Study on Relation between Social Circles and Communities in Facebook Ego Networks

Soo-jin Shin[1], Yong-jin Jeong[1], Chan-Myung Kim[1], Youn-Hee Han[1,*], and Chan Yeol Park[2]

[1] Advanced Technology Research Center,
Korea University of Technology and Education, Korea
{soojin1116,jyjin989,cmdr,yhhan}@koreatech.ac.kr
[2] Supercomputing Center,
Korea Institute of Science and Technology Information,
Daejeon, Republic of Korea, 305-806
chan@kisti.re.kr

Abstract. Community detection is a core problem in social network analysis. Strictly speaking, however, the communities does not exactly correspond to the real group, well-known as social circles. In this paper, we study on 1) how close relation between the ground-truth social circles and communities exists and 2) whether the social circles can be detected by the classical community detection algorithm or not. We use the SNAP facebook dataset to reveal the correlation between the social circles and the detected communities. We listed up the community's modularity values and the balanced accuracy values with the ground-truth circles per each level in the iterative process of divisive clustering. We analyzed the Spearman's rank correlation between the paired data. The experimental results show that there is a strong correlation between the ground-truth social circles and the communities detected by classical method.

1 Introduction

Nowadays, there are numerous online social networks services, e.g., Facebook, Twitter, Google+, etc. They form human social networks by generating new connections between nodes. In a social network, each node usually denotes a user and edge denotes a connection with a friend or an acquaintance [1].

Most social networks services also allow users to categorize their friends into social circles, e.g., 'circles' on Google+, and 'lists' on Facebook and Twitter. Currently, users usually manage their circles either by manually identifying friends sharing a common attribute. This method is time consuming and does not update automatically as a user adds more friends into his/her social network. In addition to that, it cannot manage users social circles finely when a user's profile information is missing [2]. Therefore, the problem of automatically discovering users social circles has become an important issue.

* Corresponding author.

Y.-S. Jeong et al. (eds.), *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering 280,
DOI: 10.1007/978-3-642-41671-2_72,

One of other important issues in social network analysis is the detection of community structure in social networks [3]. The most widely used and accepted definition of a community is follows: “a group of nodes of a graph which are more strongly connected to each other than with other nodes in the same graph.” Many approaches to detect communities in a social network have been proposed in the past [1]. Among them, we use the hierarchy-centric divisive clustering method that is to build a hierarchical structure of communities based on network topology [4].

In this paper, we study the relation between communities and social circles. As mentioned above, the communities are detected by the given algorithm, particularly hierarchy-centric divisive clustering algorithm, while social circles are manually managed by users identifying friends sharing a common profile attribute. If there is a strong correlation between them, we may be able to solve the social circle detection problem by using the classical community detection methods.

The rest of this paper is organized as follows. Section 2 presents preliminaries and general approach for analyzing the correlation between social circles and communities. Section 3 shows an experimental analysis, and Section 4 finally concludes this paper.

2 Relation between Social Circles and Communities

2.1 Preliminaries

In this subsection, we present the used notions for analyzing the correlation between social circles and communities.

To compare the social circle with identified community, we first use the hierarchy-centric divisive clustering algorithm that is to build a hierarchical structure of communities based on network topology [4]. More specifically, it first partitions the nodes into several disjoint sets. Then each set is further divided into smaller ones until each set contains only one node. The key here is how to split a network into several parts, and one particular divisive clustering algorithm is to recursively remove the “weakest” tie in a network. that is, The weakest tie, the higher value of edge betweenness. Newman and Girvan [5] proposed to find the weak ties based on “edge betweenness”. Edge betweenness is a measure to count how many shortest paths between pair of nodes pass along the edge, and this number is expected to be large for those between-group edges.

For calculating the correlation between social circles and identified communities by the divisive clustering algorithm, we use the community modularity, the balanced accuracy and Spearmans rank correlation coefficient. Modularity [3] is usually used to measure the strength of a community partition by taking into account the degree distribution of nodes. Accuracy [6] considers all the possible pairs of nodes and checks whether they reside in the same community and social circles. It is considered error if two nodes of the same community are assigned to different social circles, or two nodes of different communities are assigned to the same social circle. Balanced

Accuracy [6] can be defined as the average accuracy assigning equal importance to false positives and false negatives. Spearmans rank correlation coefficient assesses how well the relationship between two variables can be described using a monotonic function ((i.e., that when one number increases, so does the other, or vice-versa). A perfect Spearman correlation of +1 or 1 occurs when each of the variables is a perfect monotone function of the other.

