

J2dpathway: A Global Metabolic Pathway Viewer with Node-Abstracting Features

Eun-Ha Song¹, Seong Il Ham³, San-Duk Yang¹, Arang Rhie¹, Hyun-Seok Park^{1,2*} and Sang-Ho Lee¹

¹Department of Computer Science, Ewha Womans University, Seoul 120-75, Korea, ²Institute of Bioinformatics, Macrogen Inc., Seoul 153-023, Korea, ³Department of Architectural Engineering, Yonsei University, Seoul 120-752, Korea

Abstract

The static approach of representing metabolic pathway diagrams offers no flexibility. Thus, many systems adopt automatic graph layout techniques to visualize the topological architecture of pathways. There are weaknesses, however, because automatically drawn figures are generally difficult to understand. The problem becomes even more serious when we attempt to visualize all of the information in a single, big picture, which usually results in a confusing diagram. To provide a partial solution to this thorny issue, we propose J2dpathway, a metabolic pathway atlas viewer that has node-abstracting features.

Availability: J2dpathway is a public domain software program. You may download the source code (<http://www.sourceforge.net/projects/j2dpathway>) for GUI modules, or the executable file (<http://dream.ewha.ac.kr/~plumbum5/>). Updates to this program will be available at frequent intervals.

Keywords: drawing algorithm, edge crossings, metabolic pathway, scale-free network

Summary

The basic graph layout algorithm, one of the available visualization techniques, deals with the problem of positioning the vertices in a way that maximizes some measure of desirability in a graph. Given the ever-increasing complexity of biological networks, automatic graph layout techniques are becoming critically important for further development in integrative systems

biology. To apply automatic graph layout techniques to the genome-scale flow of metabolism domains, it is important to reduce unnecessary edge crossings (Karp *et al.*, 1994; Moritz *et al.*, 2001; Becker *et al.*, 2001; 2002, Kato *et al.*, 2005).

We previously proposed a metabolic pathway-drawing algorithm for reducing the number of edge crossings by considering the fact that the metabolic pathway graph is a scale-free network (Lee *et al.*, 2004; Song *et al.*, 2006). The scale-free network is composed of several nodes that have a high degree and many nodes that have a low degree that are connected to highly connected nodes. Generally, metabolic pathways tend to be scale-free networks (Jeong *et al.*, 2000). Our experimental results showed that the number of edge crossings was reduced about 37% to 40% by consideration of the scale-free network, in contrast with lack of consideration for the scale-free network.

Our layout algorithm (1) finds the layout of the connected pathways of an entire network, (2) finds a global layout of the nodes with respect to the key nodes of the connected component, and (3) refines a local layout of the nodes within each connected component.

However, metabolic studies, which used to be dedicated to a single pathway, recently began to focus on the entire network. For example, KEGG, the representative biological pathway database, now provides the KEGG Metabolism Atlas (Okuda *et al.*, 2008) by manually combining about 120 existing metabolic pathway maps, as shown in Fig. 1. However, the static approach to representing metabolic pathway diagrams offers no flexibility.

On the other hand, our initial attempts to visualize all information automatically in a single atlas map resulted in a confusing diagram that was difficult to interpret, as shown in Fig. 2. It should be noted that Fig. 2 differs in many aspects from Fig. 1 or conventional drawings in biochemistry textbooks. For this reason, we designed a new metabolic atlas viewing tool called J2dpathway, which has node-abstracting features.

