BIOINFORMATICS APPLICATIONS NOTE Vol. 23 no. 8 2007, pages 1040–1042
doi:10.1093/bioinformatics/btm057

Systems biology

Cerebral: a Cytoscape plugin for layout of and interaction with biological networks using subcellular localization annotation

Aaron Barsky[1], Jennifer L. Gardy[2], Robert E. W. Hancock[2] and Tamara Munzner[1,*]

[1]Department of Computer Science and [2]Centre for Microbial Diseases and Immunity Research, University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada

Received on December 21, 2006; revised and accepted on February 10, 2007

Advance Access publication February 19, 2007

Associate Editor: Limsoon Wong

ABSTRACT

Summary: Cerebral (Cell Region-Based Rendering And Layout) is an open-source Java plugin for the Cytoscape biomolecular interaction viewer. Given an interaction network and subcellular localization annotation, Cerebral automatically generates a view of the network in the style of traditional pathway diagrams, providing an intuitive interface for the exploration of a biological pathway or system. The molecules are separated into layers according to their subcellular localization. Potential products or outcomes of the pathway can be shown at the bottom of the view, clustered according to any molecular attribute data—protein function—for example. Cerebral scales well to networks containing thousands of nodes.

Availability: http://www.pathogenomics.ca/cerebral

Contact: tmm@cs.ubc.ca

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

1 INTRODUCTION

Systems biology is characterized by a focus on interaction networks—the biomolecules involved in a particular biological system or process, as well as the relationships between these components. Network diagrams are used for visualizing and understanding these interactions, interpreting high-throughput experimental data, generating hypotheses and sharing results. These diagrams can be difficult for a user to explore with currently available network display tools—the networks are often too large, on the order of thousands of nodes, and many tools do not provide biological context to the diagram. We have therefore created Cerebral (Cell Region-Based Rendering and Layout), a Cytoscape plugin that automatically arranges networks of up to several thousand nodes using subcellular localization information.

Cerebral generates layouts in the style of traditional biological pathway/cellular overview diagrams. Subcellular localization annotation is used to position nodes in layers, resulting in a cross-sectional view of the cell. Potential products or outcomes of the network can be placed at the bottom of the diagram and clustered according to molecular function or any other attribute of interest.

Cerebral can be used on any network for which subcellular localization annotation exists. Localization information can be gathered from a number of resources (Gardy *et al*., 2006), including databases such as HPRD (Mishra *et al*., 2006) or prediction methods such as PSORTb (Gardy *et al*., 2005) for bacteria or Proteome Analyst (Lu *et al*., 2004) for eukaryotes. Localization annotation does not need to be complete, as reasonable results can be achieved with coverage as low as 30%. Nodes without annotated localization sites are positioned according to the location of their first neighbors. Interacting with a large network is facilitated by panning and zooming, intelligent labeling at multiple zoom levels, highlighting a node's first neighbors upon mouseover, category-based highlighting, automatic navigation that frames node neighbors in the view, and the grouping of edges into bundles to reduce clutter.

2 IMPLEMENTATION AND PERFORMANCE

Cerebral was implemented with the Prefuse Information Visualization toolkit (Heer *et al*., 2005). Cerebral has been tested with Cytoscape versions 2.3 and 2.4 (Shannon *et al*., 2003) on all supported platforms, including Windows, Mac and Unix.

For the small network shown in Figure 1A, with 57 nodes and 74 edges, Cerebral computed a layout in 4.5 s. For a large network of 760 nodes and 1269 edges, shown in Figure 1B, computation time was 204 s. This benchmark was run on a standard desktop computer (3 GHz Pentium 4 with 2 GB of memory running Windows XP).

3 LAYOUT

Laying out large network diagrams is a challenging and well-studied problem (Herman *et al*., 2000). Many techniques use only the topology of the network—the pattern of links amongst nodes—to guide the layout. However, the layout of biological network diagrams can be improved by including contextual information, for example subcellular localization and functional annotation.

*To whom correspondence should be addressed.

