Int. J. Advance Soft Compu. Appl, Vol. 7, No. 3, November 2015
ISSN 2074-8523

Social Network Analysis for Political Blogosphere dataset

Nor Amalina Abdul Rahim[1]and Sarina Sulaiman[1]

UTM Big Data Centre,
Ibnu Sina Institute for Scientific and Industrial Research,
Department of Computer Science, Faculty of Computing,
Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
e-mail: haura.malina@gmail.com, sarina@utm.my

Abstract

Social network analysis is used to analyze networks for better understanding of the network structure. Visualizations in social network analysis show interesting relationships and interactions between actors in the network, while measures of centrality identify the important actors in the network. The purpose of this study is to visualize and measures the centrality of betwenness, closeness, degree and eigenvector for Political Blogosphere dataset by using three different social network analysis tools; ORA, NodeXL and UCINET. The results visualize the actors in a network and indicate which nodes provide better performance for each centrality measure.

Keywords: *Centrality Measurement, Network Visualization, Social Network, Social Network Analysis.*

1 Introduction

Social network analysis (SNA) has been one of the most interesting topics for researchers to analyse and explore the world inside the networks. The concept of SNA helps researchers to study the social network for better understanding of the network structure, the behavior of actor and the relationships between actors in a network. A network is a group of social structures that made up of nodes which some of the nodes are connected to each other by links or also called as edges or ties. Nodes in the network are the actors while links show relationships or connections between the nodes [1].

SNA not only focusing on social networks, but it also focus on other fields such as marketing [1], business [2], medical [3], education [4], community structure [5] and many more. Through SNA, the importance actors, crucial links and network behavior could be determine, and meaningful questions about the structures of relationship can be answer. SNA presume that relationships are important [2]. In [2], they defined SNA as the mapping and measuring of relationships and flows between people, teams, organizations, computers, web sites, and other information or knowledge processing entities. In the early of 1970s, SNA become much more popular with researchers when improvement in computer technology made it possible to study large groups. SNA which consists in generating patterns that allow identifying the underlying interactions between users of different platforms, has been an area of high impact since years ago [4].

In recent years, a number of SNA tools has been developed such as Pajek [5], ORA [6], UCINET [7], NodeXL [8], NetDriller [9] and many more. These tools show a visualized network to the user and user can use the provided functions to perform some general analyses on the network [9].

The results of a social network analysis might be used to identify vital individuals, teams, and units that play the main roles, to make out opportunities to accelerate the flow of knowledge across functional and organizational boundaries, to strengthen the efficiency and effectiveness of existing formal communication channels, to raise awareness of and reflection on importance of informal networks and ways to enhance their organizational performance and to improve innovation and learning [2].

The remainder of this paper is organized as follows. Section 2 describes SNA and related works. Section 3 presents the experimental results. Section 4 discusses related to big data. Finally, section 5 summarizes the most important conclusions of our work.

2 Social Network Analysis and Related Works

Social network analysis (SNA) is a useful tool for studying relations or connections. It is a collection of graph analysis methods that researchers developed to analyze networks in social sciences, computer networks, economics, communication studies, political science, and others [10]. In the early 1930s, Jacob Morene introduced the dynamics of social interaction, and widely credited as the founder of social network analysis. Moreno established the foundations of sociometry, a field of study that later became SNA [11].

In [12], the authors used Social network analysis (SNA) to increase the awareness of leaders about the power of networks, to further catalyze relationships and connections, and to strengthen the capacity of the network to act collectively.

In 2013, the authors in [13] uncovers hidden relationships in a Facebook network. This study aims to explore the following concepts: a) representation of the Facebook network, b) identify the high degree nodes in the network, c) behavior of high degree nodes in the Facebook network. They used a dataset collected in April of 2009 through data scraping from Facebook. A sub-graph consisting of high-degree nodes is obtained from a Facebook social graph. The attributes of these high degree nodes were analyzed using the SNA tool called GEPHI [13].

In [14], the authors proposed the analysis of co-authorship in a specific conference and relation of these co-authors with paper proceedings. They used two different SNA tools which are UCINET and ORA. UCINET is used to calculate centrality measurements statistics, while ORA is used to visualize the data in order to simplify SNA and to express the analysis more clearly.

In 2014 in [5], the authors studying the community structure of Flight MH370 to identify the patterns that emerge from that structure which can lead to demystify some of the many ambiguous aspects of that flight. The aim of their study is to analyze the mesoscopic and macroscopic features of that community by using SNA. They used Pajek to generate a series of social networks that represent the different network communities.

