Yu, Han; Hageman Blair, Rachael
Scalable module detection for attributed networks with applications to breast cancer.
(English) J. Appl. Stat. 49, No. 1, 230-247 (2022)

Summary: The objective of network module detection is to identify groups of nodes within a network structure that are tightly connected. Nodes in a network often have attributes (aka metadata) associated with them. It is often desirable to identify groups of nodes that are tightly connected in the network structure, but also have strong similarity in their attributes. Utilizing attribute information in module detection is a major challenge because it requires bridging the structural network with attribute data. A Weighted Fast Greedy (WFG) algorithm for attribute-based module detection is proposed. WFG utilizes logistic regression to bridge the structural and attribute spaces. The logistic function naturally emphasizes associations between attributes and network structure accordingly, and can be easily interpreted. A breast cancer application is presented that connects a protein-protein interaction network gene expression data and a survival outcome. This application demonstrates the importance of embedding attribute information into the community detection framework on a breast cancer dataset. Five modules were significant for survival and they contained known pathways and markers for cancer, including cell cycle, p53 pathway, BRCA1, BRCA2, and AURKB, among others. Whereas, neither the gene expression data nor the network structure alone gave rise to these cancer biomarkers and signatures.

MSC: 62Pxx Applications of statistics

Keywords: module detection; community; attribute; survival; gene expression

Full Text: DOI

References: