Author's response to reviews

Title: Hybrid coexpression link similarity graph clustering for mining biological modules from multiple gene expression data sets.

Authors:

Saeed Salem (saeed.salem@ndsu.edu)
Cagri Ozcaglar (cagriozcaglar@gmail.com)

Version: 2
Date: 5 August 2014

Author's response to reviews: see over
The response for the reviewer’s comments is shown below. For reference, the original reviewer comments are provided in quotes.

Reviewer 1:
Major Compulsory Revisions

“1) Please move the details of the data set analyzed and the analysis details to the methods section.”

We have moved the details of the datasets as requested. Moreover, we revised the description of how the coexpression graphs were obtained.

“2) Please indicate whether your software is available as open-source.”

The current code is not in a state to be released as open source. There are many pieces of the software that needs to be integrated first. We use the MCL as the clustering algorithm. Our code prepares the hybrid link similarity matrix and then runs the MCL clustering on the graph representing the similarity matrix. The hybrid similarity matrix can be calculated easily in languages, like matlab, Python, and R.

If required, we can put a link for the proposed approach where users can download binary for calculating the hybrid similarity matrix and some links to installing the binary for the MCL from the authors of the MCL algorithm.

Minor Essential Revisions

“1) Why was a p-value cutoff of 0.01 used? Please add a sentence to justify this type I error rate.”

We added statements in the Datasets section explaining that we only use significant correlation.

“2) I didn’t see any mention of phenotypes. Would be nice to see some details about how relevant these results are to biological endpoints.”

The summary graph is an aggregate graph and thus the results are not guaranteed to have any density or occurrence in any particular graph in the coexpression graph.

The phenotypes or label of the coexpression graph was not used in the mining process. Therefore, we didn’t analyze the phenotypes.

We are currently working on an enumeration algorithm that guarantees some characteristics of modules in individual coexpression graphs.