Social Network Modeling Approach for Brand Awareness

Andry Alamsyah
School of Electrical Engineering and Informatics
Institut Teknologi Bandung
Telkom Economy and Business School
Telkom University
andrya@telkomuniversity.ac.id

Fatimah Putri
Telkom Economy and Business School
Telkom University
fatimahfapud@gmail.com

Osa Omar Sharif
Telkom Economy and Business School
Telkom University
osaomarsharif@telkomuniversity.ac.id

Abstract— **The advance of social computing study is grown very fast to the extents that influence many aspects of our daily life. One of those aspects is implementation in marketing / advertising. We are always looking for the effective way to understand our market, but the absent of powerful metrics has cause some problem. The availibility of Big Data all around us has triggered a new perspective on how we approach those problems. Online social network disseminate information much faster than before, supporting highly exposure of brand awareness. The network behavior of the information spreading can be explained by the study of complex network. In this paper, we propose the social network modelling approach using graph theoritic to understand on how brand information travels in online social network and how it can benefit for business. In prior study in marketing, it is uncommon to approach phenomenon using social network model and online social data, they are mostly using questionnaire from population sample. Our paper will enrich effort in marketing study. Our experiment use conversation data from Indonesian Twitter user contains specific brand keyword**

Keywords—social computing; social network analysis; graph theory; brand awareness; marketing; Twitter; complex network; Big Data

I. INTRODUCTION

Advances in computer technology, particularly the development of the Internet, have affected virtually aspect of society. Online data production and consumation has taken us to the level where we can gain many insights from our daily activities. This insight can describe, explain and even predict human behavior individually and socially. Big Data is one of the recent important issues that show us the opportunies as well as the challenge of taming large-scale data. Some advantages of having Big Data are the abilility to have accurate pattern from such big volume data, the ability to view comprehensive issue from variety type of data, and the ability to process on-the-fly data stream / real-time data.

Social Computing study is concern about social behavior and social context using computational system. Social network plays important rules on the development of social computing. The others factor support the development such as pervasive use of mobile devices, sophisticated social network services and cheap Internet connection. Human leaves digital trace that actually has economic value that we can analyze. Some interesting topics in social computing such as: network topology aspect (social network analysis or complex network), content mining, semantic analysis, sentiment analysis/opinion mining, recommender system, and trust/reputation system. We are interested in network aspect of the information spreading on specific topics. This approach can be explained by study of complex network, which will be explained in part ll.

Brand awareness is the condition on which brand will recognized by potential customer and it is correctly identified to particular product. Business objective is to get their brand at top-of-the-mind on the market. In order to achieve the objective, it needs hard work and long time continuous effort. The question is how we can make those efforts more effective and efficient, considering today market competition is very high thus it is getting difficult for market to associates our brand with product at their mind. In prior study, to answer those questions, companies arrange questionnaire to public that prove to be very expensive and takes long times.

By understand how the market behave, we can explain and predict how and why our product are failed in the market and competitors product are more succesful than ours. These markets are represented as dynamic and complex social network, where online “word-of-mouth” can travel faster. The advantage is much greater in a network of close-knit friendship or communities, they can intensify attitude towards a product from only brand recognition to the dominant brand at their mind. This attitude can works both in positive and negative way. The carefull strategy to introduce brand awareness in social network is needed.

Today, social network data can help business to predict and understand their market behavior. The methods are started from the development of data mining technique, which now developped into content analysis and mining. Reasoning the objective and action, connecting different types of data, learning from the pattern are some of the activities needed [1]. Decisions and negotiations can also be taken based on

978-1-4799-3580-2/14/$31.00

comprehensive understanding empirical and analytical network model [2].

II. COMPLEX NETWORK AND BRAND AWARENESS

Networks that display substantial non-trivial topological features are called as complex network [3]. The patterns of complex network are both not fully regular and not fully random, which is exactly found in our empirical observations from real-world network such as social network. The research of complex network is very active recently with the advanced of computing power and availability of data set.

Some important properties of complex network are scale-free networks and preferential attachment. The scale-free network shows that the degree distribution of nodes is following power-law function. It implies that node tends to groups together forms a big component inside the network. The network with this characteristic is sensitive to attack that will spread very fast, especially when the target are higher nodes in the network. It also has characteristics of small-world network [4], where information spread very quickly to the whole network. Scale-free network degree distributions are illustrated in Figure 1 from [4].

