Detecting modules in multiplex networks – an application for integrating expression profiles across multiple species

Koon-Kiu Yan  
*K Yale University, koon-kiu.yan@yale.edu*

Daifeng Wang  
*K Yale University*

Joel Rozowsky  
*K Yale University*

Henry Zheng  
*K Yale University*

Baikang Pei  
*K Yale University*

*See next page for additional authors*

Follow this and additional works at: [http://elischolar.library.yale.edu/dayofdata](http://elischolar.library.yale.edu/dayofdata)

Part of the Bioinformatics Commons, Computational Biology Commons, and the Genomics Commons

[http://elischolar.library.yale.edu/dayofdata/2013/Posters/6](http://elischolar.library.yale.edu/dayofdata/2013/Posters/6)

This Event is brought to you for free and open access by EliScholar – A Digital Platform for Scholarly Publishing at Yale. It has been accepted for inclusion in Yale Day of Data by an authorized administrator of EliScholar – A Digital Platform for Scholarly Publishing at Yale. For more information, please contact elischolar@yale.edu.
Detecting modules in multiplex networks – an application for integrating expression profiles across multiple species

Koon-Kiu Yan*1,2, Daifeng Wang*1,2, Joel Rozowsky1,2, Henry Zheng1,2, Baikang Pei1,2, Mark Gerstein1,2,3
1Program in Computational Biology and Bioinformatics, 2Department of Molecular Biophysics and Biochemistry, 3Department of Computer Science, Yale University

Abstract

Multiplex network, a set of networks linked through interconnected layers, is a useful mathematical framework for data integration. Here, we present a general method to detect modules in multiplex networks and apply it in a specific biological context: to simultaneously cluster the genome-wide expression profiles of C. elegans and D. melanogaster generated by the ENCODE and modENCODE consortia. The method revealed modules that are fundamentally cross-species and can either be conserved or species-specific.

In general, the method could be applied in various contexts like the integration of different social networks.

Motivation: Integrating Big Data in Genomics

A spin model to detect modules in a multiplex network

Modules are densely connected regions in a network. Every node i is assigned with a spin value σi (labels of modules 1, 2, ..., q).

\[ H = \sum_{i,j} (-W_{ij}^{A} + p_{ij}^{A}) \delta_{\sigma_i\sigma_j} + \sum_{i,j} (-W_{ij}^{B} + p_{ij}^{B}) \delta_{\sigma_i\sigma_j} + K \sum_{i} \delta_{\sigma_i0} \]

- reward a co-expression pair with the same value
- punish a non co-expression pair with the same value
- reward an orthologous pair with the same value
- punish an orthologous pair with the same value
- punish a non orthologous pair with the same value

Input

Orthologous pairs between species
Expression profiles of species A
Expression profiles of species B

Output

A set of cross-species modules, with different number of shared orthology

Conserved module
B-specific module
A-specific module

Potts Clustering

Simultaneous clustering of expression profiles

worm genes (20377) fly genes (13623)

Further Application and Outlook

- Modules can be used to infer the functions of proteins as well as non-coding RNAs based on “guilt-by-association”. ncRNAs from the worm and fly anchored to a module may potentially have analogous functions.
- Applying other optimization techniques to speed up the algorithm.
- Various social networks (Facebook and LinkedIn) can be integrated into a multiplex network. The method presented could be used to detect communities in a social setting.