ISSN: 2477-3328

The 3rd International Conference on Science, Technology, and Humanity

A Comparison of Community Clustering Techniques: Fruchterman-Reingold and Wakita-Tsurumi

Irma Yuliana*, Sukirman, Sujalwo
Universitas Muhammadiyah Surakarta, Department of Informatics Education
Jl. Ahmad Yani Tromol Pos 1 Surakarta
irma.yuliana@ums.ac.id

Abstract

The study of social network analysis depends on the relationships of people and online community. The relationships define who they are and how they act. Personality, educational background, race, and ethnicity, all of these interact with the patterns of relationship. Thus, by observing and analyzing such patterns, people can reveal and answer many questions about society. The relationship can be visualized in many ways, e.g. online community. Fruchterman-Reingold is a standard method force-directed algorithm or spring embedders place vertices by assigning forces according to the edges connecting the vertices. Meanwhile, Wakita-Tsurumi is a clustering algorithm used for cluster detection. It uses the metric of modularity (Q) as a quality measure of division in a network, based on the idea that networks with inherent community structure deviate from random networks and that networks with high modularity have denser connections inside a community, but fewer connections between nodes of different communities. The comparison shows the ability of both techniques to recognize the community based on sociometric and its contents. Both of the graph detected the same number of vertices (363), edges (5814) and density (0.02975511).

Keyword: Social network analysis, Fruchterman-Reingold, Wakita-Tsurumi

1. Introduction

First of all, graph representation was very simple when it was formulated by Leonhard Euler to solve Immanuel Kant's problem—who regularly took long walks and passed one of the Konigsberg bridge more than once (Tsvetovat & Kouznetsov, 2011). Currently, such a representation becomes more complex to be implemented in online network as well as in a social media. The growing impact of social media, including Facebook, Twitter, blogs, wiki, Youtube, and Flickr, has changed the way people work, relate, think, and react [1][2][3]. It has great potential to resolve national priorities list, such as disaster response, health/wellness, community safety, energy sustainability, citizen science, open government, etc. Ensuring universal access in usability, huge scalability, and reliability during peak usage is very challenging, therefore it requires an interdisciplinary research. A lot of information in online crowd should be meticulously analyzed, particularly to obtain the trusted and useful ones.

Social Network Analysis (SNA) focuses on the relationships among social entities. It is used widely in the social and behavioral sciences, as well as in political science, economics, organizational science, animal behavior, and industrial engineering. The social network perspective used in many studies has been developed over the last sixty years by researchers in psychology, sociology, statistics, and anthropology (John Scott, 2000)(Nemiche, Cavero, & Pla Lopez, 2012)(Tsvetovat & Kouznetsov, 2011). The goal of this research was to map the online crowd of social media in Fanpage Facebook of Greenpeace Indonesia by detecting some actors with strong influences to the group (Greenpeace Indonesia). Furthermore, community detection is used to identify more specifically actors found by SNA.

2. Literature Review

Social network analysis is a method to map and measure relationships and communication among people, groups, organizations, computers or entity that processes the information (Song, Kim, Kim, & Kim, 2010)(Aggarwal Charu, 2011). The relationships are visualized by graph so it will be easier to be analyzed. Social networks can be employed to analyze soccer games patterns in a video (Park & Yilmaz, 2010). It constructs two social networks for both team based on the interactions between the players. The strong connections between the players of different positions belong to the winner team. In culinary, social network analysis methods are used even in recipes around the world to link the reviews and the historical background of a region through the hallmark of traditional culinary offerings (Kular, Menezes, & Ribeiro, 2011). Subsequently, another studies applied SNA to identify terrorists from a particular community and websites visited by the criminals (Tayebi & Glasser, 2012)(Yang & Ng, 2007). Other studies utilized the connectivity between the graph theory and SNA in disaster management (Zelenkauskaite, Bessis, Sotiriadis, & Asimakopoulou, 2012)(Cheong & It, 2011), political campaign (Ya-ting, 2011), trip planning travel (Ahmedi, Rrmoku, & Sylejmani, 2012), traffic congestion monitoring (Dimokas, 2010), economy and business (Alamsyah & Rahardjo, 2013)(Jackson, 2010)(Wei & Wen, 2011).

