

How many communities are there in a network?

Shujie Ma

University of California, Riverside

ABSTRACT

Advances in modern technology have facilitated the collection of network data which emerge in many fields including biology, bioinformatics, physics, economics, sociology and so forth. Network data often have natural communities which are groups of interacting objects (i.e., nodes); pairs of nodes in the same group tend to interact more than pairs belonging to different groups. Community detection then becomes a very important task, allowing us to identify and understand the structure of a network. Thus, the development of methods for community detection has attracted much attention in the past decade, and as a result, different efficient approaches have been proposed in literature.

A fundamental limitation of most existing methods is that they divide networks into a fixed number of communities, i.e., the number of communities is known and given in advance. However, in practice, such prior information is typically unavailable. Determining the number of communities is a challenging yet important task, as the following community detection procedure relies upon it. In this talk, I will introduce a convenient and effective solution to this problem under the degree-corrected stochastic block models (DC-SBM). The proposed method takes advantages of spectral clustering, likelihood principle and binary segmentation. Determining the number of communities is essentially a model selection problem, and we therefore establish the selection consistency of our proposed procedure under a mild condition on the average degree. We demonstrate the approach on different networks. At the end of my talk, I will briefly talk about our other on-going and future research projects in this line of work.