Systems Biology

CellNetVis: a web tool for visualization of biological networks using force-based layout constrained by cellular components

Henry Heberle[1], Hugo Henrique Slepicka[2], Guilherme P. Telles[3], Rosane Minghim[1#], Gabriela Vaz Meirelles[4#*]

[1]Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, SP, Brazil; [2]Brookhaven National Laboratory, Upton, New York, USA; [3]Instituto de Computação, Universidade Estadual de Campinas, São Carlos, SP, Brazil; [4]Laboratório Nacional de Biociências, Centro Nacional de Pesquisa em Energia e Materiais, Campinas, SP, Brazil

[#]These authors contributed equaly to this work

Received on August 31, 2015

ABSTRACT

Summary: Proper visualizations of networks allow a deeper understanding of biological systems. Some pathways are related to specific functions that occur in specific regions of the cell. Splitting networks across cellular components helps to quickly identify where network elements are located and where they concentrate. Here we introduce CellNetVis, a web tool to display biological networks that restricts nodes to their cellular components. The layouts are built through force-based scheme, a ~~very~~ used technique to draw a network according to its topology. CellNetVis offers diverse interactive features that help users in identifying specific nodes, their interaction partners and values stored in their attributes.

Availability and implementation: CellNetVis (http://1.cellstructurenetwork.appspot.com/) is an open-source software written in HTML and JavaScript, integrated in the IIS platform (http://www.lge.ibi.unicamp.br/lnbio/IIS/).

Contact: gabriela.meirelles@lnbio.cnpem.br

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

1 INTRODUCTION

With the advent of "omics" science, analyses performed from the screening of a wide range of physical, genetic and chemical-genetic interactions have brought new perspectives in contemporary biology, as ~~it~~ provides new clues in protein/gene function, helps to understand how the metabolic, regulatory and signaling pathways are organized and facilitate the validation of therapeutic targets and potential drugs. An appropriate display of the data is crucial for understanding biological interaction networks, particularly regarding high-throughput analysis.

Visually organizing network nodes into cellular components can help understanding the distribution of network components over the cellular structure. For instance, the position of nodes can unveil the relation among cell components in a specific network. Also, it is common to query just a subnetwork of an entire interactome, so when users ~~query some~~ specific pathways by a list of their components (e.g. gene symbols) they ~~can~~ easily see where these pathways may occur in the cell if such a layout ~~is~~ available.

*To whom correspondence should be addressed.

Currently, only a few tools provide the capability of visually organizing networks by cellular components. T~~he~~ Cerebral (Barsky *et al.*, 2007), a Cytoscape (Shannon *et al.*, 2003) plugin, can divide the network into subcellular regions represented by parallel rectangles, one over the other. Kojima et al. (Kojima *et al.*, 2008) developed a grid layout over a cellular diagram, representing the cellular components properly. Despite the capability of drawing over a cell representation, this tool cannot handle big and dense networks because of limitations of the grid layout approach.

In this context, we developed a web tool for displaying biological networks in order to transform static into interactive data. The tool accept as input networks generated by queries in the IIS platform (Carazzolle *et al.*, 2014) in the XGMML format or any network containing an unique cellular component annotation for each protein in the same file format. The tool presents a cellular diagram showing the main partitions and organelles according to the Gene Ontology (GO) cellular component database (Ashburner et al., 2000). It also has interactive features such as search, selection, drag and drop of organelles and nodes, count/percentage of nodes in each cellular compartment and the capability of displaying node~~s~~ annotation information ~~from the XGMML network file~~.

2 METHODS

~~We have developed~~ CellNetVis, an open-source software for visualization of biological networks ~~through~~ a web browser. A network is drawn over a cell representation using the force-based layout algorithm implemented in D3 (Bostock *et al.*, 2011).

The ~~force based~~ layout of D3 does not implement constraints for the nodes of a network to be placed on a specific region of a two-dimensional display. We implemented th~~e~~ constraint for each node so that they are placed in their respective cellular component.

As a force-based layout is computationally expensive, the cell diagram ~~was~~ drawn using only circles. With this approach, our constraint method ~~had~~ only to verify if a node was inside or outside the correct circle (cellular component). If a node ~~was~~ outside the correct circle, the algorithm subtract~~ed~~ the difference between the current radius (with respect to the position of the node to the center of the circle) and the maximum allowed

radius. For some components, such as plasma membrane, the algorithm has to check the cellular component boundaries as well.

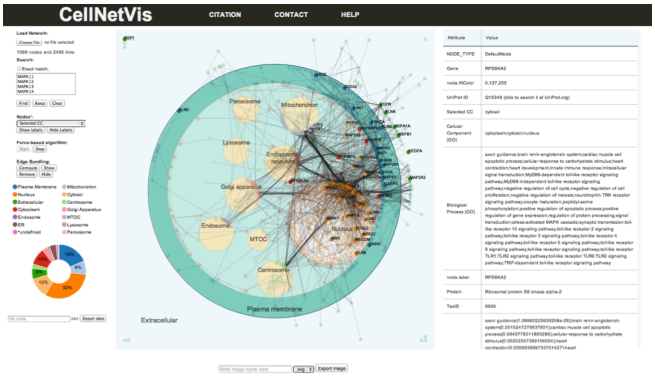

Fig. 1. CellNetVis interface. Visualization of a first neighbors network queried from IIS platform using as input 2097 proteins from the HPA supportive IH and IF data. ~~The nodes~~ colors were set to be displayed according to the "Selected CC" attribute. Edge bundling is also ~~displayed~~. The highlighted nodes correspond to selected proteins involved in the MAPK signaling pathway. IIS: Integrated Interactome System; HPA: Human Protein Atlas; IH: immunohistochemistry; IF: immunofluorescence; Selected CC: selected cellular component.

