Jurnal
Teknologi

Full Paper

VISUALIZATION AND CENTRALITY MEASUREMENT OF SOCIAL NETWORK ANALYSIS

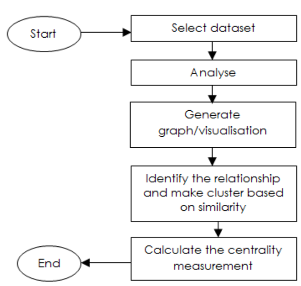
Nor Amalina Abdul Rahim, Sarina Sulaiman*, Siti Zaiton Mohd Hashim and Nor Bahiah Ahmad

UTM Big Data Centre, Ibnu Sina Institute for Scientific and Industrial Research, Faculty of Computing, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia

Article history
Received
16 November 2015
Received in revised form
23 February 2016
Accepted
22 March 2016

*Corresponding author
sarina@utm.my

Graphical abstract



Abstract

Social networks have increased in popularity and play an important role in people's life nowadays. Hundreds of millions of people participate in social networks and the number is growing day by day. Social networks have become a useful tool and help people in every field of life such as in education, politics and business. Social networks give people the idea of knowing and interacting with each other, experiencing the power of sharing and being connected with people from different places and countries. The purpose of this study is to analyse the behaviour of actors in a network, the graph and the relationship between actors in social networks. The researcher expects to use the technique of Social Network Analysis with Organisation Risk Analyser (ORA) tool to analyse the data. Three different types of dataset are analysed in the form of network visualisation and centrality measurement. The results reveal the hidden relationships and clusters in the network, and indicate which nodes provide better performance for each centrality measure.

Keywords: Social network, social network analysis, network visualisation, centrality measurement

Abstrak

Dewasa ini, rangkaian sosial telah meningkat dari segi populariti dan memainkan peranan yang penting dalam kehidupan manusia. Hari demi hari, berjuta-juta orang melibatkan diri dalam rangkaian sosial dan jumlah itu semakin meningkat. Rangkaian sosial telah menjadi alat yang berguna dan membantu manusia dalam setiap bidang kehidupan seperti dalam bidang pendidikan, politik, dan perniagaan. Rangkaian sosial memberi idea untuk mengetahui dan berinteraksi antara satu sama lain, berpeluang merasai kuasa perkongsian dan berhubung dengan orang lain dari tempat dan negara yang berbeza. Tujuan kajian ini adalah untuk menganalisis tingkah laku pengguna dalam rangkaian, graf dan hubungan antara pengguna dalam rangkaian sosial tersebut. Penyelidik dalam kajian ini menjangka untuk menggunakan teknik analisis rangkaian sosial menggunakan alat *Organisation Risk Analyser* (ORA) untuk menganalisa data. Tiga jenis data yang berbeza dianalisa dalam bentuk visualisasi rangkaian dan ukuran keutamaan. Hasil analisis mendedahkan hubungan dan kelompok yang tersembunyi dalam rangkaian, dan menunjukkan nod yang manakah yang memberikan prestasi yang lebih baik bagi setiap ukuran keutamaan.

Kata kunci: Rangkaian sosial, analisa rangkaian sosial, visualisasi rangkaian, pengukuran keutamaan

1.0 INTRODUCTION

Social networks (SN) have grown to be popular sites that contain a lot of information from all over the world. It has become a concern for several researchers to investigate the world inside SN. The World Wide Web, people's activities on the internet such as chatting, email exchanges, group interactions, professional citations, consumer behaviour in e-commerce and e-learning are some examples of social networks [1].

A social network is a group of actors and their interconnections [2]. There are two basic elements in a network which are called nodes (actors) and links that connect the nodes (the interconnections). The number of links that the nodes have, emphasises the node has more or less connections with other nodes in a network. Mostly, SN are visualised by graphs, where nodes represent individuals or groups and links represent the connection or relation between them. Fig.1 shows an example of a social network in the form of a graph. Graphs can be directed or undirected depending on the type of the connection between the linked nodes. Links between the nodes can be weighted or dichotomous (unweighted) to label different interaction strengths. There are many SN visualisers widely used for network visualisation such as ORA, NetDraw, NetVis, NodeXL, Gephi, Pajek and many others [3].

The emergence and growth of SN users has encouraged researchers to analyse social networks and explore the world inside the networks. Social Network Analysis (SNA) is developed to understand the behaviour of actors in a network, the graph and the relationship between actors in social networks, in which the term actor could be a person or individual, organisation, event or object.

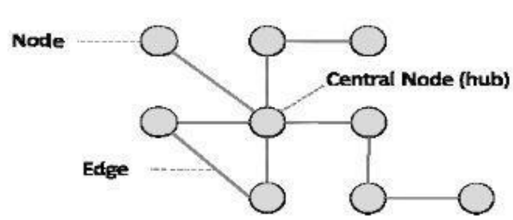


Figure 1 Example of a social network as a graph [5].

In addition, the goals of SNA are to determine vital actors, identify the crucial links, roles, subgroups, network behaviour, to answer meaningful questions about structures and many others [4].