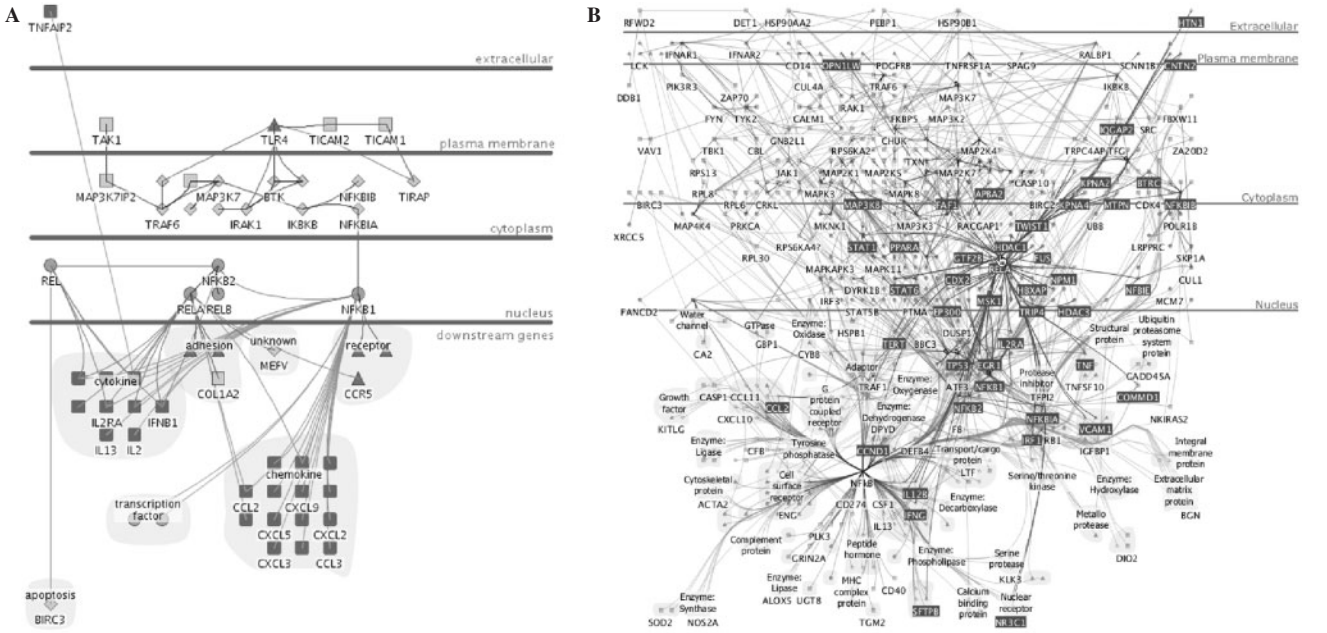

Fig. 1. Cerebral layouts illustrating the TLR4 innate immune signaling pathway (left) and the MAP kinase-NFκB pathway (right). (**A**) Small TLR4 network of 57 nodes and 74 edges, laid out in 4.5 s. (**B**) Larger MAP kinase-NFκB network of 760 nodes and 1269 edges, laid out in 204 s. A colour version of this figure is available as supplementary material online.

Recently, Li and Kurata (2005) used the simulated annealing technique of Davidson and Harel (1996) to lay out biological networks on a discrete grid. Kato *et al.* (2005) then improved the scoring function to reduce edge crossings and node-edge overlap, as well as to localize proteins to cellular compartments. We similarly employed a search-based layout algorithm, but used a stochastic approach to searching and an optimized scoring function to make the layout of large networks tractable.

Our network layout algorithm was modeled after hand-drawn pathway diagrams. Nodes are restricted to a regular lattice grid which provides room for labels and eliminates overlapping nodes. From an initial, random layout of nodes on the grid, neighboring layouts are stochastically sampled and evaluated until the layout stops improving. A neighboring layout is generated by moving a node to a new position within its subcellular compartment.

A position is evaluated by measuring: edge-edge and node-edge overlaps, the distance between nodes with similar molecular function, the distance between topological neighbors, and the empty space between unrelated nodes.

The efficiency of evaluation functions has previously been limited by the expense of scoring node-edge and edge-edge overlaps. We use a modified version of Bresenham's line drawing algorithm to increase efficiency by two orders of magnitude.

Once the layout has been computed, the edges of high degree nodes are curved into bundles to reduce visual complexity (Holten, 2006). A light blue convex hull is also drawn around pathway products that share molecular function (Fig. 1).

4 INTERACTION

With a small network (e.g. Fig. 1A), Cerebral produces a sufficiently organized layout to be understood as a static diagram. For large networks (Fig. 1B), the inherent complexity requires interaction techniques to understand the interactions.

Passing the mouse over a node of interest highlights its first neighbors in red and brings the labels to the foreground (Fig. 1B). As nodes of interest are found, they can be assigned to groups with a simple right-click. All members of the group can later be recalled and selected by clicking on the group name in the plugin's information panel (not shown). Node labels remain legible at different zoom levels and do not overlap. Edges are drawn on top of the nodes where they start and end, but drawn underneath any other node that they cross. Finally, Cerebral has several navigation helpers allowing the user to frame the view around nodes of interest.

5 CONCLUSION

We present here Cerebral, an open source Cytoscape plugin to automatically create network layouts in the style of pathway diagrams. Through a combination of layout efficiency and interaction techniques, Cerebral provides an intuitive, biological context-based method of visualizing and interacting with networks of up to several thousand nodes. Cerebral is freely available under the BSD license.

ACKNOWLEDGEMENTS

J.L.G. is a Michael Smith Foundation For Health Research trainee. R.E.W.H. has a Canada Research Chair. We gratefully acknowledge the support of Genome BC and Genome Prairie for the Pathogenomics of Innate Immunity research program. Funding to pay the Open Access publication charges was provided by Agilent Technologies.

Conflict of Interest: none declared.

REFERENCES

Davidson,R. and Harel,D. (1996) Drawing graphs nicely using simulated annealing. *ACM Trans. Graph.*, **15**, 303–331.

Gardy,J. *et al.* (2005) PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics*, **21**, 617–623.

Gardy,J. *et al.* (2006) Methods for predicting bacterial protein subcellular localization. *Nat. Rev. Microbiol.*, **4**, 741–751.

Heer,J. *et al.* (2005) prefuse: a toolkit for interactive information visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM Press, New York, pp. 421–430.

Herman,I. *et al.* (2000) Graph visualisation in information visualisation: a survey. *IEEE Trans. Vis. Comp. Graph.*, **6**, 24–44.

Holten,D. (2006) Hierarchical edge bundles: visualization of adjacency relations in hierarchical data. *IEEE Trans. Vis. Comp. Graph.*, **12**, 741–748.

Kato,M. *et al.* (2005) Automatic drawing of biological networks using cross cost and subcomponent data. *Genome Inform.*, **2**, 22–31.

Li,W. and Kurata,H. (2005) A grid layout algorithm for automatic drawing of biochemical networks. *Bioinformatics*, **21**, 2036–2042.

Lu,Z. *et al.* (2004) Predicting subcellular localization of proteins using machine-learned classifiers. *Bioinformatics*, **20**, 547–546.

Mishra,G.R. *et al.* (2006) Human protein reference database—2006 update. *Nucleic Acids Res.*, **34**, D411–D414.

Shannon,P. *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **11**, 2498–2504.