SNIIFY : SOCIAL NETWORK INFLUENTIAL PEOPLE IDENTIFICATION ON ONTOLOGY

Rolly Maulana Awangga[1], Jimmy Tirtawangsa[2], Setia Dwi K M.eng.[3]

[1]Magister Teknik Informatika, Fakultas Teknik Informatika, Universitas Telkom

---

Abstrak

Most important things on Social Customer Relationship Management field is focusing on managing engagement between industry and customers. The engagement created by industry-customer relationship based on social activity. However, the industry have limitation to interact with every single customer, because high cost and maintenance. Furthermore, they choose other methods to keep the relationship with effective and efficient resources.

The customer have social activity, interact each other and tied each other. This social activity creating a graph called social network. The graph consisting nodes as central point and edges as relation between nodes. The social network with customer as nodes and activity as edges will be have same behavioral with graph, have influential nodes or influential people as customer. This study was to propose a method of industry-customer engagement to find influential people in social network based on their relation activities.

Influential people identification starts with model design of social network. The model uses Ontology as knowledge representation design. The calculation conducted on the model with centrality measurement. The measurement uses in friendship relationships and activities data. The study have influential people by the result of measurement not only by their relationship status, but also use activity data.

Kata Kunci : Social Customer Relationship Management; Influential People; Social Network; Ontology;Centrality

---

Abstract

Most important things on Social Customer Relationship Management field is focusing on managing engagement between industry and customers. The engagement created by industry-customer relationship based on social activity. However, the industry have limitation to interact with every single customer, because high cost and maintenance. Furthermore, they choose other methods to keep the relationship with effective and efficient resources.

The customer have social activity, interact each other and tied each other. This social activity creating a graph called social network. The graph consisting nodes as central point and edges as relation between nodes. The social network with customer as nodes and activity as edges will be have same behavioral with graph, have influential nodes or influential people as customer. This study was to propose a method of industry-customer engagement to find influential people in social network based on their relation activities.

Influential people identification starts with model design of social network. The model uses Ontology as knowledge representation design. The calculation conducted on the model with centrality measurement. The measurement uses in friendship relationships and activities data. The study have influential people by the result of measurement not only by their relationship status, but also use activity data.

Keywords : Social Customer Relationship Management; Influential People; Social Network; Ontology;Centrality

---

CHAPTER I

INTRODUCTION

1.1. Rationale

During 1990s, CRM introduce himself focuses mainly on implementing and automating processes of customer data [1]. In Internet era, CRM was transforming to Web 2.0 concept. The concept called Social Customer Relationship Management, as know as SCRM. Greenberg [1]defined SCRM as "a philosophy and a business strategy, supported by a technology platform, business rules, processes and social characteristics, designed to engage the customer in a collaborative conversation in order to provide mutually beneficial value in a trusted & transparent business environment. It is the company's response to the customer's ownership of the conversation". Mix and match between customer and business is the core of SCRM Strategy. The SCRM Strategy is not only managing customer data but also make customer engagement.

To make customer engagement, an integrated system between a company and customer activity is needed. The Company needs relationships data such as friendships, social status, and others meaningful relationships between customers on Social CRM. The system also needs a way to get the social activities customer data. One solution to obtain social activity data from customers is by interacts directly with them. However, the company has limited quality interaction with all customer; therefore, the social media is one many tools to build interactions between a company and its customers. Customers are already in social networks and form a community themselves based on their interest or hobbies. The company must have engagement to this community rather than to every user. In the network community, the customers themselves control the community. The Company has to manage the community to control them as the primary goal of Social CRM. One form of managing over the community is by taking the influential people on the network community.

Social Network Analysis (SNA) was used to identifying influential people on the social network [2]. The method of SNA calculation dividing into three major ways of calculation or searches, i.e. Connection, Distributions, and Segmentation. SNA Connection used to look for similarities and differences in content, the amount of content, familiarity and closeness tendencies. SNA Distribution lead to the inside cover mathematical calculations between points and reconnection is in the graph. Moreover, the last is SNA Segmentation used to gathering nodes then identified a group.

