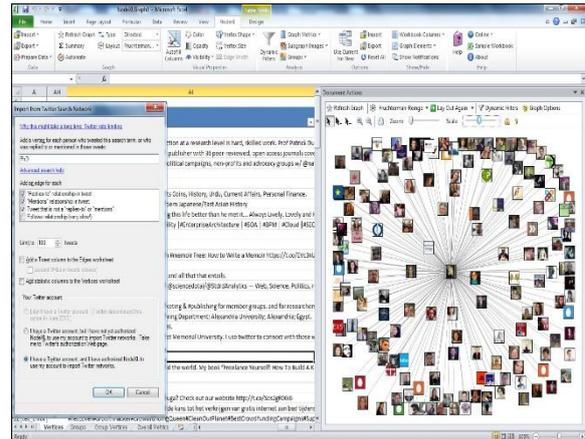


Figure 15. Users network with the search (hashtag) PhD

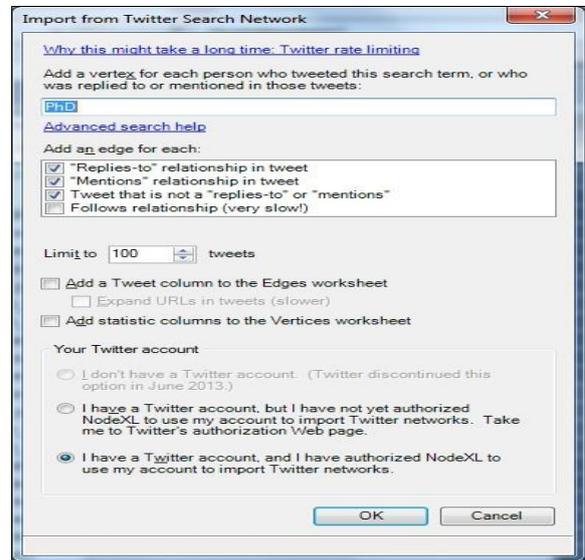


Figure 16. Import tweet from username network

The step-by-step guideline explains how we select data from large volume of social data that relate 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 map to assist their decision-making. In other case, domain

experts might want to use different tool to capture this data and the process to map relevant data might be different.

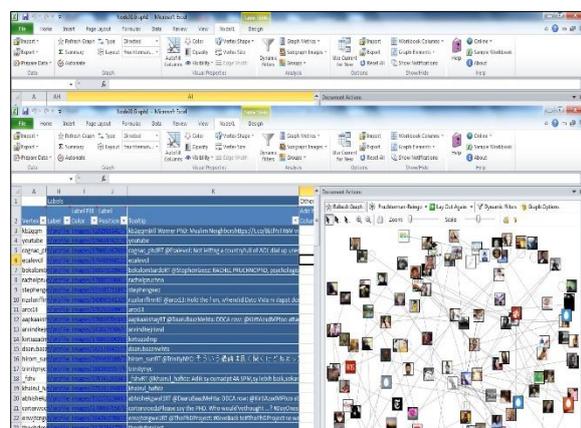


Figure 17. Import tweet from username network

5. Discussion

There are some limitations that need to be considered when dealing with social data. Twitter users create high volume of data and make it difficult to analyse this data for certain query. The amounts of tweet are increasing every day. A reliable analysis approach for this data is a big challenge for entrepreneurs. When it comes to collecting, computing, analyzing, and acting on social data, technical challenges are quite different because number of social data always increase and make it difficult to evaluate. For example, today data might be important but tomorrow this data might not be important anymore. As a result, it is difficult to capture relevant data from social media.

A unique contribution of this paper is its perspective on how to develop the relationship between social data and goal. The relationship helps the process to identify the goal, identify the relevant data that relate to the goal and analyse this relationship to retrieve relevant information to support decision-making.

In this paper, NodeXL has been found to deal better with complex and unstructured social data, in which it can import social data to be analysed in a basic spreadsheet platform. However, this paper limited to social data only. Therefore, future work can be conducted to analyse this data. In the future, an ontology can be used to develop the relationship for social media data. The relationship will filter the amount of social media data. For example, despite the challenges in capturing relevant data from social media, filtering this data would be a better solution to store and analyse the data. As large volumes of social data are unstructured, data becomes more complex to analyse. At the same time, we can examine this data to see if data from social media can support

organizational data for better decision-making. We will apply an ontology to incorporate social media and organizational data.