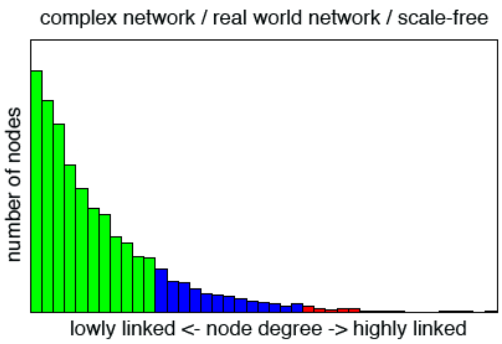

Figure 1. Scale-free network nodes degree distributions [4]

Preferential attachment stated that network during their evolution and growth; a great majority of new edges are connected to nodes with an already high degree, thus the degree of these nodes increase disproportionately [4]. In real world, this is make sense since people want to be associated with popular one to increase their popularity, or want to be connected to the higher quality brands, or every people/brand that reach "critical mass" will become stars with many friends or followers.

Brand equity in marketing is a value of having a well-known brand [5]. There are five steps in forming brand equity: brand awareness, brand association, perceived quality, brand loyalty, and proprietary brand assets. The first important step is brand awareness, which is defined as how a product is recognized by potential customer and associated with correct product [5]. Brand awareness usually generated using advertising or word-of-mouth in spreading the information. The processes of brand awareness [6] are following: unaware of the brand, brand recognition, brand recall, and top-of-mind. It is shown in Figure 2.

Our first hypothese is we can model the spreading process of brand awareness in online social network which objective similar to previous effort to disseminate awareness through advertising and word-of-mouth. The second hypothese if we use social network modelling approach, then we can achieve brand awareness process more effective and efficient by understanding network formation that leads to social behavior of market.

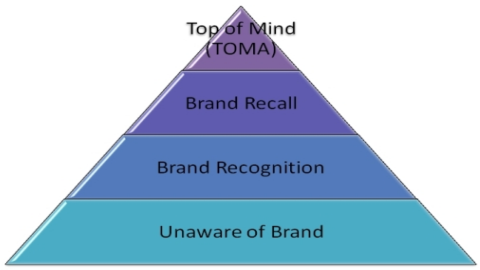

Figure 2. The process of brand awareness [6]

III. SOCIAL NETWORK ANALYSIS

A. The Model

Social Network Analysis (SNA) is a method of modelling relationship between users and represents it into graph [7]. By analyzing interactions or relationship, we will understand better peoples and the groups / communities. The central concept is by looking at our relationship, and they are all taken together, it will define who we are and how we act. The relationship in SNA could be anything from friendship, organizations, family relations, common interest, financial exchange, colleagues, knowledge flows, and others.

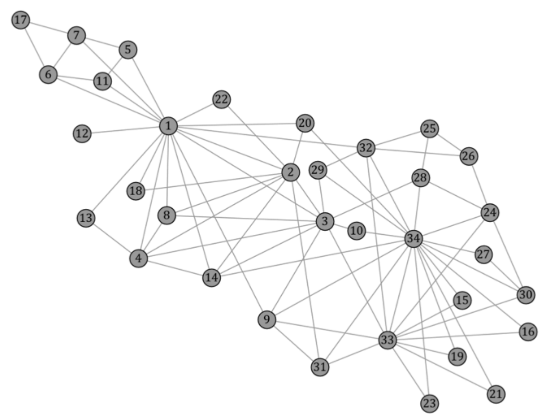

Figure 3. The visualization of social network with 34 actors and 78 relations between them [8]

SNA models use graph theory [9] of mathematics to describe and visualize the network. In this model actors / users represented by nodes and relationship represented by edges. In the formal methods we can state a network as a set of graph *G(N, E)*, where *G* is graph contains a set of node *N* and a set of edge *E* that connect some $n \in N$. An example of visualization social network consists of 34 actors and 78 relationships / friendship among the actors are shown in Figure 3 is taken from [8]. From the visualization we can see the obvious measurement such as actor 1 and actor 34 have more friends than other actors. The less obvious measurements from visualization are discussed in part lll.B below.

B. Metrics and Measurement

Graph theoritic [9] helps SNA to provide usefull properties that later can be used as metric and measurement of an arbitrary social network. Two identical networks size might not have the same measurement result as it depends on the their internal structure or network topology. This leads to the richness of measurement combination between network formation and the metrics used.