Focusing on SNA Distributions theory calculation. The method consists of six pieces sub methods, the bridge, i.e. centrality, density, distance, structural holes, and tie strength. Influential people identification uses centrality sub methods. It used centrality because it is a special method to categorize and quantify the importance and influence point or user node relative to other user nodes that are connected to the network, either directly or indirectly. Calculation methods in centrality consist of degree, betweeness and closeness that is suitable method applied to this thesis research.The result can be categorized into other definition of influential people. Lisa Barone [3] identified influential people as the social butterflies, the thought Leaders, the trendsetters, the reporters, and the everyday customer. So, the result can be categorized in the definition above by looking the object content or profile manually.

1.2. Statement of the Problem

Based on rationale above the problem based on question: "How to identify influential people on social networks with activity relationships attributes?" The problem of the research is influential people identification as parts of Social CRM Input. The influential people engaging with the company to attempt high quality relationships with the company as the community. Influential people identification proposed with many classifications. The classification is using subjective parameters; they do not have exactly parameter to calculate them automatically by math or calculation algorithm, they are based on the user profile surveyed.

1.3. Theoretical Framework

Nowadays, since W3C proposing new concept of web communication, semantic web has become popular. Ontology chosen as knowledge based semantic description for basic formalization based on ontology as knowledge based to communicate between website in Internet. Ontologies give more power on conceptual with domain and relation in the concept. Ontology is modeling language for formalization design in software engineering. The modelling language communicating between remote team member design [1] to avoid miss-configuration of software and mediating mechanism for accessing heterogeneous data and information sources [2]. Ontology is a solution for software engineering knowledge representation [1]. Ontologies are classified into software development life cycle phase [3]. Ontology presenting metadata schemas, offering a controlled vocabulary of concept richer than UML class/object model [4].

Ontology can improve SNA measurement [5]and software engineering knowledge [6]. Within graph theory and network analysis, centrality is used to determine the relative importance of a node within the graph [7]. Centrality researchs in different aspects [8] to get influential people. Beside for political issues [9] and terrorism information [10] on social network [11], in online market business with higher values for three degrees the most popular (degree, betweeness, closeness) of centrality represent users who can targeted as productive user to selling product [12].

Semantic description on ontology to identify influential people will conduct in this thesis. Ontologies are an effective tool for modeling, sharing and reusing knowledge [6]. Facilitates open model to identify influential people for better description in future works.

This thesis will apply semantic description on ontology to identify influential people. Ontologies are an effective tool for modeling, sharing and reusing knowledge (Hilera & Fernández-Sanz, 2010). Ontology serving model to identify influential people for better description in future works. Ontologies are closely related to modern object oriented software engineering with richer information model compared to UML and as a shared conceptualization, naturally easy to develop in future work [4].

Within the scope of graph theory and network analysis [13] [14] by using centrality [15]. A person can be considered influential from how many listed friend he has, his ability to connect two networks of friends, or a person who known a lot of others influential person. Degree, betweeness

and closeness centrality as popular centrality concept is adapted in measuring graph. A graph consist a node and an edge. A node is define person, edge define relation between nodes. Influential generates by friendship graph, which only has node of people connect each other with friend status relation as edge. Activity graph has user with other object node in social media considered as the one of source measuring influential people.

Influential people is classified by activity data like group, works, education and the others data. Several classification is proposed to identify type of influential people. Lisa Barone categorized influential people in Five aspect. There are the social butterflies, the thought leaders, the trendsetter, the reporters and the everyday customer. Klout.com Categorized influential people in Sixteen type of them. Lastly, Richard Millington classifies influential people in four aspects from their activities. i.e. expertise, fame, authority, relationship.