When J2dpathway is initially executed, a window frame appears, as shown in Fig. 3. The screen consists of views and editors. The tool-bar menu at the top lists various tool icons, including zoom-in, zoom-out, cliques, highly connected nodes, and obtaining cycles. The Map Repository View on the left side lists a preinstalled data source that has many example pathways to explore, which are arranged in a tree view of the components of

*Corresponding author: E-mail neo@ewha.ac.kr
Tel +82-2-3277-2831, Fax +82-2-3277-2306
Accepted 11 June 2008

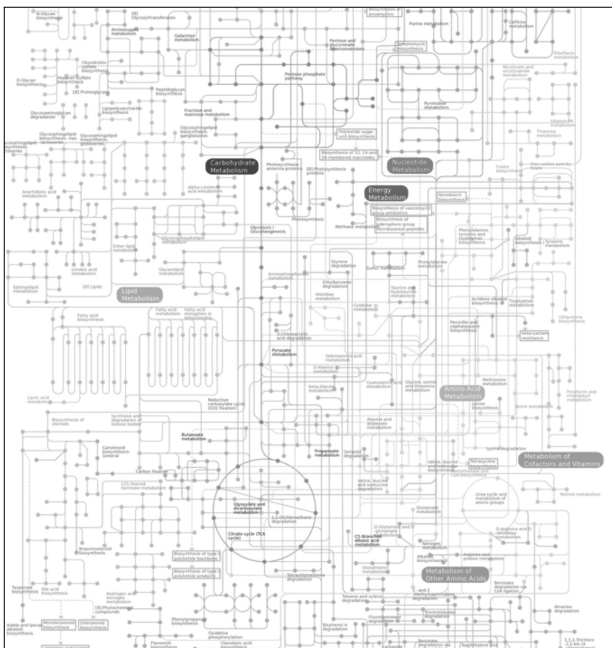


Fig. 1. KEGG metabolism atlas (<http://www.genome.jp/kegg/atlas/metabolism/>) is a new graphical interface to the KEGG suite of databases.

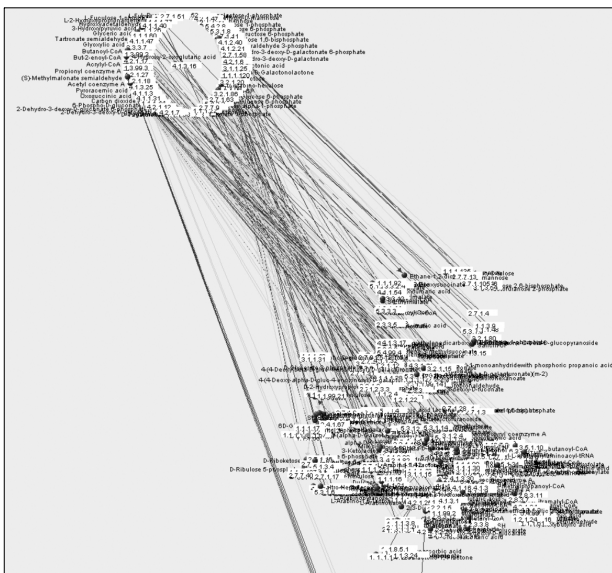


Fig. 2. Automatically Drawn Metabolism Atlas.

the pathways. The Editor window in the center is the main entry point for displaying a chosen pathway. Either a single pathway or a multi-pathway can be displayed.

Fig. 4A shows an example of a multi-pathway map composed of arbitrarily chosen 13 pathways, opened in the Editor window. Among these pathways, a user can select a specific pathway to display further details as in

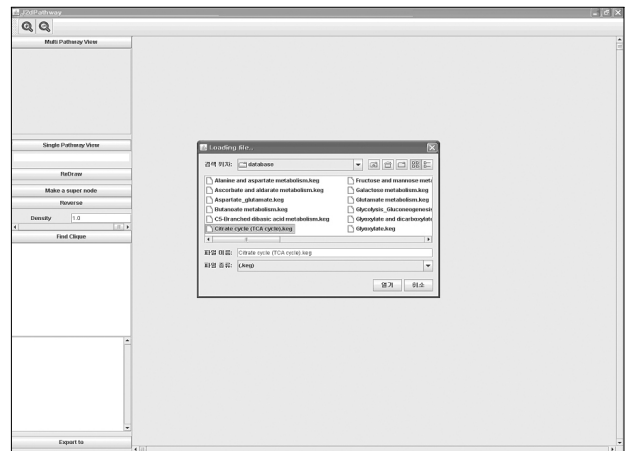


Fig. 3. System windows and toolbar.