Recently in 2015 in [15], the authors explore some of the many ways in which SNA can be applied to the field of security. They investigated what information someone could infer if they were able to gather data on a person’s friend-groups or device communications such as email and whether this could be used to predict the “hierarchical importance” of the individual. This research could be applied to various social networks to help with criminal investigations by identifying the users with high influence within the criminal gangs on Dark Web Forums, in order to help identify the ring-leaders of the gangs. In this study, they conducted an initial investigation on the Enron email dataset, and investigated the effectiveness of existing SNA metrics in establishing hierarchy from the social network created from the email communications metadata. They tested the metrics on a fresh dataset to assess the practicality of their results to a new network [15].

SNA provides both a visual and a mathematical analysis of human relationships. SNA techniques have been applied to a variety of problems and they have been successful in uncovering relationships that cannot be seen with any other traditional method [6]. Also, visualization techniques are important aids in helping researchers to understand social and conversational patterns in online interactions. Visualizations of social networks can show whether interactions are occurring between all members of a group or whether some group members are communicating more (or less) with other specific individuals [15].

Visualization in SNA represents the network visually, showing interesting relationships and interactions between actors in the network, whether interactions are occurring between all actors or whether some actors are communicating more or less with other specific individuals which those situation of interaction may be analyzed and have a depth explored [16, 17]. In 1997, Alfred Crosby in 1997 asserted that besides measurement, visualization is one of the two factors accountable for the evolution of modern science [18]. Visualization is such a great help for SNA researchers in understanding new methods to produce and portray images that contains information inside it and effectively convert those information into meaningful information. [18, 19, 20].

In [21], Freeman mentioned that in order to identify certain nodes within the network, various shapes and colors according to social variables or other node properties are useful. He suggests different icons to distinguish gender, age classes or ethnic groups.

Figure 1 depicts example of visualization in social network.

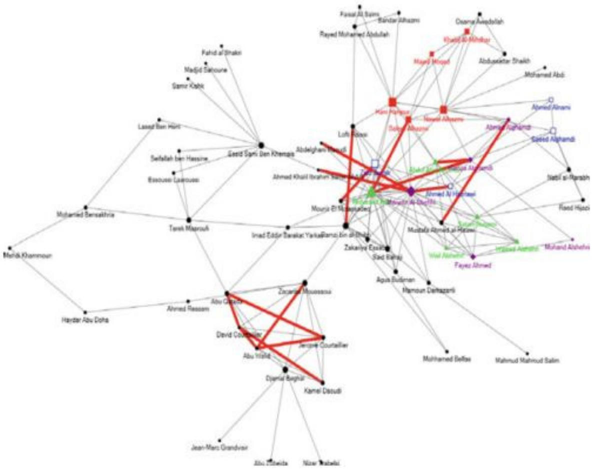


Fig. 1: Example of visualization in social network [22]

Even small networks are hardly articulate when their nodes are highly interlinked especially when the links of the relations have different meanings. It is important to layout the nodes of the network in a clear way to make the structure of a network easily intelligible [23].

In SNA, centrality is an important concept [24]. Measures of centrality is used to identify how importance the actor is in a network. As stated by Freeman in [25], it seems that people agree that centrality is a vital structural attribute of social network.

In 1948, Bavelas came out with the idea of centrality as applied to human communication. The follow-up studies concluded that centrality was related to group efficiency in problem-solving, personal satisfaction of participants and the perception of leadership. The idea of centrality is alive and has been applied in extensive range of applications. Many centrality measures have been proposed to estimate the prominence of a node in a network.

As mentioned by Palazuelos et al., (2013) in [4], the concept of centrality raised the question “Which are the most important nodes in a social network?” Although there are many possible definitions of importance, prominent nodes are supposed to be those that are extensively connected to other nodes.

Four main types of centrality measures in network analysis which include the following: betweenness centrality, closeness centrality, eigenvector centrality and degree centrality. These four measures of centrality are widely used in SNA.

3 Experimental Results

In this study, we analyzed Political Blogosphere (PolBlogs) dataset on two types of SNA tools. This dataset is represented in different format according to the tools. This section divides data representation in two parts; the first part covers the obtained results from ORA, the second part describes the results reveal from NodeXL. Each part discusses the visualization of the network in the dataset and the results of centrality measurement.

3.1 ORA

For ORA, we have used two types of format for PolBlogs dataset. Both format are compatible for ORA tool. It is a directed network of hyperlinks between weblogs on US politics recorded in 2005 [26]. The dataset that obtained from CASOS website is in xml format [27] and from [28] in GraphML format. In xml format, the nodes are the URLs of the blogs and the edge connects the URLs. Nodes and edges of this dataset are represent by numbers. Both contains 1490 nodes. Both format are imported into ORA. Table 1 reports the statistic network for political blogosphere dataset is analyzed by using ORA.

