Organization Mapping Using Social Network Graph Analysis

Edward Benedict B. Suyu
Luisa B. Aquino, MIT

Abstract—**This paper studied the utilization and weight factors of graph metrics on visualization and mapping of sub-networks that exist in a given network of connected people. It is utilized in this study the use of algorithms such as the Clauset-Newman Moore Algorithm and the Harel-Koren Fast Multiscale Layout, and graph metrics such as degrees and centralities. The study showed that using such methods yields results that show network visualization that is useful for investigative purposes.**

Keywords— **Networks, Social Connections, Mapping, Graph**

I. Introduction

SNA (Social Network Analysis) is a very powerful tool in developing knowledge maps and in analyzing knowledge flows for an organization [2]. Certain studies have been conducted for large-scale measurement study and analysis of the structure of multiple online social networks [11], and types, behaviors and properties of Social Networks [9].

Certain factors can affect how people in a network can be clustered upon visualization of a network, this includes how or what posts or topics they have commented on, their similar posts liked, how common they interact with mutual topics, or their mere connectivity to each other and anyone else in the network itself. Another factor is based on the theories on mere belonging [14], in which they stated that whatever relations and behaviors you are exposed to, it is likely the same with your peers. Given the different factors that can affect the visualization and meaning of the graph, this study tries to answer the question how can this metrics affect the network.

This paper tries to introduce a new way of organizational mapping, using only the mere connections a person has to other person in a given community. Formulation of methods and steps are made to show sub-organizations can exist in a community of network, and the ability to pinpoint certain people of that organization that can have an impact between the community and the sub-organizations in terms of continuous connectivity between persons and flow of information.

Edward Benedict B. Suyu
University of Saint Louis
Philippines

Luisa B. Aquino, MIT
University of Saint Louis
Philippines

This is done by plotting three network graphs from the same data set using the same layout and grouping algorithm, and alter the size of each node of each graph based on the Degree, Closeness Centrality and Betweenness Centrality Metrics[17][18].

II. Review of Literature

Different studies have been made to introduce the use of SNA to pinpoint different organizations and network, one of the most prominent is the study conducted by Valdis E. Krebs in 2002, which utilized SNA techniques of network mapping to review the networks of the hijackers in the 9/11 incident. Through their paper, they showed that SNA is also applicable not only in showing the network of the crime suspects, but also in predicting crime-related incidents [13].

In a study conducted by Jodi Blomberg in 2012, she pointed out that SNA is not only applicable to Marketing, but also to crime fighting. By using SNA on Facebook and Twitter, she retrieved valuable information like messages, post, tags and other data that can be used to discover if a crime is being discussed on a social network platform [15]. Use of Ambient Information Systems have allowed others to keep track and be awre of others activities in social networks [5].

Also on a similar study on networks, a study by Hoppe and Reinelt in 2010 offered a framework for conceptualizing different types of leadership networks and uses case examples to identify outcomes typically associated with each type of network [6]. Study on two different types of social networks as exemplar platforms for modeling and predicting group stability dynamics has recently been done [12]. A study also suggested that the use of Social Network Analysis (SNA) as part of Extension's evaluation agenda is promising [1].

The World Wide Web has been evolving into a read-write medium permitting a high degree of interaction between participants, and social network analysis (SNA) seeks to understand this on-line social interaction, for example by identifying communities and sub-communities of users, important users, intermediaries between communities. Semantic web techniques can explicitly model these interactions, but classical SNA methods have only been applied to these semantic representations without fully exploiting their rich expressiveness. The representation of social links can be further extended thanks to the semantic relationships found in the vocabularies shared by the members of these networks. These enriched representations of social networks, combined with a similar enrichment of the semantics of the meta-data attached to the shared resources, will allow the elaboration of shared knowledge graphs [4]. Social Network Analysis have been linked with the analytic hierarchy process for knowledge mapping in organizations [8].

