Advanced Science Letters
Vol. 4, 400–407, 2011

Social Network and Sentiment Analysis for Social Customer Relationship Management in Indonesia Banking Sector

Andry Alamsyah, Anindia Aulia Indraswari
School of Economics and Business, Telkom University, Bandung, Indonesia

The increasing number of social media users affects both individual and corporation user. Banking sector, for example, use social media to support their Social Customer Relationship Management activity. We investigate the dynamics and evolution of conversation network between bank customer using Social Network Analysis methodology. Measurement is conducted by calculating its network properties to see the characteristic and how active the network is. Customers talking about banks' services can also express their opinion on social media. Therefore, we perform sentiment analysis to classify customer's opinion into positive, negative and neutral class. This research was performed on Twitter's conversation about Bank Mandiri, Bank Central Asia (BCA) and Bank Negara Indonesia (BNI). The result of this research is beneficial for business intelligence purpose to support decision making.

Keywords: Social Network Analysis, Customer Relationship Management, Sentiment Analysis, Network Property, Banking Industry, Business Intelligence

1. INTRODUCTION

Indonesia is one of the country with highest internet user. With 72.7 million internet users in January 2015 and 72 million active social media account[11]. Twitter is one of the most popular social media platform in Indonesia with approximately 50 million individual and corporate users in early 2015 [19]. Companies from various sector use twitter as a tool to introduce their product, interact with customers and listen to what their customer need.

Using social media, companies not only can listen to the conversation between their customer about their products and services but can also measure how they use social networking[15]. Companies also use it to implement their customer relationship management strategy and engage with their customer better [9,29]. We then call this activity as social customer relationship management (SCRM) [29].

Social media offers benefits to companies in all sectors[29], unexceptionally banking sector. Presence of the bank in social media is considered as a positive factor leading to engagement[9]. But according to previous research, most of banks in emerging country still haven't satisfy their customer on social media. Only 21% of the bank develop and implement social media strategy, but only 13% of the bank have a comprehensive social media strategy. And only 38% of the bank monitor customer's conversation on social media [24].

The purpose of this research is to investigate how bank customer conversation network evolve and the dynamics of information spreading, including the information of relationship between banks and their customers in social media. To see the dynamics and evolution of conversation network, we compare the network properties between three banks in Indonesia. We also perform sentiment analysis to discover which of the

*Email Address: andrya@telkomuniversity.ac.id

 1936-6612/2011/4/400/008 doi:10.1166/asl.2011.1261

three bank has the most positive sentiment based on twitter conversation about the banks. We try to adopt the method from[5] to understand better conversation in Bahasa Indonesia using appraisal theory[5]. As explained in former work, the appraisal theory says that emotions are consequences of how an author appraised a particular situation [21]. Therefore, this theory is used in sentiment analysis to discover what does an author feels when writing on microblog and what is the situation that caused the feeling appears[21].

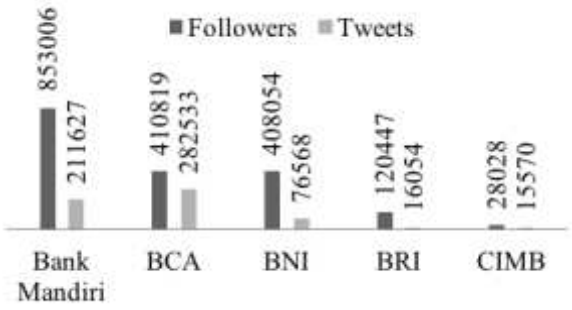

Figure-1. Number of Accumulated Followers & Tweets of Top 5 Bank in Indonesia with Largest Asset**s**

As we see from the figure above, Bank Mandiri, BCA and BNI has the largest number of followers and tweets. Therefore, we choose them to get significant result of conversation traffic data.

In this research, we use Social Network Analysis (SNA) approach to measure SCRM network properties and sentiment analysis to classify positive, negative and neutral sentiment for each bank. We compare eight network properties for each bank, which are network size, density, modularity, diameter, average path length, average degree, reachability and connected component.

