

ModuLand plug-in for Cytoscape: determination of hierarchical layers of overlapping network modules and community centrality

Máté Szalay-Bekő^{1,†}, Robin Palotai^{1,†}, Balázs Szappanos², István A. Kovács^{1,3,4}, Balázs Papp² and Péter Csermely^{1,*}

¹Department of Medical Chemistry, Semmelweis University, Budapest 1444, ²Institute of Biochemistry, Biological Research Centre, Hungarian Academy of Sciences, Szeged 6726, ³Department of Physics, Loránd Eötvös University and ⁴Institute for Solid State Physics and Optics, Wigner Research Centre, Hungarian Academy of Sciences, Budapest 1525, Hungary

Associate Editor: Mario Albrecht

ABSTRACT

Summary: The ModuLand plug-in provides Cytoscape users an algorithm for determining extensively overlapping network modules. Moreover, it identifies several hierarchical layers of modules, where meta-nodes of the higher hierarchical layer represent modules of the lower layer. The tool assigns module cores, which predict the function of the whole module, and determines key nodes bridging two or multiple modules. The plug-in has a detailed JAVA-based graphical interface with various colouring options. The ModuLand tool can run on Windows, Linux or Mac OS. We demonstrate its use on protein structure and metabolic networks.

Availability: The plug-in and its user guide can be downloaded freely from: <http://www.linkgroup.hu/modules.php>.

Contact: csermely.peter@med.semmelweis-univ.hu

Supplementary information: Supplementary data are available at *Bioinformatics* online.

Received on November 21, 2011; revised on June 13, 2012; accepted on June 14, 2012

1 INTRODUCTION

Nodes of biological networks often belong to multiple network communities. Recently, a number of methods were published to determine tightly or extensively overlapping network modules (Adamcsek *et al.*, 2006; Ahn *et al.*, 2010; Fortunato, 2010; Kovács *et al.*, 2010; Mihalik and Csermely, 2011; Palla *et al.*, 2005). Our ModuLand framework (Kovács *et al.*, 2010) introduced community landscapes. The *x*–*y* plane of a community landscape is a conventional 2D visualization of the network, whereas the *z*-axis represents community centrality. Community centrality of a given edge (or node) was defined as the sum of local influence zones of all network edges (or nodes) including the given edge (or node; Supplementary Fig. S1). Thus community centrality represents an integrated measure of the whole network's influence to one of its edges or nodes. Hills of the community landscape correspond to network modules (Supplementary Fig. S1) yielding extensive overlaps. This concept led to the development of the ModuLand family of network modularization methods (Kovács *et al.*, 2010).

*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 the joint first Authors.

The widely used Cytoscape program (Shannon *et al.*, 2003) has several very useful clustering plug-ins (Bader and Houge, 2003; Morris *et al.*, 2011; Rhirissorakrai and Gunsalus, 2011; Rivera *et al.*, 2010; Su *et al.*, 2010). However, these methods do not focus on extensive modular overlaps, and do not build a modular hierarchy, where meta-nodes of the higher level represent modules of the lower level. Moreover, existing plug-ins do not provide measures identifying the centre of the module, as well as key nodes bridging two or multiple modules (see Supplementary Table S9, and Supplementary Discussion). Here, we introduce the Cytoscape plug-in of the most widely applicable version of the ModuLand method family (Kovács *et al.*, 2010). We demonstrate its ability to determine biologically relevant, extensively overlapping network modules, hierarchical layers of modules, module cores and key inter-modular nodes using protein structure and metabolic networks.

2 SOFTWARE OVERVIEW

The ModuLand Cytoscape plug-in uses the LinkLand influence zone determination method and the ProportionalHill module assignment method of our formerly published ModuLand network module determination method family (Kovács *et al.*, 2010). These two methods provide a good trade-off between the fast (but rather inaccurate), and accurate (but rather slow) other ModuLand methods.

The installation of the ModuLand plug-in follows Cytoscape procedures. This is much easier than the setup required for the earlier version (Kovács *et al.*, 2010). The program can be distributed as a single .jar file. Moreover, the current implementation works on Linux, Windows and Mac OS, thereby extending the options of the former version.

The plug-in determines extensively overlapping modules using any undirected network type and weight description of Cytoscape. Moreover, the plug-in calculates a set of hierarchical modules. In the modular hierarchy, modules of the lower level become meta-nodes of the upper level, and modular overlaps of the lower level become weights of the meta-edges at the upper level. The plug-in creates, automatically re-loads and visualizes the higher and higher level hierarchies (with lower and lower number of meta-nodes and meta-edges, see Fig. 1), until the whole network coalesces into a single meta-node.