1.4. Conceptual Framework / Paradigm

The existence or relation of people seen from Internet especially on the social media. However, the influential factor of a person to his population cannot be measured. Knowing who is having influential factor on the community can make easy step market infiltration. The search engine showing only about attributes data of user, and it cannot see the most influential people on the community. Therefore, developing influential people engine that is able to search and find the influential people on the community is required. Influential people ranks from customer of a public service company data or ranks table from user communities or population can be generated.

1.5. Objectives and Hypotheses

The objective of this research is to look for influential people by building user data collector and Calculation engine to get influential people ranks by some calculation. The data collector engine collects the attributes, relationships and activities of user data. The calculation engine calculates the user data to generate the ranks table from the data collector output. On the other hand, Ontology builds the model on semantic based. Data collector interacts with source of the user data, which is the data on the Internet especially on the social media.

Based on theoretical framework, on graph theory there is a node in the graph as important node relatively with the graph. Social Network have same behavioral with graph will be use same method to find the important user in the network. Hypothesis of this research is the influential people can be identify from relationship and activities data with same method to get influential nodes in the graph theory.

1.6. Assumption

There are some variables that considered in general problem becausehe result needs people data taken from social network :

1. Subjective evaluation for influential people is conducted by checking user profile who has one of the aspects [17]; fame, authority, relationship and expertise and then the results are compared to thode of the others.
2. Social media have a user with non-personal account.
3. User always use real names in social network because some people have a same name or even their full name. It is quite difficult to identify each of them.
4. To obtain full personal data, user privilege is required.
5. Not every people shared their personal data.

1.7. Scope and Limitation

The scope and limitation of this study are:

1. Semantic description in this thesis designed for the purpose of software engineering, with ontology to describing the user relationship scheme in social network based software. Intelligent system or data mining fields is not include.
2. This thesis select Facebook as social media because it has a required model with the data collector. The social media must have relationships user data.
3. Subjective verification will be used to confirm user profile based from the calculation process.
4. Data population are collected from community by means of data collector application. The application integrated into community registration portal.

1.8. Importance of the Study

The study of the research has some important advantages in the real world, such as

1. It can be used to develop software for influential people identification.
2. In the field of social customer relationship management, it helps to identify influential people.

CHAPTER V

CONCLUSION AND FUTURE WORK

5.1. Conclusion

Influential People Identification system had two main sub systems, the data collector and the calculation engine. The data collector interacted with the data source. The data source was built many applications and it implemented with different of web portal. The applications was integrated in the web community portal to populate the data source for influential people identification. Every single web portal had one key for identification purpose of the data source. All off user data were pulled out in one time with single action in the program of the application. The data collector using online web server consisted of web application, social media SDK and API as a core of data collector. To interact and pull out data from the data source, data collector used social media API on each community web portal. Several algorithms tried to manage timed out of bulk user data from social network or media server to the data collector database.

The calculation engine worked for generate rank table of influential people. Influential people identification using centrality measurement was successfully classifying influential people by their activity and not only by a more tangible of their friendship relation. Activity data measurement by degree, betweeneess, and closeness centrality gave different people with different aspect of influential people.

Several influential people appeared in rank of table sort by their populairity in the network. This was the result of calculation with one of graph theory to get the influence people from different data of relationship and activity. Each data source of relationship and activity showed that the person had influential aspect from the data. Another data source was used as comparation result of the influential people. The other source had the same method to get people web ranks. It was used as comparator of the influential people result.

The ranks table has verified with people web ranks data. It was successfully identified influential people in population data. Every ranks table showed Influential People from different aspect. Influential people identification was classified people whom significant on the network. The result from this research shows people, who had influential aspect of influential people from human friendship relation and their activities.

Influential people on technology social media have no relation with conventional social media. Not every user who has popularity on technology social media has popularity on conventional social media. Ontology gave the meaning of relationship between the domains. The Calculation was conduct on the domain section. Changing ontology design affected to meaning of every calculation result. This study was successfully proposed one method obtainen influential people not only by their relationship but also with their activity in technology social media.