The result of this research is beneficial for business intelligence purpose. Business can understand their competitors and customer deeper and better with this insight to support decision making.

2. LITERATURE REVIEW

2.a Social Network Analysis

Social network analysis (SNA) was first introduced as a basic approach of studying social relations to understand relationship among people within a social group [8]. SNA is described as a study of human relationship by means of graph theory [17]. The study of social network defined as a set of nodes or network member, and connected by different types of relation called links [8].

More than just a methodology, SNA gives a perspective on how society functions [12]. SNA assumes that people are interdependent, thus it differs with other traditional approaches which assumes that what people do, think and feel is independent of who they know [14]. SNA can also be applied to a wide range of business problem such as strategy, sales and marketing, human resource, team-building, also knowledge management and collaboration [14].

In this research, we use SNA as an approach to calculate network properties for each SCRM bank network. But before we jump to the measurement, we need to understand the definition of each network properties one by one. The network properties are shown in Table-1.

Table-1. Network Properties Description

Network Properties	Description
Size	Number of nodes in network [23].
Density	The fraction of number edges in network to the maximum edges possible [4].
Modularity	The fraction of edges within communities minus the expected value of that fraction if the position of the edges are randomized [16].
Diameter	The largest distance recorded between any pair of nodes [1].
Average Path Length	The average distance between all pairs of nodes in the network [1].
Average Degree	The average number of links a node has to the other nodes [1].
Reachability	The fraction of node pairs that are connected [27].
Connected Component	Measure of maximal subset of nodes such that each node is reachable by some paths from each other [4].

[4, 16, 1, and 27].

2.b Sentiment Analysis

Nowadays, people use social media to express their opinion about products or services. Opinions can be positive or negative depends on customer's experience and satisfaction. Opinions posted on social media becomes valuable data resource [7]. Customers want to know opinions of existing user before purchasing a product. In the other hand, businesses and organizations need to know public opinions about their products and services to help decision making process.

The field of study that analyzes people's opinions, sentiments, evaluations, appraisals, attitudes and emotions toward entities such as products, services, organizations, individuals, issues, events, topics, and their attributes is called sentiment analysis [7].

The main task of sentiment analysis is to classify text into positive or negative class. In this research, we adopt the method used by [5] to classify sentiment in Bahasa Indonesia using appraisal theory[5] because the methods are proven applicable to social media conversation in Bahasa Indonesia. Appraisal theory describes how authors use language to communicate with other[5] and it shifts sentiment classification further and considers the appraisal expression. Appraisal expression is a basic grammatical unit by which an opinion is expressed[21].

Sentences written in Bahasa Indonesia contains three basic attributes of appraisal theory and target [5]. According to [21], those three attributes are:

a. Attitude, expresses the current state of a person at the time she/he wrote a text. Attitude has

three subcategories: affect, which represents the feeling of the author (happy, sad); appreciation, which talks about the opinion that a person has about the inner or outer qualities of an object (ugly, beautiful); and judgement, which describes the behavior of somebody in a social context (heroic, feeble-minded).

b. Graduation, expresses that the meaning of a term is gradated by some adjective (or other part of speech) and its meaning is strengthened or weakened (very, few).

c. Engagement, determines the position of text proposal. It reflects probability or possibility (perhaps, seems, maybe) in most cases.

While target of a post or sentence can be an object, an event, or a person. In this case, the target is the keyword related to brand of bank (Bank Mandiri, BCA, and BNI and/or its feature) [21].

2.c Social Media

Social media is a group of new kinds of online media, which share characteristic such as participation, openness, conversation, community and connectedness [2]. Social media can also be defined as the production, consumption and exchange of information across platforms for social interaction[27]. Basically, there are six kind of social media. Those are social networks, blogs, wikis, podcasts, forums, content communities, and microblogging. But there's no limitation that there will be another kind of social media in the future [2].

Social media enables us to connect and interact with each other anywhere and anytime. It is also allowing us to observe human behavior in an unprecedented scale with a new lens that enables us to understand individuals at scale and to mine human behavioral patterns that were impossible before [25].