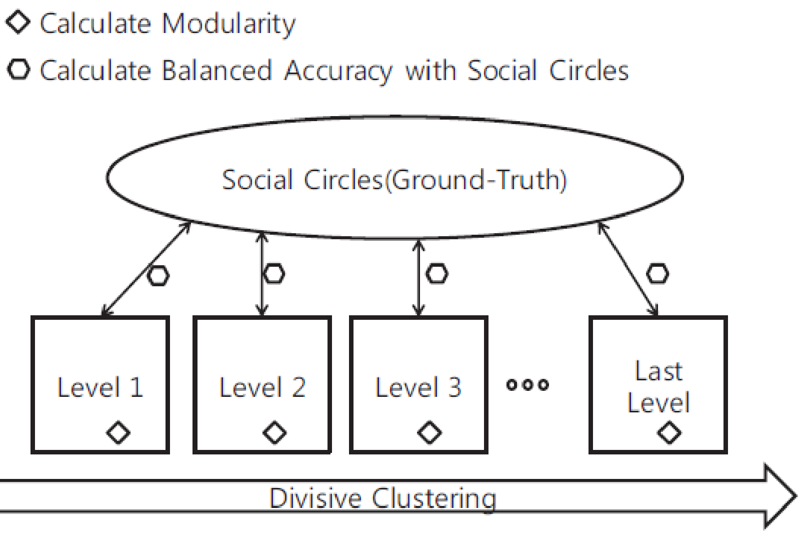


Fig. 1. Experimental Process

2.2 Social Circle and Community

It would be nice if there were a direct link connecting social circle and community. The general approach is simple. If there is a strong correlation between them, we may be able to solve the social circle detection problem by using the classical community detection methods. For finding this correlation, the social circle for each user can be known. Unfortunately, this is not possible in general. It is not easy to see a scenario that hardly presents itself in real-world large-scale networks. For the study, so, we use real Facebook ego network datasets provided by SNAP (Stanford Network Analysis Project) [4].

The figure 1 shows the experiment process for study. In advance of analysis, First we generate an ego network for each node based on Facebook ego network datasets. An ego network consists of a local node ("ego") and the nodes ("alters") to whom the ego is directly connected to plus the edges, if any, among the altars [7]. For each ego network dataset, we first 1) executed the community detection by using the hierarchy-centric divisive clustering algorithm, 2) listed up the communitys modularity and the balanced accuracy between community and the social circles (i.e., ground truth) per each level in the iterative process, and 3) analyzed the Spearmans rank correlation between the two datasets, i.e., the modularity and the balanced accuracy.

3 Experimental Results

We conducted the above experiment on a Facebook ego networks. We randomly choose the seven ego networks with 0, 348, 414, 686, 690, 3437 and 3980 node. For each ego network we increase the level of divisive clustering from 0 to 2500.

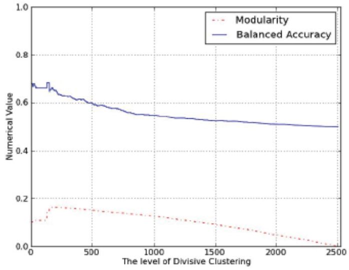
Figure2 depicts the two values, the modularity and the balanced accuracy, per the modularity and the balanced accuracy. In figure2, the level of Divisive clustering is plotted along the X axis, and the numeric value is plotted along the Y axis. As can be seen from the figure2, in general, the two values are highly correlated, meaning that there is high strength of a linear relationship between the paired data. For getting exact value of correlation coefficient we also obtained the value of Spearman correlation. The results are shown in Table 1 and the average of Spearman's Rank is 0.852.

In case of 348 and 3980 node, Spearman's Rank is relatively low compared with other node's cases. The most widely used and accepted definition of a community is follows: "a group of nodes of a graph which are more strongly connected to each other than with other nodes in the same graph". However, there are many nodes weakly connected to each other than with other nodes unlike above definition in those cases. Those weak ties make Spearman's Rank low. Actors in real worlds tend to form closely-knit social circles. That is, individuals interact more frequently with members within social circles than those outside the social circles. Therefore, it is very unusual cases in a real worlds.

