Network Market Analysis using Large Scale Social Network Conversation of Indonesia's Fast Food Industry

Andry Alamsyah, Yahya Peranginangin
School of Economics and Business, Telkom University
Bandung, Indonesia
andrya@telkomuniversity.ac.id, yahyaperanginangin@telkomuniversity.ac.id

***Abstract*—The high competitiveness of the Indonesia Fast Food market has forced the industry to find the new way to understand market behavior. The new challenge should include faster data collection and analytical process, preferably time delivery needed close to real-time. The common practice of gathering market data using questionnaires and interviews are considered expensive and time-consuming process compared to mining online conversation with brand community respected. With the availability of large-scale data from online social network services (oSNS), we can extract valuable information represent dynamic behavior of the market. Many brands have their presence in oSNS as a part of their customer relationship management (CRM) effort. The social interactions formed in oSNS can be modeled using Social Network Analysis (SNA) methodology. In this paper, we compare two brand communities of head to head competitive product in the fast food industry, they are *McDonald's* and *Burger King*. The SNA model constructs large-scale network, its size, reaching close to a million of nodes and edges. The result will give us insight about what is important in understanding the dynamic market beside the market size represented by the community conversations.**

Keywords*—social network analysis; fast food industry; customer relationship management; community; market; online conversation; large-scale data; online social network services;

I. INTRODUCTION

The characteristic of Indonesia market shown by the formation of natural and artificial online communities built around the brands. The rise of Internet users and industry awareness of cheap and effective media for communicating with their market has increased the activities of online social network conversation. The contents of such conversations can be in the form of exchange experiences, products reviews, questions and answers, and other socializing-related activities. Most big industries present their channel in social media through popular social network services in Indonesia such as *Facebook, Twitter* and *Youtube* as part of their customer relationship management (CRM) effort, which ultimately leads to brand promotions and repurchase products.

The more complex CRM effort, called social CRM have objectives such as the capabilities to listen what the market wants, to understand market competition, and to understand market segmentation [1]. This paper will show how a business organizations mine online conversational in *Facebook* related to their brand issue and analyse them in the context of CRM to extracrt several insight regarding their network market.

Fast food industries are integral part of Indonesia society life style reflected by its large market size. It is strongly present in urban area, where also the basis of Indonesia Internet activities. We argue that mining market information through social media activities can represent the real behavior. The common practice of gathering market data using questionnaires and interviews are considered expensive and time-consuming process. With the availability of large-scale data from online activities, the idea of mining network market from online community activities can be done [2][3].

The Fast food industry is rarely in touch with the development of ICT outside their internal information management system. This is due to the lack of knowledge on how to mine and analyze online community information into business related predictions. The abilities to listen and understand voices from social media are strategically crucial to view the scale of a multidimensional aspect of their market, whether as main resources or as an additional information [3][4].

To increase community member participation in discussions, we need to provide topics and contents that generate lots of information from the audience. Designing good content also crucial to maintain engagement value from the community member.

In this paper, we compare the large-scale community network between two head-to-head competitors in the fast food industry in Indonesia. The brands are *McDonald's* and *Burger King*, both serves the similar menu, and often considered as a rival by their own customers. We analyze their official communities page in *Facebook* using SNA methodology. The first research question is which one of those two communities is more engaging by measure engagement value of the contents discussed in the page. The second question is which communities are more robust by measure their respective social network metrics.

II. SOCIAL NETWORK AND COMMUNITY ANALYSIS

The social network modeled by graph theory consists of nodes and edges. The nodes represent actors and the edges represent relations between actors. The advantages of model the network using graph theory are intuitive, easy to visualize and provide several metrics based on graph theory characteristics. The model and set of metrics are called *Social Network Analysis* (SNA) [5][6].

Community measured by the natural interaction pattern between the nodes. The decision whether a node is a part of one particular community or other communities are depending on which community detection algorithm is used. Community measurement is considered to be *NP-Hard* problem in computation. One of the most sophisticated measurements to detect community is *Modularity*. In practice, there are many possible sub-groups inside global community. By knowing which sub-groups are dominant, we can strategically apply decision whether to encourage sub-group development or not.

