Anindita Varshneya

BIOL 598-04: Advanced Systems Biology Research

**Test-driven development and new features improve GRNsight: a web application and service for visualizing small- to medium-scale gene regulatory networks**

Introduction

GRNsight is a web application and service for visualizing small- to medium-scale gene regulatory networks (GRNs). A GRN diagrams various genes, transcription factors, and their regulatory relationships (MacNeil et al., 2011). These interactions define the types and amounts of mRNA and protein produced by the cell. Our sister project, GRNmap, is a MATLAB program that performs parameter estimation and forward simulation of GRNs. GRNmap takes in an Excel file containing various sheets of necessary information to complete the estimation. Typically, GRNs are presented in adjacency matrices, which include source genes and target genes listed across columns and rows respectively, and the elements of the matrix indicate the relationship between these genes. GRNmap input files include unweighted adjacency matrices in which a “0” indicates no relationship between the corresponding genes, and a “1” indicates a regulatory relationship of some unknown magnitude and direction. Using this data, GRNmap outputs a weighted adjacency matrix where each “1” is replaced by a more descriptive number which indicates the magnitude of the relationship between the gene and its transcription factor, and the sign of the number (positive or negative) indicates the direction of the relationship (activation or repression).

While these network models do not provide comprehensive evidence about *how* changes in one gene or regulator in a network may affect other genes and regulators, and instead purely indicate the estimated interactions between existing genes, it is an essential first step for a systems-level understanding of genomic interactions (Kitano, 2002). In order to most effectively and efficiently analyze biological data of this type, network visualizations are often used to demonstrate relationships between GRNs (Pastrello et al., 2013). These visualizations are often shown as graphs where different nodes represent genes, and the edges between them represent their relationships. Currently, several different programs exist that visualize network interactions including Cytoscape, Gephi, and yED. While each of these programs are effectively diagram GRN visualizations, none of them effectively address the niche cases for which GRNsight was created.

Cytoscape is one of the best-known GRN graphing programs, and has tools that allow users to view and create large networks (Shannon et al., 2003). While Cytoscape is compatible with all of the file types GRNsight is also compatible with, Cytoscape does not take in adjacency matrices as a data structure, and is therefore not compatible with GRNmap generated output sheets. In fact, because adjacency matrices are the most common format through which GRN data is presented, many users uploading GRN data produced in microarrays will need to reformat their data in order to successfully use Cytoscape. Furthermore, Cytoscape is a stand-alone program that a user must install onto their machine. While this may not always raise issues, in the context of a university-based lab, installing new software onto a university machine is often difficult and requires admin login information that is not available to students. Because Cytoscape can run as a stand-alone application, it comes with several options and features that allow users to customize their visualizations and get more information about their networks than other web-based programs. While this is beneficial in some cases, this complicates the learnability of the software, so it is not the best tool for novice users and researchers. Finally, Cytoscape is optimized to present visualizations for large-scale networks. While this is useful for understanding genome-level interactions, it is not as useful when analyzing particular processes, which produce small- to medium-scale networks.

Gephi is another popular software aimed at visualization and explanation of large-scale networks (Bastian et al., 2009). While Gephi boasts its learnability and user interface, it still faces several of the problems previously addressed in regards to Cytoscape. Gephi also needs to be installed onto a user’s machine in order to be accessed, and therefore does not address the niche target audience that GRNsight is tailored towards. Despite simplifying the program and making it more accessible to the average user, Gephi still provides several customization options that would only be beneficial when understanding large-scale networks. Ultimately, like Cytoscape, Gephi was created for researchers analyzing large amounts of data in large-scale networks, which does not address the key concern that led to the development of GRNsight.

Another common program that allows users to create custom network visualizations is yEd (<https://www.yworks.com/products/yed>). yEd is another GRN visualization software that allows users to upload Excel files or XML files containing adjacency matrices and produce directed graphs of that data. This program allows users to customize the type of graph created (flowchart, family tree, semantic network, etc.) and also allows users to arrange and export their graphs easily. Unlike the previously mentioned software, yEd has a browser-enabled version of the application in addition to the stand-alone version that must be installed onto a users machine. However, the online version of the program is limited to only a certain number of nodes and edges that would prevent diagramming of some medium-scale networks. In order to access the full version of the program, the user must install the program onto their machine and pay for the program. While yEd comes closest to addressing the chief concerns that led to the development of GRNsight, it does not provide a solution for each of the goals our research team had for creating GRN visualizations.

In order to address the concerns that none of the previously mentioned programs adequately solved, we created GRNsight, More specifically, we wanted GRNsight to fulfill the following requirements (Dahlquist et al., 2016):

* Exist as a web application that does not require installation.
* Have high learnability so that any novice user could quickly understand the program.
* Accept Excel files containing weighted or unweighted adjacency matrices similar to those produced by GRNmap.
* Automatically produce small- to medium-scale, weighted or unweighted, directed network graphs that allow for easy and efficient analysis of biological data.

GRNsight effectively addresses these goals while also following open-source and software development best practices. Most recently, improvements to GRNsight include the integration of Test-Driven Development (TDD), descriptive error and warning messages, and other front-end options that allow users to customize the appearance of their network.