

Detection of overlapping communities in a friendship network

Yoshie OTSUKA and Kazue KUDO

Department of Comp. Sci., Ochanomizu University, 2-1-1 Otsuka, Bunkyo-ku, Tokyo 112-8610 Japan

E-mail: otsuka.yoshie@is.ocha.ac.jp

Keyword: Twitter, Community, Network

1 Introduction

Recently a lot of people are using Social Networking Services (SNS). In many cases, people have connections on SNS with their real friends. Most of SNS suggest “Recommended User” who may be your friend. However, sometimes, the recommended user may not be your real friend or acquaintance. In this study, we make a network from following-follower relationships in Twitter and analyze the network by using an algorithm of community detection.

2 Methods

In a real social network, a person belongs to more than one communities such as family, co-workers, classmates, club team, and hobby friends. In this study, we use the Newman’s algorithm [1] which can detect overlapping communities. This algorithm is an EM algorithm to infer link communities in a network. We first detect overlapping communities by using this algorithm. Next, we evaluate how proper the result is and how likely it is to have potential friends in the community. Then, we optimize the number of communities by calculating modularity [2].

3 Results

We make a network from a following-follower relationship in Twitter. This network has 213 nodes and 2,610 edges. The clustering coefficient of the network is 0.502, and the mean distance is 1.884. This indicates that the network is a small-world network. We detect communities by using the algorithm and find that the modularity is high when the network is divided into 3 or 7 communities. Next, we evaluate “correct answer rate” and “friend possibility”. The correct answer rate indicates whether the members of a community are correctly assigned, compared with the real relationship. The friend possibility is the proportion of unlinked pairs that have a real relationship among all the pairs in a community. We find that the correct answer rate and the friend possibility have positive correlation. Since both of them depend on how proper the communities are partitioned, we improve the algorithm by optimizing the number of communities by calculating modularity.

References

- [1] B. Ball, B. Karrer and M. E. J. Newman, “Efficient and principled method for detecting communities in networks”, *Phys. Rev. A* **84**, 043607 (2011).
- [2] M. E. J. Newman, and M. Girvan, “Finding and evaluating community structure in networks”, *Phys. Rev. E* **69**, 026113 (2004).