A. Metrics

We use several metrics below to compare the two network observed:

- *Average Degree* is the average connection of all nodes with their respective neighboring nodes. A higher average degree means construct a dense network and increase probability of shortcut paths between the nodes, thus it follows *scale-free* and *small-world* properties of social network [7].
- *Network Density* is defined by the ratio of the actual numbers of edges E to the maximum number of possible edges. The dense network is more desirable, since it gives the alternative route for traversing information and increase network strength. Several dense part of the network indicates a presence of community. Density formulated as:

$$D = \frac{2|E|}{|N|(|N|-1)} \tag{1}$$

  D is the density, E is the number of edges and N is the number of nodes. The D value is between [1,0]

- *Average Path Length* is defined as the average of the distance along the shortest path for all pairs of nodes in the network. The smaller distance represents the efficiency of information to travel
- *Network Diameter* is defined as the longest of all calculated shortest paths in a network. This metric represents the furthest / worst scenario distance needs to be taken for information to travel.
- *Modularity* is used to detect communities in a network. This value computes the actual number of edges inside a community and the expected value of two nodes in random network fall into the same community [8]. This metric work similar to generic probabilistic heuristic when decide whether a node is a member of community or not. It is the formulated as:

$$Q = \frac{1}{2m}\sum_{ij}(A_{ij} - \frac{k_i k_j}{2m})\delta(c_i, c_j) \tag{2}$$

  Q is modularity value represent the network partition quality, with value between [-1/2,1], where m is number of edges, A_{ij} is the adjacency value between node i and node j, k_i and k_j is degree of node i and node j.

B. Bipartite Networks

According to graph theory, a bipartite network is a network whose nodes can be divided into two disjoint sets U and V, where U and V are each independent set, such that every edge connects a node in U to one in V. We apply this formulation to our topic, where the two sets of nodes are sets of users who respond to contents or posts in the community wall and sets of posts who generate responds or conversation among the users. A visualization of bipartite network is shown at Figure 1 [9].

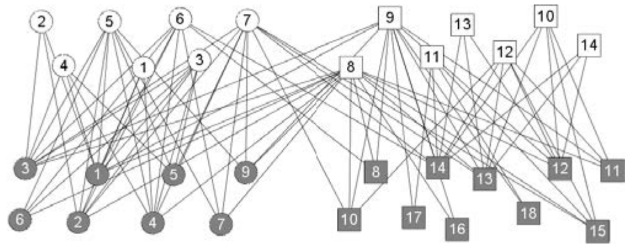

Fig. 1. Example of bipartite network taken from [Liu]

Fig. 2.

C. Engagement Value

In order to quantify social interactions, we propose a metric called *engagement value*, based on users activity, which spark engagement in the network [10]. There are three kinds of interactions between users and posts in *Facebook*; they are comments, shares and likes. Based on how *Facebook* works, the more users make comments, sharing and likes a post will increase the probability that the post will disseminate across a *Facebook* user's timeline, thus it generates more conversations.

We define post engagement value based on total number of comments, shares and likes that a post has. The formal definition is:

$$EV = \sum c_i + s_i + l_i \tag{3}$$

,where EV is a post engagement value, c_i is total number of comments, s_i is total number of shares and l_i is total number of likes.

III. Network Comparison and Discussion

We crawl data on *McDonald's* (MD) and *BurgerKing* (BK) Indonesian official *Facebook* pages. The data descriptions is as follows: The MD page as per January 2015 have 55,423,739 likes, while BK page has only 107,102 likes. The MD bipartite network of posts and users, consist of 438,568 nodes and 909,748 edges, while BK network consist of 220,289 nodes and 266,337 edges.

We use *R language* to compute the metrics and *Gephi* to visualize the network. However due to large-scale data, we limit our comparison to only 5 indidual metrics; *Average Degree, Network Density, Average Path Length, Diameter, Density* and *Modularity.* The computation results shown in Table 1 and the value ratio between both networks is in Table 2. Further network information regarding the engagement value and user sex shown in Table 3. Figure 2 and Figure 3 show maps of global MD and BK network by post type. The node data structure of our crawling data is shown in Figure 4.

By number of members, or number of likes MD network is much higher than BK network. By number of nodes BK network is 0.5 times MD network. We can see even though smaller in size, BK network generated more conversations, with the ratio between the number of members and number posts and members involved in conversation reach 0.46, while the same ratio in MD network is 126.37. The huge difference in ratio shows us that members BK network is much more active and form close community network [11].

TABLE I. SNA METRICS COMPARISON

No	Metric	MD Network	BK Network
1	Size (Number of Likes)	55,423,739	107,102
2	Size (Bipartie Network)	438,568 nodes 909,748 edges	220,289 nodes 226,337 edges
3	Average Degree	2.418	0.116
4	Density	0.00001097725	0.000003890513
5	Average Path Length	3.42506	3.245614
6	Diamater	8	8
7	Modularity	0.670 (13 communities)	0.527 (10 Communities)

TABLE II. RATIO METRICS BETWEEN BK AND MD NETWORK

No	Metric	Ratio BK / MD Network
1	Size (Number of Likes)	0.0019
2	Size (Bipartie Network)	Nodes: 0.5022 Edges: 0.2488
3	Average Degree	0.0478
4	Density	0.3544
5	Average Path Length	0.9476
6	Diamater	1
7	Modularity	0.7865

