Detectability threshold in the spectral method for graph partitioning

Tatsuro KAWAMOTO[1] and Yoshiyuki KABASHIMA[1],

[1]Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology, 4259-G5-22, Nagatsuta-cho, Midori-ku, Yokohama, Kanagawa, 226-8502, Japan

E-mail: [1]kawamoto@sp.dis.titech.ac.jp

Abstract

Graph partitioning is one of the important tools for the investigation of structure in real data. While there exist many methods and objective functions to optimize, the spectral algorithm is applicable to several major methods and the analytical study for its performance is tractable. We consider a graph of two loosely connected clusters, each of which is consists of a regular random graph, and calculate the so-called detectability threshold in the spectral method. That is, the method cannot perform clustering above that threshold.

Keyword: Community detection, Spectral method, Detectability

The fundamental goal of graph partitioning is to find densely connected subgraphs, or communities, in a graph data. While detecting the community structure in a graph is a mathematically interesting problem, it is also an important task. It is often the case that such densely connected subgraphs are not just mathematical objects, but are physically meaningful. In the case of a social network, for example, such subgraphs are expected to indicate social groups. Motivated by these reasons, a number of methods for graph partitioning (and community detection) have been developed by computer scientists and physicists [1].

While the methods of graph partitioning is widely applicable because the only ingredients are the adjacency matrix and the number of modules [2], in many cases, we are not very sure about the result we obtain, or the result may not be the partition we really wished to see. It happens for a variety of reasons. It may be due to the **NP**-completeness of the problem that we cannot reach the true optimum by an approximation algorithm, the theoretical restriction that the method has, or simply that the desired communities are densely connected subgraphs in a different sense. We do not face to the problem of **NP**-completeness here and assume that we wish to obtain the partition in the sense of spectral clustering [3]. We here investigate the theoretical restriction called the *detectability threshold* in a method of spectral clustering. That is, the method cannot perform clustering above that threshold.

For that purpose, we consider the following synthetic random graph. We consider N nodes, each of which has degree κ and belongs to either of two clusters so that the sizes of clusters are $p_1 N$ and $p_2 N$ ($p_1 > 0$, $p_1 + p_2 = 1$). The connection between nodes are determined at random under the constraint that total number of links between the clusters is γN ($0 < \gamma < 1$). For this graph, we calculate the second-smallest eigenvalue and eigenvector of the graph Laplacian using the replica method [4], which gives the exact solution in $N \to \infty$. As a result, we see that a phase transition occurs at a certain value of γ. It means that, above this parameter, the method of spectral clustering is unable to detect clusters correctly. Since this problem had been considered by several authors with various approaches [5, 6, 7], we will also discuss how the present result is related to the others.

References

[1] S. Fortunato, *Phys. Rep.*, **486**, 75 - 174 (2010).

[2] In the case of community detection, the number of modules is determined automatically, and this is the very difference between graph partitioning and community detection.

[3] U. Luxburg. *Statistics and Computing*, **17**, 395 - 416 (2007).

[4] Y. Kabashima and H. Takahashi, *J. Phys. A*, **45**, 32 (2012).

[5] A. Decelle, F. Krzakala, C. Moore, and L. Zdeborová, *Phys. Rev. Lett.*, **107**, 065701 (2011).

[6] R. R. Nadakuditi and M. E. J. Newman, *Phys. Rev. Lett.*, **108**, 188701 (2012).

[7] F. Radicchi, *Phys. Rev. E*, **88**, 010801 (2013).