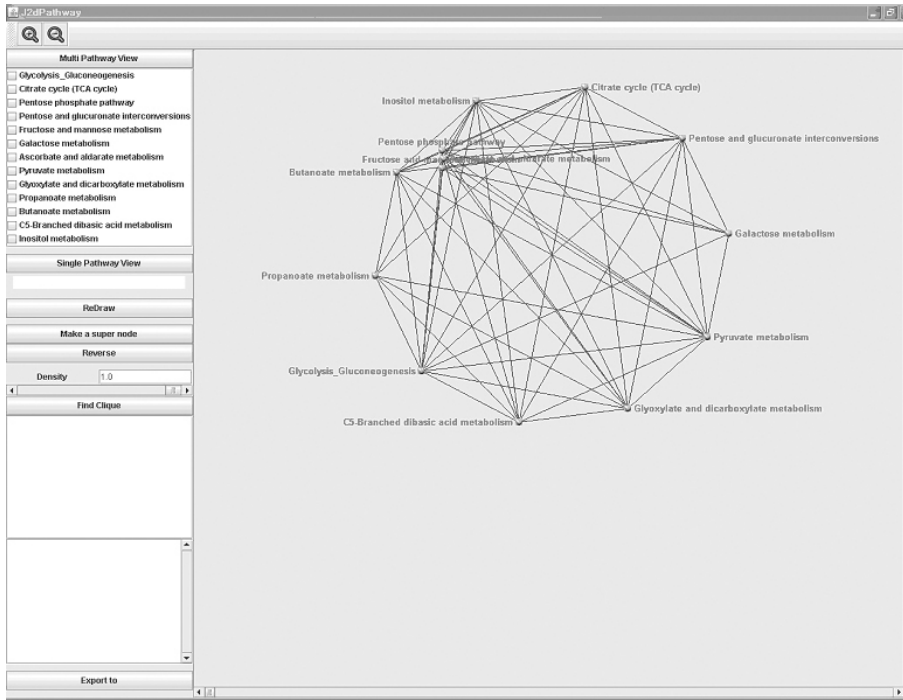
Fig. 4B. Or, they can be abstracted into a single node again. The function icons in the toolbar allow one to select a portion or a subset of nodes, as in Fig. 4. J2dpathway also has a feature that selects arbitrary nodes in an extended pathway and abstracts them into a single node. One can select several nodes by dragging the mouse, as in Fig. 5A, and turn them into a single supernode, as in Fig. 5B. Also, one can find cliques, as in Fig. 5C. A clique in an undirected graph $G=(V,E)$ is a subset of V , each pair of which is connected by an edge in E . Finally, Fig. 5D shows a function that finds highly connected nodes. In metabolic pathways, the number of nodes that have a degree that is higher than 6 is low. Therefore, we consider these nodes to be highly connected nodes. When they are selected, the edges that are selected within the high connectivity area are deleted.

For layout, we place nodes that have high connectivity to the center of the connected components, following Holme's approach (Holme *et al.*, 2003). After that, each component is collapsed into a supernode. Then, the remaining graph is grouped into connected components, and the connected components are placed by using graph layout algorithms according to their topologies.

