

Subject Section

GenePlus: a software built on electron for hierarchical visualization and analysis of biological networks

A portal, scalable tool for hierarchical visualization and analysis of biological networks

Corresponding Author^{1,*}, Co-Author² and Co-Author^{2,*}

¹Department, Institution, City, Post Code, Country and

²Department, Institution, City, Post Code, Country.

Zhiwen Qiang¹, Hao Chen², Xiaopeng Zhu³, Shikui Tu^{1*}

1. Department of Computer Science and Engineering, School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China

2. Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA 01605, USA

3. School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213

*To whom correspondence should be addressed.

Associate Editor: XXXXXXXX

Received on XXXXX; revised on XXXXX; accepted on XXXXX

Abstract

Motivation: In system biology, analyzing the structure and relationship of interacting genes and proteins in biological networks is one of the major job, numerous kind of tools have therefore been invented. Electron is a framework for building cross-platform desktop applications. It suits well for our need to build a lightweight data visualization system.

Results: We present GenePlus, a software built on electron that provides a suite of interaction and navigate tools to help visualize and analysis the hierarchical information of biological networks.

Availability: The installation package and source code can be accessed at <https://github.com/QLightman/GenePlus>.

Contact: name@bio.com

Supplementary information: Supplementary data are available at [Bioinformatics](#) online.

Motivation:
It is very important to analyze the network structure of interacting genes, proteins, RNAs, etc. in large scale biological networks which represent complex biological systems. Many tools have been made for this purpose. However, they are either too complicated, or not easy to build in cross-platforms.

Contact: tushikui@sjtu.edu.cn

1 Introduction

The biological networks (protein-protein interactions, genetic interactions, etc.) are of vital importance in nearly all biological processes. But over the years, as interactions contain in an organism are increasing, analyzing the relationship of interacting genes and proteins is a challenging task. Currently, there is a shortage of software tools which can effectively analyze and display such network. Some tools only offer some simple graph layouts. (Breitkreutz, B.J *et al.*, 2003; Goldovsky, L *et al.* 2005), these tools can only display the network locally, therefore cannot fully reveal the structure and hierarchies of the entire network. Also, for a normal biological network with tens of thousands interactions, this kind of layout cannot reveal any valuable information. So we need more sophisticated design.

Other tools which allow the users to visualize the biological networks at different levels like CNplot (Batada, N.N. 2004) which displays the biological network pre-clustered, but does not allow the user to further analyze each cluster within the network, GenePro (James Vlasblom, Samuel Wu *et al.*) facilitates the visualization and analysis of protein networks and provides methods for parsing the network into meaningful

functional modules. But it cannot customize the result and may lack order when the network is very large.

In this paper, we present GenePlus, a software that provides hierarchical ...
hierarchical interactive visualization and analysis for biological networks.
It contains three main features:

- It is able to visualize network which contains tens of thousands interactions in a **readable** manner. **structural details?**
- It provides a suite of interactive methods for the user to analyze **every** detail of the network.
- Users can set some peremeters to customize the layout to satisfy their demand.

2 Methods

We use ...
In our design, we use a top-down approach to cluster the biological networks into a multilevel **node-link** graph. As is shown in Figure 1a for the visualization of STRING interaction data (Szklarczyk, Nucleic Acids Res *et al.* 2015). It is well-known that many biological networks

We demonstrate the main features in Figure 1 through a dataset

but does not allow the user to zoom in the cluster for deeper visualization and analysis.
GenePro (Vlasblom *et al.*, 2007) provides methods for ..., but its results could be customized, and it may also lack order (什么order? ?) when the network is very large.

© The Author 2015. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

1

Users don't need to go through the tedious installation process, and bothered by installing a lot dependent packages.