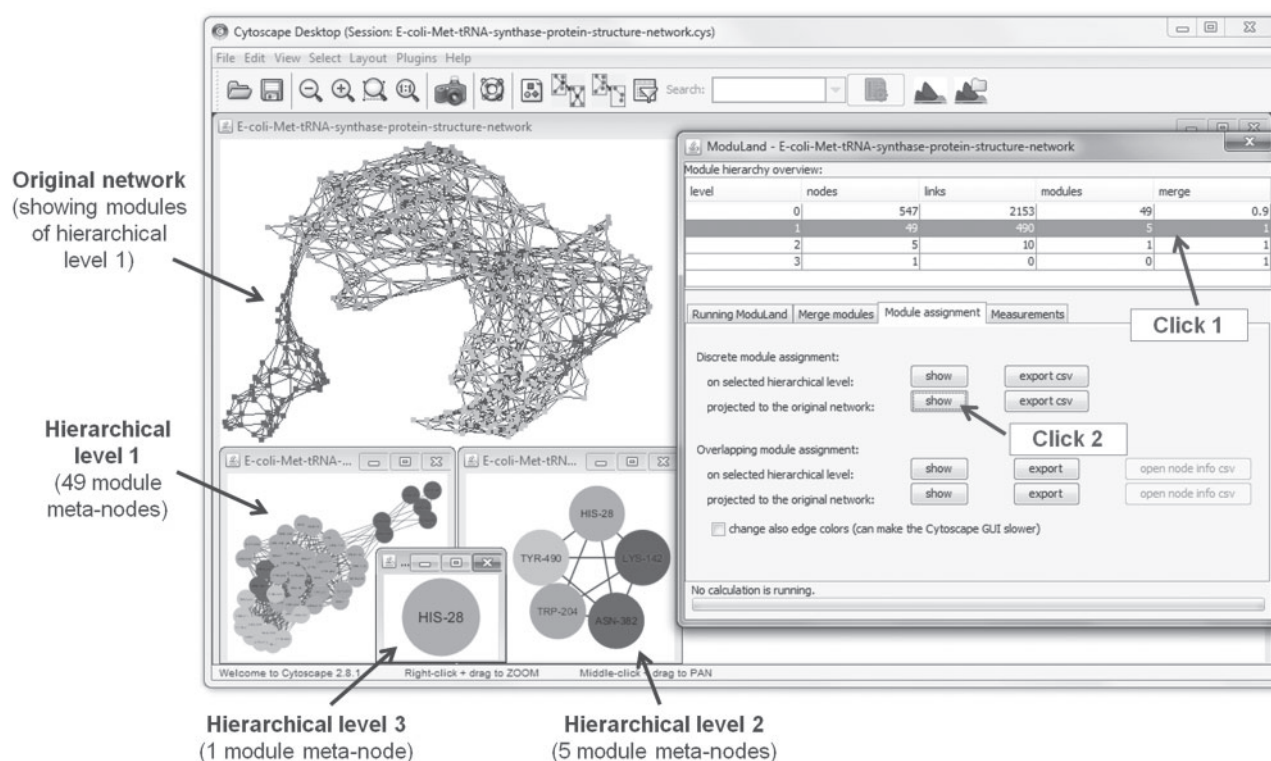


Fig. 1. The hierarchical modular structure determined by the ModuLand Cytoscape plug-in. The left side shows the protein structure network of *E. coli* Met-tRNA synthase and its three hierarchical levels as determined by the plug-in. Each meta-node of a higher hierarchical level represents a module of the level right below. All networks are coloured according to the five modules identified on hierarchical level 1. This colouring option can be performed by the two clicks of the plug-in main dialog box shown on the right

The lower number of modules at higher hierarchical levels may be visualized either using the meta-nodes of the higher hierarchical level itself, or projecting this higher level modular structure to the nodes of the original network. On any level of module hierarchy, nodes or meta-nodes can be visualized assigning them the colour of the module they mostly belong to. This shows a non-overlapping assignment of nodes to modules. Nodes can also be marked by blending the colours of the modules proportional to the overlapping module assignment of the given node. Edges may be optionally visualized as a mix of the colour of their two nodes. The plug-in sets meta-node labels on the higher hierarchical level based on the modules of one level below in the hierarchy. The meta-node on the higher hierarchical level representing the module has the name of the node having the highest modular assignment value for the corresponding module at one level below in the hierarchy.

