Node and link roles in protein-protein interaction networks

Sumeet Agarwal, Charlotte Deane, Nick Jones, Mason Porter

University of Oxford, Oxford, United Kingdom

April 3, 2009

Abstract

A key question in modern biology is how the complexity of protein-protein interaction networks relates to biological functionality. One way of understanding the set of proteins and their interactions (the interactome) is to look at them as a network of nodes connected by links. By studying the structure of this network, we may hope to learn something about the interactome's organisation. Here we attempt to look at different approaches for using network models to assign structural and functional roles to proteins and protein interactions. It has been proposed that highly connected nodes, or hubs, in the interactome fall into two classes, 'date' and 'party' [1], and that these play a key role in the modular organisation of the yeast interactome. This classification was made on the basis of the extent to which hubs are co-expressed with their interaction partners, but was then used to impute to them specific topological roles. We attempt to use purely topological statistics to examine the extent to which these hubs really fall into the roles thus attributed. We use a community detection approach based on maximising modularity [2] to partition the interaction network into functionally coherent modules. We then assign roles to proteins based on how their interactions are distributed within their own module and across other modules [3]. Based on a study of multiple yeast and human datasets, our results suggest that there is little evidence for a clear date/party distinction, but rather nodes in the protein interaction network seem to perform a variety of roles falling along a continuum, and there is no strong correlation between these roles and co-expression. We also examine alternative approaches to studying topological roles. So far, most work has focused on node-centric measures; here we attempt using a betweenness metric [4, 5] to quantify the centrality of links rather than nodes. We show that this measure relates to protein functional similarity as assessed by annotation overlap in the Gene Ontology [6], and may also be relevant to understanding how the interactome works as a system.

References

[1] Jing-Dong J. Han, Nicolas Bertin, Tong Hao, Debra S. Goldberg, Gabriel F. Berriz, Lan V. Zhang, Denis Dupuy, Albertha J. Walhout, Michael E. Cusick, Frederick P. Roth, and Marc Vidal. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature*, 430(6995):88–93, July 2004.

[2] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69:026113, 2004.

[3] Roger Guimerà and Lúis A. Nunes Amaral. Functional cartography of complex metabolic networks. *Nature*, 433:895–900, 2005.

[4] Linton C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, 40(1):35–41, 1977.

[5] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proc. Natl Acad. Sci. USA*, 99(12):7821–7826, 2002.

[6] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin, and G. Sherlock. Gene Ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet*, 25(1):25–29, May 2000.