Microblogging is one of commonly used social media nowadays. It is a combination between social network and blog but with shorter and smaller amount of content accessible online [2]. There are 255 million worldwide active users [26] and 50 million users in Indonesia as of 2014 [3], we can say that twitter is the leader in microblogging platform.

2.d Social Customer Relationship Management

Customer relationship management defined as a process of carefully managing detailed information about individual customers and all customer "touch point" to maximize loyalty. CRM enables companies to provide excellent real-time customer service through the effective use of individual account information [22]. But with the exponential growth of web 2.0 technology and social media, communication shifts into customer centralized service. Thus, a new strategy called social customer relationship management appear as a response to this new technology [6].

Social customer relationship management (SCRM) is 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 to provide mutually beneficial value in a trusted and transparent business environment. It's the company's response to the customer's ownership of the conversation [20].

3. METHODOLOGY

In this research, we describe the characteristics of SCRM network from three banks in Indonesia, which are Bank Mandiri, BCA and BNI. We use SNA methodology and approach to measure network properties of each bank. SNA enables us to study about human interaction using graph theory. We also use sentiment analysis to classify whether a sentence has a positive, negative or neutral sentiment.

We use twitter's conversation as data source in this research. We only use tweets that include three kinds of keyword, such as "mandiri", "bca" and "bni". Where "mandiri" keyword represents Bank Mandiri, "bca" keyword represents Bank Central Asia (BCA) and "bni" keyword is for Bank Negara Indonesia (BNI). We take tweets sample containing referred keywords starting from February 27th 2015 to March 5th 2015. Below are several processes of data analysis:

Data Collection

We crawl twitter data using twitter API (Application Programming Interface) functionality that support twitter data collection. We only use tweets which contain keywords "mandiri", "bca" and "bni".

The result of this process is raw tweet in CSV format containing 16 columns. But we only use two columns. The first column is "text" which contains the conversation and actors who join the conversation. The second is "screenName" column which contain the username of tweet writer.

Although we take the data of communication between brand and their customers, we are confident that this activity doesn't violate any privacy since we only take public data of twitter conversation and the data is accessible by anyone worldwide.

Data Preprocessing

In this process, we split the data into two parts. The first part, we only clean unnecessary conversation for sentiment analysis. In the second part, we remove conversation content and only left actors or username included in conversation. Username of actor is a term started with '@'.

Sentiment Analysis

The task of sentiment analysis in this research is to classify conversation into positive, negative and neutral sentiments. These activities were conducted using appraisal theory. The process of classifying the sentiment is explained on the Figure-2.

As an output, a sentence will be labelled as positive if it contains more positive appraisals. If a sentence contains more negative appraisals, it will be labelled as negative sentiment. While, a sentence will be labelled as neutral if it doesn't contain appraisal or has an

 1936-6612/2011/4/400/008 doi:10.1166/asl.2011.1261

equal positive and negative appraisal.

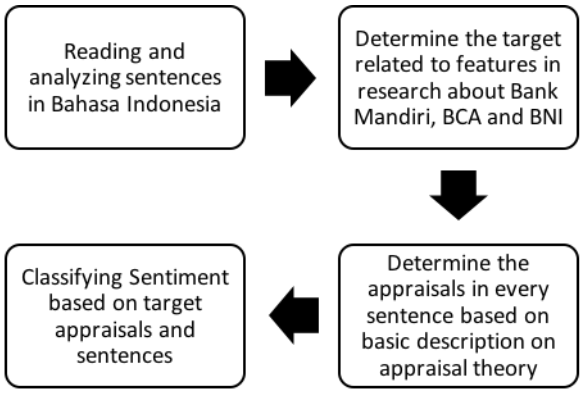

Figure-2. Workflow for Sentiment Analysis Process. Adapted from [5].