Table 1: Statistic network for PolBlogs dataset

Number of Nodes	Number of Edges
1000	10238

The result shows that it contains only 1000 nodes and 10238 edges. However, in previous study done by Kósa et al, (2015) in [29], they claimed that it contains 1490 nodes and 19090 edges.

Fig. 2 (a) and Fig. 3 (b) illustrate the centralized effect in a 2D mode. Fig. 2 (a) shows the visualization of Polblogs dataset in xml format which the nodes are label by the name of the URLs. For Fig. 2 (b), in GraphML format, nodes are label by number.

mediaprima.com/ny1962
tomdaschle.com
charlineandjamie.com/dotnetweb01a/blogdisplay.aspx
americanworldview.tripod.com
tesibria.typepad.com
xanga.com/kerry4president
takebackthemedia.com
obamablog.com
alternateworlds.blogspot.com
lonelybored.typepad.com
feminist_hellcat
commonprejudice
makinglight
wiserblog

Fig. 2(a): 2D visualization in XML format

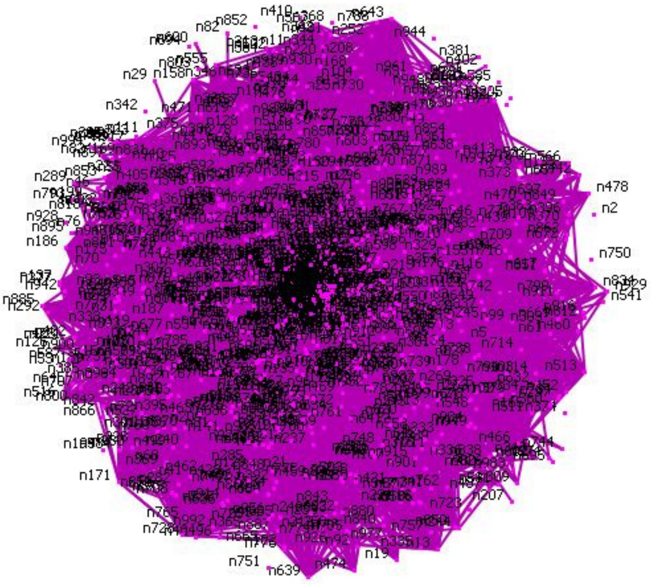


Fig. 2(b): 2D visualization in GraphML format

The clusters cannot be seen clearly since there are more than 10000 links and some of them are overlapping on each other. However, it can be seen clearly that the most connections happen in the center. We zoom into the center of the network to see the connections between the nodes as shown in Fig.3 (a) and other parts of the network as shown in Fig.3 (b), and the results reveal that there are many cases which one node (URL) has connection with many URLs.

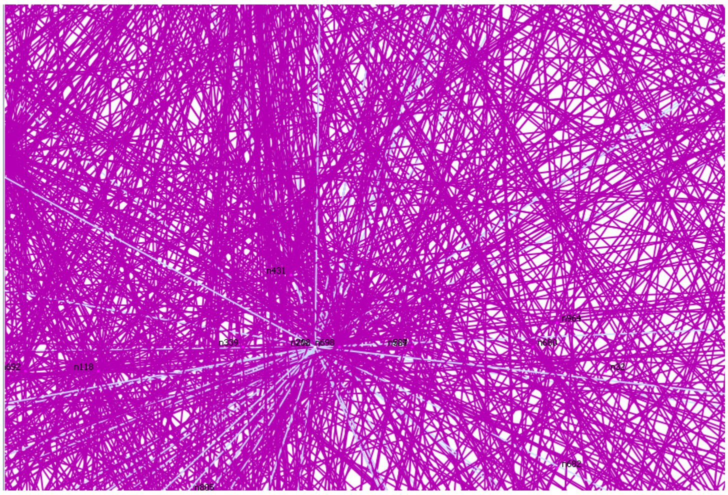
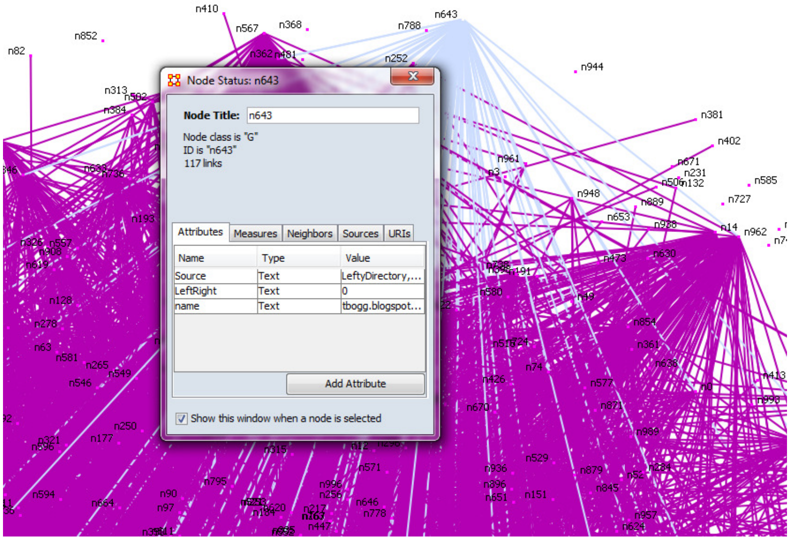