We divide the measurement into three big parts with several metrics on each every measurement. Those parts are Connections, Distributions, and Segmentation. Connections measures how actors connected to each other's, type of connection, network structure resulted from connections between actors. Distributions measures how actors distributed across the network (location), how informations walks through paths in network, what can we add or remove in order to make distributions across network more efficient. Segmentation measures network segmentation into groups / communities, cohesive bonds between actors / strenght-ties inside groups. Several important metrics from the three general measurements can be details explained in Table 1.

TABLE I. METRICS AND MEASUREMENTS

No	Metric	Function	Measurement
1	Homophilly	Measure how actors form ties with similar one rather than dissimilar one, example peoples tends to form grup with peoples from same ethnicity	Connections
2	Multiplexity	Measure the number of relationship between two actors, example two peoples that are friends, are also colleagues at work.	Connections
3	Community Identification	Detect / Identify how many communities / groups inside the network or how clustered network is?	Segmentation
4	Structural Holes	Network that are clustered into several groups but there are no actors that connect those groups.	Distributions
5	Bridge	An actor with weak ties that fill structural holes in network or an actor that connect disconnected groups	Distributions
6	Centrality	In terms of the relationship, measure how important an actor in a network. There are several metrics in centrality explained further in part lll.C	Distributions
7	Density	The proportions of direct ties in a network relative to the total number possible ties.	Distributions
8	Distance, Diameter and Size	Measures distance between actors, measures diameter of a given network, and measures how many actors and their relations of a given network	Distributions
9	Clustering Coefficient	Measures of the likelihood relation between actor A and actor B, when both actors are connected to actor C	Segmentation

C. Centrality and Community Identification

Having handfull metrics and measurement (as in Table 1 above) is one of the advantage using network modelling for brand awareness case. However, the two most interesting metrics are *centrality* and *community identification* for the reason they complement each other in describing how the information flows in market. Both metrics are the most asked questions about network formations.

Centrality is a group of metrics that measure the importance of actors in a network [10]. Several most important centrality metrics are betweenness, closeness, degree, and eigenvector. Betweeness and closeness measure the importance of actor location in network. Betweenness centrality measures how often the informations flow through the path of the actor location. Closeness centrality measures how fast an actor reaches all other actors in the whole network. Degree centrality measures total number connection that an actor has. Eigenvector centrality measures the influence of an actor in network by assigning relative weight to all actors in network based to the concept that an actor connected to high-weight actors contribute more to the weight of that actor. In this paper, we focus on betweenness centrality, as it is suited to measure the activity of one node passed by information flow.

Community identification measures whether there are group or community inside a network [10]. Modularity [11] is a metric proposed to measure the quality of a division of a network. It views whether a node belong to one group or another. The higher value of modularity means network have dense connections between the nodes, and otherwise network have sparser connections. Several question in market network regarding this metric, such as knowing how many communities, what is the size of each community and their proportional to network size, how the communities behave, how the communitiies interact, which one the most influential community and several others.

IV. TWITTER AND EXPERIMENT

Twitter is a social network services that provide us a platform for analyzing social network models and its quantification. Twitter platform simplifies the most difficult works in mining online conversational data, which is standardisation of data source and data format. Our methods construct social network model based on those conversational data in specific context / brand. The result of this method should be better in terms of efficiency and effectiveness than conventional approach such as questionnaire and snowball sampling.

In order to have satisfactory result, data size should be big enough so that we can extract pattern or insight of the network. Twitter *application programming interface (API)* offer simpler way of gathering information / issue by providing uniform format. Twitter API is much more efficent than using previous technique such as web mining, which proved difficult regarding variety formats dan data types available on the web.

Twitter has 500 million users worldwide users in 2012. Indonesia is the 5th largest twitter user in the world and Jakarta is the busiest city in the world making conversation on twitter [12]. With this data in mind, we are confidant about the accuracy of our approach using Indonesian Twitter data conversation. Indonesian likes to discuss many things on their daily life such, it reflected from the variations of local Twitter trending topics, some examples are topics in entertainment, politics, hot news, including product review, product recommendations, and product advertisement. Those last activities support product branding. With this data in mind, our experiment is conducted specifically to Indonesian user.

Our experiment uses real-world online conversation in Twitter. We limit conversation within users in Indonesia geolocation area. In our experiment we investigate brand awareness of PT. Telekomunikasi Indonesia, which is the biggest Information and Communication Technology Company in Indonesia owned by government. "Telkom" brand has been labeled to many products and services. We would like to know how the brand "Telkom" perceived by market and which product or services that dominantly discussed in the market. Thus for the experiment, we use keyword "Telkom" as the label of conversation data on Twitter. The process of our experiment is shown in Figure 4.