Graph in SNA represents the interaction between individuals and their role in the dissemination of information and innovation. One of the SNA method developments is community analysis. Basically, online community in social media is formed from a large number of small groups or meta-community. Identification of a small group which is also a meta-community can be based on many things, such as occupation, educational background, preferences and even groceries (Berger-wolf & Kempe, n.d.)(Clauset, Newman, & Moore, n.d.). Identification of this group is very useful to determine the structure of the community, which ultimately is able to identify individual profiles.

3. Method and Material

The descriptive-correlational research design was employed and the data were obtained from Facebook Fan Page of "Backpacker Dunia" online community by using SocialNet Importer, in March 2017. The online crowd consisted of many opinions that might not be entirely true since it depends on the actor who spread the content. SNA detected the relations of each member in a group even the content of the topic.

Several stages were carried out after the literature study was completed. The stages are illustrated on Fig. 1. They were performed on each method and the comparison was carried out.

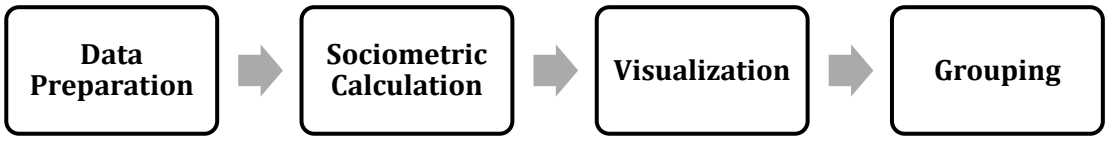

Fig. 1. Research Method.

Data preparation is very important to clean the data from multiple and redundant relationships and node. NodeXL as an open source adds in Excel has brought many advantages in the process

(Smith, et. al. 2010); Node is representation of an actor in real world. Unique node indicates that there is only one actor in the online community. While SNA is a series of process of sociometric calculation (Analysis & Raya, 2013) and visualization by Fruchterman-Reingold and Wakita-Tsurumi (Kobourov, 2013).

3.1 *Social Network Analysis - Sociometry*

Sociometry is a quantitative method to measure social relationships among the people, which was developed by psychotherapist Jacob L. Moreno in his research of the relationship between social structures and psychological well-being (John Scott, 2000). In this section, we introduce three types of centrality measures that are commonly used for SNA, namely degree, closeness, and betweenness. The centrality assumes that relationships are directed and weighted through social networks.

Degree Centrality, in case of weighting and directing network in a graph, an actor's degree is defined by the number of other actors who directly have connection with her/him as in-degree, vice versa out-degree. The size of node formed by an actor depends on the metric. The most influence actor can be observed from the highest value of in-degree centrality (Rachman & Maharani, 2013)

$$C_d(v_i) = d_i^{in} \qquad \textit{(prestige)} \qquad \textbf{(1)}$$

Where d_i is the degree (number of adjacent edges) of node v_i.

Closeness centrality directly relates to the geodesic distance (or the cardinality of the shortest path) between two actors. It is the reflection of the whole connectivity in the network structure (Analysis & Raya, 2013).

$$C_c(i) = \frac{n-1}{\sum_{j-1}^{n} d(i,j)} \qquad \textbf{(2)}$$

Where d (i,j) defines the geodesic distance between two actors (i and j), while n is the number of nodes/actors.

Betweenness centrality depicts the position of an actor in between, so she/he can control other two actors and does not have direct connectivity with both of those two actors. The metric is important in global centrality measure to investigate the strength of connectivity between the actors in the network (Analysis & Raya, 2013).

$$C_B(i) = \sum_{j<k} \frac{p_{jk}(i)}{p_{jk}} \qquad \textbf{(3)}$$

Where, *pjk* is the number of geodesic paths between two actors (j and k), and *pjk(i)* is the number of geodesic paths between *j* and *k* that contains actor *i*.

3.2 *Fruchterman-Reingold*

Fruchterman-Reingold is a standard method of force-directed algorithm or spring embedders place vertices by assigning forces according to the edges connecting the vertices. The edges are treated like springs that move vertices closer or further from each other in an attempt to get the equilibrium. The strength of the repulsive force between vertices shows the higher numbers

make vertices push away from each other more, but sometimes it has to be done to reduce vertex overlap. While the iteration per layout determines the number of the layout by Fruchterman-Reingold performed when the graphs is refreshed from the current state. It is often important to run the re-layout to get to a desirable one. It makes the process slow and computationally intensive. This research performed 100 iterations to refresh the spigot.

