Indentifying sub-network functional modules in protein undirected networks

Identifying sub-network functional modules in protein undirected networks

Massimo Natale[1], Alfredo Benso[1], Stefano Di Carlo[1] and Elisa Ficarra[1]
[1] *Department Control and Computer Engineering, Politecnico di Torino, Corso Castelfidardo 34, Turin, Italy*
{massimo.natale, alfredo.benso, stefano.dicarlo, elisa.ficarra}@polito.it

Keywords: Systems Biology, Network Analysis, Pathways Analysis, Network Modules, Cytoscape, Budd-Chiari Syndrome.

Abstract: Protein networks are usually used to describe the interacting behaviours of complex biosystems. Bioinformatics must be able to provide methods to mine protein undirected networks and to infer sub-networks of interacting proteins for identifying relevant biological pathways.
Here we present FunMod an innovative Cytoscape version 2.8 plugin able to identify biologically significant sub-networks within informative protein networks, enabling new opportunities for elucidating pathways involved in diseases. Moreover FunMod calculates three topological coefficients for each sub-network, for a better understanding of the cooperative interactions between proteins and discriminating the role played by each protein within a functional module.
FunMod is the first Cytoscape plugin with the ability of combining pathways and topological analysis allowing the identification of the key proteins within sub-network functional modules.

1 INTRODUCTION

Systems biology broadly uses networks to model and discover emergent properties among genes, proteins and other relevant biomolecules.

Theoretical studies have indicated that biological networks share many features with other types of networks, as computer or social networks. The biological networks allow the application of several mathematical and computational methods of the graph theory to biological studies (Barabasi, 2004)(Huber, 2007). The computational analysis of biological networks has therefore become increasingly used to mine the complexity of cellular processes and signalling pathways (Spirin, 2003). Many types of biological networks do exist, depending on the information associated to their nodes and edges. In general, they can be classified as directed or undirected networks (Pieroni, 2008). In directed networks the nodes are molecules and edges indicate causal biological interactions between them, such as transcription and translation regulations (Li, 2012). Instead, in undirected networks, an edge indicates a shared property, such as sequence similarity (Kuchaiev, 2011), gene co-expression (Prifti, 2010), protein-protein interaction (Chen, 2013), or term co-occurrence in the scientific literature (Gatti, 2012)(Lee, 2010)(Gabow, 2008). In order to extract relevant biological insights from undirected networks, also referred to as informative networks (Lysenko, 2011), it is useful to complement the topological information with independent biological information retrieved from Gene Ontology or pathway databases. Often the goal is to identify densely interconnected areas and correlate them to a biological function (Bu, 2003)(Weatheritt, 2012). Several algorithms and bioinformatics tools have been proposed for partitioning the network into structural modules, or for clustering sub-network modules within an informative network (Thomas, 2010)(Shen, 2012).

Several bioinformatics researchers developed Cytoscape plugins (Smoot, 2011) to mine functional modules in a varieties of networks types, such as: Clust&See (Spinelli, 2013), clusterMaker (Morris, 2011), CyClus3D (Audenaert, 2011), GLay (Su, 2010), and Enrichment map (Merico, 2010). These methods infer biological functions on the basis of topological properties following a "first topological clustering" strategy aimed at partitioning complex networks into modules and assigning to these sub-

network a certain biological function (Milo, 2002). However these strategies are heavily influenced by the topological structure of the network itself, and by the way the network is constructed (Dong, 2007).

In practice, the connectivity of informative networks is established by experimental methods, which can lead to sample a subnet of the real biological network (Aittokallio, 2006)(Heatha, 2009). Often, biases in the sampling strategies lead to apparent scale-free topologies not corresponding to the actual complete network topology. As an alternative to the "first topological clustering" methods, some authors used a "first biological assignment" strategy capable of analysing gene networks and extracting functional modules starting from the biological enrichment analysis. Thanks to this approach several Cytoscape plugins have been implemented such as BiNGO (Maere, 2005), ClueGO (Bindea, 2009), ClusterViz (Cai, 2010), JEPPETTO (Glaab, 2012) and Reactome (Jupe, 2012).

