Mining protein complexes from protein-protein interaction networks using quasi-cliques

Wong Lin Kit Daniel[1], Sriganesh Srihari[2], Hon Wai Leong[2]
[1]CenTaD Research Programme, Hwa Chong Institution
[2]School of Computing, National University of Singapore

Summary

Proteins complexes play key roles in the functioning of the cell. Identifying them is thus an important goal for cell and molecular biology. Besides detecting them through wet lab experiments, it is useful to be able to predict their existence using the large amounts of protein interaction data that are available. In this paper, they are modelled using a structure known as 'quasi-cliques'. We designed a framework for analysing and evaulating different algorithms for protein complex prediction, propose improvements to an existing method for protein complex detection and show experimental results with protein interaction data.

ABSTRACT

Protein complexes are groups of proteins that play a crucial role in many cellular processes. Large PPI datasets have been amassed using high-throughout experimental techniques. In this project, we focus on computational methods of deriving protein complexes from PPI network data. We set out to study the efficacy of the CMC algorithm for the automatic extraction of complexes from PPI network data. Thereafter, we sought to propose improvements to the algorithm and evaluate the modifications using protein interaction data. Our experiments affirmed that CMC has been well-tuned for complex prediction, and that its default parameters had been optimised to deliver a good performance on yeast PPI data. We propose improvements to the merging methods and criteria of the algorithm and have evaluated our modified algorithms, CMC-BF and CMC-BFA, on yeast PPI datasets and manually curated reference complex datasets. We noted that CMC-BF and CMC-BFA are able to achieve higher precision at the expense of a lower recall. Moreover, a slightly better recall-precision curve was observed. Furthermore, we developed a new workflow, including a framework for the evaluation and comparison of complex prediction results. We also created a new utility to enable us to examine the data more efficiently and effectively.

Keywords: protein-protein interaction networks, protein complexes, quasi-cliques, cliques, merging
[1] Email: wonglkd@gmail.com
[2] Dept of Computer Science, School of Computing, National University of Singapore, 13 Computing Drive, Singapore 117590. E-mail: {leonghw,srigsri}@comp.nus.edu.sg.