TABLE III. Further Information About The Networks

No	Description	MD Network	BK Network
1	Post type	99.7% user comments, the rest are video, status, photo, link	99.55% user comments, the rest are video, status, photo, link
2	5 highest engagement value	137609, 132372, 113299, 99668, 97973	129423, 25625, 25053, 21295, 17046
3	Post type of 5 highest engagement value	Photo, status, status, link, photo	All photo
4	User sex	55.27% male, 41.50% female	55.68% male, 43.83% female

Another topology metric is average degree, MD network have significantly bigger value than BK network, it means more posts or more user comments are replied / shared / liked by other members, thus it creates more new connection. Diameter and average path length are almost the same between both networks, this means regardless the network size, both network have a same network topology with different scale. MD network has a slightly denser network, this shows the opportunity of new connections between members and post is higher than BK network [12]. In case of modularity metric, MD network has slightly the tendency to form communities, it shows that MD network form more communities than BK network.

From Table 3, we see in both networks that user comments dominate all the conversation comparing page posts, which consists of photo, video, status, and link. This is as expected in an active social network as it follows *scale-free, small-world* and *preferential attachment* rules [13]. Male users dominate the conversation above female users in both networks.

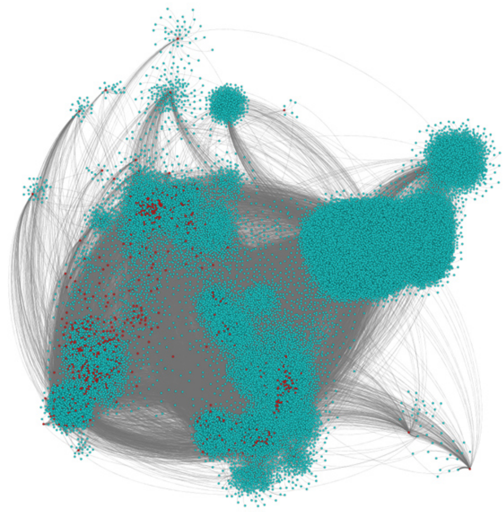

Fig. 3. Large scale MD network, green colour are user posts, while other colour are page posts.

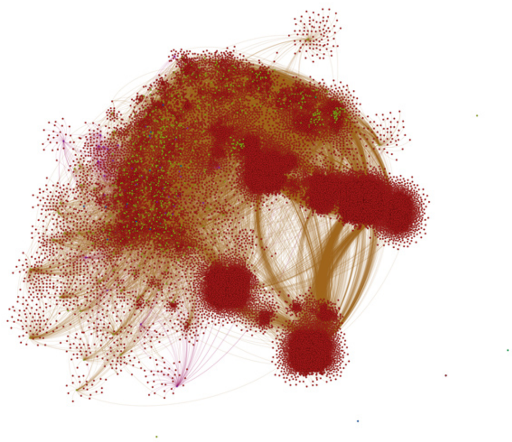

Fig. 4. Large scale BK network, red colour are user posts, while other colour are page posts.

CONCLUSION

Market size is not the only factor that guarantees the success of brand campaign / awareness. MD network size that has hundreds times bigger than BK network turn out to have similar network properties (robustness) and engagement value. These properties support network mechanism to attract new member to join the network. Although smaller in size, BK network will grow faster while it is hard for the MD network to keep its growth when its size has already very big. It is shown that some types of page posts generate higher engagement value, this pattern can become guidance in order to grow market network size.

We have analyzed two large-scale market network data in the Indonesia fast food industry. The network resulted from social interactions give us the way to observe the underlying complex process in a dynamic market, which looks like randomly changing each time, but actually follows common social network properties, such as *preferential attachment*, *scale-free* and *small-world* [13]. The current practice of understanding market are dominated by size and frequency. Set of SNA metrics helps to explain issues such as *disemminating information, community detection, tie-strength,* and some others.

Due to limitations on our computer power, we cannot do more analytical process, thus this research can be developed further by using a more powerful computer using several approaches such as; 1). Approach to deeper analytical process in *community detection* 2). Approach by including more SNA metrics such as *centrality, global and local clustering, components, structural holes* 3). Approach by include node properties and edge weights in metric computations. Using the same data, we also can build model (learning) based on data mining algorithm such as classification, clustering, association and some others to predict market behaviour.