Example 1
Tlp ke mandiri call udah 15 menit cuma dpt jawaban
Target
"maaf seluruh staff km sibuk". Sangat memalukan.
Appraisal
Kecewa.... @mandiricare
Appraisal Target

The example of tweet above is classified as negative, because the sentence contains negative appraisal "Sangat memalukan" (very shameful) and "kecewa" (disappointed) that associated with "mandiri call" (Bank Mandiri's call center service) and "@mandiricare" (Bank Mandiri's twitter account).

Example 2
@newsplatter mandiri om piring. Mobile banking
Target Target
dan internet bankingnya juara.
Target Appraisal

While the tweet above is an example for positive sentiment because the targets "Mandiri" (Bank Mandiri's brand name), "Mobile banking" (refers to Bank Mandiri's mobile banking service), and "internet bankingnya" (refers to Bank Mandiri's internet banking service) is associated with positive appraisal "Juara" (the best).

Example 3
@AciiDR Utk pengajuan CC BCA Bpk/Ibu bs mengisi
Target
formulir disertakan KTP,NPWP serta slip gaji di cabang
Target
terdekat cc:@KartuKreditBCA ^Lady
Target

We use the same method to classify neutral sentiment. After we finished labelling all the tweets, we calculate the total of positive, negative and neutral sentiment of each Bank. The number of positive sentiment is divided by the total tweets about the Bank. For example, Bank Mandiri has 1724 tweets classified as positive sentiment from the total of 6082 tweets. Then the percentage of Bank Mandiri's positive sentiment is:

$$\frac{1724 tweets}{6082 tweets} x100\% = 28.35\%$$

SCRM Network Visualization

The next step after preprocessing the data is to make a visualization using only username of actor in the conversation. We already extract all of username in conversation and put it into two columns. First column contains the tweet writer's username and the other contains the username of people in the conversation. For example, if a conversation involves two accounts @budiman and @bankmandiri, then they will appear in a same row, but in separated column. The account @budiman will be in the first column, and the account @bankmandiri in the second column.

We uploaded the CSV file containing the list of nodes to Gephi software to make a visualization for our data. Gephi is an open source network graph and analysis tool [13]. We make visualization for each SCRM network bank using Yifan Hu Proportional layout. This layout enables us to make visualization in a short time despite the large amount of data.

SCRM Network Properties Calculation

We calculate the value of network properties for each SCRM network bank using Gephi, NodeXL and manual calculation. Gephi also provides us a feature for network properties measurement. NodeXL is similar with Gephi, except NodeXL provides more decimal characters than Gephi does. Network size, modularity, diameter, average path length and average degree and connected component are measured using Gephi. Density value is calculated using NodeXL. While reachability is measured using manual calculation using the the ratio of number connected node pairs and total number node pair combinations in network[27].

SCRM Network Comparison

The result of network properties measurement will be compared for each bank. The more the bank get best value, it means the network is better and more active. Then we can determine which bank has the best interaction with their customer based on network property values.

4. RESULT AND ANALYSIS

We perform sentiment analysis on the three banks by classifying tweets containing "mandiri", "bca" and "bni" keywords into three classes of sentiment. Tweets containing positive appraisal such as satisfaction statements, supportive statements and representing a good attitude were classified into positive sentiment. While the opposite statement such as dissatisfaction statements, complains, and disappointment were classified into negative sentiment. The rest tweets that didn't have such statements above were classified into neutral sentiment. The result of sentiment analysis is shown in the Table-3.

Table-3. Percentage of Sentiment Analysis Result Between Three Banks

Sentiments	Bank Mandiri	BCA	BNI
Positive	1724 *tweets*	861 *tweets*	272 *tweets*
	28,35%	23,05%	13,51%
Negative	727 *tweets*	967 *tweets*	306 *tweets*
	11,95%	**25,88%**	15,20%
Neutral	3631 *tweets*	1908 *tweets*	1435 *tweets*
	59,70%	51,07%	**71,29%**
Total	6082 *tweets*	3736 *tweets*	2013 *tweets*
	100%	100%	100%

From the Table-3, we can summarize that the highest percentage of positive sentiment is achieved by Bank Mandiri with 28,35% when BCA only got 23,05% and BNI got 13,51%. The highest negative sentiment percentage comes from BCA with 25,88% of the total tweets, leaving BNI with 15,20% and Bank Mandiri with 11,95% behind. While bank with the most neutral sentiment is BNI with 71,29% of total tweets when Bank Mandiri got 59,70% and BCA only got 51,07%.