In their study in 2009, Michiel, Poorthuis and Dugundji, described a multi-method approach to delineate a "real world" community of practice from a large N dataset derived from the social networking site Twitter. Their starting point is previous qualitative research of a virtual community of independent ("indie") developers who create software for Apple's Macintosh and iPhone platforms. They used the publicly available Twitter API to mine a network consisting of several million edges, which is sized down to a large network containing roughly 1 million edges through several pruning methods. The fast greedy algorithm is then used to detect sub graphs within this large network. Triangulation with qualitative data proves that the fast greedy algorithm is able to distill meaningful communities from a large, noisy and ill-delineated network. The proposed multi-method approach allows micro level inferences from a macro dataset of which the individual Twitter user might be completely unaware. The results could have consequences for the anonymity of key persons behind the scenes of social and political movements or any other communities whose members are active on Twitter or other social networks[10].

III. Methods

A. *Data Set*

On this study, the data set that will be used is gathered from Facebook® using the Facebook SNA Application NameGenWeb®. The data set will consist of connections of friends-of-friends of 1 sample Facebook® Account. For the sample facebook account, the network used was the friend network of one of this paper's author. The graph metrics will be calculated and analyzed to reveal how these metrics affect the grouping and clustering [7] of different vertices/nodes.

B. *Layout and Grouping*

For the sample data, the study make use of the friend and friend-of-friend network sample from the data set and presented it using the Harel-Koren Fast Multiscale Layout [3]. The Grouping was done by using the Clauset-Newman-Moore Algorithm [16].

C. *Metrics and the network graph*

Three network graphs will be plotted using the Harel-Koren Fast Multiscale layout, grouped by cluster using the Clauset-Newman-Moore Algorithm. Each one of the graph will have the sizes of the vertices are altered based on the Degree, Closeness Centrality and Betweenness Centrality.

IV. Results

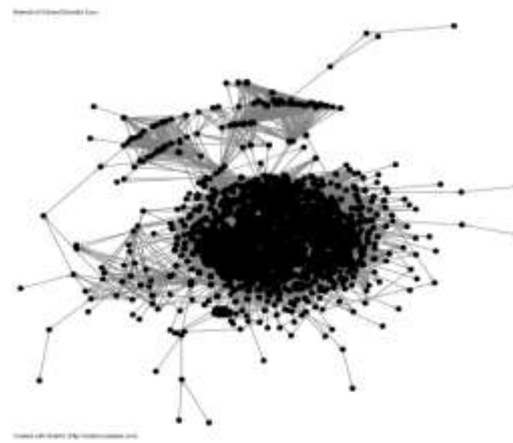
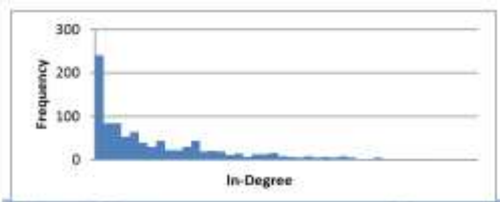


Figure 1. Using Harel-Koren Fast Multiscale Layout.

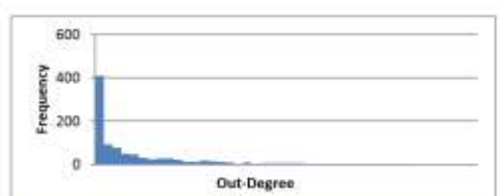
In figure 1, the graph shows the visible outer groups that exist in the network. The inner group appears to be bigger and much more clutter than the rest, this suggests that inner individuals involved in the network are much more connected than those of the outside groups.



Minimum In-Degree	0
Maximum In-Degree	190
Average In-Degree	30.909
Median In-Degree	18.500

Figure 2. Network In-Degree Statistics

Fig. 2: In a directed graph, a vertex's in-degree is the number of incoming edges incident to the vertex. In an undirected graph, in-degree is undefined and is not calculated. A self-loop in a directed graph is counted once as an incoming edge (in-degree) and once as an outgoing edge (out-degree).



Minimum Out-Degree	0
Maximum Out-Degree	304
Average Out-Degree	30.909
Median Out-Degree	12.000

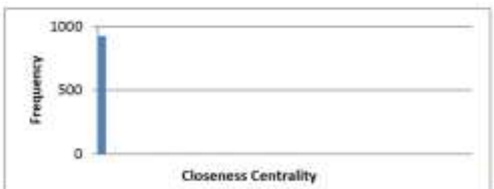
Figure 3. Network Out-Degree Statistic

Fig. 3: In a directed graph, a vertex's out-degree is the number of outgoing edges incident to the vertex.