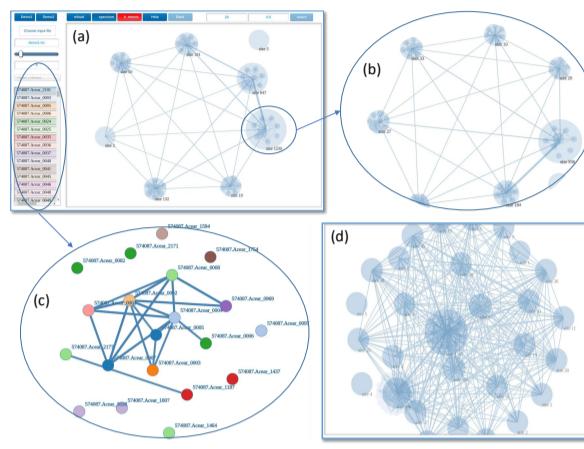


Fig. 1. Illustration of features in GenePlus, a software built on electron for hierarchical visualization and analysis of biological networks. (a) Overview of the software's interface, the software displays the network using k -means in each hierarchy, with each node representing one cluster, there are smaller nodes inside the bigger node which represent the structure within. (b) The full scope of the nodes inside (a). (c) Displaying the gene, its most related genes and their interactions when choosing a gene in (a). (d) The software displays the network using modified spectral clustering in each hierarchy, with each node representing one cluster.

contain clique and hub genes which is highly clustered and contains self-similar structure. So clustering them together into one node can reveal their hierarchical information and make the visualization readable. We use the force-directed algorithm for the layout of each hierarchy. It can generate good-quality results timely with less than hundreds of nodes, so it suits well for our situation.

Currently we provide two cluster algorithms, k -means clustering and modified spectral clustering. k -means algorithm is one of the most popular clustering methods, it has reached a balance in its speed and performance. However, the user has to know how many groups to cluster at each stage, which is not always possible. Therefore, we introduce the spectral clustering method, which performs better than k -means and can determine the sorting groups in advance. Therefore it is now widely used for large-scale data mining problems, but this clustering method has limited applicability due to its high computational complexity, so we implemented the k -means-based approximate spectral clustering method (Donghui Yan, Ling Huang *et al* 2009). When comparing the results with k -means, studies have shown that the result obtained by k -means didn't yield significant improvements in accuracy.

3 Implementation

As we have said before, it is difficult to display the biological network meaningfully in one scope, GenePlus solve it by cluster them into a

hierarchical tree and only reveal the nodes in the current view and beneath it. To be more specific, here we present a example illustrated in Figure 1, after loading a file into GenePlus, it clusters them into k groups(k can be adjusted by user), the genes or proteins inside the group are further divided, this procedure goes on until it reaches certain boundaries.

Figure 1a is the overview of GenePlus, the genes are clustered using k -means and each node represents a group of them, the user can also choose different number of sorting group for each hierarchy to evaluate the result. It can also switch the algorithm to modified spectral clustering, the result is shown in Figure 1d. GenePlus offers smooth animation methods to further analyze the computed clusters. When selecting one cluster, it will show a full scope of its inner structure, as shown in Figure 1b, users can also return to the previous stage, or they can further analyze the current cluster with the same procedure. Apart from that, GenePlus will show each genes within the cluster, as is shown in the left part of Figure 11, when choosing one gene, GenePlus will display the gene and its most related genes in a network, as is shown in Figure 1c (the number of genes and the minimum value of interaction to be shown can be adjusted by the user). Thus it will help to analyze the network locally.

Users can choose to download the genes or proteins within the cluster, they can use the data to do further analysis like gene ontology via other tools.

4 Conclusion

In this paper we present GenePlus, a software built on electron for hierarchical visualization and analysis of biological networks. We implemented the k -means clustering, modified spectral clustering algorithm and force-direct layout algorithm to let GenePlus capable of handling large-scale biological data. It is worth noting that this kind of displaying technique is very general and can be applied to many other field like social network, security network, etc. Based on our previous discussion, we can say that GenePlus can give user a full scope of biological data and can help reveal the structure and hierarchies of the network.

References

- Breitkreutz, B.J *et al.* (2003) Osprey: a network visualization system. *Genome Biol.*, **4**, R22.
- Batada, N.N. (2004) CNplot: visualizing pre-clustered networks. *Bioinformatics*, **20**, 1455-6.
- Szklarczyk *et al.* Nucleic Acids Res. (2015) 43(Database issue):D447-52
- James Vlasblom, Samuel Wu *et al.*, (2006) GenePro: a cytoscape plug-in for advanced visualization and analysis of interaction networks. *Bioinformatics*, **22**, 2178-9.
- Goldovsky, L. *et al.*, (2005) BioLayout(Java): versatile network visualisation of structural and functional relationships. *Appl. Bioinform.*, **4**, 71-74.
- Donghui Yan, Ling Huang *et al* (2009) Fast Approximate Spectral Clustering. *Technical Report*.