After we calculate each network properties for each bank, we put the result on a table for easier comparison. Result for each bank will be further explained in Table-5, while Table-4 shows the network size.

Table-4. Network size of Three Banks

	Bank Mandiri	BCA	BNI
Nodes	4779	2794	2272
Edges	4694	2879	2116
Graph Type	*Undirected Graph*		

The Table-4 represents the number of nodes and edges of the three banks. We choose undirected graph to visualize the networks because twitter's conversation doesn't consider who is connected to whom. But it only giving the node a link to show that one node is connected to another.

The Table-5 are the result of measurement using Gephi, NodeXL and manual calculation. It shows the comparison of network property values between three banks:

Table-5. Comparison of Network Properties Between Three Banks

Network Properties	Bank Mandiri	BCA	BNI
Size	**4779**	2794	2272
Density	0,0004	0,0007	**0,0008**
Modularity	**0,903**	0,880	0,790
Diameter	16	16	**14**
Average path length	4,938	5,267	**2,684**
Average degree	1,964	**2,061**	1,863
Reachability	0,0003	0,0004	**0,0008**
Connected Component	372	332	**268**

As we can see from the table above, the highest value comes from Bank Mandiri with a total of 4779 nodes followed by BCA and then BNI. Density of network gives us the idea of how dense the connection of every node in a network. The higher the density value of a network means the more connections appear in every node in the network. It is simply like "everyone knows each other" in the network. From the three banks above, BNI gets the highest density value, followed by BCA and Bank Mandiri.

Modularity measure the strength of communities within a network. The higher the modularity value means the communities within the network is denser and stronger [1]. The highest modularity value of 0,903 is achieved by Bank Mandiri, followed by BCA and then BNI. Previous research states that high modularity value indicates a good community structure [10].

The next property is diameter of network, which is the longest shortest path in a graph or the distance between the two furthest away nodes[1]. The best network is a network with the smallest diameter because it indicates small-world phenomenon[10]. BNI has the smallest network diameter value of 14, followed by Bank Mandiri and BCA with the same diameter.

Average path length is the average distance between all pairs of nodes in the network[1]. Lower value of average path length indicates a stronger connection between nodes within a network[18]. BNI has 2,684 which is the lowest value of average path length, followed by Bank Mandiri and BCA. It means that information dissemination in BNI network will be much faster than the rest of other banks.

Average degree represents the average number of links of nodes has[1]. The higher value of average degree is better, because it means that the actor or nodes "know" more actors within the network. BCA has the higher average degree value of average degree, followed by Bank Mandiri and BNI. Hence, the actors or nodes in BCA network is more connected than the two other banks.

An actor is reachable by another if there exist any set of connections by which we can trace from the

1936-6612/2011/4/400/008 doi:10.1166/asl.2011.1261

source to the target actor, regardless how many others fall between them[23]. Reachability of a network is the fraction of node pairs that are connected[27]. In this case, we are measuring the percentage of node pairs connected in the network. BNI has the highest reachability value, meaning BNI has more connected node pairs. Higher value of reachability in networks is better, because information dissemination will be much easier and faster.

Connected component is simply described as a maximal set of nodes such that there is a path between every pair of nodes. The component are separate "pieces" of the graph such that there is no connection between the pieces[28]. From the three networks above, BNI has the smallest value of connected component, followed by BCA and Bank Mandiri. Smaller number of connected component is better, it means that the network doesn't separated into many tiny groups.

Three figures below are visualization of SCRM network for Bank Mandiri (Figure 3), BCA (Figure 4) and BNI (Figure 5). All of them were processed with Gephi software using Yifan Hu Proportional layout. The graphs show the relationship between nodes in the network. Each graph represents actors in network. For example, the first figure represents the actors joined in the conversations about Bank Mandiri using "mandiri" keyword.

