

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

Máté Szalay-Bekő<sup>1,‡</sup> and Robin Palotai<sup>1,‡</sup>, Balázs Szappanos<sup>2</sup>, István A. Kovács<sup>1,3</sup>, Balázs Papp<sup>2</sup> and Péter Csermely<sup>1,\*</sup>

<sup>1</sup>Department of Medical Chemistry, Semmelweis University, Budapest, Hungary

<sup>2</sup>Institute of Biochemistry, Biological Research Centre, Hungarian Academy of Sciences, Szeged, Hungary

<sup>3</sup>Department of Physics, Loránd Eötvös University, Budapest, Hungary and Research Institute for Solid State Physics and Optics, Budapest, Hungary

\*To whom correspondence should be addressed. E-mail: [csermely.peter@med.semmelweis-univ.hu](mailto:csermely.peter@med.semmelweis-univ.hu)

‡ equal contributions

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Associate Editor: XXXXXX

---

## ABSTRACT

**Summary:** The ModuLand plug-in provides Cytoscape users an algorithm determining a.) extensively overlapping network modules; b.) several hierarchical layers of modules, where meta-nodes of the higher hierarchical layer represent modules of the lower layer; c.) module cores predicting the function of the whole module; and d.) key nodes bridging two or multiple modules in complex networks. The plug-in was written in C++, has a detailed JAVA-based graphical interface with various colouring options, can be installed as a single file and 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](mailto:csermely.peter@med.semmelweis-univ.hu)

**Supplementary information:** Supplementary information is available at *Bioinformatics* online.

The widely used Cytoscape program (Shannon *et al.*, 2003) has several very useful clustering plug-ins (Bader and Houge, 2003; Morris *et al.*, 2011; Rhissorakrai and Gunsalus, 2011; Rivera *et al.*, 2010; Su *et al.*, 2010). However, these methods do not focus on extensive modular overlaps; do not build a modular hierarchy, where meta-nodes of the higher level represent modules of the lower level, and do not provide measures identifying the centre of the module, as well as key nodes bridging two or multiple modules (see Suppl. Table 9, and Suppl. Discussion). Here we introduce the Cytoscape plug-in using the most widely applicable version of the ModuLand method family (Kovacs *et al.*, 2010), and demonstrate its ability to determine biologically relevant, extensively overlapping network modules, hierarchical layers of modules, module cores and key inter-modular nodes on protein structure and metabolic networks.

## 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; Kovacs *et al.*, 2010; Mihalik and Csermely, 2011; Palla *et al.*, 2005). Our ModuLand framework (Kovacs *et al.*, 2010) introduced community landscapes, where the  $x$ - $y$  plane is a conventional 2D visualization of the network, while 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; Suppl. Figure 1). 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 (Suppl. Figure 1) yielding extensive overlaps. This concept led to the development of the ModuLand family of network modularization methods (Kovacs *et al.*, 2010).

## 2 SOFTWARE OVERVIEW

The ModuLand Cytoscape (Shannon *et al.*, 2003) plug-in uses the LinkLand influence zone determination method and the ProportionalHill module determination method of our formerly published ModuLand network module determination method family (Kovacs *et al.*, 2010), since these two methods proved to be very good compromises of the fast (but rather inaccurate), and accurate (but rather slow) other ModuLand methods.