Minimum Betweenness Centrality	0.000
Maximum Betweenness Centrality	60789.277
Average Betweenness Centrality	1281.168
Median Betweenness Centrality	271.679

Figure 4. Network Betweeness Centrality Statistics

Fig. 4: A vertex that occurs on many shortest paths between other vertices has a larger betweenness centrality than vertices that do not.



Minimum Closeness Centrality	0.000
Maximum Closeness Centrality	1.000
Average Closeness Centrality	0.008
Median Closeness Centrality	0.000

Figure 5. Network Closeness Centrality Statistics

Fig. 5: The closeness centrality of a vertex is the inverse of the sum of the shortest distances between the vertex and all other vertices reachable from it.

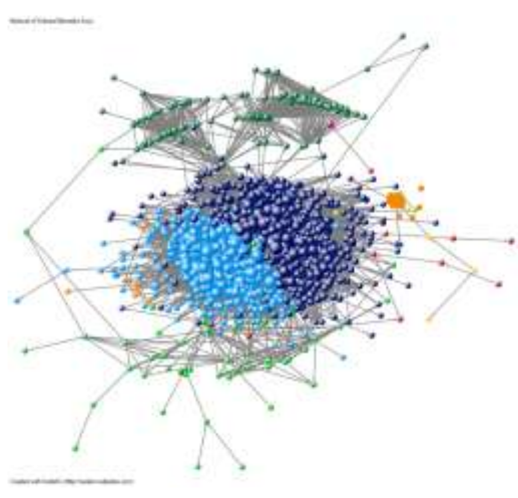


Figure 6.

In fig. 6, the shape of each node is defined by its Degree. When using this metric, the importance of each vertex is defined by how many connections it has. The more the connection, the more it is important. With this metric, it can be observed that the most connected and influential people belongs the two inner and bigger groups of the network. Less connected sub-groups are positioned on the outer ring of the graph.

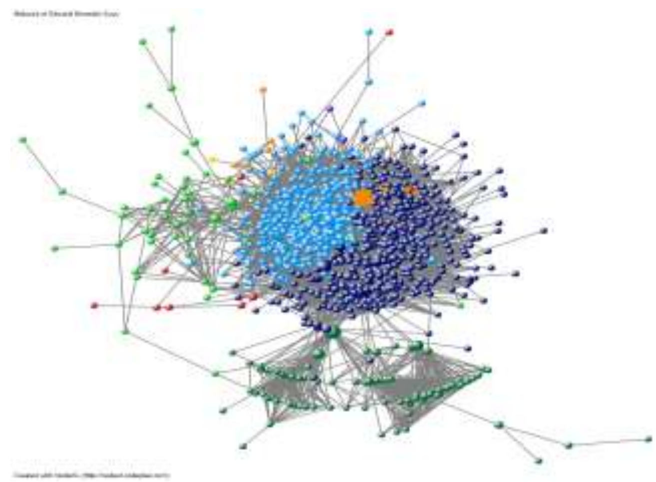


Figure 7.

Fig. 7 shows the network with the sizes of each vertex based on betweenness centrality. With this, the importance of each node is not defined by how many connections it have, but depending on how many times it occurred on the shortest paths of the other vertices. The interpretation of this is that a node is important if, for this particular network, acts as the bridge of connection that connects one node from the other. Bigger nodes on this graph with this metric shows individual that connects sub-groups to the inner part of the network.