Fig. 3(a): 2D visualization – zoom in



Node Status: n643

Node Title: n643

Node class is "G"
ID is "n643"
117 links

Attributes | Measures | Neighbors | Sources | URIs

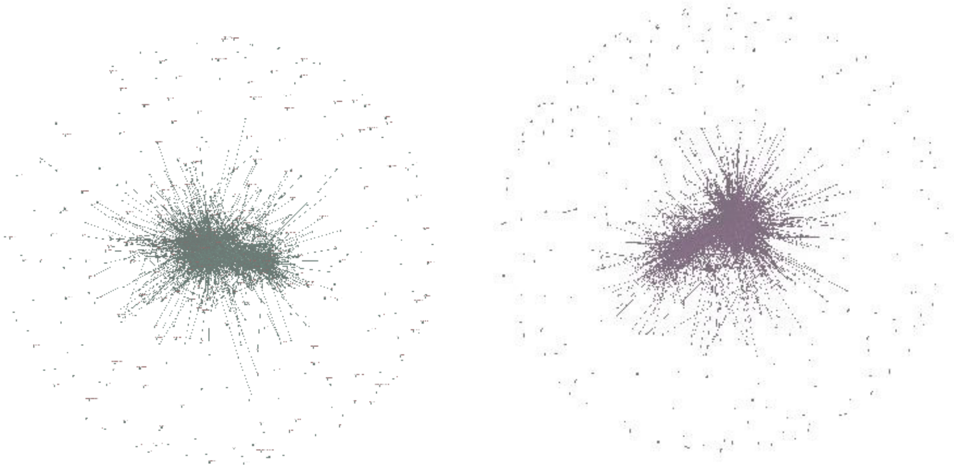
Name	Type	Value
Source	Text	LeftyDirectory,...
LeftRight	Text	0
name	Text	tbogg.blogspot...

Add Attribute

☑ Show this window when a node is selected

Fig. 3(b): 2D visualization – zoom in

Then, we visualize the data in 3D mode. It is quite slow for ORA to visualize the data in 3D mode for both, xml and GraphML format. The visualizations exhibit an interesting result as shown in Fig. 4. The data is pulled to the center of network.



(a) XML format (b) GraphML format

Fig. 4: 3D visualization

By using centrality measurement that provided in ORA, we can identify which nodes have the highest value for betweenness centrality, closeness centrality, eigenvector centrality and degree centrality for PolBlogs dataset.

Both dataset obtained the same results for each centrality measures. Table 2 indicates the result for betweenness centrality. Blog atrios.blogspot.com is on the top of the rank, which means this blog occurs on many of the shortest paths between other blogs in the network, so that it has the highest betweenness compared to others. The maximum value of betweenness centrality in this network is 0.061.

Table 2: Top 5 scores node for betweenness centrality

Node	Value
atrios.blogspot.com	0.061
blogsforbush.com	0.051
dailykos.com	0.042
newleftblogs.blogspot.com	0.031
23madkane.com/notable.html	0.028

Eigenvector centrality calculates the influences of a node in a network. Table 3 demonstrates the results of eigenvector centrality. Blog atrios.blogspot.com is on the top of the rank with maximum value 0.253. We could say that this blog is the most popular and the most influential since it has the highest eigenvector value compared to others. Blog washingtonmonthly.com has the lowest value of eigenvector centrality for top five ranking which shows that this blog has low influences and less popular among the network.

Table 3: Top 5 scores node for eigenvector centrality

Node	Value
atrios.blogspot.com	0.253
dailykos.com	0.252
talkingpointsmemo.com	0.220
liberaloasis.com	0.199
washingtonmonthly.com	0.196

The average closeness of a node to the other nodes in a network is called as closeness centrality. In this study, the value of average closeness of a node to the other nodes is 0.002. The nodes in the network is really closed to each other.