The structure of this paper is organised as follows: in Section 1, the background and introduction of SN is introduced. The related works and important concepts of social network analysis are reviewed in Section 2. Section 3 presents the proposed experimental method for this research. The experimental and analysis results are discussed in Section 4. In Section 5, this paper is concluded with suggestions for further research.

2.0 SOCIAL NETWORK ANALYSIS

SNA is the study of relationships among individuals in a group, including the analysis of social positions, social structures, role analysis, and many more. In short, SNA is the study of social networks for a better understanding of the network structure and behaviour. In the early 1970s, SNA become much more popular with researchers when improvements in computer technology made it possible to study large groups. Within the last ten years, SNA has increased to prominence in a number of fields, including anthropology, sociology, organisational behaviour, and medicine.

In [2], Sun and Qiu used the SNA method to explore and study the link structure of the Sina's VIP Blogosphere and the behaviour patterns of its members. They used a network visualisation tool called Graphviz to draw the structure of the Blogosphere. They focused on degree centrality: out-degree and in-degree for centrality measurement. The experimental result shows that the larger the out-degree of a blog, the more attention is paid to other bloggers by the owner of this blog.

In [6], Wu et al. used UCINET to analyse the database of social networking websites by applying the techniques of SNA and web mining. They wanted to discover the social relationship of members in the blogs and the association between members, and to find the interest groups in the blogspace. They proposed a methodology to combine the techniques of social network analysis and web mining to discover the interest groups in the blogspace. The interest groups are used to develop a mechanism and to construct a product recommendation system based on the network of consumers. They used degree centrality and closeness centrality, as measures of SNA.

In 2010, Mansur et al. carried out the analysis of social learning networks and revealed the hidden behaviour during the interaction inside E-Learning@UTM wiki. They modelled the relationship in the form of a graph representation and then analysed it by using SNA through the UCINET tool. Their study also focused on degree centrality: out-degree and in-degree. The experimental result shows one of the authors is very active in editing, updating or creating the wiki [7].

In [8], Sathik and Rasheed analysed the blog responses that were posted by AIDS patients over a period of time. The dataset only contained 146 nodes. They used NetDraw to show the visualisation of the network. Two different centrality measures; betweenness centrality and closeness centrality are used as measures of SNA. Based on the results, they found that the node that has the maximum betweenness centrality, has the lowest score of closeness centrality. However, the node that has a low betweenness centrality can achieve a high closeness centrality. They concluded that a vertex which has the highest betweenness centrality has a lesser score in closeness centrality [8].

Recently in 2013, Akhtar et al. [9] uncovered hidden relationships in a Facebook network. This study aimed to explore the following concepts: a) representation of the Facebook network, b) identification of the high-degree nodes in the network, c) the behaviour of high-degree nodes in the Facebook network. They used a dataset collected in April 2009 through data scraping from Facebook. A sub-graph consisting of high-degree nodes was obtained from a Facebook social graph. The attributes of these high-degree nodes were analysed using the SNA tool called GEPHI [9].

In [10], Raca and Cico proposed the analysis of co-authorship in a specific conference and the relation of these co-authors with paper proceedings used separate SNA tools, which are UCINET and ORA. UCINET is used to calculate the centrality measurements statistics, while ORA is used to visualise the data in order to simplify SNA and to express the analysis more clearly.

2.1 Visualisation

Visualisation plays an important role in improving the understanding of SNA. Visualisation represents the social network visually, showing interesting relationships between points in the social network which may be analysed and have their depth explored [11]. In 1997, Alfred Crosby mentioned that besides measurement, visualisation is one of the two factors accountable for the evolution of modern science [12]. Visualisation of SN is more than generating impressive images; it is about producing images that contain information inside: "images of social networks have provided investigators with new insights about network structures and have helped them to communicate those insights to others" [12, 13].

There are two most common techniques to display the images of networks. The first technique is generating a graph made up of points, called nodes, and connecting lines; the second technique uses matrices, in which rows and columns represent individuals and cell entries represent the connections [12]. However, the first technique could be classified as the primary technique since the majority of social network applications focus on graph representation [12, 13].

2.2 Centrality Measurement

In 1948, Bavelas came out with the idea of centrality as applicable to human communication. The follow-up studies concluded that centrality was related to group efficiency in problem-solving, the personal satisfaction of participants and the perception of leadership. The idea of centrality is alive and has been applied in an extensive range of applications. According to Freeman in [14], it seems that people agree that centrality is a vital structural attribute of SN [14].

The four important concepts used in network analysis are betweenness, network density, centrality, and centralisation. Within SNA, centrality is an important concept [15]. Degree, betweenness and closeness are all measures of centrality [16]. As in [17], Lee chosen two different centrality measures in his study which are degree and betweenness centrality. This is because two different centrality measures represent two different groups of centrality measures [17].

This paper will focus more on 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 the basic and widely used in SNA [5, 18]. Some previous studies that used these centralities are in [19], [20], [10] and in [21].

Betweenness centrality is a measure of a vertex within a graph. Vertices that occur in many of the shortest paths between other vertices have a higher betweenness compared to others. Fig. 2 depicts an example of a social network diagram in which the node marked in yellow has the highest betweenness centrality.

