Statistical Inference on Membership Profiles in Large Networks

Jianqing Fan

Department of Operations Research and Financial Engineering

Princeton University

ABSTRACT

Network data is prevalent in many contemporary big data applications in which a common interest is to unveil important latent links between different pairs of nodes. The nodes can be broadly defined such as individuals, economic entities, documents, or medical disorders in social, economic, text, or health networks. Yet a simple question of how to precisely quantify the statistical uncertainty associated with the identification of latent links still remains largely unexplored. In this talk, we suggest the method of statistical inference on membership profiles in large networks (SIMPLE) in the setting of degree-corrected mixed membership model, where the null hypothesis assumes that the pair of nodes share the same profile of community memberships. In the simpler case of no degree heterogeneity, the model reduces to the mixed membership model and an alternative more robust test is proposed. Under some mild regularity conditions, we establish the exact limiting distributions of the two forms of SIMPLE test statistics under the null hypothesis and their asymptotic properties under the alternative hypothesis. Both forms of SIMPLE tests are pivotal and have asymptotic size at the desired level and asymptotic power one. The advantages and practical utility of our new method in terms of both size and power are demonstrated through several simulation examples and real network applications. (Joint work with Yingying Fan and Jinchi Lv)