Table 4 demonstrates the top five blogs for in-degree and out-degree centrality. In-degree centrality refers to the number of links that the node receives from other nodes. The higher the in-degree of a blog, the more attention that the blog receives from other blogs, meaning large number of blogs interacts with that blog. Blog dailykos.com received highest number of connections from other blogs while juancole.com received least connections from other blogs. The maximum value for in-degree centrality is 0.309. In short, many blogs communicate to dailykos.com.

Table 4: Top 5 score nodes for degree centrality

In-Degree centrality		Out-Degree centrality	
Node	Value	Node	Value
dailykos.com	0.309	newleftblogs.blogspot.com	0.137
atrios.blogspot.com	0.249	politicalstrategy.org	0.129
talkingpointsmemo.com	0.242	madkane.com/notable.html	0.124
washingtonmonthly.com	0.175	liberaloasis.com	0.115
juancole.com	0.154	corrente.blogspot.com	0.106

Out-degree centrality refers to the number of links that the node sends to other nodes. Blog newleftblogs.blogspot.com sends the most connections to other blogs. This blog is actively interact to the other blogs. The maximum value for out-degree centrality is 0.137. The corrente.blogspot.com is less active since it is the least blog that sends connections to other blogs.

3.2 NodeXL

For NodeXL, we used the PolBlogs dataset in GraphML format since xml format is not compatible for NodeXL. Table 4 reports the statistic network for PolBlogs dataset analyse by using NodeXL. It contains 1490 nodes and 19090 total edges. The results of statistic network is same as in [30].

Table 5: Statistic network for PolBlogs dataset

Number of Nodes	Number of Edges
1490	19090

Fig. 5 depicts the visualization of the network in NodeXL. The most connections happen in the center but some of the nodes on the side looks like not linked to any other nodes.

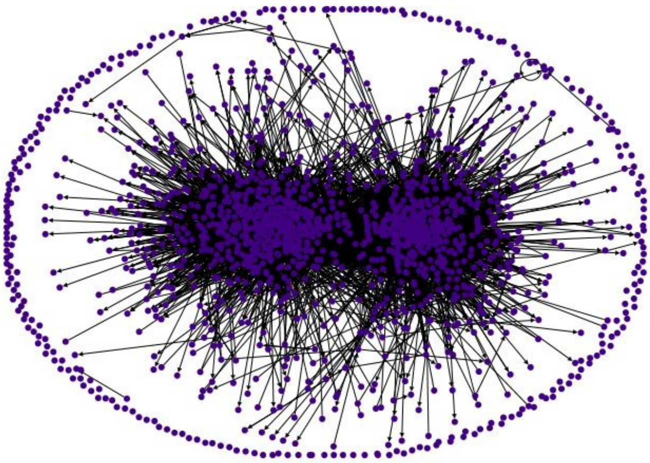
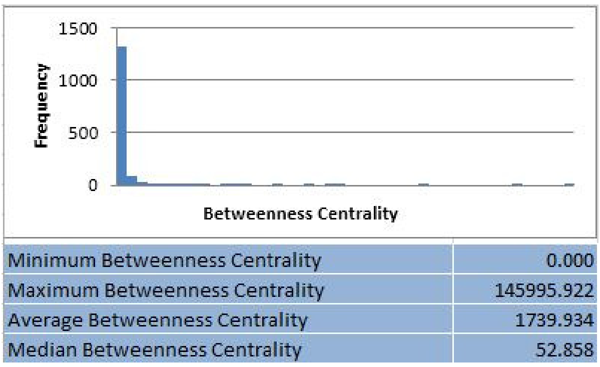


Fig. 5: Visualization in GraphML format

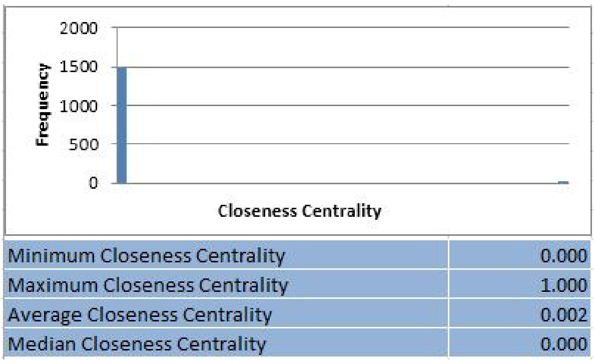
In NodeXL, the Analysis | Graph Metrics option was used to generate the centrality measures. Fig. 6 indicates the result for betweenness centrality. The maximum value of betweenness centrality is 145995.922 which is own by blogsforbush.com. This result is different compared to result that obtained from ORA.