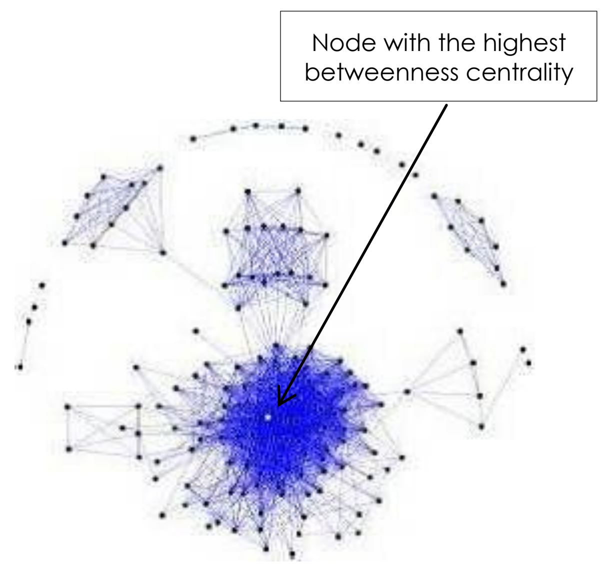


Figure 2 An example of social network diagram. The node with the highest betweenness centrality is marked in yellow [22]

The between centrality of node k (for example, p_k) is formulated as in (1), where g_{ij} is the geodesic distance (shortest paths) linking p_i and p_j and $g_{ij}(p_k)$ is the geodesic distance linking p_i and p_j that contains p_k [23].

$$C_B(p_k) = \sum_{i<j}^{n} \frac{g_{ij}(p_k)}{g_{ij}}; i \neq j \neq k \quad (1)$$

Closeness centrality is preferred in network analysis to mean the shortest-path length as it gives higher values to more central vertices. Vertices that tend to have short geodesic distances to other vertices have a higher closeness [22]. Closeness centrality is defined by Freeman as in (2), where $d(p_i, p_k)$ is the geodesic distance (shortest paths) linking p_i and p_k [23].

$$C_C(p_k) = \sum_{i-1}^{n} d(p_i, p_k) \quad (2)$$

Eigenvector centrality is a measure of the influences of a node in a network. Eigenvector centrality is calculated as in (3), where $M(i)$ is the set of nodes that are connected to i^{th} node, N is the total number of nodes and λ is a constant [22].

$$x_i = 1/\lambda \sum_{j \in M(i)} x_j = \frac{1}{\lambda} \sum_{j=1}^{N} A_{i,j} x_j \quad (3)$$

Degree centrality is the simplest centrality. The degree centrality of node k (for example, p_k) is calculated as in (4), where n is the number of nodes in the network and $a(p_i, p_k) = 1$ if and only if node i and k (for example, p_i and p_k) are connected; $a(p_i, p_k) = 0$ otherwise [23].

$$C_D(p_k) = \sum_{i-1}^{n} a(p_i, p_k) \quad (4)$$

For undirected networks, degree centrality is defined as the number of ties or links that the node has. For directed networks (the ties in a network have direction), the measures of degree centrality are separated into two, which is in-degree and out-degree [24]. In-degree refers to incoming links, the number of links that the node receives from the other nodes, while out-degree links are outgoing links, the number of links that the node sends to others [24, 17].

3.0 THE PROPOSED METHOD

This study proposed an analysis of datasets by using the Organisation Risk Analyser (ORA), version of NetScenes 3.0.0.2. This tool is downloaded from [25]. The ORA tool generates a graph based on data relationship and clusters the data based on their node similarity. In ORA, an agent, resource, or knowledge is used to represent the node [20].

According to Yin and Chen in [26], ORA is an analysis tool and network evaluation that can track the relevant index of the group, identify the style of location and contrast the relation among networks, groups and individuals from the perspective of a dynamic network. Compared with other SNA platforms, ORA can support multiple data input forms such as DyNetML (XML), EL, and CSV and can instantly show the dynamic change of the network. Thus, ORA is prefer in this study since it can support XML dataset, and CSV format dataset that able to select the dataset based on SQL query as used in [20].

The methodology of the proposed method is shown in Fig.3. The proposed method begins by analysing the selected dataset by using the SNA tool. The graph of the visualisation will be generated. SNA will identify the relationship of the data and cluster the dataset contents. The centrality measures can be calculated and the results can be generated in the form of a table and a graph.

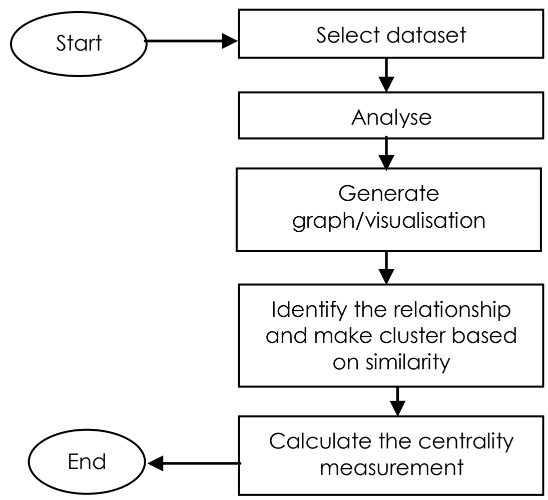


Figure 3 Methodology of proposed method

4.0 RESULTS AND DISCUSSION

In this study, the experiment consists of three datasets; AIDS dataset, Political Blogosphere dataset and Boston University (BU) dataset. AIDS and Political Blogosphere datasets are represented in XML format, while BU dataset is represented in Excel (.csv) format. We used three different datasets to compare the visualisation between three sizes of data: small data (hundreds), moderate data (thousands), and large data (more than ten thousand). In this research, the datasets are still valid and relevant to be used because other researchers had been tested the datasets for their studies [2, 8, 20, 28].