Software Release Stage

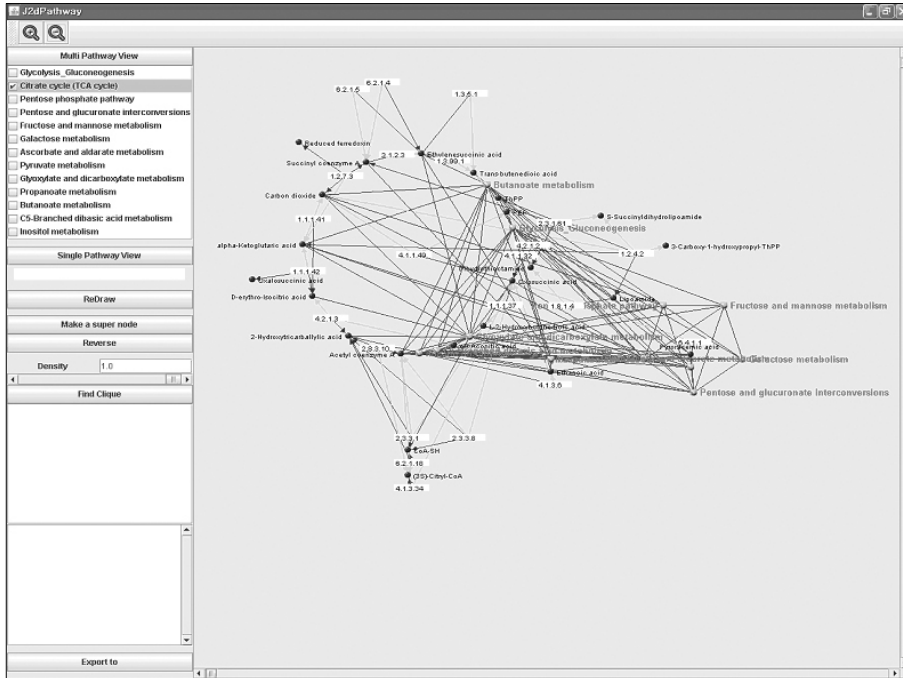
J2dpathway was implemented in Java and the Java-based graph library yFiles, an extensive Java™ class (Wiese *et al.*, 2000). Our system still awaits full testing of all of its functionalities for practical usage. The current release is not feature-complete, but we will constantly update our software.

A



Initial state of 13 multi-pathways

B



Extending a single pathway among 13 pathways

Fig. 4. Extending a pathway in the atlas map by clicking on a specific pathway. (A) shows an example of a multi-pathway map that is composed of 13 arbitrarily chosen pathways, opened in the Editor window. Among these pathways, a user can select a specific pathway to display further details, as in (B). Or, they can be abstracted into a single node again.

Acknowledgements

This work was partially supported by a grant from the Ministry of Information and Communication of Korea.

References

Becker, M.Y., and Rojas, I. (2001). A graph layout algorithm for drawing metabolic pathways. *Bioinformatics* 17, 461-467.

