

Oral presentation

Open Access

Mining expression-dependent modules in the human interaction network

Elisabeth Georgii*^{1,2}, Sabine Dietmann³, Takeaki Uno⁴, Philipp Pagel³ and Koji Tsuda¹

Address: ¹Max Planck Institute for Biological Cybernetics, Spemannstr. 38, 72076 Tübingen, Germany, ²Friedrich Miescher Laboratory of the Max Planck Society, Spemannstr. 39, 72076 Tübingen, Germany, ³GSF National Research Center for Environment and Health, Ingolstädter Landstraße 1, 85764 Neuherberg, Germany and ⁴National Institute of Informatics, 2-1-2, Hitotsubashi, Chiyoda-ku, Tokyo, Japan

Email: Elisabeth Georgii* - georgii@tuebingen.mpg.de

* Corresponding author

from Third International Society for Computational Biology (ISCB) Student Council Symposium at the Fifteenth Annual International Conference on Intelligent Systems for Molecular Biology (ISMB) Vienna, Austria. 21 July 2007

Published: 20 November 2007

BMC Bioinformatics 2007, 8(Suppl 8):S4 doi:10.1186/1471-2105-8-S8-S4

This abstract is available from: <http://www.biomedcentral.com/1471-2105/8/S8/S4>

© 2007 Georgii et al; licensee BioMed Central Ltd.

Motivation

We propose a novel method for automatic module extraction from protein-protein interaction networks. While most previous approaches for module discovery are based on graph partitioning [1], our algorithm can efficiently enumerate all densely connected modules in the network. As currently available interaction data are incomplete, this is a meaningful generalization of clique search techniques [2]. In comparison with partitioning methods, the approach has the following advantages: the user can specify a minimum density for the outgoing modules and has the guarantee that all modules that satisfy this criterion are discovered. Moreover, it provides a natural way to detect overlapping modules. Many proteins are not steadily present in the cell, but are specifically expressed in dependence of cell type, environmental conditions, and developmental state. Therefore we introduce an additional constraint for modules which accounts for differential expression.

Results

We analysed human interaction data from MINT, Intact, HPRD, and DIP in the context of tissue-specific gene expression data in human provided by Su *et al.* [3]. We discretized the expression information into binary states (expressed versus not expressed) and searched for densely

connected modules where all proteins are expressed in at least 3 tissues and all proteins are not expressed in at least 10 tissues. To deal with the fact that protein interaction data contain a high number of false positives, we computed reliability scores for each experimental source. Similarly to the work by Jansen *et al.* [4], we used for that purpose a gold standard set of known interactions as well as a gold standard set of false interactions and calculated the likelihood ratio, which was used to assign edge weights to the interaction graph. The density of a module is defined as the sum of the edge weights inside the module divided by the maximal possible weight sum for a module of that size.

Setting the minimum density threshold to 35% and removing modules that are totally contained in other modules, we obtained a set of 949 differentially expressed modules. They were ranked in descending order according to the average weight per node (see [5]), so larger and denser modules appear first. On the one hand, we discovered known complexes and modules that link strongly cooperating complexes like MCM and ORC. On the other hand, we found extensions of known complexes that confirm hypothetical functional annotation in Uniprot as well as modules which are not contained in the manually curated set of known complexes, but share the same func-

tional annotation. Finally, some modules are candidates for further biological investigation, containing proteins with unknown functional relationships.

Conclusion

We developed a general method for exhaustive dense module extraction from networks. Remarkably, it allows to determine exact P-values for the predicted modules without having to rely on any network model and can easily integrate information from different heterogeneous data sources.

Acknowledgements

We are grateful to Andreas Rupp for providing a curated set of known human complexes and to Gunnar Rätsch for his encouragement and support.

References

1. Newman MEJ: **Modularity and community structure in networks.** *PNAS* 2006, **103(23)**:8577-8582.
2. Palla G, Derenyi I, Farkas I, Vicsek T: **Uncovering the overlapping community structure of complex networks in nature and society.** *Nature* 2005, **435**:814-818.
3. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR, Hogenesch JB: **A gene atlas of the mouse and human protein-encoding transcriptomes.** *PNAS* 2004, **101(16)**:6062-6067.
4. Jansen R, Yu H, Greenbaum D, Kluger Y, Krogan NJ, Chung S, Emili A, Snyder M, Greenblatt JF, Gerstein M: **A Bayesian networks approach for predicting protein-protein interactions from genomic data.** *Science* 2003, **302**:449-453.
5. Bader GD, Hogue CW: **An automated method for finding molecular complexes in large protein interaction networks.** *BMC Bioinformatics* 2003, **4**:2.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