Minimum Betweenness Centrality	0.000
Maximum Betweenness Centrality	145995.922
Average Betweenness Centrality	1739.934
Median Betweenness Centrality	52.858

Fig. 6: Betweenness centrality

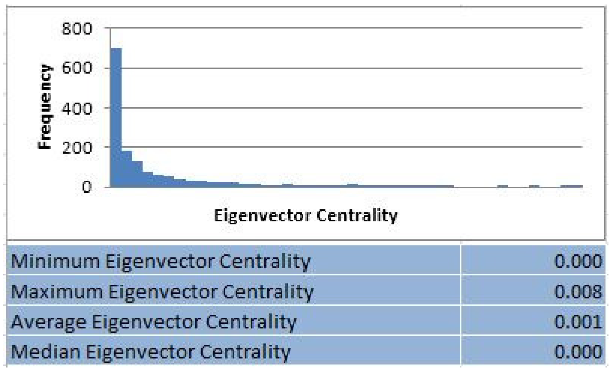
For closeness centrality as shown in Fig. 7, the average closeness of a node to the other nodes is 0.002 which is same as obtained by ORA.



Minimum Closeness Centrality	0.000
Maximum Closeness Centrality	1.000
Average Closeness Centrality	0.002
Median Closeness Centrality	0.000

Fig. 7: Closeness centrality

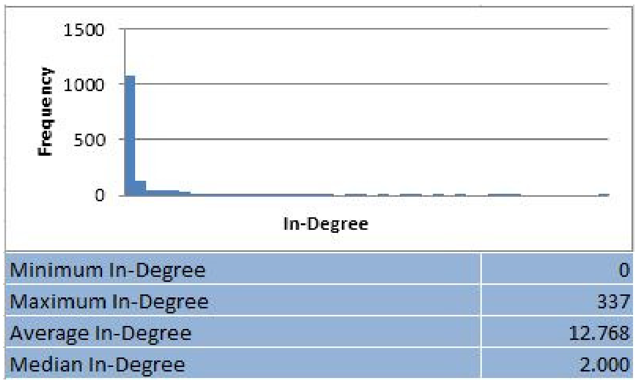
For eigenvector centrality, the results is also same as in ORA as depict in Fig 8. Blog atrios.blogspot.com has the highest eigenvector centrality.



Minimum Eigenvector Centrality	0.000
Maximum Eigenvector Centrality	0.008
Average Eigenvector Centrality	0.001
Median Eigenvector Centrality	0.000

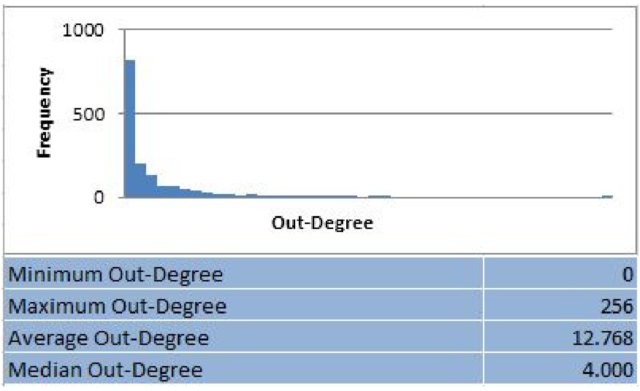
Fig. 8: Eigenvector centrality

Fig. 9 and Fig. 10 demonstrate the results for degree centrality. The maximum in-degree value is 337 which is scored by blog dailykos.com. This blog received highest attentions from other blogs. This result is same as obtained by ORA. The maximum out-degree value is 256 which is scored by blogforbush.com. Blog blogforbush.com sends the most connections to other blogs. This blog is actively interact to the other blogs. Result for out-degree centrality is different from ORA.



Minimum In-Degree	0
Maximum In-Degree	337
Average In-Degree	12.768
Median In-Degree	2.000

Fig. 9: In-degree centrality



Minimum Out-Degree	0
Maximum Out-Degree	256
Average Out-Degree	12.768
Median Out-Degree	4.000

Fig. 10: Out-degree centrality

3.3 UCINET

For UCINET, we used PolBlogs dataset in .net format. In UCINET, we only measured for betweenness, eigenvector, and degree centrality.

For betweenness centrality, the maximum value is 6.589 which is owned by blogsforbush.com. This result is same as obtained by NodeXL. For eigenvector centrality, blog dailykos.com has the highest eigenvector centrality. This result is different compared to ORA and NodeXL.

The maximum in-degree value is 337 which is scored by blog dailykos.com. This blog received highest attentions from other blogs. This result is same as obtained by ORA and NodeXL. For out-degree centrality, the maximum value is 256 which is scored by blogforbush.com. Blog blogforbush.com sends the most connections to other blogs. This result is same as obtained by NodeXL. Therefore, the blogger of blogforbush.com is the active blogger since he likes to communicate and interact with most of the blogs in this network.

