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 Michoel4,∗

1Department of Information Technology, Ghent University – IBBT, Gaston Crommenlaan 8/201, B-9050 Gent, 2Department of Plant Systems Biology, VIB, 3Department of Plant Biotechnology and Genetics, Ghent University, Technologypark 927, B-9002 Gent, Belgium and 4School of Life Sciences – LifeNet, 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 three-node network motifs using a 3D spectral clustering algorithm.

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

Supplementary Information: Supplementary data are available at Bioinformatics online.

Received on December 9, 2010; revised on March 16, 2011; accepted on April 4, 2011

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., 2003). Cellular functions are carried out by independently functioning units (Lee association network which can be clustered by traditional means multiple interaction types are overlayed to create a single integrated more realistic topological modules. In the naive Bayes approach, a multitude of algorithms have been developed to identify such DNA, protein–metabolite or genetic interactions (Zhu

Studies, University of Freiburg, Albertstrasse 19, D-79104 Freiburg, Germany

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

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 3D spectral clustering algorithm. Network motifs are frequently occurring subgraphs in regulatory (Shen-Ort et al., 2002) or integrated networks (Yeger-Lotem et al., 2004; Yu et al., 2006), which aggregate to form topological modules (Kashan 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-Ort 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 three-node network motif whose edges can be of any type, we denote the list of all motif instances as a 3D array T with T_{ijk} = 1 if the system contains a motif between 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

\[ S(X_1,X_2,X_3) = \sum_{i,j,k} T_{ijk} \left[ \frac{1}{|X_1|} + \frac{1}{|X_2|} + \frac{1}{|X_3|} \right] \],

(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

\[ R(x_1,x_2,x_3) = \sum_{i,j,k} T_{ijk} x_{i1} x_{j2} x_{k3} / \left[ |x_1| |x_2| |x_3| \right] \],

(2)

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

\[ x^2_{i1} = \sum_{j,k} T_{ijk} x_{j2} x_{k3}, \]

subject to the constraint \[ |x|_1 = 1 \] 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 threshold is the one which minimizes \[ \sum_{i,j,k} T_{ijk} - w_{ij} w_{jk} \] where \[ w_{ij} = x_{i1}^p x_{j2}^p \] for (i,x) and 0 otherwise (see Supplementary Material). Having thus found a high-scoring

To whom correspondence should be addressed.
*The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.
motif cluster, we remove it from the list of motif instances \( T \) and repeat the procedure until no more instances remain. The best rank-1 approximation to the motif index array plays the same role as the dominant eigenvector of a network adjacency matrix and our algorithm can be understood as a generalization of 2D spectral clustering algorithms (Inoue and Urahama, 1999).

2.2 Implementation
Network motifs are often invariant under the permutation of some of their nodes. Thus, motif instances need to know their inherent symmetries, e.g. to efficiently determine the equality of two instances. We generated the motif symmetry groups offline and used a code generator to generate Java classes which are equipped with optimized methods for comparing and storing motifs. To locate all motif instances, we developed a motif finder which works on the principle of motif extensions. It allows quick pruning of branches in the search tree and is significantly faster than other subgraph matching algorithms (see Supplementary Material). To solve Equations (2) we implemented a power algorithm (De Lathauwer et al., 2000). The Java classes for network motif enumeration and clustering are independent of the Cytope visualization classes and can be plugged into other network analysis and visualization environments as well.

3 APPLICATION
To illustrate the workflow of CyClus3D (postfix for 3-dimensional Clustering in Cytope), we imported an integrated network of physical, genetic and signaling interactions between kinases and phosphatases in yeast (Breitkreutz et al., 2010; Fiedler et al., 2009) (data available as Supplementary Material). In the CyClus3D control panel (Fig. 1A), a query motif and one or more input networks are selected, interaction types are assigned to each edge and a value for the resolution parameter \( r = 1/p \) (cf. Section 2) and the minimal number of motif instances in a cluster are set. An edge type is inferred to be directed if the edge in the motif it is assigned to is directed. The resolution parameter allows to vary the typical size and density of a cluster. At low \( r \), the aggregation score is maximized by large sets of loosely connected motifs, while at high values, high-scoring motif clusters are small and dense. In our experience, the intermediate value \( r = 0.5 \) balances size and density and is recommended as a starting value (see Supplementary Material).

After running the algorithm, CyClus3D opens a new network containing all clustered motifs. For instance, Figure 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 multicluster 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 toward higher dimensional motifs, with applications in the domain of network alignment and comparison.

Funding: Grants from the IWT (SBO-BioFrame); IUAP P6/25 (BioMaGNet); Ghent University (MRP ‘Bioinformatics: from nucleotides to networks’).

Conflict of Interest: none declared.

REFERENCES

Fig. 1. CyClus3D screenshot with the workflow (A), a multicluster network with a node properties menu (B) and a single-cluster network (C).