Nodes	type_p...	post_publis...	likes	likes_coun...	comment...	comments...	comment...	comment_l...	shares	engagem...	post_id
Lengkapi harimu dengan men	photo	2014-10-02T0	128859	131875	400	362	38	105	59	129423	119637636283
Ada yang HOT di Burger King!	photo	2014-04-03T0	24247	25449	883	786	97	230	265	25625	119637636283
BK Lovers_ yuk ikutan BK Snac	photo	2014-11-03T0	24945	25201	69	53	16	21	18	25053	119637636283
Berani menjawab HOT CHALL	photo	2013-06-05T1	18054	18980	1707	1421	286	605	929	21295	119637636283
Buruan cobain paket BK Bukb	photo	2014-07-07T0	14219	14503	792	679	113	1746	289	17046	119637636283
Menu terbaru dari Burger King	photo	2014-09-02T0	8461	8712	152	143	9	42	79	8734	119637636283
Treat Your Friend Only Rp. 5.(	photo	2014-09-10T0	2562	2626	207	135	72	17	92	2878	119637636283
user_119637636283	user		58		824	55	769			882	
Go Vote! Bring a friend! Show	photo	2014-07-08T1	200	201	66	46	20	1	152	419	119637636283
BIG BREAK TIME dengan man	photo	2014-01-16T0	253	258	64	52	12	6	36	359	119637636283
Masih awal bulan mau yang h	photo	2014-09-08T0	153	153	49	38	11	2	118	322	119637636283
Lapar dan malas keluar? Telp	photo	2014-07-21T1	193	198	65	45	20	1	35	294	119637636283
Buat BK Lovers yang lagi ngan	photo	2013-07-26T0	187	192	28	27	1	17	34	266	119637636283
Banyak banget yang bikin kan	photo	2014-09-18T1	206	208	18	18	0	4	25	253	119637636283
NEW FROM BURGER KING!!!	photo	2013-07-15T1	104	108	29	27	2	9	107	249	119637636283
Yang lagi liburan kumpul den	photo	2014-07-30T0	168	170	54	34	20	5	17	244	119637636283
Wooho kabar menyenangkan	photo	2013-10-14T0	115	118	31	22	9	6	80	232	119637636283
user_100003697717801	user		217		1	1	0			218	
Yang belum makan siang_ Ay	photo	2014-11-06T0	189	197	22	13	9	2	1	214	119637636283
user_690094619	user		212		0	0	0			212	
Sekarang saat yang tepat unt	photo	2014-09-09T0	122	123	47	30	17	4	29	202	119637636283
Pengen makan siang tapi duit	photo	2013-06-07T0	87	91	59	42	17	14	42	202	119637636283

Fig 4. Node data structures from crawling process

REFERENCES

[1] A.Alamsyah, Y. Peranginangin. Learning Organization using Conversational Social Network for Social Customer Relationship Management Effort. *2nd International Conference and Seminar on Learning Organizations, 2014*

[2] D. Birke. *Social Network and their Economics : Influence Consumer Choice.* John Wiley & Sons, 2013

[3] A. Alamsyah, F. Putri, O.O.Sharif. Social Network Modelling Aprroach for Brand Awareness. *2nd International Conference on Information and Communication Technology, 2014*

[4] D. Lazer, A. Pentland, L. Adamic, S. Aral, Barabasi. A.L, Brewer. D, Christakis. N, Contractor. N, Fowler. J, Gutmann. M, Jebara. T, King. G, Macy. M, Roy. D, Van Alstyne. M. Computational Social Science, *In Science. Vol. 323. No. 5915, pp. 721-723, 2009*

[5] R. Zheng, Q. Sheng, L. Yao. A Simulation Methods for Social Networks. *Journal of Social Network Analysis and Mining, Vol. 5, Issue 14, 2015*

[6] M. Goldberg, S. Kelley, M.M. Ismail. Communication Dynamics of Blog Network. *Advances in Social Network Mining and Analysis and Mining : Second International Workshop, SNAKDD, 2008*

[7] G. Caldarelli. *Scale-Free Networks: Complex Webs in Nature and Technology.* Oxford University Press, 2007

[8] V. Blondel, J.L. Guillaume, R. Lambiotte, E. Levebvre. *Fast Unfolding of Communities in Large Networks.* Journal of Statistical Mechanics: Theory and Experiment. Retreived from Cornel University Library database, 2008

[9] X. Liu, T. Murata. Community Detection in Large-Scale Bipartite Networks. *International Conference on Web Intelligence and Intelligence Agent Technology, 2009.*

[10] K. Seddon, N.C. Skinner, K.C. Postlethwaite. *Creating a Model to Examine Motivation for Sustained Engagement in Online Communities.* Journal Education and Information Technologies, Volume 12, Issue 1, pp 17-34, 2008

[11] M.E.J. Newman. *Network: An Introduction.* University Michigan and Santa Fe Institue. Oxford University Press, 2011

[12] C.C. Aggarwal. *Social Network Data Analytics*. Springer, 2011

[13] U. Brandes, T. Erlebachs. *Network Analysis: Methodological Foundations*. Springer, 1998