BiNGO and ClueGO are widely used tools to determine which biological functions are statistically over-represented in a list of genes or a network. These plugin offer the possibility to calculate enrichment by using different algorithms. However, they don't evaluate the connectivity between genes.

A recent focus of bioinformatics has been to develop computational tools able to mine the connectivity of gene networks revealing sets of molecules that participate in a common biological function (Mitra, 2013).

Reactome is based on an un-weighted human protein functional interaction (FI) network, on which the authors calculated the Pearson correlation coefficients among all functional interaction pairs in the biological data base, and then used the Markov clustering algorithm to cluster the weighted network into a series of gene interaction modules. Each module consists of a set of genes that are both connected in the protein FI network and highly correlated in biological databases (Wu, 2012). However, the protein FI network is a previously defined network and the user is not able to analyse its own protein informative networks. Moreover this approach does not consider the topology or the connectivity of gene interaction modules.

JEPPETTO identifies functional associations between genes and pathways using protein interaction networks and topological analysis. Even though JEPPETTO combines network analysis with functional enrichment, this tool requires a list of genes in input and the selection of a database of interest from which reference gene sets will be extracted.

To the best of our knowledge, there is no software available to identify sub-networks of genes that are highly connected and belonging to the same biological pathways. Moreover, the cited plugins do not allow to analyse user defined undirected networks, such as protein–protein interaction, functional association, gene co-expression and literature co-occurrence.

Using a "first biological assignment" strategy we developed a new method called FunMod. FunMod analyses informative networks using functional and pathway information. It identifies and extracts sub-network modules that are statistically significantly enriched. According to the principle that a sub-network functional module is a distinct group of interacting molecules driving a common biological process.

Moreover FunMod analyses the topological features of the identified sub-networks performing a measure of three topological scores, in order to identify specific modules that can discriminate the role of each protein within the single pathway.

FunMod is proposed to be an efficient tool for identifying relevant pathways involved in protein informative networks. The identification of sub-network functional modules in a protein informative network can be a useful method for exploring biomedical information and inferring automated functional hypotheses between biomedical contexts (Liekens, 2011). Assessment of the sub-network topological proprieties can consequently be used for the ranking of genes in the context of a research domain, such as a disease (Dolinski, 2013).

The FunMod challenge is the analysis of user defined undirected networks in order to provide more realistic networks that incorporate information on particular cellular contexts, developmental states or disease conditions.

FunMod is implemented as a Cytoscape version 2.8 plugin and compared to the current state of the art our plugin enables the analysis of any gene undirected network in Cytoscape.

2 IMPLEMENTATION

FunMod analyzes the protein informative network displayed in the Cytoscape Main Network View window. The plugin supports many standard protein annotation formats and the protein nodes can be identified (node ID) by six different dictionaries: Entrez Gene ID, Ensembl, Official Symbol (HGNC

symbol), NCBI Reference Sequence, UniGene, Uniprot Entry Name.

FunMod recursively selects all edges and assigns a functional annotation to an edge when the two linked nodes are annotated in the same biological group or pathway in the ConsensusPathDB (DB) database (Kamburov, 2013). In other words FunMod considers a network G = (V, E) with n vertices V joined with edges E, and collects, for each ConsensusPathDB pathway, pairs of linked nodes modelling a functional sub-network $G_p = (V_p, E_p)$, where $V_p \subseteq$ pathway and $E_p \in E$.

Afterwards FunMod extracts all the pairs of nodes annotated for the same pathway in a new sub-network, it tests the statistical significance, and calculates the topological properties of the sub-network. In this way FunMod is able to identify sub-networks that are statistically enriched in biological functions and that show interesting topological features.