Node colours can also visualize several node (or meta-node) measures including weighted degree, betweenness centrality, community centrality, overlap and bridgeness (see Supplementary Fig. 2).

The plug-in has an option to merge highly similar module pairs containing roughly the same nodes or meta-nodes with the same intensity. For merging of modules the plug-in offers a correlation histogram, and allows the user to select an appropriate correlation threshold. The runtime complexity of the plug-in version remained $\sim O(n^3)$, as defined earlier (Kovács *et al.*, 2010). To enhance the

performance of the plug-in for calculating the higher hierarchical layers further, we introduced a user-selected optimization. This results in the disappearance of meta-edges with very small weights at the higher hierarchical levels. These low intensity meta-edges are derived from the minor overlaps of distant modules of the lower level. This optimization allowed a speedup in running time by a factor of 7 for larger networks (Supplementary Table S10).

The plug-in is capable to generate overview reports for each hierarchical level. These reports list the number of the nodes (meta-nodes), edges (meta-edges) and modules, the effective number of modules (see Kovács *et al.*, 2010) and the size of each module. The overview also contains the list of the 10 nodes of each module having maximal module assignment value to the respective module (called as the module core). Data related to the module assignment and the calculated measures of nodes (and meta-nodes of higher hierarchical levels) can be exported in a csv or txt format.

The plug-in contains a Help function, and a detailed step-by-step User Guide can also be downloaded from the plug-in webpage: www.linkgroup.hu/modules.php.

3 RESULTS AND CONCLUSION

ModuLand-derived communities of various yeast protein-protein interaction networks gave a functionally meaningful description of the yeast interactome (Kovács *et al.*, 2010). Function of module core proteins proved to be good indicator of the function of the whole

module (Mihalik and Csermely, 2011). Here, we demonstrate the use of the ModuLand Cytoscape plug-in on the protein structure network of *Escherichia coli* Met-tRNA synthase, since an elegant study (Ghosh and Vishveshwara, 2007) showed the existence of four alternative communication paths in this enzyme. The five major sub-domains of Met-tRNA synthase were well reflected by the five modules obtained at the second hierarchical level of the protein structure network (Fig. 1; Supplementary Table S3). Key amino acids of the most frequently used communication path (Ghosh and Vishveshwara, 2007) either belonged to the module cores of the three modules involved in transmission of conformational changes, or were inter-modular nodes between these modules (see Supplementary Table S4). These observations were in agreement with earlier findings (Ghosh and Vishveshwara, 2008; Sethi et al., 2009).

We further demonstrated the use of the ModuLand plug-in by comparing the modular structures of the metabolic networks of the free-living bacterium *E.coli* and the endosymbiont *Buchnera aphidicola* (Pál et al., 2006). *E.coli* metabolic module cores had a significant overlap (Fisher's exact test $P = 1.4 \times 10^{-7}$; see Supplementary Information for more details) with the modules determined earlier by Guimera and Amaral (2005).

Both visual inspection (see Supplementary Figs S4–S7) and numerical values (see Supplementary Table S7) suggested a more differentiated modular structure of the *E.coli* metabolic network than that of *B.aphidicola*. This finding is in agreement with earlier findings (Kreimer et al., 2008; Mihalik and Csermely, 2011; Parter et al., 2007; Samal et al., 2011; Tamames et al., 2007). The difference in modular structure was not likely to be caused by the difference in the size of the *E.coli* and *B.aphidicola* networks (see Supplementary Figs S8 and S9, and Supplementary Tables S7 and S8).

E.coli module cores corresponded to significantly less metabolic functions than those of *B.aphidicola* (0.53 versus 0.67 functions per module core reactions, respectively; bootstrap method $P = 0.0392$). This difference remained even when we used an ensemble of 1000 randomly selected sub-networks of the *E.coli* metabolic network having the same number of nodes or edges as found in the smaller *B.aphidicola* network (see Supplementary Material for more details). Moreover, additional tests suggested that the large twin-modules forming the centre of the *B.aphidicola* network were not responsible for the differences observed in the number of metabolic functions (see Supplementary Material). These results indicated that modules of the metabolic network of an organism from a variable environment (*E.coli*) are more specialized than metabolic network modules of a symbiont having a constant environment (*B.aphidicola*). It is noteworthy that our result is in agreement with earlier findings using non-overlapping modularization (Parter et al., 2007), which is a further indication that the module cores of the plug-in capture well the biologically relevant function of modules.