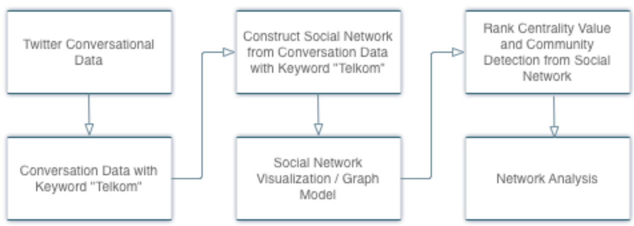

Figure 4. The process in the experiment

We crawl Twitter data for period 18 hours in December 4th, 2013 between 5.57 until 23.59. We get 2024 posts that mention keyword "Telkom" and 1728 among those posts are marked as conversational posts. In result we have graph model with 2024 nodes and 1728 edges. The whole network graph model is in Figure 5.

From Figure 5, we have a sparse network, the characteristics we expect from a social network. This network clearly has scale-free and preferential attachment based on very few nodes with high degree of connections. This implies that our network contains many communities / groups.

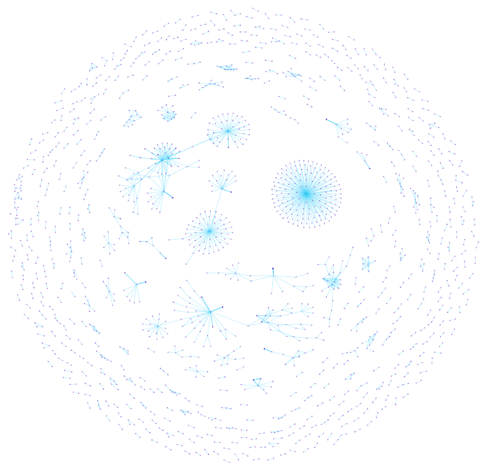

Figure 5. The graph model of twitter keyword "Telkom"

We apply community detection algorithm from [13] to identify how many communities in the network and how large their size. This algorithm using modularity metric for community quality and stopping criterion in iteration step. As we expected, we found around hundreds communities, with the majority is very small communities. We only interested in some biggest communities. We rank the communities based on their size. Table ll shows rank of biggest communties and Figure 6 shows map of communties where nodes with the same colors indicate they belong to the same communty.

TABLE II. COMMUNITIES RANK

No	Community Number	Number of Node	Size (Percentage)
1	7	153	7.55
2	1	133	6.57
3	0	79	3.90
4	9	78	3.85
5	41	55	2.71
6	62	37	1.82
7	72	34	1.67
8	35	28	1.38

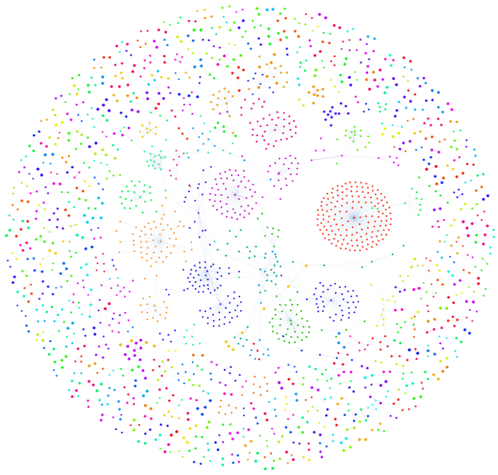

Figure 6. Map of communities, nodes with the same colors indicate they belongs to the same community

The biggest community is community number 7, which contains 153 posts and conversations amongs its member. We investigate their conversationss and it turns out that this community is about an event sponsored by PT. Telkom Indonesia. Second and third biggest talks about different events and activities in University Telkom, which is part of education foundation of PT. Telkom Indonesia. Fourth and fifth biggest talks about product and services of Telkom mobile services and Internet.

For centrality metric, we implement Brandes algorithm [14] to measure betweenness centrality. The summary of the results is shown at following Table lll. DC means Degree Centrality, and BC means Betweenness Centrality.