When cellular components constraints are added, edges cross at higher rates in the layout. Also, the distances among cellular components make nodes pull each other with larger force. The result is a poor network layout, with many overlapping nodes and many edge crossings. To reduce this effect, we added two more features to the algorithm. The first is a constraint that prevents nodes to overlap. The second is to allow users to improve the layout by moving the organelles and the nodes. If an organelle has no nodes, it is automatically removed from the view.

Cell contour lines are drawn using cold colors, as they serve only as a reference. In contrast, network nodes are displayed with hot colors by default, ~~but if some~~ nodes are selected the remaining ones are shown in colder colors to improve contrast. The cell structure diagram was colored using a ColorBrewer (Harrower and Brewer, 2003) scheme. The colors of this scheme ~~were~~ easily differentiated, even by ~~colorblind people.~~

~~Despite the fact that~~ CellNetVis ~~was developed for networks~~ constructed using the XGMML format (Punin and Krishnamoorthy, 2001) ~~generated by the IIS platform, any network in this format can be used.~~ The only requirement is that it has to have an attribute called "Selected CC", which corresponds to a unique selected cellular component (CC) from GO database for each node in the network. The majority of proteins are described as acting in more than one subcellular compartment in GO, so IIS applied a priority filter to assign each protein only one most specific cellular component, which is then used by CellNetVis to position the nodes in the cell diagram. For instance, if a node has values cytoplasm or no value (empty attribute) it will be placed in the cytosol region of the cell diagram.

~~Furthermore,~~ selection of nodes, highlight of neighbors, display of labels and the possibility to change the attribute related to the nodes colors in the network ~~were implemented~~. Network topology measures (degree, betweeness centrality and clustering coefficient) were also added to the nodes attributes. ~~To improve network readability, we integrated Edge Bundling (https://github.com/upphiminn/d3.ForceBundle) to the system~~. Counting of nodes per cellular component was also implemented as a donut chart, based on the NVD3.js library (http://nvd3.org/examples/pie.html). ~~The user~~ may also export the entire diagram, including the network, as an SVG file (Ferraiolo *et al.*, 2000) or a table containing the donut chart information as a CSV file. The diagram can be easily edited in a vector graphics software.

CellNetVis was integrated in the IIS platform using a special parameter named "file" which describes the URL for the selected XGMML file. When this parameter is used, an asynchronous call is executed and after the successfull download the file is parsed and processed the same way as a regular input.

3 DISCUSSION

CellNetVis is capable of displaying information related to networks, nodes and edges as well as their relations with cell partitions. Fig.1 shows one example of a high-throughput network visualization. The first neighbors network constructed from 2097 proteins from the Human Protein Atlas (Uhlen *et al.*, 2010) supportive data (Supplementary Information S1) was loaded from IIS. As expected, comparing the donut chart information to the HPA data, the GO annotations ranking by the percentage of nodes distributed in each cellular component was similar to the HPA annotations ranking, particularly concerning the top (nucleus followed by cytoplasm, including cytoskeleton and cytosolic proteins) and bottom (microtubule organizing center) terms of the ranking (Supplementary Information S2). ~~In addition, besides~~ being very useful to connect the network to information regarding subcellular compartments, CellNetVis is also useful to analyze their interactions and pathways by setting nodes colors to be displayed according to e.g. the GO biological processes or KEGG (Kanehisa *et al.*, 2002) pathways, or highlighting only the nodes annotated for a particular process/pathway, such as the MAPK signaling pathway (Supplementary Information S3) highlighted in Fig.1.

ACKNOWLEDGEMENTS

The authors acknowledge the financial support of Conselho Nacional de Desenvolvimento Científico e Tecnológico.

REFERENCES

Barsky,A. *et al.* (2007) Cerebral: a Cytoscape plugin for layout of and interaction with biological networks using subcellular localization annotation. *Bioinformatics*, **23**, 1040–2.

Bostock,M. *et al.* (2011) D^3: Data-Driven Documents. *IEEE Trans. Vis. Comput. Graph.*, **17**, 2301–9.

Carazzolle,M.F. *et al.* (2014) IIS--Integrated Interactome System: a web-based platform for the annotation, analysis and visualization of protein-metabolite-gene-drug interactions by integrating a variety of data sources and tools. *PLoS One*, **9**, e100385.

Ferraiolo,J. *et al.* (2000) Scalable vector graphics (SVG) 1.0 specification iuniverse.

Harrower,M. and Brewer,C.A. (2003) ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. *Cartogr. J.*, **40**, 27–37.

Kanehisa,M *et al.* (2002) The KEGG databases at GenomeNet. *Nucleic Acids Res,* **30**(1), 42-6.

Kojima,K. *et al.* (2008) Fast grid layout algorithm for biological networks with sweep calculation. *Bioinformatics*, **24**, 1433–41.

Punin,J. and Krishnamoorthy,M. (2001) XGMML (eXtensible Graph Markup and Modeling Language) 1.0 Draft Specification.

Shannon,P. *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–504.

Uhlen,M. et al. (2010) Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol.*, **28**(12), 1248–50.