Author's response to reviews

Title: A survey of visualization tools for biological network analysis.

Authors:

  Georgios A Pavlopoulos GAP (pavlopou@embl.de)
  Anna-Lynn Wegener ALW (wegener@embl.de)
  Reinhard Schneider (rschneid@embl.de)

Version: 2 Date: 7 October 2008

Author's response to reviews: see over
Dear reviewers,

We herewith submit a thoroughly revised version of our manuscript entitled “A survey of visualization tools for biological network analysis”. We addressed the comments of the reviewers as detailed in the appended response.

We would like to thank the reviewers for their helpful suggestions, and tried to take them into account wherever possible.

I hope that you will find the revised manuscript now acceptable for publication in “BioData Mining”, and we look forward for your decision.

Thank you very much for your interest.

With best regards
Detailed response to reviewers’ comments and suggestions

We thank both referees for their thorough review of our paper and for their constructive input. We have revised our manuscript by the points that were raised, as detailed below.

Reviewer 1

The graphical illustration of the section “Networks represented as graphs and their data structures” is a good suggestion. In accordance with point 1) raised by reviewer 2 we decided however to remove the whole section from the paper entirely for reasons specified below (Reviewer 2, point 1)).

Reviewer 2

1) This is a fair comment. The explanation of graph theory and mathematical definition of its components might be of limited interest to the users of the tools. Their interest will rather focus on the functionalities of tools and how they can be applied to biological questions. We have shifted the emphasis of the review towards this kind of information and have removed section 2 from the paper and instead expanded the tool survey. The information contained in former section 2 are standard definitions of single unweighted or weighted, undirected or directed graphs, that interested readers can look up in references [2,3].

2) Our review intends to provide a broad overview of the best existing visualization tools. The audience are users, mostly but not exclusively biologists and novices to the field of data visualization. In accordance with the reviewer’s comment we have revised the tool survey to provide maximal value for the specified audience. In a first step we have recategorized the information presented. We are now describing the tools in terms of visualization, compatibility, functionalities and strength. We consider these categories as more informative, application oriented and thus more practical than the previous format. It allows the reader an at-a-glance assessment of the crucial properties of a visualization tools that determine its usefulness and applicability to a given problem. It now clearly addresses the major questions: What type of data can a tool handle? What quantity of data can it cope with? What type of visualization does it use? And what kinds of problems is this visualization useful for? What other tools and data sources is the tool compatible with?
With the newly introduced category called ‘strength’ we also introduce a comparative element into the survey that facilitates the choice of one tool over another depending on the project. At the end of the survey a summarizing paragraph expands this comparative perspective and states those tools best suited to address the major challenges facing visualization (increasing data quantity, heterogeneity of data and the display of multiple connections).

In a second revision step we have expanded the information provided by the survey to include more facts about the functionalities of the various tools. We believe that with the described changes we have addressed the reviewer’s comment and provide a useful resource for users.

3) The unifying theme of the review is the visualization of biological data and networks. We have deliberately chosen those tools that are broadly applicable to a wide range of data and questions to provide the broad readership of 'BioData Mining' with an overview of currently available tools. The reviewer’s comment is valid in that the study of protein-protein interactions and evolutionary relationships requires different analyses, but the analyses of both types of data starts with visualizing them to detect connections and patterns. Any additional analytical functionalities that come with individual tools might be more suited to one type of data than the other, but the analytical features of the surveyed tools are of secondary importance for this review. The focus lies on their visualization capacities.

Furthermore, in the age of systems biology where different approaches are chosen to study complex, dynamic systems, the experimental results are often highly heterogeneous datasets that needs to be integrated. So, while we agree that gene networks, protein-protein interactions and evolutionary relationship are quite different data often they have to be integrated and analyzed together to achieve a full picture. This is why we have chosen this selection of very general tools.

Moreover, the different tools presented are in many respects complementary since they support different types of data, networks of different sizes, layout algorithms and file formats. Often a combinatorial analysis of data applying several tools provides the most complete answer to a research question, which is why we intended to offer a broad range of tools covering the diversity of functionalities that exists.

We have, however, excluded tree-viewers from our analysis. These are tools exclusively concerned with the analysis of evolutionary relationships, that we agree do not fit the scope of this review.
4) We agree with this comment and have moved the sentence from Section 3 to the Discussion. See p8.

Since the review intends to give an overview of existing visualization tools and is aimed at biologist users and especially novices to the field of visualization we did not consider it helpful to go into technical and mathematical details of algorithms. The algorithms have been discussed in depth in a recent review that we reference repeatedly. Furthermore, we have suggested outsourcing of computational efforts (layout algorithms) to web services and libraries with access to powerful machines that can run several jobs in parallel as a possible solution to the problem of visualizing large networks.

5) In the discussion we suggest several different approaches that could help overcome the current limitations of existing visualization tools. To avoid compromising computational speed when dealing with large-scale networks we propose to outsource computational efforts to web services and libraries with access to powerful machines that can run several jobs in parallel. Furthermore, we discuss the extension of existing layout algorithms to encompass a third dimension as a possible way to reduce cluttered views and minimize the possibility of crossovers between connections, and such render visualization and navigation of networks more efficiently. Powerful filters, that reduce the noise in a dataset and restrict the user’s attention to a core set of nodes of a particular interest, are another proposition to improve the performance of existing visualization tools. Similarly, we propose that more efficient and interactive graphical user interfaces (GUIs) would allow the user to visualize and explore relevant sub networks or limited areas of a whole network without having to sieve through vast data masses. Finally, autostereoscopic 3D displays are mentioned as a way to improve visualization. Please refer to page 16-19 for more detail.

To make these conclusions and our concrete suggestions more obvious to the reader we have added a paragraph to the end of the discussion, which summarizes them in a bullet point format (p. 19).

Propositions that go beyond this level require considerable primary research and technology development that would go beyond the scope of a review intended to give an overview of existing visualization tools.