3.3 *Wakita Tsurumi*

Wakita-Tsurumi is a clustering algorithm was used for cluster detection. It uses the metric of modularity (Q) as a quality measure of division in a network based on the idea that networks with inherent community structure deviate from random networks while networks with high modularity have denser connections inside a community, but fewer connections between nodes of different communities

4. Result and Discussion

At the end of the research, the findings indicated variant results based on sociometric calculation.

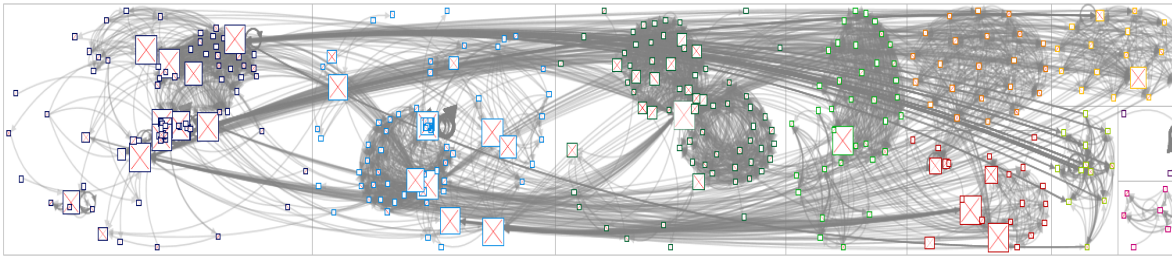

Fig. 2. Graph generated by Fruchterman-Reingold Method.

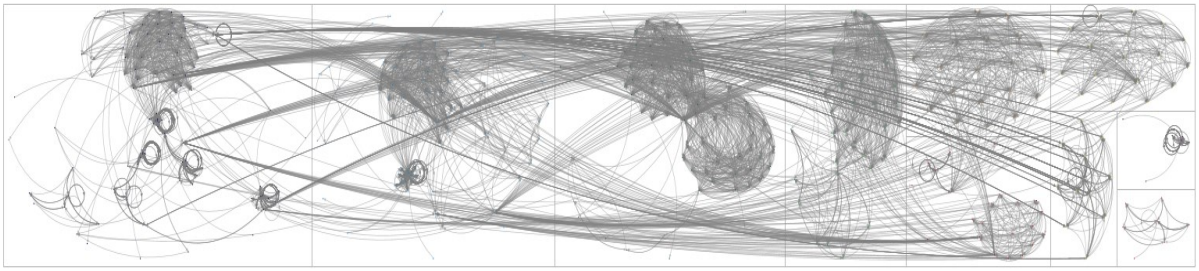

Fig. 3. Graph generated by Wakita-Tsurumi Method.

As demonstrated on Fig. 2 and 3, the graphs show identical visualization in which both of them mapped the data into 10 (ten) communities or groups with the same proportion on each group. Both of the graphs detected the same number of vertices (363), edges (5814) and density (0.02975511). More detail, visualization shown by Fruchterman-Reingold Method which has been generated from 100 iterations seems to be a little bit different on the second block of the graph.

This research was a comparison between two methods of community clustering that can be used to detect communities in online crowd. Mapping the online communities could be done by Social Network Analysis. The application of methods in SNA should be supported by other parametric, for example by eigenvalue analysis and page rank, as well as depictions of the

graph and the community detection method. In addition, the samples should be extended in order to obtain more accurate data.

References

Aggarwal Charu (Ed.). (2011). *Social Network Data Analytics*. New York: Springer.

Ahmedi, L., Rrmoku, K., & Sylejmani, K. (2012). Tourist Tour Planning Supported by Social Network Analysis. *2012 International Conference on Social Informatics*, (SocialInformatics), 295–303. https://doi.org/10.1109/SocialInformatics.2012.89

Alamsyah, A., & Rahardjo, B. (2013). Financial Fraud Detection using Social Network Analysis, (July). https://doi.org/10.13140/2.1.2696.9285

Analysis, D. D., & Raya, V. (2013). Social Network Analysis , Methods and Measurements Calculations, 2–5.