This section divides the data representation in three parts; the first part covers the obtained results from the AIDS dataset, the second part describes the results revealed from the Political Blogosphere, and the third part discusses the obtained results from the BU dataset.

For the AIDS and Political Blogosphere dataset, each part discusses the visualisation of the network in the dataset and the results of the centrality measurement. For the BU dataset, we only discussed the visualisation of the network because the aim is to compare the visualisation for different sizes of dataset.

4.1 AIDS Dataset

This dataset is collected by Gopal [8] to analyse the blog responses on social networks that were posted by AIDS patients over a three-day period in August 2005. The dataset contains 146 unique blogs related to AIDS, patients, and their support networks. It is a directed network. However, the vertices (blog posts) and the edges (responses) are represented only by numbers to preserve the privacy of the patients. There are repeated responses from the same user.

We converted the dataset into Extensible Markup Language (XML) format. Sathik and Rasheed [8] analysed this dataset by using NetDraw to show the visualisation of the network. They used two centrality measures; betweenness centrality and closeness centrality, as measures of SNA. In this study, we used ORA to analyse this dataset and four different centrality measures, as measures of SNA. We identified 183 links in the entire network as shown in Table 1.

Table 1 Statistic network for AIDS dataset

Number of Nodes	Number of Edges
146	183

Fig.4 and Fig.5 illustrate the visualisation of the network in 2D mode. 2D mode visualisation clearly shows the cluster and the numbering of each node. We can see clearly that node 143 and node 7 have more links or connections and form big clusters compared to others.

Fig.6 and Fig.7 illustrate the visualisation in 3D mode. The cluster can be seen clearly as shown in 2D mode, but the number of each node cannot be seen.