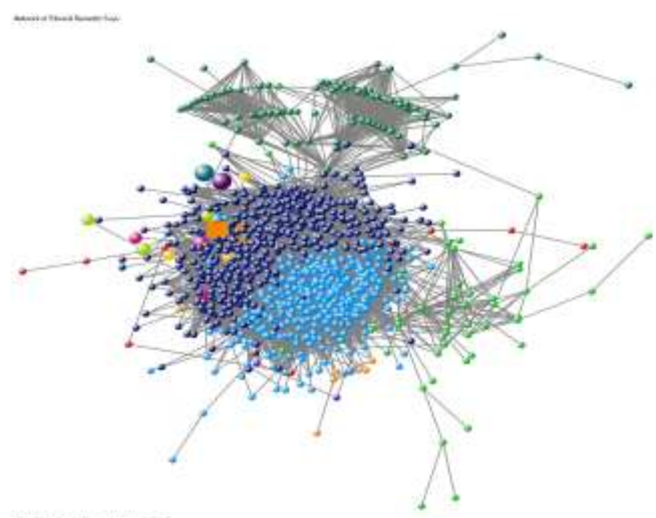


Figure 8.

Fig. 8 shows the nodes sizes altered based on their closeness centrality. As this is the inverse of the measure of its shortest paths, nodes that appear bigger are theoretically the nodes that are likely the center of their local cluster. High closeness centrality individuals tend to be important influencers within their local network community. They may often not be public figures to the entire network of a corporation or profession, but they are often respected locally and they occupy short paths for information spread within their network community.

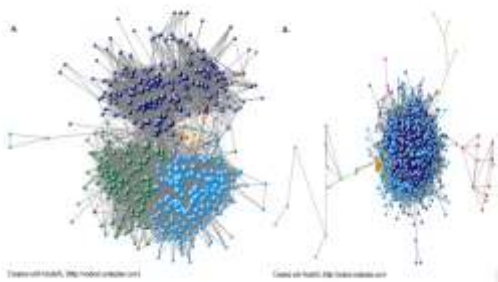


Figure 9.

For the purpose of checking the validity of the used methods, another two (2) sets of network graph belonging to two test individuals has been made. This graphs (Fig. 9.A and Fig. 9.B) has been plotted using the same exact methods and the sizes of the vertices are altered based on the Degree metric. It is obviously visible that in the two graphs, major groups also exist, just like in the first test graph. This groups are identified by the colors Dark-Blue and Light-Blue, although this time, fig. 9.A has a third major group, in color Dark-Green, while fig. 9.B only has two.

V. Discussion

Using a Fast Multi-scale Layout like the one used in this yielded a good visualization of the network, it also showed some major groups that exist in the network(light-blue and dark-blue), the visualization of the test network showed several sub-groups existing in the main network and the outer as well. Furthermore, by applying and integrating the different metrics to the graph, it showed several important nodes highlighted by bigger size; these nodes represent data points that need to be considered when conducting investigative analysis of such networks. By grouping the nodes using clustering of the Clauset-Newman-Moore Algorithm, 12 distinct groups were produced. These groups were later on verified and identified by the test network's owner. People on the network with great influence on both information dissemination and continuous connection can be revealed using the total degree metric that calculates how many interaction that person have in the actual network. This is relatively important on use for investigative analysis due to the fact that whoever is most connected in a given network is more likely the bearer of the most information that network has. The Betweenness centrality showed persons that hold the connection of other persons and links it to other in the shortest possible way, they are the ones that links other subgroups shown in the graph to the main network or other subgroups. The last metric used is the closeness centrality. It is shown in the graph that it showed a few nodes that are colored uniquely and with less other nodes in that particular color. These nodes represent persons that, although less active and participative in the network, tend to be the center of their local cluster. This persons are highly influential to other nodes in their local cluster.

Overall, this study laid out important and useful steps in utilizing friend networks that can be useful for investigative purposes. It devised a set of steps that can be used on data containing networks of people and extract information from it.

VI. Future Work

Further work is being encouraged by the authors to increase the accuracy on network and organizational mapping. Furthermore, study on the application of the methodologies used are encouraged to be tested on both uni-modal and bi-modal user-interaction network.