The statistical significance is determined by calculating the p-value preforming a hypergeometric test, a well-established method used in gene enrichment analysis (Maere, 2005).

The hypergeometric probability is based on the following formula:

$$h(x; X, n, N) = [XCx][N - XCn - x]/[NCn] \quad (1)$$

where x is the number of nodes of the sub-network (the items in the sample that are classified as successes), n is the number of genes in the network (items in the sample); X is the number of genes annotated in the DB with that pathway (items in the population that are classified as successes); and N the number of all genes annotated in DB (items in the population). FunMod preserves the sub-networks with a p-value < 0.05.

For a better understanding of the systemic functions and the cooperative interactions between genes within the functional modules, FunMod checks whether the sub-network topology fits into a specific module. Network modules represent patterns in complex networks occurring significantly more often than in randomized networks. They consist of sub-graphs of local interconnections between network elements. FunMod calculates a fitting score of each sub-network for three modules: clique, star, and path; the most common motifs that are found in various networks (Pavlopoulos, 2011).

A clique is a sub-network in which all genes are connected to each other. Cliques are the most widely used modules for assigning a biological function to a topological sub-network. FunMod calculates the clique score using the Graph Density (GD) formula:

$$GD = \frac{2E}{n \times (n - 1)} \quad (2)$$

where E is the number of edges in the sub-network and n the number of genes in the sub-network.

The star module, particularly interesting for identifying drug targets, is characterized by a central gene with a high degree (the hub) connected to a set of first-degree neighbours loosely connected among each other. In a star sub-network the hub gene has influence on its neighbourhood genes and possibly on the whole network. In order to identify a star module, FunMod calculates the sub-network centralization (CE) using the formula:

$$CE = \frac{n}{(n - 2)} \times \left[\frac{\max(k)}{n - 1} - GD\right] \quad (3)$$

where max(k) is the highest degree in the sub-network.

The path module corresponds to a real pathway where the genes contribute to a signal transduction. The path score is calculated as the sub-network diameter (D), the maximum length of all shortest paths between any two connected nodes, using the formula:

$$D = \max_{i,j} \delta_{min}(i, j) \quad (4)$$

where δ_{min} is the minimum path between two nodes i and j of the network.

FunMod displays into the Cytoscape Results Panel the identified pathways, ranking them, according to their p-value. For each pathway FunMod displays its clique, star and path coefficient.

By clicking the pathway button FunMod selects the corresponding nodes in the network, and by clicking the “View subnet” button, it creates a new network containing only those genes and edges annotated with that pathway. Finally FunMod enables to save the results into a tab-delimited file. FunMod plugin, user guide, screenshot and demo networks can be freely downloaded from the SourceForge project page at:
http://sourceforge.net/projects/funmodnetwork/.

3 RESULTS AND DISCUSSION

To showcase FunMod we analyzed a bibliometric network of proteins related Budd-Chiari syndrome, an uncommon condition characterized by obstruction of the hepatic venous outflow tract (Ferral, 2012) (MacNicholas, 2012). We chose this syndrome because it is a rare disease whose diagnosis and management may be difficult. Often Budd-Chiari syndrome is fatal if not treated optimally and none of the current medical therapy are based on a high-level of evidence due to the difficulty in making appropriate clinical studies on a sufficient number of patients with this rare disease (Hefaiedh, 2013). In this context, a meta-analysis can be effective to provide valuable information for researchers and to improve the quality and usefulness of literature knowledge (Walker, 2008).

In this meta-analysis we used ProteinQuest (PQ), an advanced text-mining tool that resorts to the web services offered by PubMed to identify all proteins co-cited in the same abstract or figure caption (Giordano, 2013), and to import the literature co-occurrence network into Cytoscape. We submitted a query to PQ using as starting search terms "Budd-Chiari Syndrome" and its NCBI MeSH-Terms alias. PQ retrieved more than 2200 papers.