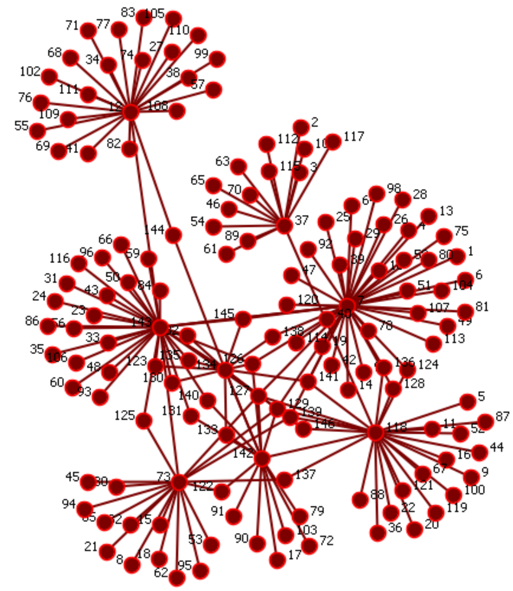


Figure 4 2D visualisation – zoom out

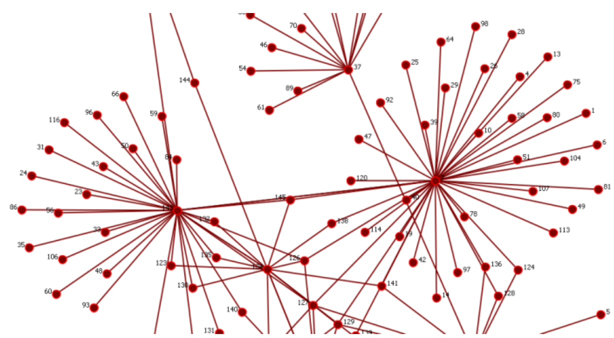


Figure 5 2D visualisation – zoom in

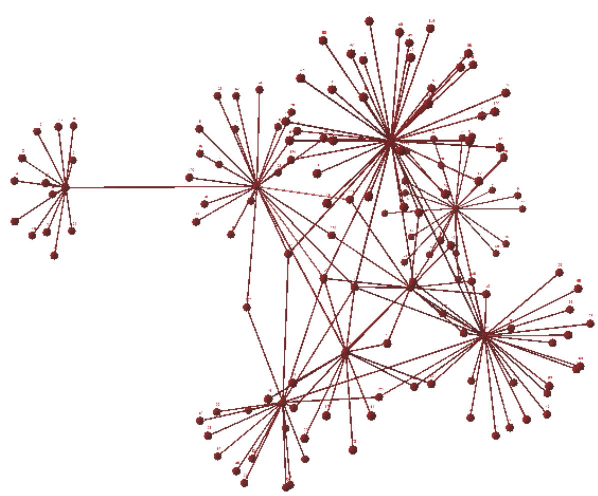


Figure 6 3D visualisation – zoom out

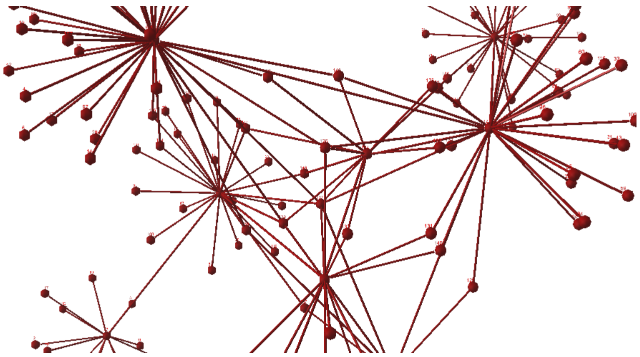


Figure 7 3D visualisation – zoom in

Then, we analyse the centrality measure. In this study, we only present the results for the five top score nodes. Table 2 represents the results for betweenness centrality, closeness centrality, and eigenvector centrality, while Table 3 represents the results for degree centrality.

The highest betweenness centrality is obtained by node 143, which is the same as obtained by Sathik and Rasheed [8]. The cluster of node 143 can be seen in Fig.5. However, the results for the next four top scores are different. In [8], the second top score of betweenness centrality is obtained by node 7, however in this study the second top score is obtained by node 134. When we look carefully at Fig.5, node 134 is a bit hidden between a few clusters, and same goes for node 142.

For closeness centrality, the highest value is obtained by node 37. In fact, it shows that node 125 that has the highest betweenness centrality has a lower value of closeness centrality. Results also show that node 146 that has the lowest betweeness centrality also has a less value of closeness centrality. These results are different when analysed using NetDraw in [8].

Node 143 also has the highest value for eigenvector centrality. It means that node 143 is highly connected to the other nodes in the network; hence, it has the highest influences compared to other nodes.

Table 2 Top 5 scores node for the betweenness centrality, closeness centrality, and eigenvector centrality

Betweenness Centrality		Closeness Centrality		Eigenvector Centrality	
Node	Score	Node	Score	Node	Value
143	0.020	37	0.232	143	0.655
134	0.015	118	0.057	7	0.613
142	0.011	143	0.032	134	0.371
7	0.008	7	0.031	127	0.307
118	0.006	134	0.031	118	0.273

For degree centrality, the highest in-degree centrality value is obtained by node 127. The higher the in-degree of a node (blog post), the more attention that the node receives from other nodes. That means node 127 is famous since it receives a lot of attention from the other nodes. The highest out-degree centrality value is obtained by node 7, which means Node 7 is the top node that interacts with other nodes.

Table 3 Top 5 score nodes for degree centrality

In-Degree Centrality		Out-Degree Centrality	
Node	Value	Node	Value
127	0.041	7	0.288
129	0.034	143	0.226
126	0.027	118	0.171
139	0.021	12	0.151
141	0.021	73	0.123

4.2 Political Blogosphere Dataset

This Political Blogosphere dataset is obtained from the CASOS website [27]. This dataset is a directed network of hyperlinks between weblogs on US politics, recorded in 2005. In this data, the nodes are the URLs of the blogs and the edge connects the URLs. Table 4 reports the statistic network for the Political Blogosphere dataset.

Table 4 Statistic network for the Political Blogosphere dataset

Number of Nodes	Number of Edges
1000	10238

In [28], this dataset consists of 1494 blogs in total. However, in this experiment, this version of ORA only manage to visualize 1000 nodes. Although it contains only 1000 nodes, it has more than 10000 edges. Hence, it becomes quite slow for ORA to visualise the network in the form of 3D. Fig. 8 and Fig. 9 illustrate the centralised effect in 2D mode.

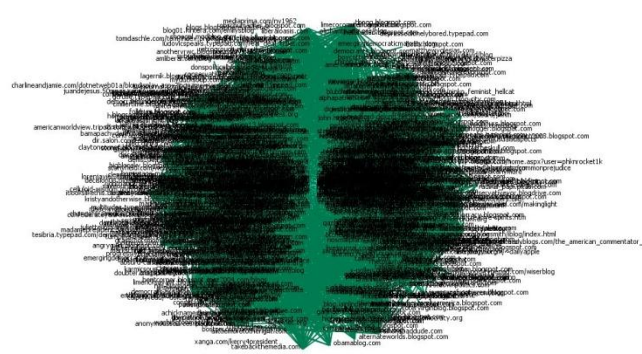


Figure 8 2D visualisation – zoom out

These two figures show that the clusters cannot be seen clearly when visualised in 2D mode. We zoom into some of the groups in the dataset to see the connections between the nodes as shown in Fig.9, and the result shows that there are many cases in which one node (URL) has a connection with many URLs. We cannot identify which node has the highest connection based on this 2D visualisation since the dataset contains about 1000 nodes and more than 10000 links.