We summarized the results for betweenness, eigenvector, and degree centrality for ORA, NodeXL and UCINET in Table 6.

Table 6: Results of centrality measures for PolBlogs dataset

	ORA	NodeXL	UCINET
Betweenness	atrios.blogspot.com	blogsforbush.com	blogsforbush.com
Eigenvector	atrios.blogspot.com	atrios.blogspot.com	dailykos.com
In - Degree	dailykos.com	dailykos.com	dailykos.com
Out - Degree	newleftblogs.blogspot.com	blogsforbush.com	blogforbush.com

4 Discussion

Burkholder (1992) stated that big data not only refers to very large data sets, the tools and the procedures used to manipulate and analyze the data, but also to a computational turn in thought and research. Big data has been used in the sciences to refer to data sets large enough to require supercomputers, but what once required such machines can now be analyzed on desktop computers with standard software. There is little doubt that the quantities of data now available are often quite large, but that is not the defining characteristic of this new data ecosystem.

Data science arises out of the big data world and complexity science. Social network analysis and visualization is a part of data science. Data science is the extraction of knowledge from data. However, Jeff Lekk posted in [30], the key word in data science is not "data"; it is "science". Data science is only useful when the data are used to answer a question. He also added that the problem with this view of data science is that it is much harder than the view that focuses on data size or tools. Data size of 100Gb or only 3Kb are useful if it is able for answering the real question.

5 Conclusion

This study looked at three different SNA tools and the results for the basic centrality measures. We used two types of format for ORA, xml and GraphML format, and both data contains 1000 nodes and 10238 edges when analysed in

ORA. This is different from the previous study and the original data itself. Both data in different format produced the same visualization for 2D and 3D mode.

For NodeXL and UCINET, the dataset contain 1490 nodes and 19090 edges. We have found that for trial version of ORA, it is limited only to 1,000 nodes per node set. This is the reason why it only contain 1000 nodes when analysed in ORA. The visualization of data in NodeXl is slightly different from ORA especially on the side part. However, most connections happened in the center same as visualized by ORA.

For betweenness centrality, NodeXL and UCINET obtained the same results. Blog blogsforbush.com has the highest betweenness centrality. For eigenvector centrality, ORA and NodeXL have the same results. Blog atrios.blogspot.com has the highest eigenvector centrality. This blog is popular and most influential.

For closeness centrality, the average is 0.002 as acquired by ORA and NodeXL which means nodes in the network is really closed to each other.

For in-degree centrality, all three SNA tools gained the same results. Blog dailykos.com received highest number of connections from other blogs. Many blogs communicate to dailykos.com. However for out-degree centrality, results from ORA is different from NodeXL and UCINET. Blog blogforbush.com has the highest out degree centrality. Blog blogforbush.com sends the most connections to the other blogs. The blogger of blogforbush.com is very active blogger.

Even though the data that has been used only contains 1490 nodes, but it has more than 10000 links. Through this analysis, we managed to answer which is the importance actors in the network. Various types of SNA tools has been provided for different purpose. How big the data is, it still can be analysed by using SNA tools and the importance actors in the network can be identified by using the measures of centralities.

ACKNOWLEDGEMENTS

This research is supported by The Ministry of Higher Education (MOHE) and collaboration with Research Management Centre (RMC), Universiti Teknologi Malaysia. This research is financial supported by University Grant (GUP) Program, Tier 1 (Q.J130000.2528.03H00), UTM Flagship Grant (Q.J130000.2428.02G70), FRGS Grant (R.J130000.7828.4F634) and E-Science Fund (R.J130000.7928.4S117). The authors would like to express their deepest gratitude to the Smart Digital Community Research Alliance, UTM and Soft Computing Research Group (SCRG) for their encouragement.

References

[1] Zeng, W., Huang, Y., and Jiang, L. 2011. The study of microblog marketing based on social network analysis. *Proceedings - 2011 4th International Conference on Information Management, Innovation Management and Industrial Engineering, ICIII 2011*, Vol.3, 410–415.

[2] Zhu, M., Liu, W., Hu, W., and Fang, Z. 2009. Social Network Analysis in IT Company. *2009 International Conference on E-Learning, E-Business, Enterprise Information Systems, and E-Government*, 305–307.

[3] Sulaiman, N. S., and Shamsuddin, S. M. 2011. Feature granularity for cardiac datasets using Rough Set. *2011 IEEE International Conference on Computer Science and Automation Engineering*, 346–352.

[4] Palazuelos, C., Garc, D., and Zorrilla, M. 2013. Social Network Analysis and Data Mining : An Application to the E-Learning Context, 651–660.

