

CancerLinker: A bottom-up exploration tool of protein network

Vinh The Nguyen (R-11497542)
Md. Yasin Kabir (R-11516715)

February 2017

1 Introduction

Visualization is the process of representing data visually and is an important task in network analysis, especially in biological network where scientists can see the complex structure of the protein. Hence, there are plethora visualization tools available on the internet in these years to utilize this capability to help better understanding of protein network. In particular, biological scientists when studying protein interaction of some cancers are interested in the following requirements of a visualization tool to answer some of a specific questions

1. Viewing side-by-side of how many proteins contained in each cancer network, and a protein belongs to how many cancer networks? (this is many-to-many relationship)
2. Which proteins are found in all cancer networks? Which protein is the most important along all pathways? Which protein is closet to all other proteins
3. Filter proteins in the network without losing its structure.
4. Showing all pathways (shortest paths) between any pairs of proteins

Cytoscape [4] is an open source, standalone Java application. One clear advantage of the software is the user-friendly format which allowed us to visualize and merge our data quickly with simple integrated buttons. The software also has a vast number of freely available apps which allow the user to configure the imported data in many different custom styles and themes, which was a greatly useful for us since we used the app “KEGGParser” to import our data. One of the downsides of the software would have to be the small amount of visualization packages available for different data interpretation. Since we worked with 15 total original networks, we wanted to be able to color nodes via user’s selection parameters, however, this was not possible with Cytoscape. In addition, the creation of our large merged network of approximately 300 nodes led

to a messy appearance of the network since the display window had processing troubles rendering all the data represented in each node, this sometimes leading to “ghost” nodes appearing and disappearing at random times. Lastly, one aspect of such network creation is the overall research reach, meaning that a researcher would want the highest number of people looking at his research. The most significant disadvantage of Cytoscape for our research was the inability to easily render our network data without the need of special software (i.e., Cytoscape and Java).

Medusa[3] is another open source system written in Java programming language and available as an applet. It is highly interactive, searchable nodes and text but it lacks of viewing side-by-side function that allow user to view two networks. Gephi [1] is an interactive visualization tool for all kind of complex networks, it is a Java application that support multiple data files. but it does not support viewing two different networks together.

Networkx [2] is a Python language software package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks. This network is mainly used by mathematicians, physicists, biologists, computer scientists, and social scientists who have expertise in network analysis. It uses command line interface and it lacks of interactive visualization.

From the above literature review, there is no “one fit all” best visualization tools that answer all specific questions of the above requirements. In this project, we design a visualization tool that helps biological scientists answer meet their requirements

References

- [1] Mathieu Bastian and Sebastien Heymann. Gephi: An Open Source Software for Exploring and Manipulating Networks. *International AAAI Conference on Weblogs and Social Media*, 2009.
- [2] P Swart DA Schult. Exploring network structure, dynamics, and function using NetworkX . *Proceedings of the 7th Python in Science*, 2008.
- [3] Bork P Hooper SD. Medusa: a simple tool for interaction graph analysis. *Bioinformatics*, 21(24):4432–4433, 2005.
- [4] Ozier O Baliga NS Wang JT Ramage D Amin N Schwikowski B Ideker T Shannon P, Markiel A. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 13(11):2498–2504, 2013.

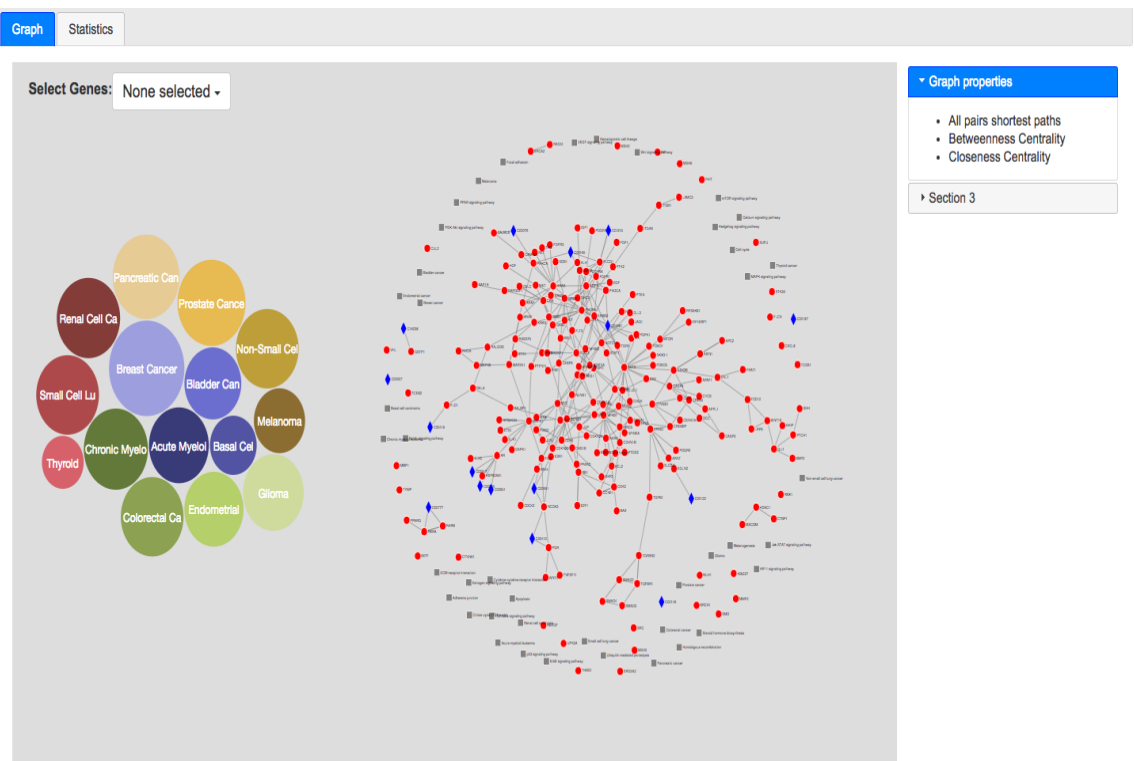


Figure 1: Tentative Network Layout