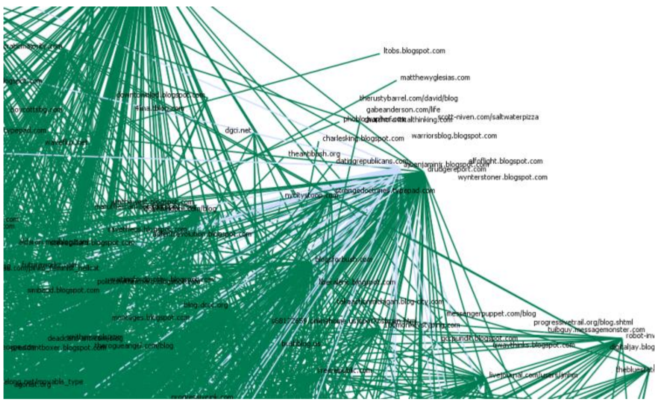


Figure 9 2D visualisation – zoom in

Then, we visualise the data in the form of 3D mode. It shows an interesting result as shown in Fig. 10. The data is pulled to the centre of the network. We zoom into the centre as shown in Fig. 11 and it reveals that there are many clusters that have been formed and some of them are hidden between those clusters.

Compared to the AIDS dataset that contains only 146 nodes, we cannot identify which nodes have the highest value in centrality only by seeing it through the visualisation of 2D and 3D modes because this dataset contains thousands of nodes and more than ten thousand links. By using the centrality measurement that is provided in ORA, we can identify which nodes have the highest value for betweenness centrality, closeness centrality, eigenvector centrality and degree centrality.

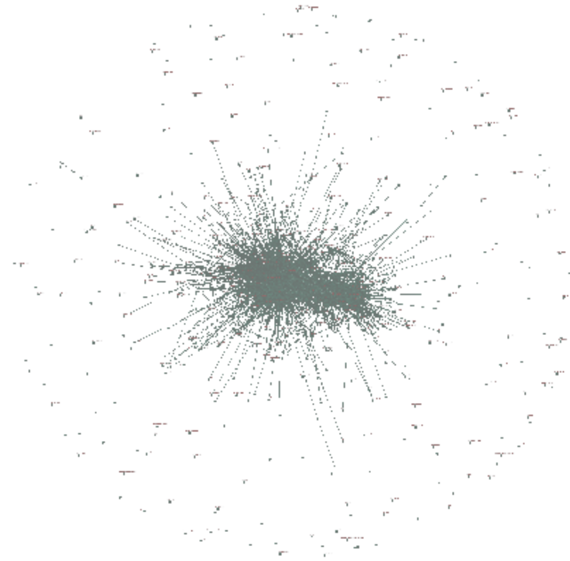


Figure 10 3D visualisation – zoom out

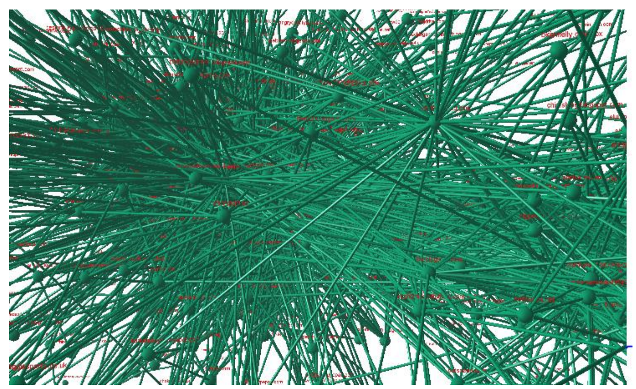


Figure 11 3D visualisation – zoom in

Table 5 indicates the result for betweenness centrality. Blog atrios.blogspot.com is at 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 5 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

The average closeness of a node to the other nodes in a network is called closeness centrality. In this study, the obtained results for closeness centrality show that the value of average closeness of a node to the other nodes is same which is 0.002 as shown in Table 6.

Table 6 Top 5 scores node for closeness centrality

Node	Value
itlookslikethis.blogeasy.com	0.002
bushmisunderestimated.blogspot.com	0.002
etherealgirl.blogspot.com	0.002
michaelphillips.blogspot.com	0.002
lennonreport.blogspot.com	0.002

Eigenvector centrality calculates the influences of a node in a network. Table 7 demonstrates the results of eigenvector centrality. Blog atrios.blogspot.com is at the top of the rank with a maximum value of 0.253, which means this blog is highly connected to others in the network, so that it has the highest influences compared to other blogs. Blog washingtonmonthly.com has the lowest value of eigenvector centrality for the top five ranking which shows that this blog has low influences in the network.

Table 7 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

Table 8 demonstrates the top five ranking blogs for in-degree and out-degree centrality in descending order. In-degree centrality of a node refers to the number of connections or links that the node receives from other nodes. The higher the in-degree of a blog, the more attention the blog receives from other blogs, meaning large numbers of blogs interact with that blog.

Blog dailykos.com is the top blog that receives the highest number of connections from other blogs while juancole.com receives the least amount of connections from other blogs. The maximum value for in-degree centrality is 0.309. The out-degree centrality of a node is the number of connections or links that the node sends to other nodes. Blog newleftblogs.blogspot.com is the top blog that sends connections to other blogs and corrente.blogspot.com is the blog that sends the least amount of connections to other blogs. The maximum value for out-degree centrality is 0.137.