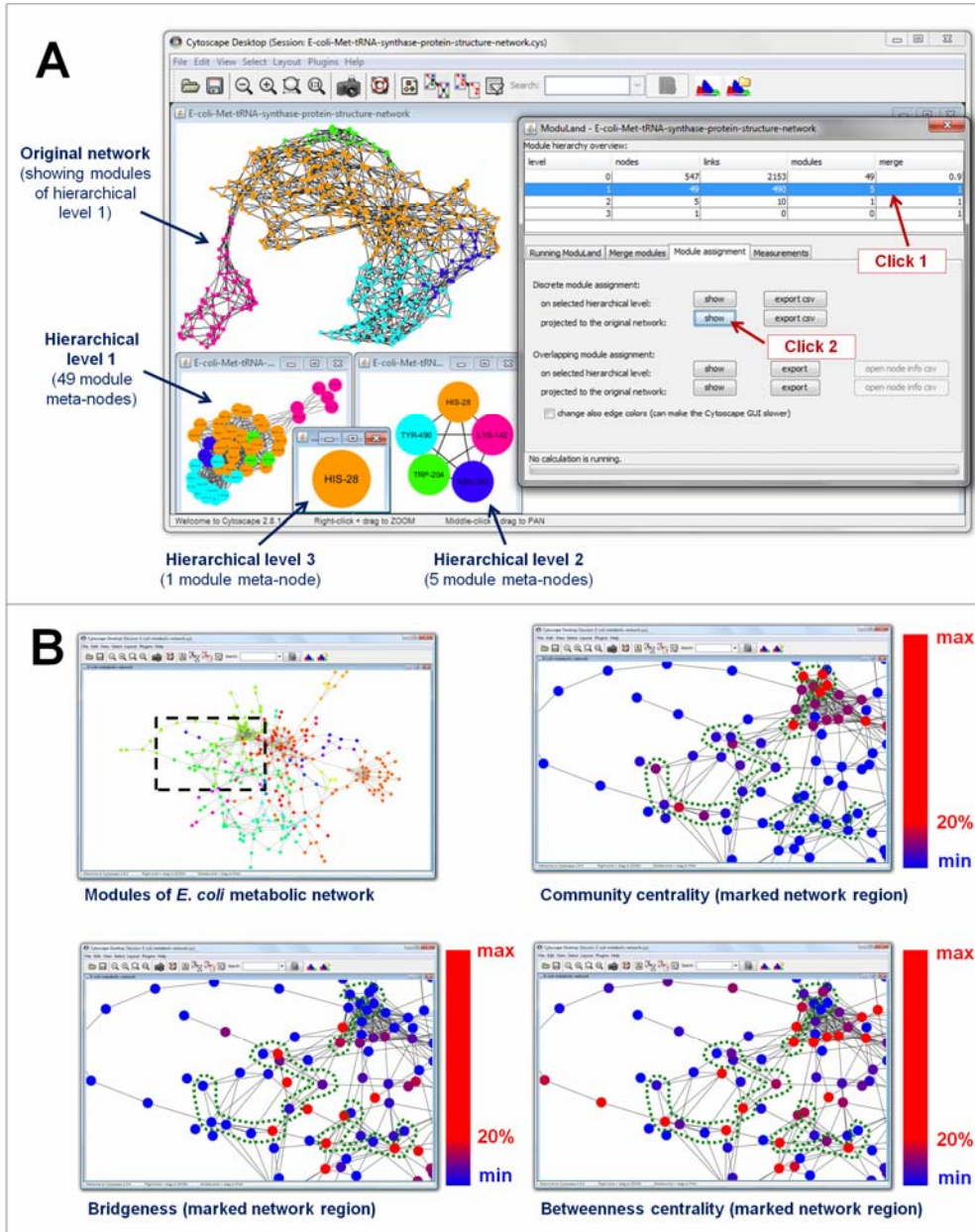
The installation of the ModuLand plug-in corresponds to Cytoscape procedures, which is much easier than the setup required for the earlier version (Kovacs *et al.*, 2010). The program can be distributed as a single jar file. Moreover, the current implementation works on Linux, Windows and Mac OS extending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

The plug-in determines extensively overlapping modules using any Cytoscape undirected network type and weight attribute. It is possible to use any integer, float or Boolean type attribute as edge 'weight' for the plug-in. Moreover, default values can be set, if an

edge does not have the selected attribute (or in case the attribute is zero or a negative number).

The plug-in calculates a set of hierarchical modules, where 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 Figure 1A), until the whole network coalesces into a single meta-node.



**Fig. 1.** Key features of the ModuLand Cytoscape plug-in. The left side of **Panel A** shows the protein structure network of *E. coli* Met-tRNA synthase and its 3 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 5 modules identified on hierarchical level 1. This colouring option can be performed by the 2 clicks of the plug-in main dialog box shown on the right. **Panel B** shows 4 screenshots of the *E. coli* metabolic network obtained by the plug-in. The top left screenshot shows the 23 modules of the original network. The top right, bottom left and bottom right screenshots show the community centrality, bridgeness and betweenness centrality measures of the smaller region of the network marked by dashed black square on the top left screenshot. Measures were visualized by continuous colour mapping using blue-to-red colour scales on the VizMapper panel of Cytoscape from 0 to 20 percentages. Green dots highlight the top 5 core metabolites of the 4 major modules present in the marked network region. All networks were visualized using the Cytoscape Organic yLayout.