The network obtained by PQ was an undirected protein-protein network, consisting of 76 nodes, proteins cited in at least one document, 174 edges and co-occurrence of two proteins in at least one abstract or caption. A screenshot of the plugin is shown in Figure 1, whereas a quick guide of the Budd-Chiari syndrome network can be downloaded at the SourceForge project page.

Using FunMod we identified 33 different biological pathways that are significantly enriched in the Budd-Chiari syndrome network. Table 1 shows the most relevant pathways. Identifying sub-network functional modules in protein undirected networks allows to reduce network complexity, to cluster proteins on the basis of common biological functions and to discover the mechanisms behind disease (Royer, 2008).

The most significant pathways, showed in Table 1, are related to the coagulation cascades, and this finding is in agreement with the knowledge that Budd-Chiari is a disorder frequently characterized by the thrombotic obstruction of hepatic venous outflow (Wang, 2013) (Akbulut, 2013). This result proves that FunMod, relying on the ABC principle (Frijters, 2013)(Li, 2013) to establish new relationships between proteins from literature co-occurrence network, can be effectively used to combine information from the scientific literature with information from pathways databases, and to support the discovery of new knowledge. Moreover, thanks to FunMod we calculated the three topological coefficients of each pathway sub-network, and in Figure 2 we show the three sub-network functional modules with the higher coefficient for each class of topological module.

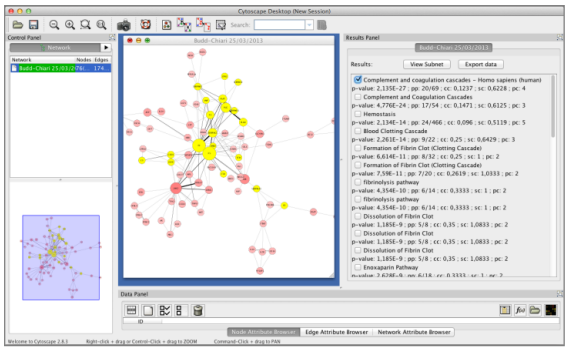

Figure 1: Screenshot of the plugin in action

In Figure 2 we show the Fibrinolysis pathway as example of a cliques module (panel A), the JAK-STAT pathway as example of a star module (panel B), and the Platelet activation, signalling and aggregation as example of path module (Panel C).

Studying the topological feature of a sub-network functional module we can extract information about the state of the art of the knowledge of that pathways in the context of Budd-Chiari syndrome.

The cliques represent a highly connected sub-network where the proteins of the pathway are highly co-studied and co-cited. The Fibrinolysis pathway is a central pathway in Budd-Chiari syndrome since prothrombotic tendency is caused by abnormalities in the coagulation of fibrinolysis pathways (Smalberg, 2011). So it is not surprising that the Fibrinolysis pathway functional sub-netwok is represented by a clique.

A star module represents a functional sub-network where a protein plays a central role in the pathway, indeed the JAK2 is the hub (the central node) in the JAK-STAT pathway sub-network. Moreover, a mutation in JAK2 tyrosine kinase (JAK2V617F) is frequently found in patients with Budd-Chiari syndrome (Qi, 2011)(Patel, 2012).

Finally, the path module represents a significant pathway in the context of the disease, which is not studied completely. The Platelet activation, signalling and aggregation is a therapeutic target for the treatment of Budd Chiari syndrome (Chaudhuri, 2012), VWF (Alkim, 2012) and SERPINA1

(Hourigan, 2012), two peripheral sub-network proteins, recently described as markers of abnormalities of coagulation playing some role in provoking thrombosis.

4 CONCLUSION

Using a "first topological assignment" strategy to identify sub-network functional modules, such as stars and cliques, can be tricky because informative networks are known to have a huge number of edges that are not always pertinent to biological functions (Narayanasamy, 2004).