Table 8 Top 5 scores node 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

4.3 BU Dataset

BU dataset is obtained from [20] consists of 17225 URLs. This dataset contains five elements: URL, size, retrieval time, number of hits, and cache. However, in this study, retrieval time and cache is not included. URL and size are set as agents, while the number of hits is set as events. This data is imported into ORA from a SQL query of database configured via ODBC. We select the dataset by using the query statement,

Select * from BU_table

and the relationship is analysed only between URL x SIZE and URL x NUMBER OF HITS. Nevertheless, based on this setting, we only managed to discover about 1187 nodes and only 904 links in the network as stated in Table 9.

Table 9 Statistic network for Political Blogosphere dataset

Number of Nodes	Number of Edges
1187	904

Fig. 12 and Fig. 13 demonstrate the visualisation of the network in 2D mode. It is quite difficult to see the clusters in the network. In 3D mode visualisation, as shown in Fig. 14, the result shows that there are some clusters formed that are centralised in the middle of the network. Some of the groups are surrounded by links that form a big cluster. Fig. 15 zooms into one of the groups in the network and it shows that the data is centralised in some locations similar to the cluster model.

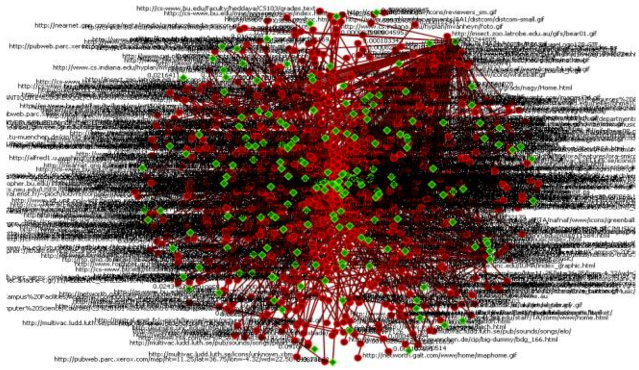


Figure 12 2D visualisation – zoom out

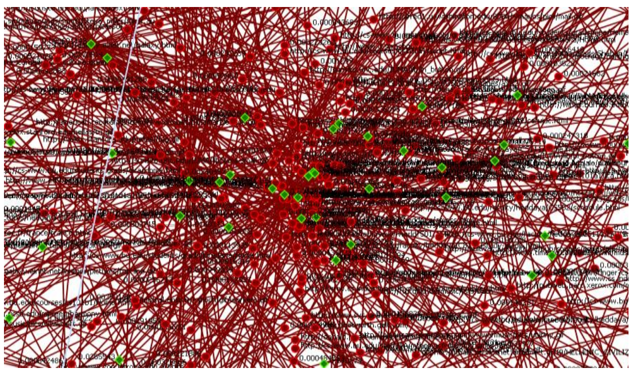


Figure 13 2D visualisation – zoom in

There is three datasets used in this study. These datasets are in different format that supported by ORA. However, threats to validity are that setting of ORA to visualize the data may not be representative of other datasets or different format of datasets. The experimental setting of dataset has been stated in this stud. Thus, repeated and replicated studies are easy to perform increasing the generalizability of results. Nonetheless, visualization of datasets using other SNA tools have not been tested and different SNA tools are likely to have different setting and supported format.

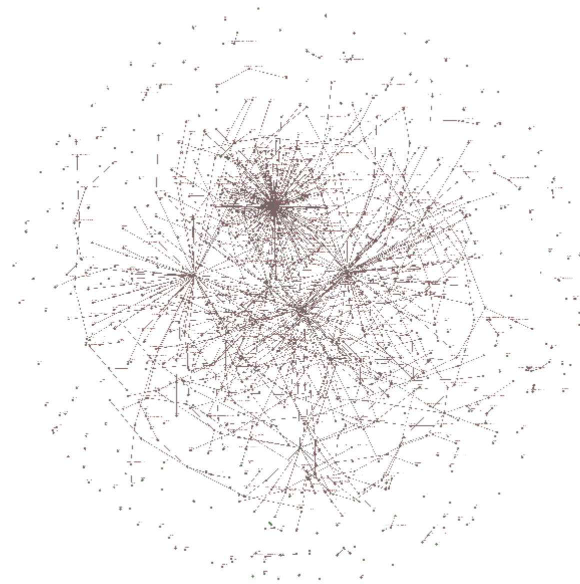


Figure 14 3D visualisation – zoom out

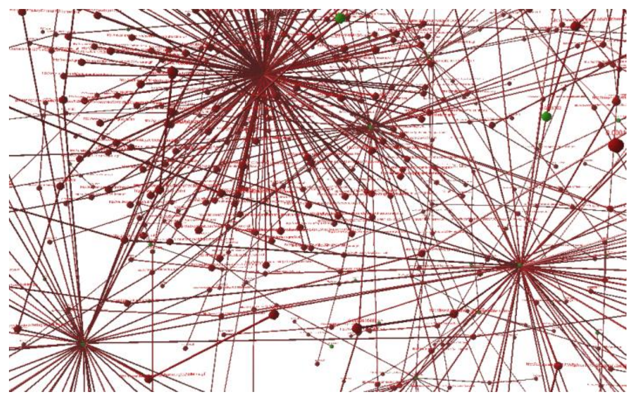


Figure 15 3D visualisation – zoom in

5.0 CONCLUSION

This study used three different sizes of datasets for analysis. The objective for using three different datasets is to compare the visualisation for different sizes of data. We analysed these datasets by using ORA tool. Based on the results, we can easily notice the clusters in a small network and identify the nodes, but it is quite hard to perceive with the eye the nodes and the clusters that are forming in large networks. We computed betweenness centrality, closeness centrality, eigenvector centrality and degree centrality.