5.2. Future Work

The study showed that influential People Identification engine have several component system. There is several future study that may can do for enhance the process for the result of the system. In the data collector, communication data between the data source and the application should

use effective efficient algorithms to manage bulk user data from timed out. The other research and experiment about data collector algorithm need to focus on data flow management between the servers. Because it takes a long time to gather all data of user. The data should be parted with different time gathering. Its not in one time pulled out data but parted in several different times for efficiency waiting time of user. The future study can enhance API (Application Programming Interface) authorization for different community web portal. The data collector uses several social networks API to interact with many social networks as the data source. Developing other method and algorithm research in calculation engine application to generate the rank table. Moreover, the calculation engine can use intelligence application. Output from calculation engine can communicate with other third party apps with API who used ranks table.

Bibliography

[1] P. Greenberg, CRM at The Speed of Light, Fourth Edition : Social Strategies, Tools, and Tecniques for Enganging Your Customer, McGraw-Hill, 2010.

[2] M. K. a. D. J. B. Ajay Mehra, "The Social Networks of High and Low Self-Monitors: Implications for Workplace Performance," *Administrative Science Quarterly,* vol. 46, no. Sage Publications, Inc., pp. 121-146, 2001.

[3] L. Barone, "Small Business Trends," Small Business Trends LLC., 15 July 2010. [Online]. Available: http://smallbiztrends.com/2010/07/the-5-types-of-influencers-on-the-web.html. [Accessed 12 12 2012].

[4] P. Wongthongtham, N. Kasisopha, E. Chang and T. Dillon, "A Software Engineering Ontology as Software Engineering Knowledge," in *Convergence and Hybrid Information Technology*, Perth, 2008.

[5] T. Dillon, E. Chang and P. Wongthongtham, "Ontology-Based Software Engineering- Software Engineering 2.0," in *19th Australian Conference on Software Engineering*, Perth, 2008.

[6] Y. Zhao, J. Dong and T. Peng, "Ontology Classification for Semantic-Web-Based Software Engineering," *IEEE JOURNALS & MAGAZINES,* vol. 2, no. 4, pp. 303-317 , 2009.

[7] W. Siricharoen, "A Software Engineering Approach to Comparing Ontology Modeling with Object Modeling," in *International Symposium on Computer Science and its Applications*, Univ. of the Thai Chamber of Commerce, Bangkok, 2008.

[8] T. Li, H. Yang, J. He and Y. Ai, "A Social Network Analysis methods based on ontology," in *Knowledge Acquisition and Modeling (KAM)*, 2010.

[9] J. Hilera and L. Fernández-Sanz, "Developing Domain-ontologies to Improve Sofware Engineering Knowledge," in *Software Engineering Advances (ICSEA)*, 2010.

[10] "Social Network Analysis Theory and Applications," 3 January 2011. [Online]. Available: http://train.ed.psu.edu/WFED-543/SocNet_TheoryApp.pdf. [Accessed 30 August 2012].

[11] M. Jamali and H. Abolhassani, "Different Aspects of Social Network Analysis," in *IEEE/WIC/ACM International Conference on Web Intelligence*, 2006.

[12] L. Ya-ting and C. Jing-min, "The social network analysis of political blogs in people: Based on centrality," in *Consumer Electronics, Communications and Networks* , 2011.

[13] B.-x. Li, M.-j. Li, D.-y. Sun, J. Li and W.-j. Li, "Evaluation method for node importance based on node condensation in terrorism networks," in *Intelligence and Security Informatics (ISI)*, 2012.

[14] W. P. O, "Ontology Based Knowledge Discovery in Social Networks," in *Ispra : Joint Research Center*, 2005.

[15] P. Kumar and K. Zhang, "Social Network Analysis of Online Marketplaces," in *e-Business Engineering*, 2007.

