

Improving identification of hub genes and gene sub-networks through data integration with the stochastic block model

The study of gene networks associated with a given disease area often involves analysis of multiple sources of network data that feature weak and widespread signal and low signal-to-noise ratio. In this setting, interest commonly lies in detection of possible latent community structure, such as unknown hub genes and sub-networks within a given gene network. The stochastic block model (SBM) is a flexible statistical model with the ability to perform community detection. However, the standard SBM still suffers from poor performance in the case of weak and widespread signal. With improvement of genomic profiling technologies, multiple relevant experiments for a given disease area have become more abundant providing unprecedented opportunity for integrative analyses. Examples of such complimentary data sources are gene expression, DNA methylation, and more recently, literature mining data, each of which are concerned with measuring associations among a similar sets of genes and often feature low signal to noise ratio. To address the issue of weak and widespread signal in network data, we propose a data integration framework for the SBM, whereby multiple available data sources are combined into a unified network model. We show through simulation studies that our proposed method offers improved detection of hub genes and gene sub-networks for a variety of settings when compared to SBMs fit to single data sources or under alternative data integration approaches. In future work, we plan to extend our data integration method to the Bayesian setting to allow for inclusion of prior biological knowledge to guide detection of hub genes and gene sub-networks, and further improve SBM performance in the case of weak and widespread signal. In addition, we also plan to apply the proposed framework to the genomic studies of twins with systemic sclerosis (SSc) to help investigation of the genetic causes of SSc.