References

[1] Bartholomay T., Chazdon S., Marczak M.S., et al. 2011. *Mapping Extension's Networks: Using Social Network Analysis to Explore Extension's Outreach*. Journal of Extension www.joe.org/joe/2011

[2] Chan, K., Liebowitz J., 2006. *The synergy of social network analysis and knowledge mapping: a case study.* Int. J. Management and Decision Making, Vol. 7, No. 1,

[3] David Harel , Yehuda Koren. 2000, *A Fast Multi-scale Method for Drawing Large Graphs*. Proceedings of the 8th International Symposium on Graph Drawing, p.183-196,

[4] Erétéo, G. Et al. 2011. *Semantic Social Network Analysis, a concrete case*. Author manuscript, published in "Handbook of Research on Methods and Techniques for Studying Virtual Communities: Paradigms and Phenomena. 122-156" DOI : 10.4018/978-1-60960-040-2.ch007

[5] García R.C. 2010. *Strengthening Elder's Social Networks through Ambient Information Systems and SNS*. CSCW 2010, February 6–10, 2010, Savannah, Georgia, USA. ACM 978-1-60558-795-0/10/02

[6] Hoppe, B. Reinelt C. 2010. *Social network analysis and the evaluation of leadership networks*. 1048-9843/$ © 2010 Elsevier Inc. doi:10.1016/j.leaqua.2010.06.004

[7] Hristo N. Djidjev and Melih Onus. 2006. *Scalable and Accurate Algorithm for Graph Clustering?* Algorithms and Models for the Web-Graph (WAW 2006), Lecture Notes in Computer Science, vol. 4936, pp. 117-128.

[8] Liebowitz J. 2005. *Linking social network analysis with the analytic hierarchy process for knowledge mapping in organizations*. Journal of Knowledge Management. VOL. 9 NO. 1 2005, pp. 76-86, *Q*, ISSN 1367-3270 DOI 10.1108/13673270510582974

[9] Machová, K. Rakuščinec T. 2012. *Dynamic Analysis of Social Networks*. American Journal of Intelligent Systems 2012, 2(6): 148-156 DOI: 10.5923/j.ajis.20120206.02

[10] Michiel V.A., Poorthuis A., Dugundji E. 2009. *Mapping communities in large virtual social networks Using Twitter data to find the Indie Mac community*. Proceedings of Social Network 2009 Conference

[11] Mislove, A. Et al. 2007. *Measurement and Analysis of Online Social Networks.* IMC'07, October 24-26, 2007, San Diego, California, USA.Copyright 2007 ACM 978-1-59593-908-1/07/0010

[12] Patil A., Liu J., Gao J. 2013. *Predicting Group Stability in Online Social Networks.* C*WWW 2013*, May 13–17, 2013. ACM 978-1-4503-2035-1/13/05. (IW3C2)

[13] Valdis E. Krebs. 2002. *Mapping Networks of Terrorist Cells.* CONNECTIONS 24(3): 43-52 http://www.sfu.ca/~insna/Connections-Web/Volume24-3/Valdis.Krebs.web.pdf

[14] Walton, et al, 2011. *Mere Belonging: The Power of Social Connections*. Journal of Personality and Social Psychology, 2012, Vol. 102, No. 3, 513–532

[15] Jodi Blomberg. 2012. *Twitter and Facebook Analysis: It's Not Just for Marketing Anymore.* SAS Global Forum 2012

[16] Clauset, A., Newman, M.E.J., Moore, C.: *Finding community structure in very large networks*. Physical Review E 70, 066111 (2004), http://www.citebase.org/abstract?id=oai:arXiv.org:cond- mat/0408187

[17] Freeman, L. C., 1978. *Centrality in social networks: Conceptual clarification. Social Networks* 1, 215-239.

[18] Opsahl, T., Agneessens, F., Skvoretz, J. (2010). *Node centrality in weighted networks: Generalizing degree and shortest paths*. Social Networks 32, 245-251.

About Author (s):

Edward Benedict B. Suyu is an undergraduate student at the University of Saint Louis, currently taking a course in Bachelor of Science in Computer Science. Currently living at Tuguegarao City, Cagayan, Philippines.

Luisa Baquiran Aquino is the Academic Dean of the ITE programs of the University of Saint Louis (USL), Tuguegarao City, Cagayan, Philippines. Her research interests cover software engineering, project management, e-learning practice and studies related to educational and curriculum development.