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 coloured according to the colour of the module they mostly belong to (giving a non-overlapping assignment of nodes to modules), or by blending the module colours proportional to the overlapping module assignment of the given node. Edges may be optionally coloured 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 like the weighted degree, betweenness centrality, community centrality, overlap and bridgeness (for definitions of the latter 3 measures see Kovacs *et al.*, 2010; see Figure 1B: note that 5 of 6 red nodes of high community centrality overlap the top 5 module core metabolites marked, while 7 of 11 red nodes of high bridgeness are located outside the 5 marked core positions).

The plug-in has an option to merge highly similar module-pairs (containing roughly the same nodes or meta-nodes with the same intensity) offering a correlation histogram, and allowing the user to select an appropriate correlation threshold. The runtime complexity of the plug-in version remained  $\sim O(n^3)$  as defined earlier (Kovacs *et al.*, 2010). To enhance the performance of the plug-in for calculating the higher hierarchical layers further, we introduced a user-selected optimization resulting in the disappearance of meta-edges with very small weights on the higher hierarchical levels derived from the minor overlaps of distant modules of the lower level. This optimization allowed a 7 times faster analysis of a network with more than 10,000 nodes (see Suppl. Table 10).

The plug-in is capable to generate overview reports for each hierarchical level, listing the number of the nodes (meta-nodes), edges (meta-edges) and modules, the effective number of modules (for definition see Kovacs *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](http://www.linkgroup.hu/modules.php).

### 3 RESULTS AND CONCLUSION

In former studies ModuLand-derived communities of various yeast protein-protein interaction networks gave a functionally meaningful description of the yeast interactome (Kovacs *et al.*, 2010). Function of module core proteins proved to be good indicator for 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 *E. coli* Met-tRNA synthase, since an elegant study (Ghosh and Vishveshwara,

2007) showed the existence of 4 alternative communication paths of this enzyme. The 5 major sub-domains of Met-tRNA synthase were reflected well by the 5 modules obtained at the second hierarchical level of the protein structure network (Figure 1A; Suppl. Table 3). Key amino acids of the most frequently used communication path (Ghosh and Vishveshwara, 2007) either belonged to the module cores of the 3 modules involved in transmission of conformational changes, or were inter-modular nodes between these modules (see Suppl. Table 4). These observations were in agreement with earlier findings (Ghosh and Vishveshwara, 2008; Sethi *et al.*, 2009).

We demonstrated the use of the ModuLand plug-in further on metabolic networks comparing their modular structure in the free-living bacterium *E. coli* and in the endosymbiont *B. 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 Suppl. Information for more details) with the modules determined earlier by Guimera and Amaral (2005).

Both visual inspection (see Suppl. Figures 3 to 6) and numerical values (see Suppl. Table 7) suggested a more differentiated modular structure of the *E. coli* metabolic network than that of the *B. aphidicola*. This finding is in agreement with our former results (Mihalik and Csermely, 2011) and earlier findings (Kreimer *et al.*, 2008; Parter *et al.*, 2007; Samal *et al.*, 2011). 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, since the difference remained, when either two ensembles of 1000 randomly selected, node or edge size-matched sub-networks of the *E. coli* network, or the networks of the common 103 nodes of the two metabolic networks were compared (see Suppl. Figures 7 and 8, and Suppl. Tables 7 and 8).

*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 Suppl. Information 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 Suppl. Information). 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 reflect 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). Plug-in modules correspond to biologically meaningful groups, module cores help the identification of biological function, and inter-modular nodes have a key functional 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](http://www.linkgroup.hu)) for their discussions and help. *Funding*: This work was supported by the EU [grant numbers FP6-518230, TAMOP-4.2.2/B-10/1-2010-0013, FP7-264780], by the Hungarian Scientific Research Fund [grant numbers OTKA K-83314 and PD-75261], by the International Human Frontiers Science Program Organization (BP), by the Momentum Program of the Hungarian Academy of Sciences (BP) and by a 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.
- Feist, A.M. et al. (2007) A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol. Syst. Biol.*, **3**, 121.
- 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.
- Thomas, G.H. et al. (2009) A fragile metabolic network adapted for cooperation in the symbiotic bacterium *Buchnera aphidicola*. *BMC Syst. Biol.*, **3**, 24.