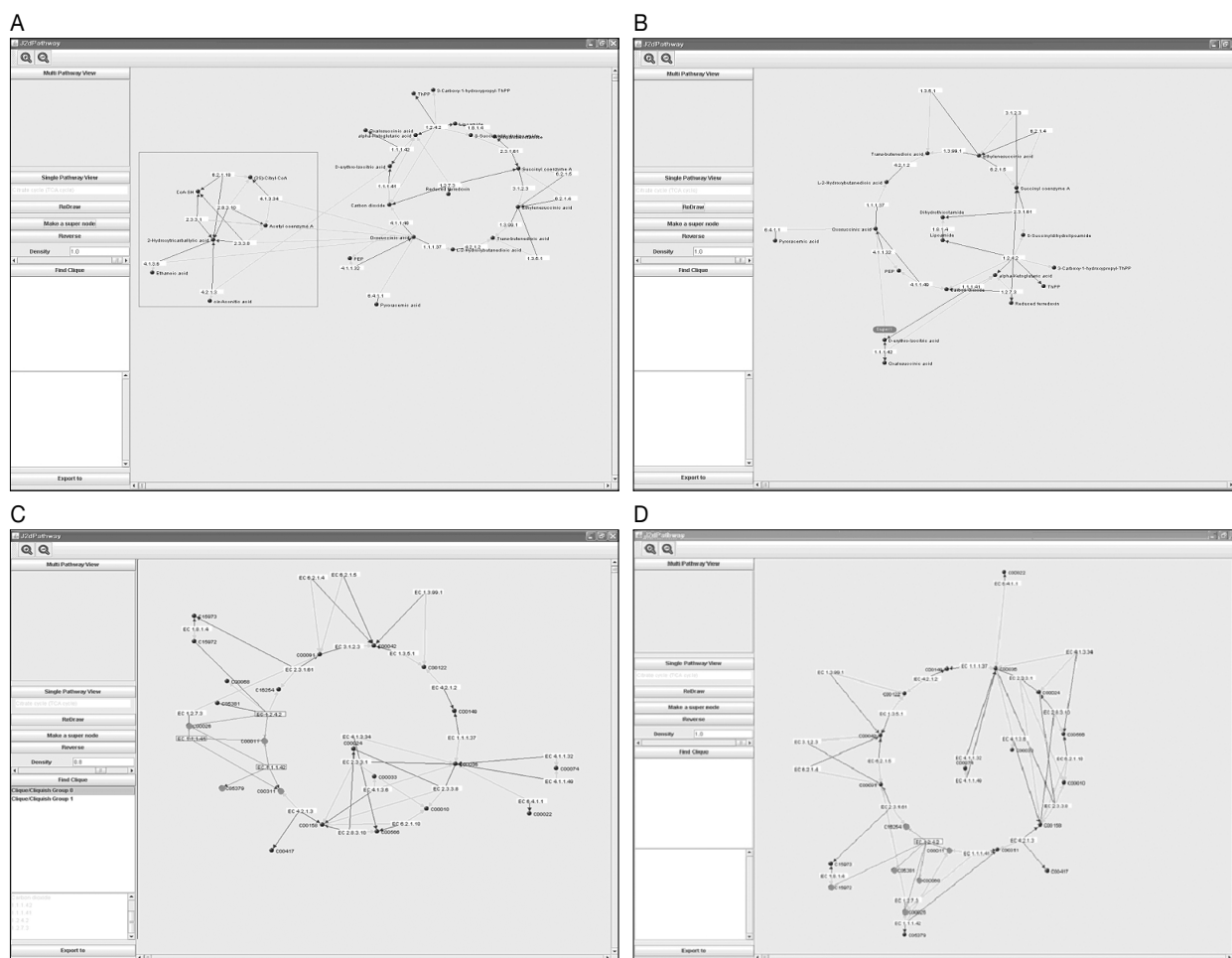


Fig. 5. Various functions. (A) Selecting nodes. (B) Making supernodes and redisplaying. (C) Finding cliques. (D) Finding highly connected nodes.

- Holme, P., Huss, M., and Jeong, H. (2003). Subnetwork hierarchies of biochemical pathways. *Bioinformatics* 19, 532-538.
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., and Barabasi, A.L. (2000). The large-scale organization of metabolic networks. *Nature* 407, 651-654.
- Karp, P.D., and Paley, S.M. (1994). Automated drawing of metabolic pathways. Proc. of the 3rd Intl. *Conference on Bioinformatics and Genome Res.* 225-238.
- Kato, M., *et al.* (2005). Automatic drawing of biological networks using cross cost and subcomponent data. *Genome Informatics* 16(2), 22-31.
- Lee, S.H., *et al.* (2004). An algorithm for drawing metabolic pathways based on structural characteristics. *Journal of*

- Korea Information Science Society* 31, 1266-1275.
- Moritz, Y.B., and Isabel, R. (2001). A graph layout algorithm for drawing metabolic pathways. *Bioinformatics* 17, 461-467.
- Okuda, S., Yamada, T., Hamajima, M., Itoh, M., Katayama, T., Bork, P., Goto, S., and Kanehisa, M. (2008). KEGG Atlas mapping for global analysis of metabolic pathways. *Nucleic Acids Res.*
- Song, E.H., Kim, M.K., and Lee, S.H. (2006). A metabolic pathway drawing algorithm for reducing the number of edge crossings. *Genomics & Informatics* 4, 118-124.
- Wiese, R., Iglsperger, M., and Schabert, P. (2000). The Y-files graph library: documentation and code. <http://www-pr.informatik.uni-tuebingen.de/yfiles>.