In conclusion, the ModuLand Cytoscape plug-in provides a user-friendly and efficient method to identify and visualize a hierarchy of extensively overlapping modules, and determines key network positions (like module cores and bridges). As shown by several case studies, modules identified by the plug-in correspond to biologically meaningful groups, module cores help the identification of biological functions and inter-modular nodes have a key role in a variety of biological networks.

ACKNOWLEDGEMENTS

The authors thank the Editor and the anonymous Reviewers for their comments and suggestions, Amit Ghosh (Lawrence Berkeley National Laboratory, Berkeley, CA, USA) and Saraswathi Vishveshwara (Indian Institute of Science, Bangalore, India) for the 3D coordinates of *E.coli* Met-tRNA-synthase, and members of the LINK-Group (www.linkgroup.hu) for their discussions and help.

Funding: EU [FP6-518230, TÁMOP-4.2.2/B-10/1-2010-0013 and FP7-264780]; Hungarian Scientific Research Fund [OTKA K-83314 and PD-75261]; International Human Frontiers Science Program Organization (BP); Momentum Program of the Hungarian Academy of Sciences (BP); and residence at the Rockefeller Foundation Bellagio Center (PC).

Conflict of interest: none declared.

REFERENCES

- Adamcssek, B. et al. (2006) CFinder: Locating cliques and overlapping modules in biological networks. *Bioinformatics*, **22**, 1021–1023.
- Ahn, Y.Y. et al. (2010) Link communities reveal multiscale complexity in networks. *Nature*, **466**, 761–764.
- Bader, G.D. and Houge, C.W.V. (2003) An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics*, **4**, 2.
- Fortunato, S. (2010) Community detection in graphs. *Phys. Rep.*, **486**, 75–174.
- Ghosh, A. and Vishveshwara, S. (2007) A study of communication pathways in methionyl-tRNA synthetase by molecular dynamics simulations and structure network analysis. *Proc. Natl Acad. Sci. USA*, **104**, 15711–15716.
- Ghosh, A. and Vishveshwara, S. (2008) Variations in clique and community patterns in protein structures during allosteric communication: investigation of dynamically equilibrated structures of methionyl tRNA synthetase complexes. *Biochemistry*, **47**, 11398–11407.
- Guimera, R. and Amaral, L.A.N. (2005) Functional cartography of complex metabolic networks. *Nature*, **422**, 895–900.
- Kovács, I.A. et al. (2010) Community landscapes: a novel, integrative approach for the determination of overlapping network modules. *PLoS ONE*, **5**, e12528.
- Kreimer, A. et al. (2008) The evolution of modularity in bacterial metabolic networks. *Proc. Natl Acad. Sci. USA*, **105**, 6796–6981.
- Mihalik, Á. and Csermely, P. (2011) Heat shock partially dissociates the overlapping modules of the yeast protein-protein interaction network: a systems level model of adaptation. *PLoS Comput. Biol.*, **7**, e1002187.
- Morris, J.H. et al. (2011) clusterMaker: a multi-algorithm clustering plugin for Cytoscape. *BMC Bioinformatics*, **12**, 436.
- Pál, C. et al. (2006) Chance and necessity in the evolution of minimal metabolic networks. *Nature*, **440**, 667–670.
- Palla, G. et al. (2005) Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, **435**, 814–818.
- Parter, M. et al. (2007) Environmental variability and modularity of bacterial metabolic networks. *BMC Evol. Biol.*, **7**, 169.
- Rhrissorakrai, K. and Gunsalus, K.C. (2011) MINE: module identification in networks. *BMC Bioinformatics*, **12**, 192.
- Rivera, C.G. et al. (2010) NeMo: Network Module identification in Cytoscape. *BMC Bioinformatics*, **11**, S61.
- Samal, A. et al. (2011) Environmental versatility promotes modularity in genome-scale metabolic networks. *BMC Syst. Biol.*, **5**, 135.
- Sethi, A. et al. (2009) Dynamical networks in tRNA:protein complexes. *Proc. Natl Acad. Sci. USA*, **106**, 6620–6625.
- Shannon, P. et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
- Su, G. et al. (2010) GLay: community structure analysis of biological networks. *Bioinformatics*, **26**, 3135–3137.
- Tamames, J. et al. (2007) Modular organization in the reductive evolution of protein-protein interaction networks. *Genome Biol.*, **8**, R94.