TABLE III. CENTRALITY RANK

No	Account	DC (value/rank)	BC (value/rank)	Community
1	@ouvalresearch	152 / 1	0.006 / 1	7
2	@detikcom	55 / 2	0.008 / 8	9
3	@infobdg	51 / 3	0.003 / 2	0
4	@EndankSoekamti	43 / 4	0.002 / 3	1
5	@SID_Merch	42 / 5	0.002 / 4	1

From Table lll, we see 5 biggest degree centrality nodes, with betweenness centrality value and its rank. It is clear that the biggest centrality nodes belong to the biggest communities. It shows the network is build around certain peoples or communities. This peoples or communities will attract more people to join according preferential attachment model.

In different case, the result could be different for some network. The result depends on their network formation or topology. The denser network will certainly give more complex computation. It could be nodes with highest degree centrality do not belong to the dominant community, in this case the roles of nodes in community approximately equal or we can say they have equal degree in conversations (two-way communication).

By understanding the measurement result and network formation, companies know what market network they are dealing with. They can identify number of communities and the dominant communites; it leads to the opportunity knowing their market segmentation. They can identify influence peoples in their networks; it leads to the opportunity to assign them as brand ambassador, others advertisement effort, and customer relationship management effort.

With this experiment we can answer first and second hypothese of the research. The first is the conventional way of spreading information using word-of-mouth can certainly be modelled by social network. The second is whether brand awareness process from unaware of the brand to top-of-mind can be achieved more effective and effiecient by understanding network formation that leads to social behavior of market. The answer is yes, for the reason by knowing number of communities, dominant communities and influential persons in network, we can save more resources to other efforts rather than blindly attack the market without knowing the specific target.

V. CONCLUSIONS

In this paper we have shown how brand awareness effort can be modelled using social network approach. Our contributions are the choice of online conversation data, which is cheap and pervasive, modelling those data into social network / graph model. The qualitative effort has to be made to interpret the metrics and measurement results. As stated in brand equity effort, brand awareness is not the only important piece; there are four more steps to be done in order to have succesful brand equity, but brand awareness is the fundamental base for other efforts. Our approach enrich effort in marketing study, and for many situations it will be the best choice considering its cheaper cost and shorter time to get data from the market.

This experiment gives coarse grain approach, it can be enhanced to get more details / accurate predicitons and patterns. Those enhancements can be done by one or more the following effort; using longer periods of data capture, using larger data involve in the computations, more metrics to compute, and different time periode to get dynamic network formations. In Big Data sense, we can combine different sources/varieties of data or different metric to get complete view about brand awareness phenomenon on our products.

ACKNOWLEDGMENT

NoLimitID who provide us with Indonesian Twitter data conversations to supports this work.

REFERENCES

[1] M.O. Jackson, Social And Economic Networks. Princeton University Press, 2008.

[2] D. Easley and J. Kleinberg, Network Crowd And Markets : Reasoning About Highly Connected World. Cambridge University Press, 2010

[3] P. Svenson. "Complex Networks And Social Network Analysis in Information Fusion". In *Fusion, page 1-7*, 2006

[4] M.E.J. Newman. *Network: An Introduction*. University Michigan and Sante Fe Insitute. Oxford University Press, 2011

[5] D.A. Aaker. *Managing Brand Equity: Capitalizing on the Value of a Brand Name*. New York: The Free Press, 1991

[6] D.A. Aaker. Measuring Brand Equity Accross Product and Markets. *California Management Review 38 (3)*, 102-120., 1996

[7] J. Scott. *Social Network Analysis: a Handbook*, Sage Publications 2000

[8] W. Zachary, "An Information Flow Model for Conflict and Fission in Small Groups". In *Journal of Anthropological Research,* 33(4) p. 452-473, 1977.

[9] R. Diestel. *Graph Theory: Electronic Edition 2005*. Springer-Verlag Heidelberg, New York 1997, 2000, 2005

[10] S. Wasserman, K. Faust. *Social Network Analysis: Methods and Application.* Cambridge University Press 1994

[11] M. Girvan. M.E.J. Newman. Finding and Evaluating Community Structure in Networks. *In Phys. Rev. E 69, 026113*. 2004

[12] Semiocast (2012, July 30). *Twitter reaches half billions accounts, More than 140 millions in US.* Available: http://semiocast.com/publications/2012_07_30_Twitter_reaches_half_a_billion_accounts_140m_in_the_US

[13] V.D. Blondel, J.L. Guillaume, R. Lambiote, E. Lefebvre. Fast Unfolding Communities in Large Networks. In *Journal of Statistics Mechanics : Theory and Experiment 10, P1000*, 2008

[14] U. Brandes. A Faster Algorithm for Betweenness Centrality. In *Journal of Mathematical Sociology 25(2):163-177*, 2001