The node that has the highest betweenness centrality also has the highest eigenvector centrality, but has a lower closeness centrality. We also compared the result of the AIDS dataset for betweenness centrality and closeness centrality with a previous study that used the NetDraw tool to analyse the AIDS dataset. Surprisingly the results are different.

The nodes that have the highest value for these two centralities are different from the previous study.

Acknowledgement

This research is supported by The Ministry of Higher Education (MOHE) and collaboration with Research Management Centre (RMC), Universiti Teknologi Malaysia (UTM). This research is financial supported by UTM Flagship Grant (Q.J130000.2428.02G70), Fundamental Research Grant (FRGS) (R.J130000.7828.4F634), E-Science Fund (R.J130000.7928.4S117) and Instructional Development Grant (IDG), Tier 1 (R.J130000.7728.4J170). 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] Ehrlich. K, Carboni. I. 2005. *Inside Social Network Analysis*. IBM Technical Report 05-10.

[2] Sun. W, Qiu. H. 2008. A Social Network Analysis On Blogospheres. *2008 International Conference on Management Science and Engineering 15th Annual Conference Proceedings*. 1769-1773. doi:10.1109/ ICMSE.2008.4669144

[3] Slaninova. K, Martinovic. J, Drazdilova. P, Obadi. G, Snasel. V. 2010. Analysis of Social Networks Extracted from Log Files, in *Handbook of Social Network Technologies and Application (B. Furht, Ed.). Boston, MA: Springer US*. 115-146.

[4] Brandes. U, Wagner. D. 2003. visone - Analysis and Visualization of Social Networks, in *Graph Drawing Software (M. Junger, P. Mutzel, Eds.). Springer-Verlag*. 321-340.

[5] Landherr. A, Friedl. B, Heidemann. J. 2010. A critical review of centrality measures in social networks, in *Business & Information System Engineering* 2. 6: 371-385.

[6] Wu. H. J, Ting. I. H, Wang. K. Y. 2009. Combining Social Network Analysis and Web Mining Techniques to Discover Interest Groups in the Blogspace. *2009 Fourth International Conference on Innovative Computing, Information and Control (ICICIC)*.

[7] Mansur. A. B. F, Yusof. N, Othman. M. S. 2010. Analysis of Social Learning Network for Wiki in Moodle E-Learning. 1:4. MeIntyre, D. 2002. *Color Blindness*. Dalton Publishing.

[8] Sathik. M, Rasheed. A. A. 2011. *Social Network Analysis in an Online Blogosphere*. 117-121.

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

[10] Raca. V, Cico. B. 2013. Raca. *2nd Mediterranean Conference on Embedded Computing, MECO 2013, Budva, Montenegro*.

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

[12] 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).

[13] 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.

[14] Freeman. L. C. 1978/79. Centrality in Social Networks Conceptual Clarification, in Social Networks, 1. 215-239.

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

[16] Coulon. F. 2005. *The use of Social Network Analysis in Innovation Research: A literature review*. 1-28.

[17] Lee. V. 2012. How Firm Can Strategically Influence Open Source Communities. 111-126.

[18] JIEDDO. 2011. Social Network Analysis (SNA) Tool Comparison. Working Paper. Retrieved from https://publicintelligence.net/jieddo-social-network-analysis/

[19] Rusinowska. A, Berghammer. R, Swart. H. D, Grabisch. M. 2011. Social Networks: Prestige, Centrality, and Influence. *Springer-Verlag Berlin Heidelberg*. 22-39.

[20] Sulaiman. S, Shamsuddin. S. M, Abraham. A. 2012. Implementation of Social Network Analysis for Web Cache Content Mining Visualization, in Computational Social Networks: Mining and Visualization *(A. Abraham, Ed.). London: Springer London*. 345-376.

[21] Al Halaseh. R. 2014. Studying Learning Networks within Moodle: A Social Network Analysis Approach.

[22] Social Network Analysis: Theory and Application. http://train.ed.psu.edu/WFED-543/SocNet_TheoryApp.pdf (Last accessed: Sept 2013).

[23] Abbasi. A, Hossain. L, Leydesdorff. L. 2011. Betweenness Centrality As A Driver Of Prefential Attachment In The Evolution Of Research Collaboration Networks*. *Journal of Informetrics*. In press.

[24] Izquierdo. L. R, Hanneman. R. A. 2006. Introduction to the Formal Analysis Social Networks Using Mathematica.

[25] http://casos.cs.cmu.edu/index.php

[26] Qiuju. Y, Qingqing. C. 2012. A Social Network Analysis Platform for Organizational Risk Analysis – ORA. 2012 International Conference on Intelligent System Design and Engineering Application. 760-763. DOI 10.1109/ISdea.2012.546.

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

[28] Lada. A. A, Glance. Y. 2005. The Political Blogosphere And The 2004 US Election. *Proceedings of the WWW-2005 Workshop on the Weblogging Ecosystem*.