Prediction-based evaluation of community detection methods for maize gene expression data

Authors

- **Taylor Ferebee**

  ⓘD [XXXX-XXXX-XXXX-XXXX](#) · ⓞ [thferebee](#) · 🐦 [Ferebee2PhD](#)

  Department of Computational Biology, Cornell University · Funded by Grant XXXXXXXX

- **Jane Roe**

  ⓘD [XXXX-XXXX-XXXX-XXXX](#) · ⓞ [janeroe](#)

  Department of Something, University of Whatever; Department of Whatever, University of Something

Abstract

Visual Abstract

Introduction

Since the days of microarrays, functional genomics has been tasked with revealing true gene function from highthroughput data. Gene co-expression networks (GCNs) are an intuitive way to understand the relationships between genes with simialar gene expresssion profiles and infer the function of these related genes. In these networks, nodes represent genes and the connection between these nodes (edges) represent the degree of similarity between the nodes. Genes with similar biological function tend to result in similar expression profiles[1]. Because of this expectations, functional biologists tend to mine GCNs for "communities" of interacting genes by using theory from network science community.

Materials and Methods

Community Detection Methods

Hierarchical Clustering on a Topological Overlap Matrix One of the most common clustering methods, agglomerative heiarcarchical clustering (implemented in WGCNA R package) attempts to create a heirarchy of clusters by starting with individual clusters and building to one cluster with all data. To begin the procedure, each gene is in its own cluster and a similarity (distance) matrix is considered. If two clusters (genes) are similar (close in distance), these two genes merge. If two clusters are not close, their clusters do not merge. Once again, we consider a similarity matrix between the newly formed clusters. This repeats until all clusters have merged into a single cluster. Most often, heirarchical clustering is pressented in a dendrogram form which tracks the merges of the clusters.

Fast Greedy Fast greedy community detection uses topological features of a network and an agglomative algorithm for detecting community structure. The fast greedy algorithm utilizes a network's modularity. The modularity measures the relationship between the density of links within communities compared with the rest of the network. In other words, a particilarly modular network (as biological networks are assumed to be) have commiunityes that are very connected iwhtin the community, but only have a few edges to other communities.

Results

Discussion

References

1. **Life with 6000 Genes**
   A. Goffeau, B. G. Barrell, H. Bussey, R. W. Davis, B. Dujon, H. Feldmann, F. Galibert, J. D. Hoheisel, C. Jacq, M. Johnston, … S. G. Oliver
   Science (1996-10-25) https://doi.org/d7mhxv
   DOI: 10.1126/science.274.5287.546 · PMID: 8849441