In this work we presented FunMod, a new Cytocape 2.8 plugin able to analyze undirected protein networks, such as co-occurrence networks, and guide the discovery of sub-network functional modules. A functional module can be considered as a distinct group of interacting proteins (Mitra, 2013) pathway relevant to a condition of interest.

FunMod identifies, within an informative network, pairs of nodes belonging to the same biological pathways, and assesses their statistical significance. It then analyses the topology of the identified sub-network to infer the topological relations (motifs) of its nodes. This sub-network profiling combined with database biological knowledge extraction will help us to better understand the biological significance of the system.

In this work the network topology is influenced by the biomedical knowledge, since the link between two proteins was established when two gene symbols appear in the same Medline record. The study of the connection between biomedical concepts by using co-occurrence network extracted from MEDLINE, proved capable of guiding the discovery of novel knowledge from scientific literature (Narayanasamy, 2004).

FunMod was tested using the co-occurrence network of proteins cited in Budd-Chiary syndrome papers, identifying 33 different biological pathways that are significantly enriched, and proving to be a useful tool for a better understanding of the cooperative interactions between proteins and discriminating the role played by each protein within a functional module.

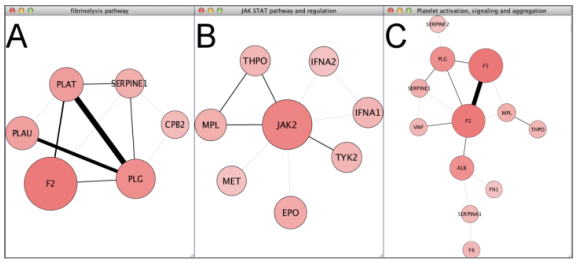

Figure 2. The three sub-network functional modules with the higher coefficient for each class of topological module: A) cluster module; B) star module; C) pathway module.

COMPETING INTERESTS

The authors declare that they have no competing interests.

Table 1: FunMod we identifyed 33 different biological pathways that are significantly enriched in the Budd-Chiari syndrome network (PCov: Pathway coverage; CC: Clique coefficient; SC: star coefficient; PC: Path coefficient)

Pathway	***PCov***	***p-value***	***CC***	***SC***	***PC***
Complement and coagulation cascades	20/69	2,13E-27	0,1237	0,6228	4
Hemostasis	24/466	2,13E-14	0,096	0,5119	5
Fibrinolysis pathway	6/14	4,35E-10	0,3333	1	2
Heparin Pathway	6/18	2,63E-9	0,3333	1	2
Vitamin B12 Metabolism	8/51	3,57E-9	0,1786	1,0952	2
Platelet degranulation	9/83	1,03E-8	0,0972	0,3571	3
Response to elevated platelet cytosolic Ca2+	9/88	1,75E-8	0,0972	0,3571	3
Folate Metabolism	8/66	2,94E-8	0,1786	1,0952	2
JAK STAT Molecular Variation 1	9/99	4,95E-8	0,125	0,8036	2
Selenium Pathway	8/77	1,01E-7	0,1786	1,0952	2
Lepirudin Pathway	5/17	1,24E-7	0,3	1,1667	2
Dicumarol Pathway	5/18	1,71E-7	0,3	1,1667	2
PI3K-Akt signaling pathway	7/334	1,29E-2	0,1429	0,7333	2

ACKNOWLEDGEMENTS

This work was conducted as a part of the project "OSDD" (CUP B15G13000010006) partially funded by Valle d'Aosta Region Government under the "Grant for creation and development of Research Units" (dgr. 1988/2011; dgr. 538/2012)" regional funding program.

REFERENCES

Aittokallio, T., Schwikowski, B., 2006. Graph-based methods for analysing networks in cell biology. *Brief Bioinform*, 7:243-55.

Akbulut, S., Yilmaz, M., Kahraman, A., Yilmaz, S., 2013. Budd-Chiari syndrome due to giant hydatid cyst: a case report and brief literature review. *J Infect Dev Ctries*, 7:489-93.