Berger-wolf, T., & Kempe, D. (n.d.). A Framework For Community Identification in Dynamic Social Networks.

Cheong, F., & It, B. (2011). Social Media Data Mining: A Social Network Analysis of Tweets During The Australian 2010 - 2011 Floods.

Clauset, A., Newman, M. E. J., & Moore, C. (n.d.). Finding community structure in very large networks, 1–6.

Dimokas, N. (2010). Social Network Analysis Concepts in the Design of Wireless Ad Hoc Network Protocols, (December), 23–29.

Hansen, D. L., Rotman, D., Bonsignore, E., Milic-Frayling, N., Rodrigues, E. M., Smith, M., & Shneiderman, B. (2012). Do You Know the Way to SNA?: A Process Model for Analyzing and Visualizing Social Media Network Data. *2012 International Conference on Social Informatics*, (SocialInformatics), 304–313. https://doi.org/10.1109/SocialInformatics.2012.26

Jackson, M. O. (2010). An Overview of Social Networks and Economic, 1–96.

John Scott. (2000). *Social_Network_Analysis_A_Handbook* (2nd Editin). London: Sage Publication Ltd.

Kobourov, S. (2013). Force-Directed Drawing Algorithms. *Handbook of Graph Drawing and Visualization*, 383–408. https://doi.org/10.1016/j.soncn.2011.02.001

Kular, D. K., Menezes, R., & Ribeiro, E. (2011). Using network analysis to understand the relation between cuisine and culture. *Proceedings of the 2011 IEEE 1st International Network Science Workshop, NSW 2011*, 38–45.

Nemiche, M., Cavero, V., & Pla Lopez, R. (2012). Understanding social behavior evolutions through agent-based modeling. *2012 International Conference on Multimedia Computing and Systems*, *1*, 980–986. https://doi.org/10.1109/ICMCS.2012.6320322

Park, K.-J., & Yilmaz, A. (2010). Social Network Approach to Analysis of Soccer Game. *2010 20th International Conference on Pattern Recognition*, 3935–3938. https://doi.org/10.1109/ICPR.2010.957

Rachman, Z. A., & Maharani, W. (2013). The analysis and implementation of degree centrality in weighted graph in Social Network Analysis. *2013 International Conference of Information and Communication Technology (ICoICT)*, 72–76. https://doi.org/10.1109/ICoICT.2013.6574552

Smith, M., Milic-Frayling, N., Shneiderman, B., Mendes Rodrigues, E., Leskovec, J., Dunne, C., (2010). NodeXL: a free and open network overview, discovery and exploration add-in for Excel 2007/2010, http://nodexl.codeplex.com/ from the Social Media Resear.

(n.d.).

Song, J., Kim, M., Kim, H., & Kim, K. (2010). A Framework: Workflow-Based Social Network Discovery and Analysis. *2010 13th IEEE International Conference on Computational Science and Engineering*, 421–426. https://doi.org/10.1109/CSE.2010.74

Tayebi, M. a., & Glasser, U. (2012). Investigating Organized Crime Groups: A Social Network Analysis Perspective. *2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 565–572. https://doi.org/10.1109/ASONAM.2012.96

Tsvetovat, M., & Kouznetsov, A. (2011). *Social Network Analysis for Startups*. *Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki*.

Wei, K., & Wen, W. (2011). Research on Emergency Information Management Based on the Social Network Analysis üü A Case Analysis of Panic Buying of Salt, 1302–1310.

Ya-ting, L. (2011). The Social Network Analysis of Political Blogs in People, 5441–5444.

Yang, C. C., & Ng, T. D. (2007). Terrorism and Crime Related Weblog Social Network: Link, Content Analysis and Information Visualization. *2007 IEEE Intelligence and Security Informatics*, 55–58. https://doi.org/10.1109/ISI.2007.379533

Zelenkauskaite, A., Bessis, N., Sotiriadis, S., & Asimakopoulou, E. (2012). Interconnectedness of Complex Systems of Internet of Things through Social Network Analysis for Disaster Management. *2012 Fourth International Conference on Intelligent Networking and Collaborative Systems*, 503–508. https://doi.org/10.1109/iNCoS.2012.25