# BMC Bioinformatics

Oral presentation

# Functional module detection by functional flow pattern mining in protein interaction networks

## Young-Rae Cho*, Lei Shi and Aidong Zhang

Address: Department of Computer Science, State University of New York at Buffalo, NY 14260, USA

Email: Young-Rae Cho* - ycho8@buffalo.edu

* Corresponding author

This abstract is available from: http://www.biomedcentral.com/1471-2105/9/S10/O1

## Background

A functional module has been defined as a group of molecules that participate in the same functional activities. Various graph-theoretic or data-mining techniques have been applied to discover functional modules from protein interaction networks [1]. However, their performance has been compromised by false-positive and false-negative interaction data and complex connectivity of the interaction networks. In our earlier study [2], we have introduced the functional flow-based approach to efficiently identify overlapping modules, which are generally large-sized, from interaction networks. In this abstract, we extend this approach by mining functional flow patterns for the purpose of detecting small-sized modules for specific functions.

## Methods

Our approach includes three steps. First, we integrate the interaction network with semantic data from Gene Ontology [3] to generate a weighted interaction network, which is functionally reliable. Next, we simulate functional flow starting from selected informative proteins and identify primary modules for general-level functions [2]. As the last step, we obtain the set of functional flow patterns for each primary module by flow simulation from all nodes within the module. A functional flow pattern is defined as a sequence of quantities of functional influence of a source protein on target proteins. The coherent patterns are then captured by a pattern-based clustering algorithm [4] as final modules for specific-level functions. The significant assumption is that if two source proteins have

**Table 1:**

| function | module size | -$log$($p$-value) |
| --- | --- | --- |
| ORI recognition and priming complex formation | 22 | 9.80 |
| extension and polymerization activity | 27 | 4.73 |
| DNA repair | 92 | 3.08 |
| DNA recombination | 42 | 3.01 |
| DNA conformation modification | 120 | 3.25 |
| Mitotic cell cycle | 111 | 8.82 |
| meiosis | 80 | 3.48 |
| cell division | 40 | 7.13 |

The detected functional modules and their accuracy by our flow-pattern clustering method. The proteins on *Cell Cycle and DNA Processing* and the interactions between them were used.

**Table 2:**

| methods | category | number of modules | average module size | *-log(p)* |
|---|---|---|---|---|
| functional flow pattern | flow-based | 14 | 11.20 | 5.41 |
| edge-betweenness | hierarchical | 43 | 9.67 | 4.62 |
| clique percolation | density-based | 16 | 6.94 | 4.63 |

Accuracy comparison of the methods for functional module detection in protein interaction networks. The flow-pattern clustering approach had approximately 20% higher accuracy than the other methods.

similar functional flow patterns across all the other targets proteins, then they are likely to have the same function.

## Results

We tested our flow-pattern clustering method using a sub-network, structured by the proteins having functions on *Cell Cycle and DNA Processing* and the interactions between them. The output modules were compared to the functional categories and their annotations from MIPS [5] using statistical *p*-value analysis (see Table 1). We assessed the performance of our algorithm comparing to two competing methods: the clique percolation method [6] as a density-based approach to find densely connected subgraphs, and the betweenness-cut method [7] as a hierarchical approach to iteratively separate a graph and find the best partition. As a result, our algorithm had higher accuracy than the others by approximately 20% (see Table 2).

## Conclusion

The modules, identified from protein interaction networks, provide an understanding of functional associations among proteins. In this study, we introduced a framework to detect functional modules in protein interaction networks. We demonstrated that our approach accurately handles the erroneous and complex networks.

## References

1. Sharan R, Ulitsky I, Shamir R: **Network-based prediction of protein function.** *Molecular Systems Biology* 2007, **3**:88.
2. Cho Y-R, Hwang W, Ramanathan M, Zhang A: **Semantic integration to identify overlapping functional modules in protein interaction networks.** *BMC Bioinformatics* 2007, **8**:265.
3. Gene Ontology Consortium: **The Gene Ontology project in 2008.** *Nucleic Acids Res* 2008, **36(Database issue):**D440-D444.
4. Wang H, Wang W, Yang J, Yu PS: **Clustering by pattern similarity in large data sets.** *Proceedings of ACM SIGMOD International Conference on Management of Data* 2002:394-405.
5. Mewes HW, Dietmann S, Frishman D, Gregory R, Mannhaupt G, Mayer KF, Munsterkotter M, Ruepp A, Spannagl M, Stumpflen V, Rattei T: **MIPS: analysis and annotation of genome information in 2007.** *Nucleic Acids Res* 2008, **36(Database issue):**D196-D201.
6. 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.
7. Dunn R, Dudbridge F, Sanderson CM: **The use of edge-betweenness clustering to investigate biological function in protein interaction networks.** *BMC Bioinformatics* 2005, **6**:39.