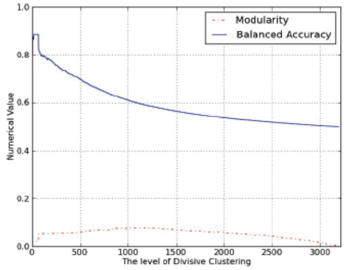
These results seem to confirm our suspicions. Clearly, we need to do more extensive testing of this hypothesis but at this stage it seems reasonable to conclude that these is evidence that classical community detection method can be used to detect social circles.

Table 1. Results of Spearman's Rank obtained in our experiment

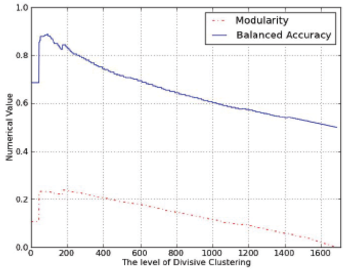
Ego network	Spearman's Rank
0	0.922
348	0.570
414	0.986
686	0.950
690	0.987
3437	0.981
3980	0.571
Average	0.852



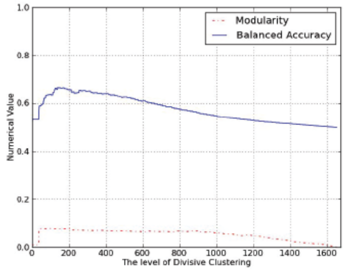
(a) 0.Ego Network Result



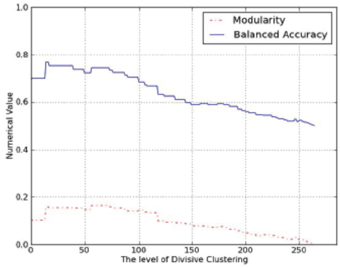
(b) 348.Ego Network Result



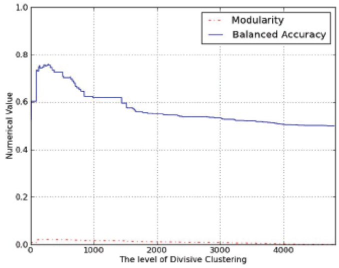
(c) 414.Ego Network Result



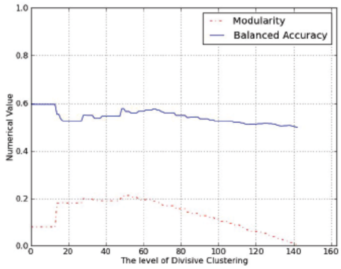
(d) 686.Ego Network Result



(e) 698.Ego Network Result



(f) 3437.Ego Network Result



(g) 3980.Ego Network Result

Fig. 2. Results of our experimental analysis on Facebook ego networks

4 Conclusions

In this paper, we study the relation between the ground-truth social circles and the communities are detected by the classical community detection Title Suppressed Due to Excessive Length 5 algorithm. From the experimental results, we obtained an average value of 0.852 of Spearman's rank correlation between them. While social circles are manually managed by users, we can conclude that classical community detection method can be used to detect social circles. However, the details on how to apply the algorithms to detect social circles are still open issue. We just disclosed the possibility of it. In future works, we will conduct experiments on various networks by using many community detection algorithms.

Acknowledgments. This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2013R1A1A2010050), and also financially supported by the Ministry of Knowledge Economy (MKE) and Korea Institute for Advancement of Technology (KIAT) through the Workforce Development Program in Strategic Technology.

References

1. Tang, L., Liu, H.: Community Detection and Mining in Social Media. Morgan & Claypool Publishers (2010)
2. McAuley, J., Leskovec, J.: Learning to Discover Social Circles in Ego Networks. NIPS (2012)
3. Newman, M.: Modularity and community structure in networks. PNAS 103(23), 8577–8582 (2006)
4. Karypis, G., Han, E., Kumar, V.: Chameleon: A hierarchical clustering algorithm using dynamic modeling. IEEE Computer 32(8), 68–75 (1999)
5. Newman, M., Girvan, M.: Finding and evaluating community structure in networks. Physical Review E (2004)
6. Brodersen, K.H., Ong, C.S., Stephan, K.E., Buhmann, J.M.: The balanced accuracy and its posterior distribution. In: Proc. of the 20th International Conference on Pattern Recognition (August 2010)
7. Marsden, P.V.: Egocentric and sociocentric measures of network centrality. Social Networks 24(4), 407–422 (2002)