겹침 비율이 커뮤니티의 수에 미치는 영향에 대한 연구

수다르산 판트, 이상돈
목포대학교 멀티미디어공학과
e-mail : {darshanz,sdlee}@mokpo.ac.kr

A study of the impact of the overlapping fraction on the number of communities

Sudarshan Pant, Sangdon Lee
Dept. of Multimedia Engineering, Mokpo National University

ABSTRACT

Several approaches have been proposed for overlapping community detection including global as well as local method. In the local seed expansion based methods, seed generation plays a crucial role in the accuracy of the community detection algorithms. While the accuracy is measured by different papers using different datasets, a common method of comparing the results is very important. The LFR benchmark [1] is a widely accepted method for evaluation of the community detection algorithms. In this paper, we study the impact of the overlapping fraction on the number of communities generated in LFR method.

1. Introduction

Community detection is an important research area mainly due to its application in analyzing complex networks. Detection of overlapping communities reveals the relationship among different communities. The objective of overlapping community detection is to find the clusters with shared nodes in a graph. Fig. 1 shows an example of the segment of the graph with overlapping communities where the circles represent the communities.

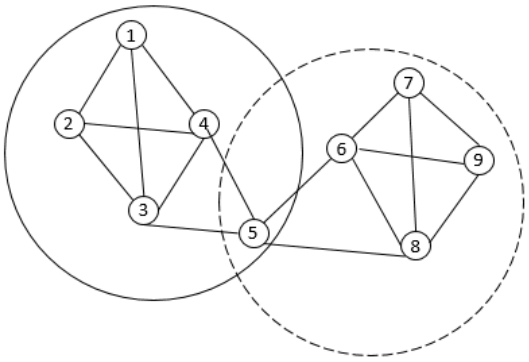

Fig. 1. Example graph with overlapping communities

Among the local expansion method, Vertex Influence Similarity Method(VISM) [2] has been proposed to have higher accuracy as compared to the state-of-the-art approaches. The results of the VISM suggest that the accuracy increases with the increase in the overlapping fraction of the communities. However, the reason behind that is unclear and the number of ground truth communities has not been mentioned. In this paper, we study the impact of the change in an overlapping fraction in the number of ground truth communities in LFR benchmark.

2. Related Work

In the past, several local methods have been proposed for overlapping community detection. Among the local methods, seed expansion methods such as NISE (Neighborhood-inflated Seed Expansion) [3] and VISM (Vertex Influence Similarity Method) focus on the local structure by selecting a set representative nodes referred to as seeds and expanding a community around them. The NISE method consists of four phases namely, filtering phase, seeding phase, seed expansion phase, and propagation phase. Removing the trivially separable nodes in the beginning and attaching them to the detected communities after detection makes this method computationally less expensive. Despite being a local expansion method, this approach has been found to be expensive for large networks in terms of time. The VISM method introduced the vertex influence based method for

seeding and thus achieves high accuracy and low computation time even in larger networks. Although the LFR benchmark method has been used to compare the results with the NISE method, the count of ground truth communities has not been mentioned.

3. Experiment

To study the impact of overlapping percentage on the generated communities, we generated a dataset using the LFR method with similar parameters as used by VISM. We, then, implemented the vertex influence based seed generation method to generate the seeds.

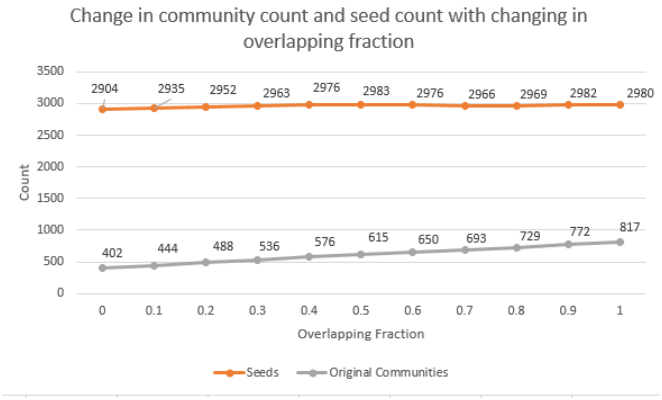

Fig 2. Comparison between generated seeds and communities with a change in the overlapping fraction

Figure 2. shows the number of ground truth communities and the seed count as obtained by vertex influence based method for seed generation.

4. Conclusion

In this paper, we studied the impact of change in an overlapping fraction in the number of communities in LFR benchmark dataset generation. With the increase of overlapping fraction, the number of ground truth communities increase. The future work includes the development of an efficient method for overlapping community detection.

참고문헌

[1] A. Lancichinetti, S. Fortunato, "Benchmarks for testing community detection algorithms on directed and weighted graphs with overlapping communities," *Physical review. E, Statistical, nonlinear, and soft matter physics*, Vol. 80, pp. 016118, 2009

[2] Y. Gao, H. Zhang, Y. Zhang, "Overlapping communities from lines and triangles in complex networks," *Physica A*, Vol.521, pp. 4551-466, 2019.

[3] J.J. Whang, D.F. Gleich, I.S. Dhillon, "Overlapping Community Detection Using Seed Set Expansion," *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, Vol. 51, pp. 2099-2108, 2013.