Alkim, H., Ayaz, S., Sasmaz, N., Oguz, P., Sahin, B., 2012. Hemostatic abnormalities in cirrhosis and tumor-related portal vein thrombosis. *Clin Appl Thromb Hemost*, 18:409-15.

Audenaert, P., Van Parys, T., Brondel, F., Pickavet, M., Demeester, P., Van de Peer, Y., Michoel, T. 2011. CyClus3D: a Cytoscape plugin for clustering network motifs in integrated networks. *Bioinformatics*, 27:1587-8.

Barabasi, A.L., Oltvai, Z.N., 2004. Network biology: understanding the cell's functional organization. *Nat Rev Genet*, 5:101-13.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pagès, F., Trajanoski, Z., Galon, J., 2009. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. 25:1091-3.

Bu, D., Zhao, Y., Cai, L., Xue, H., Zhu, X., Lu, H., Zhang, J., Sun, S., Ling, L., Zhang, N., Li, G., Chen, R., 2003. Topological structure analysis of the protein-protein interaction network in budding yeast. *Nucleic Acids Res*, 31:2443-50.

Cai, J., Chen, G., & Wang, J., 2010. ClusterViz: a Cytoscape plugin for graph clustering and visualization. *School of Information Science and Engineering,* 1.

Chaudhuri, M., Jayaranganath, M., Chandra, V.S., 2012. Percutaneous recanalization of an occluded hepatic vein in a difficult subset of pediatric Budd-Chiari syndrome. *Pediatr Cardiol*, 33:806-10.

Chen, L., Xuan, J., Riggins, R.B., Wang, Y., Clarke, R., 2013. Identifying protein interaction subnetworks by a bagging Markov random field-based method. *Nucleic Acids Res*, 41, e42.

Dolinski, K., Chatr-Aryamontri, A., Tyers, M. 2013. Systematic curation of protein and genetic interaction data for computable biology. *BMC Biol*. 11:43.

Dong, J., Horvath, S., 2007. Understanding network concepts in modules. *BMC Syst Biol*, 4:1-24.

Ferral, H., Behrens, G., Lopera, J., 2012. Budd-Chiari syndrome. *AJR Am J Roentgenol*, 199:737-45.

Frijters, R., van Vugt, M., Smeets, R., van Schaik, R., de Vlieg, J., Alkema, W., 2010. Literature mining for the discovery of hidden connections between drugs, genes and diseases. *PLoS Comput Biol*, 6:doi:10.1371/journal.pcbi.1000943.

Gabow, A.P., Leach, S.M., Baumgartner, W.A., Hunter, L.E., Goldberg, D.S., 2008. Improving protein function prediction methods with integrated literature data. *BMC Bioinformatics*, 15:9-198.

Gatti, S., Leo, C., Gallo, S., Sala, V., Bucci, E., Natale, M., Cantarella, D., Medico, E., Crepaldi, T., 2012. Gene expression profiling of HGF/Met activation in neonatal mouse heart. *Transgenic Res*, 3:579-93

Giordano, M., Natale, M., Cornaz, M., Ruffino, A., Bonino, D., Bucci, E.M., 2013. iMole, a web based image retrieval system from biomedical literature. *Electrophoresis*. 34:1965-8.

Glaab, E., Baudot, A., Krasnogor, N., Schneider, R., Valencia, A., 2012. EnrichNet: network-based gene set enrichment analysis. *Bioinformatics*, 28:i451-i457.

Hartwell, L.H., Hopfield, J.J., Leibler, S., Murray, A.W. 1999. From molecular to modular cell biology. Nature 402, C47–C52.

Heatha, A.P., Kavrakia, L.E., 2009. Computational challenges in systems biology. *Computer Science Review*, 3: 117.