[5] Al-taie, M. Z., Shamsuddin, S. M., and Ahmad, N. B. 2014. Flight MH370 Community Structure, *International Journal of Advanced Soft Computing and Applications*,Vol.6, No.2, 1–20.

[6] Sun, W. and Qiu, H. 2008. A social network analysis on Blogospheres. *2008 International Conference on Management Science and Engineering 15th Annual Conference Proceedings*, 1769–1773.

[7] Sathik, M. M., & Rasheed, A. A. 2011. Social Network Analysis in an Online Blogosphere, Vol.3, No. 1, 117–121.

[8] Lieberman, M. 2014. Visualizing Big Data: Social Network Analysis. *Digital Research Conference*.

[9] Koochakzadeh, N., Sarraf, A., Kianmehr, K., Rokne, J., and Alhajj, R. 2011. NetDriller: A Powerful Social Network Analysis Tool. *2011 IEEE 11th International Conference on Data Mining Workshops*, 1235–1238.

[10]Erlin., Yusof, N., and Rahman, A. A. 2009. Analyzing Online Asynchronous Discussion Using Content and Social Network Analysis. *2009 Ninth International Conference on Intelligent Systems Design and Applications*, 872–877

[11]Moreno, J. 1934. Who Shall Survive?, Beacon House.

[12]Hoppe, B., and Reinelt, C. 2010. Social network analysis and the evaluation of leadership networks. *Leadership Quarterly*, Vol.21, No.4, 600–619.

[13]Akhtar, N., Javed, H., and Sengar, G. 2013. Analysis of Facebook Social Network. *2013 5th International Conference on Computational Intelligence and Communication Networks*, 451–454.

[14]Raca, V., and Cico, B. 2013. Social Network Analysis , Methods and Measurements Calculations. *2nd Mediterannean Conference on Embedded Computing, MECO 2013*, 2–5.

[15]Phillips, E., Nurse R. C. J., Goldsmith, Michael., and Creese, S. 2015. Applying Social Network Analysis to Security. *International Conference on Cyber Security for Sustainable Society*, 11–27.

[16]Haythornthwaite, C. 2002. Building social networks via computer networks: Creating and sustaining distributed learning communities, 159–190. Cambridge University Press.

[17] Bertini. E. 2008. Social Networks Visualization: A Brief Survey.

[18]Freeman. L. C. 2000. Visualizing Social Network. Journal of Social Structure.(1).http://www.cmu.edu/joss/content/articles/volume1/Freeman.html (Last accessed: Oct 2013).

[19]Viegas. F. B, Donath. J. 2004. Social Network Visualization: Can We Go Beyond the Graph?,in Workshop on Social Networks (CSCW'04), Chicago. 6:10.

[20]Tantipathananandh, C., Berger-Wolf, T., and Kempe, D. 2007. A framework for community identification in dynamic social networks. *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '07*, 717.

[21]Freeman, L.C. 2005. Graphic techniques for exploring social network data. In: P.J. Carrington, J. Scott and S. Wasserman (Eds.), Models and Methods in Social Network Analysis, 248–269. Cambridge University Press.

[22]Ghali, N., Panda, M., Hassanien, A. E., Abraham, A., and Snasel, V. 2012. Social Networks Analysis: Tools, Measures and Visualization. *Computational Social Networks: Mining and Visualization*, Springer- Verlag London (2012), 3–23.

[23]Holzhauer, S. 2010. Developing a Social Network Analysis and Visualization Module for Repast Models. Kassel University Press.

[24]Dekker. A. H. 2008. Centrality in Social Networks: Theoretical and Simulation Approaches. *Proceeding of SimTecT.* Melbourne, Australia, 33-38.

[25]Freeman, L. C. 1978. Centrality in Social Networks Conceptual Clarification, *1*(1968), 215–239.

[26]Adamic, L. A. and Glance, N. 2005. The Political Blogosphere and the 2004 US Election : Divided They Blog, *Proceedings of the WWW-2005 Workshop on the Weblogging Ecosystem (2005).* 36–43.

[27]CASOS. http://www.casos.cs.cmu.edu/tools/data2.php (Last accessed: June 2013).

[28]Nexus. http://nexus.igraph.org/api/dataset_info?id=13&format=html (Last accessed: August 2015)

[29]Kósa, B., Balassi, M., Englert, P., and Kiss, A. 2015. Betweenness versus Linerank. *Computer Science and Information Systems*, *12*(1), 33–48.

[30]Jeff Leek. (December, 12). Retrieved from http://simplystatistics.org/2013/12/12/the-key-word-in-data-science-is-not data-it-is-science/