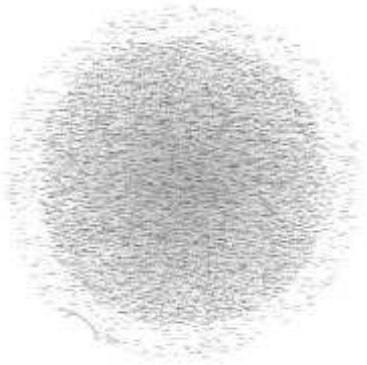

Figure-3. SCRM Network Visualization for Bank Mandiri

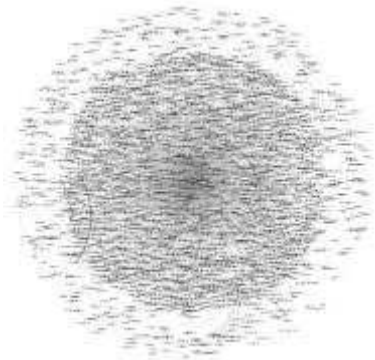

Figure-4. SCRM Network Visualization for BCA

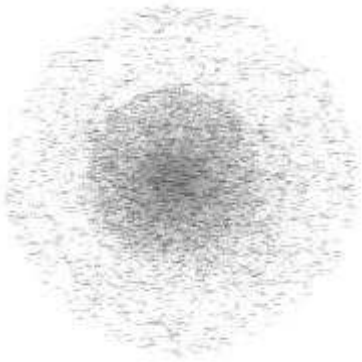

Figure-5. SCRM Network Visualization for BNI

5. Conclusions

We have shown that it is possible to measure SCRM activity using network properties. We use network size, density, modularity, diameter, average path length, average degree, reachability and connected component for SCRM network measurement. From these measurement, we find that BNI has the most active SCRM network, followed by Bank Mandiri and BCA.

We find that network SCRM measurement is not directly proportional with sentiment analysis. BNI has the most active SCRM network, but has the smallest percentage of positive sentiment. While, BCA has the highest percentage of positive sentiment but the least active SCRM network. There is a possibility for the next researcher to study about the relationship between SCRM network measurement and sentiment analysis. Next researcher can also use new tools, approach and other media for experiment.

We can use this research to support business intelligence activity. After we understand what is the strengths and weaknesses compared to the competitor, we can use this insight to formulate strategy for SCRM. Therefore, the actions and decisions related to SCRM activity is based on conversational data will be more accurate.

REFERENCES

[1] A. Barabasi. Network Science. Available: http://barabasi.com/networksciencebook/. (2012).

[2] A. Mayfield. What is Social Media?. Available: http://ebooksoneverything.com/marketing/WhatisSocialMedia.pdf . (2008).

[3] A. Wicaksono. Jumlah Pengguna Twitter di Indonesia Akhirnya Terungkap. Retrieved from CNN Indonesia: http://www.cnnindonesia.com/teknologi/20150326141025-185-42076/jumlah-pengguna-Twitter-di-indonesia-akhirnya-terungkap/. (2015).

[4] A. Alamsyah, B. Rahardjo, Kuspriyanto. Social Network Analysis Taxonomy Based on Graph Representation. *Proceeding of The 5th Indonesian International Conference on Innovation, Entrepreneurship, and Small Business.* (2013).

[5] A. Alamsyah, R. Widita. H. Irawan. Sentiment Analysis Based On Appraisal Theory For Marketing Intelligence In Indonesia's Mobile Phone Market. Journal of Theoretical and Applied Information Technology. 82(2). (2015) 335-340.