Hefaiedh, R., Cheikh, M., Marsaoui, L., Ennaifer, R., Romdhane, H., Ben Nejma, H., Bel Hadj, N., Arfa, N., Khalfallah, M.T., 2013. The Budd-Chiari syndrome. *Tunis Med*, 91:376-81.

Hourigan, S.K., Anders, R.A., Mitchell, S.E., Schwarz, K.B., Lau, H., Karnsakul, W., 2012. Chronic diarrhea, ascites, and protein-losing enteropathy in an infant with hepatic venous outflow obstruction after liver transplantation. *Pediatr Transplant*, 16:E328-31

Huber, W., Carey, V.J., Long, L., Falcon, S., Gentleman, R. 2007 Graphs in *molecular biology. BMC Bioinformatics*, 8:S8.

Jupe, S., Akkerman, J.W., Soranzo, N., Ouwehand, W.H., 2012. Reactome - a curated knowledgebase of biological pathways: megakaryocytes and platelets. *J Thromb Haemost*, doi: 10.1111/j.

Kamburov, A., Stelzl, U., Lehrach, H., Herwig, R., 2013. The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res*, 41:D793-800.

Kuchaiev, O., Przulj, N., 2011. Integrative network alignment reveals large regions of global network similarity in yeast and human. *Bioinformatics*, 27:1390-6.

Lee, I., Ambaru, B., Thakkar, P., Marcotte, E.M., Rhee, S.Y., 2010. Rational association of genes with traits using a genome-scale gene network for Arabidopsis thaliana. *Nat Biotechnol*, 28:149-56.

Li, C., Liakata, M., Rebholz-Schuhmann, D., 2013 Biological network extraction from scientific

literature: state of the art and challenges. *Brief Bioinform*.

Li, J., Hua, X., Haubrock, M., Wang, J., Wingender, E. 2012. The architecture of the gene regulatory networks of different tissues. *Bioinformatics*, 28:i509-i514.

Liekens, A.M., De Knijf, J., Daelemans, W., Goethals, B., De Rijk, P., Del-Favero, J., 2011. BioGraph: unsupervised biomedical knowledge discovery via automated hypothesis generation. *Genome Biol*. 12:R57.

Lysenko, A., Defoin-Platel, M., Hassani-Pak, K., Taubert, J., Hodgman, C., Rawlings, C.J., Saqi, M., 2011. Assessing the functional coherence of modules found in multiple-evidence networks from Arabidopsis. *BMC Bioinformatics*, 12:203.

MacNicholas, R., Olliff, S., Elias, E., Tripathi, D., 2012. An update on the diagnosis and management of Budd-Chiari syndrome. *Expert Rev Gastroenterol Hepatol*, 6:731-44.

Maere, S., Heymans, K., Kuiper, M., 2005. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, 21:3448-9.

Merico, D., Isserlin, R., Stueker, O., Emili, A., Bader, G.D., 2010. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One*, 5: e13984.

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., Alon, U.. 2002 Network motifs: simple building blocks of complex networks. *Science*, 298:824-7

Mitra, K., Carvunis, A.R., Ramesh, S.K., Ideker, T.. 2013. Integrative approaches for finding modular structure in biological networks. *Nat Rev Genet*. 14:719-32.

Morris, J.H., Apeltsin, L., Newman, A.M., Baumbach, J., Wittkop, T., Su, G., Bader, G.D., Ferrin, T.E., 2011. clusterMaker: a multi-algorithm clustering plugin for Cytoscape. *BMC Bioinformatics*, 12:436

Narayanasamy, V., Mukhopadhyay, S., Palakal, M., Potter, D.A., 2004. TransMiner: mining transitive associations among biological objects from text. *J Biomed Sci*. 11:864-73.