[16] Wikipedia The Free Encyclopedia, "Centrality," Wikipedia, 29 June 2012. [Online]. Available: http://en.wikipedia.org/wiki/Centrality. [Accessed 7 July 2012].

[17] Creative Commons Attribution-Share Alike 3.0 Unported, Social Network Analysis Theory and Applications, Creative Commons.

[18] J. Warmbrodt, H. Sheng and R. Hall, "Social Network Analysis of Video Bloggers' Community," in *Proceedings of the 41st Hawaii International Conference on System Sciences*, 2008.

[19] R. Millington, "The Online Community Guide," 03 May 2011. [Online]. Available: http://www.feverbee.com/2011/05/influencers.html. [Accessed 26 June 2012].

[20] L. Chen, S. Wei and Z. Qingpu, "Semantic Description of Social Network Based on Ontology," in *E-Business and E-Government (ICEE)*, 2010.

[21] T. S. Z. M. e. a. Deng ZH, "Overview of Ontology," in *Acta Scicentiarum Naturalum Universitis Pekinesis*, 2002.

[22] B. W. N, "Construction of Engineering Ontologies for Knowledge Sharing and Reuse," in *University of Twente*, 1997.

[23] S. a. F. K. Wasserman, Social Network Analysis: Methods and Applications. Cambridge, ENG and New York: Cambridge University Press, 1994.

[24] J. Scott, Social network analysis: a handbook, London: Sage, 1991.

[25] J. B. Bavelas, On "naturalistic" family research, 1984.

[26] G. Sabidussi, The centrality index of a graph. Psychometrika, Psychometrika 31, 581--603., 1966.

[27] L. C. Freeman, Centrality in social networks: Conceptual clarification, Social Networks, 1:215-239., 1979.

[28] B. P. S. P. L. P. H. D. E.-E. K. a. R. J. Kitchenham, "Preliminary Guidelines for Empirical Research in Software Engineering," *National Research Council of Canada,* 2001.

[29] S. A. Catanese, P. D. Meo, E. Ferrara, G. Fiumara and A. Provetti, "Crawling Facebook for Social Network Analysis Purposes," in *ACM*, Sogndal, 2011.

[30] U. Khurana, V.-A. Nguyen, H.-C. Cheng, J.-w. Ahn, X. Chen and B. Shneiderman, "Visual Analysis of Temporal Trends in Social Networks Using Edge Color Coding and Metric Timelines," in *Privacy, security, risk and trust (passat), 2011 ieee third international conference on and 2011 ieee third international conference on social computing (socialcom)*, 2011.

[31] D. Hansen, M. Smith and B. Shneiderman, "EventGraphs: Charting Collections of Conference Connections," in *System Sciences (HICSS), 2011 44th Hawaii International*, 2011.

[32] "Web People Search Using Ontology Based Decision Tree," *International Journal of Data Mining & Knowledge Management Process (IJDKP)* , vol. 2, 2012.

[33] Wikipedia, "Guttman Scale," Wikimedia Foundation, Inc.,, 16 August 2011. [Online]. Available: http://en.wikipedia.org/wiki/Guttman_scale. [Accessed 05 July 2012].

[34] S. F. a. L. Lee, "Getting Brand Communities Right," Harvard Business Review, April 2009. [Online]. Available: http://hbr.org/web/special-collections/insight/marketing-that-works/getting-brand-communities-right. [Accessed 01 July 2012].

[35] P. Mika and A. Gangemi, "Descriptions of Social Relations," Peter Mika, 2002. [Online]. Available: http://www.cs.vu.nl/~pmika/research/foaf-ws/foaf-x.html. [Accessed 01 July 2012].

[36] "Wikipedia," Wikimedia Foundation, Inc.,, 16 August 2011. [Online]. Available: http://en.wikipedia.org/wiki/Guttman_scale. [Accessed 05 July 2012].