[6] A. Alamsyah, Y. Peranginangin, G. Nurhadi. *Learning Organization using Conversational Social Network for Social*

Customer Relationship Management Effort.2nd International Conference and Seminar on Learning Organization. (2014).
[7] B. Liu. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publisher. (2012).
[8] C. A. Pinheiro. *Social Network Analysis in Telecommunications*. New Jersey: John Wiley & Sons, Inc. (2011).
[9] C. Giannakis-Bompolis, C. Boutsouki. Customer Relationship Management in the Era of Social Web and Social Customer: An Investigation of Customer Engagement in the Greek Retail Banking Sector. *Procedia Social and Behavioral Sciences*, (2014) 67-78.
[10] E. Ferrara. A large-scale community structure analysis in Facebook. *EPJ Data Science* 1.1. (2012) 1-30.
[11] E. Lukman. The Latest Numbers on Web, Mobile and Social Media in Indonesia (INFOGRAPHIC). Retrieved from Tech In Asia: https://www.techinasia.com/indonesia-web-mobile-data-start-2015/ (2015).
[12] G. Cheliotis. Social Network Analysis. Available: http://haroldliu.weebly.com/uploads/1/5/8/1/15810196/socialnetworkanalysis.pdf . (2010).
[13] K. Chevren. *Mastering Gephi Network Visualization.* Birmingham: Packt Publishing Ltd. (2015).
[14] K. Ehrlich, I. Carboni. Inside Social Network Analysis. Available: http://ppr.cs.dal.ca/sraza/files/social%20networks(1).pdf (2005).
[15] M. J. Mosadegh, M. Behboudi. Using Social Network Paradigm For Developing A Conceptual Framework In CRM. *Australian Journal of Business and Management Research*, (2011) 63-71.
[16] M. Newman. Communities, Modules and Large-Scale Structure in Networks. *Nature Physics*, 8, (2012) 25-31.
[17] M. Tsvetovat, A. Kouznetsov. *Social Network Analysis for Startups*. Sebastopol: O'Reilly Media Inc. (2011).
[18] M. W. Hadley, et al. A New Measure Based On Degree Distribution That Links Information Theory And Network Graph Analysis. *Neural Systems & Circuits*, 2(7). (2012) 1-15.
[19] N. A. Hasibuan, Jumlah Pengguna Twitter di Indonesia Akhirnya Terungkap. CNN Indonesia. Available: http://www.cnnindonesia.com/teknologi/20150326141025-185-42076/jumlah-pengguna-twitter-di-indonesia-akhirnya-terungkap/ (2015).
[20] P. Greenberg. *CRM at the Speed of Light*. The McGraw-Hill Companies (2010).
[21] P. Korenek, M. Simko. Sentiment Analysis on Microblog Utilizing Appraisal Theory, Springer, 17(4), (2013) 847-867.
[22] P. Kotler, K. L. Keller. *Marketing Management 14th Edition*. New Jersey: Pearson Education, Inc. (2012).
[23] R. A. Hanneman, M. Riddle. *Introduction to Social Network Methods*. Available at: http://faculty.ucr.edu/~hanneman/nettext/C7_Connection.html. Department of Sociology, University of California, Riverside. (2005).
[24] R. Wicaksono. Beberapa Bank Masih Kesulitan Memenuhi Harapan Nasabah. Retrieved from http://swa.co.id/business-research/beberapa-bank-di-negara-berkembang-masih-kesulitan-memenuhi-harapan-nasabah . (2012).
[25] R. Zafarani, M. A. Abbasi, H. Liu. *Social Media Mining: An Introduction*. New York: Cambridge University Press. (2014).
[26] S. Bennett. *Facebook, Twitter, Instagram, Pinterest, Vine, Snapchat – Social Media Stats 2014 [INFOGRAPHIC]*. Retrieved from Adweek: http://www.adweek.com/socialtimes/social-media-statistics-2014/499230. (2014).
[27] S. Perur, S. Iyer. Reachability: an alternative to connectivity for sparse wireless multi-hop networks. *Poster at IEEE infocom. Barcelona, Catalunya, Spain, 2329*. Available: http://www.it.iitb.ac.in/ srinath/rch.pdf . (2006).
[28] S. S. Skiena. *The Algorithm Design Manual Second Edition*. New York: Springer. (2008).
[29] V. Dutot, A New Strategy for Customer Engagement How Do French Firms Use Social CRM? *International Business Research, 6*(9), (2013) 54-67.