Patel, R.K., Lea, N.C., Heneghan, M.A., Westwood, N.B., Milojkovic, D., Thanigaikumar, M., Yallop, D., Arya, R., Pagliuca, A., Gäken, J., Wendon, J., Heaton, N.D., Mufti, G.J., 2006. Prevalence of the activating JAK2 tyrosine kinase mutation V617F in the Budd-Chiari syndrome. Gastroenterology, 130:2031-8.

Pavlopoulos, G.A., Secrier, M., Moschopoulos, C.N., Soldatos, T.G., Kossida, S., Aerts, J., Schneider, R., Bagos, P.G., 2011. Using graph theory to analyze biological networks. *BioData Min*, 4:10.

Pieroni, E., de la Fuente van Bentem, S., Mancosu, G., Capobianco, E., Hirt, H., de la Fuente, A. 2008. Protein networking: insights into global functional organization of proteomes. *Proteomics*, 8:799-816.

Prifti, E., Zucker, J.D., Clément, K., Henegar, C., 2010. Interactional and functional centrality in transcriptional coexpression networks. *Bioinformatics*, 26:3083-9.

Qi, X., Yang, Z., Bai, M., Shi, X., Han, G., Fan, D., 2011. Meta-analysis: the significance of screening for JAK2V617F mutation in Budd-Chiari syndrome and portal venous system thrombosis. *Aliment Pharmacol Ther*, 33:1087-103.

Royer, L., Reimann, M., Andreopoulos, B., Schroeder, M., 2008. Unraveling protein networks with power graph analysis. *PLoS Comput Biol*, 4:e1000108.

Shen, R., Goonesekere, N.C., Guda, C., 2012. Mining functional subgraphs from cancer protein-protein interaction networks. *BMC Syst Biol*, 6:S2.

Smalberg, J.H., Kruip, M.J., Janssen, H.L., Rijken, D.C., Leebeek, F.W., de Maat, M.P., 2011. Hypercoagulability and hypofibrinolysis and risk of deep vein thrombosis and splanchnic vein thrombosis: similarities and differences. *Arterioscler Thromb Vasc Biol*, 31:485-93.

Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L., Ideker, T., 2011. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics*, 27:431-32.

Spinelli, L., Gambette, P., Chapple, C.E., Robisson, B., Baudot, A., Garreta, H., Tichit, L., Guénoche, A., Brun, C., 2013. Clust&See: A Cytoscape plugin for the identification, visualization and manipulation of network clusters. *Biosystems*, 113:91-5.

Su, G., Kuchinsky, A., Morris, J.H., States, D.J., Meng, F., 2010. GLay: community structure analysis of biological networks. *Bioinformatics*, 26:3135-7.

Thomas, S., Bonchev, D., 2010. A survey of current software for network analysis in molecular biology. *Hum Genomics*, 4:353-60.

Walker, E., Hernandez, A.V., Kattan, M.W., 2008. Meta-analysis: Its strengths and limitations. *Cleveland Clinic Journal of Medicine*, 75:431-439.

Wang, R., Meng, Q., Qu, L., Wu, X., Sun, N., Jin, X., 2013. Treatment of Budd-Chiari syndrome with inferior vena cava thrombosis. *Exp Ther Med*, 5:1254-1258.

Weatheritt, R.J., Jehl, P., Dinkel, H., Gibson, T.J., 2012. iELM-a web server to explore short linear motif-mediated interactions. *Nucleic Acids Res*, 40:W364-9.

Wu, G., Stein, L., 2012. A network module-based method for identifying cancer prognostic signatures. *Genome Biol*, 13:R112.

Spirin, V., Mirny, L.A., 2003. Protein complexes and functional modules in molecular networks. *Proc Natl Acad Sci USA*. 100:12123-8.

APPENDIX

Operating system(s): Platform independent
Programming language: Java
Other requirements: Cytoscape 2.8.4
License: FunMod is freely available for non-commercial purposes, (GNU GPL, FreeBSD etc.)
Any restrictions to use by non-academics: e.g. licence needed