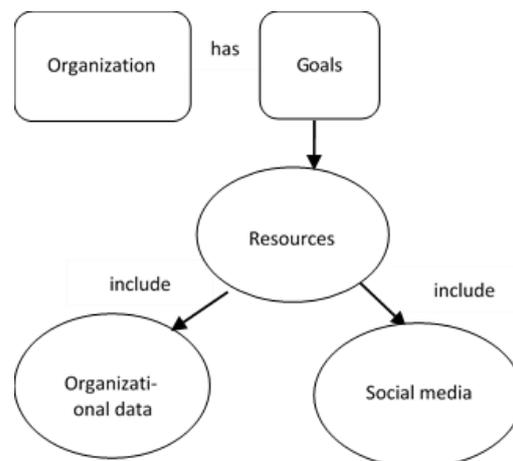


Figure 18. Future work

6. Conclusion

Social media such as Facebook and Twitter have become a necessary to our lives. In this paper, we have described the main features of the proposed framework when we 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 approach. This paper enables researchers to classify and evaluate existing research, to design scientific research, to identify the gaps and weaknesses prior to future research directions.

7. References

- [1] K. McIntyre, 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 (2014) 6-25.
- [2] D. Krause, M. Smith, 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 (2014) 8-27.
- [3] N.F.F.d. Silva, E.R. Hruschka, E.R.H. Jr, Tweet sentiment analysis with classifier ensembles, *Decision Support Systems*, 66 (2014) 170-179.
- [4] W. He, S. Zha, L. Li, Social media competitive analysis and text mining: A case study in pizza industry, *International Journal of Information Management*, 33 (2013) 464-472.
- [5] N.A. Ghani, S.S.M. Kamal, A Sentiment-based filtration and data analysis framework for social media,

5th International Conference on Computing and Informatics ICOCIstanbul, Turkey, 2015, pp. 632-637.

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

[7] L.E. Domianos, D. Cuomo, J. Griffith, D.M. Hirst, J. Smallwood, Exploring the adoption, utility and social influences of social bookmarking in a corporate environment, *International Conference on System Sciences Hawaii*, 2007.

[8] L.E. Damianos, D.L. Cuomo, S. Drozdetski, Handshake: A case study for exploring business networking for enterprise, inside and out, *International Conference on Human Computer Interaction*, 2011.

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

[10] P. Doreian, Causality in social network analysis, *SOCIOLOGICAL METHODS & RESEARCH*, 30 (2001) 81-114.

[11] T.A. Pempek, Y.A. Yermolayeva, S.L. Calvert, College students' social networking experiences on Facebook, *Journal of Applied Developmental Psychology*, 30 (2009) 227-238.

[12] K. Subrahmanyama, S.M. Reich, N. Waechter, G. Espinoza, Online and offline social networks: Use of social networking sites by emerging adults, *Journal of Applied Developmental Psychology*, 29 (2008) 420-433.

[13] V. Venkatesh, M.G. Morris, G.B. Davis, F.D. Davis, User Acceptance of Information Technology: Toward a Unified View, *MIS Quarterly*, 27 (2003) 425-478.

[14] K.A. Saeed, S.A. Helm, Examining the effects of information system characteristics and perceived usefulness on post adoption usage of information systems, *Information & Management*, 45 (2008) 376-386.

[15] C.-C. Shen, J.S. Chou, The impact of perceived ease of use on Internet service adoption: The moderating effects of temporal distance and perceived risk, *Computers in Human Behavior*, 26 (2010) 42-50.

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

[17] T.H. Davenport, J. Dyché, *Big data in big companies*, International Institute for Analytics, 2013.

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

[19] M. Hilbert, P. Lopez, The world's technology capacity to store, communicate and compute information, *Science*, 332 (2011) 60-65.

[20] D. Baum, *Big data, big opportunity*, Oracle.

[21] T.H. Davenport, *Information management and big data- Reference architecture*, in: O.W. Paper (Ed.), Oracle, California, USA, 2014.

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

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

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

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

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

[27] A.M. Kaplan, M. Haelin, Users of the world, united! The challenges and opportunities of Social Media, *Business Horizons*, 53 (2010) 59-68.

[28] P.H. Keckley, *Social networks in health care: Communication, collaboration and insights*, 2010.

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

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

[31] G. Governatori, R. Iannella, A modeling and reasoning framework for social networks policies, *Enterprise Information Systems*, 5 (2011) 145-167.

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

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

[34] A. Ritter, S. Clark, Mausam, O. Etzioni, Named entity recognition in tweets: an experimental study, *Proceeding of the Conference on Empirical Methods in Natural Language Processing Stroudsburg, PA, USA 2011*, pp. 1524-1534.

[35] A.S. Veenstra, N. Iyer, M.D. Hossain, J. Park, Time, place, technology: Twitter as an information source in the Wisconsin labor protests, *Computer in Human Behavior*, 31 (2014) 65-72.