CyClus3D: a Cytoscape plugin for clustering network motifs in integrated networks

Pieter Audenaert1, Thomas Van Parys2,3, Florian Brondel1, Mario Pickavet1, Piet Demeester1, Yves Van de Peer2,3 and Tom Micheöl1†

1Department of Information Technology, Ghent University – IBBT, Gaston Crommenlaan 8/201, B-9050 Gent, Belgium; 2Department of Plant Systems Biology, VIB and 3Department of Plant Biotechnology and Genetics, Ghent University, Technologiepark 927, B-9052 Gent, Belgium; *Freiburg Institute for Advanced Studies, University of Freiburg, Albertstraße 19, D-79104 Freiburg, Germany

ABSTRACT
Summary: Network motifs in integrated molecular networks represent functional relationships between distinct data types. They aggregate to form dense topological structures corresponding to functional modules which cannot be detected by traditional graph clustering algorithms. We developed CyClus3D, a Cytoscape plugin for clustering composite 3-node network motifs using a 3-dimensional spectral clustering algorithm.

Availability: Via the Cytoscape plugin manager or http://bioinformatics.psb.ugent.be/software/details/CyClus3D.

Contact: tom.michel@frias.uni-freiburg.de

1 INTRODUCTION
In systems biology, the cell is modeled as an integrated network with multiple types of interactions, e.g. protein-protein, protein-DNA, protein-metabolite or genetic interactions (Zhu et al., 2007). Cellular functions are carried out by independently functioning units called modules (Hartwell et al., 1999), which, in graph-theoretical terms, correspond to clusters of densely connected nodes, and a multitude of algorithms have been developed to identify such clusters in undirected graphs (Fortunato, 2010). A major problem remains how to harness the multi-layered information contained in different interaction networks in order to identify biologically more realistic topological modules. In the naive Bayes approach, multiple interaction types are overlayed to create a single integrated association network which can be clustered by traditional means (Lee et al., 2004). While SAMBA has the advantage of preserving the identity of each interaction type, information is inevitably lost by representing complex networks as bipartite graphs.

We developed CyClus3D, a Cytoscape (Shannon et al., 2003) plugin for the identification of modules in integrated networks which uses network motifs to query a 3-dimensional spectral clustering algorithm. Network motifs are frequently occurring subgraphs in regulatory (Shen-Orr et al., 2002) or integrated networks (Yeger-Lotem et al., 2004; Yu et al., 2006), which aggregate to form topological modules (Kashtan et al., 2004; Zhang et al., 2005). Each network motif defines a relationship between heterogeneous data types, with a distinct information-processing role or functional interpretation (Shen-Orr et al., 2002; Zhang et al., 2005; Zhu et al., 2007). Hence, CyClus3D identifies modules composed of multiple interaction types which reflect regulatory, signaling or compensatory pathway mechanisms in addition to the stable protein complexes found by traditional clustering algorithms.

2 METHODS
2.1 Network motif clustering algorithm
We consider a system modeled by N types of pairwise interactions which may be directed or undirected. For a given 3-node network motif whose edges can be of any type, we denote the list of all motif instances as a 3-dimensional array $T$ with $T_{ijk}$ being the number of instances of the motif consisting of nodes $(i,j,k)$, and 0 otherwise. We define a motif cluster by three sets of nodes $(X_1, X_2, X_3)$ with an aggregation score

$$S(X_1, X_2, X_3) = \sum_{i<j<k} T_{ijk} \frac{1}{|X_1|^{1/p}|X_2|^{1/p}|X_3|^{1/p}},$$

(1)

where $|X|$ is the number of nodes in $X$ and $p > 1$ will act as an (inverse) resolution parameter. To maximize $S$, we first determine the best rank-1 approximation to $T$, i.e. find real-valued vectors $(x_1, x_2, x_3)$ maximizing

$$R(x_1, x_2, x_3) = \sum_{i<j<k} T_{ijk} x_i x_j x_k \frac{1}{\|x_1\|_p \|x_2\|_p \|x_3\|_p},$$

where $\|x\|_p = (\sum_i x_i^p)^{1/p}$ is the $p$-norm of $x$. A maximizer of $R$ is found by solving the Euler-Lagrange equations

$$\lambda x_i^{p-1} = \sum_{j<k} T_{ijk} x_j x_k,$$

(2)

subject to the constraint $\|x_i\|_p = 1$ and similarly for the other dimensions (De Lathauwer et al., 2000). The solutions $(x_1, x_2, x_3)$ are interpreted as cluster membership weight vectors and converted to a motif cluster by taking a suitable threshold on the weights. It can be shown that the optimal

© The Author (2011). Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

1These authors contributed equally to this work
2To whom correspondence should be addressed
After running the algorithm, CyClus3D opens a new network containing all clustered motifs. For instance, Fig. 1B shows all clusters of genetically interacting, copointing kinases (with the settings of Fig. 1A). By right clicking on a node of interest, we can create new networks for the clusters containing this node, while through the CyClus3D entry in the Plugins menu, new networks can be created for all clusters. By default, edges in multi-cluster networks are colored by their cluster membership (‘Cluster View’, Fig. 1B), while in single-cluster networks they are colored by interaction type, with the colors matching the edge assignments in the control panel (‘Interaction View’, Fig. 1C). Via the VizMapper panel, the user can easily switch between these two visual styles. Multiple motifs can be clustered sequentially and newly found clusters either are added to or replace the existing clustered network (to add them, all query motifs must be formed from subsets of the same three edge types and the Interaction View will be updated to the latest edge assignment).

By integrating heterogeneous types of molecular interaction data, CyClus3D identifies modules which reflect regulatory, signaling or compensatory functions which are not found by clustering each network in isolation (Zhang et al., 2005). The underlying algorithms are highly efficient and allow further extension. In particular, future versions will extend CyClus3D towards higher-dimensional motifs, with applications in the domain of network alignment and comparison.

ACKNOWLEDGMENT

This research was supported by grants from the IWT (SBO-BioFrame), IUAP P6/25 (BioMaGNet) and Ghent University (MRP “Bioinformatics: from nucleotides to networks”).

REFERENCES


De Lathauwer, L. et al. (2000). On the best rank-1 and rank-((r_1, r_2, \ldots, r_n)) approximations of higher-order tensors. SIAM J Matrix Anal Appl, 21, 1324.

De Lathauwer, L. et al. (2000). On the best rank-1 and rank-((r_1, r_2, \ldots, r_n)) approximations of higher-order tensors. SIAM J Matrix Anal Appl, 21, 1324.


