Mihir Samdarshi

March 1, 2019

BIOL 588 – Advanced Systems Biology Research

**Refactor of code to the Model-View-Controller paradigm improves** **GRNsight: a web application for visualizing small- to medium-scale gene regulatory networks**

Abstract

GRNsight is a web application for the visualization of small- to medium-scale gene regulatory networks that was first developed and released in 2015 by the Dahlquist lab at Loyola Marymount University. Since the initial creation and release of GRNsight, the application has undergone many revisions and added new features. The features that have been added since v1.0.0 create more options for users to better share and modify the graphs. Nodes can now be colored to visualize timecourse gene expression data, edges can be normalized against user-input values, users can determine thresholds at which to set edges to appear gray, and the application is more accessible to colorblind with a transition to a new color scheme and the option to set gray edges to appear dashed, among a plethora of other new features. The constant revision and update to the application drove the need to refactor the application’s codebase to a far more sustainable architecture because of the code disorganization and various bugs that arose from rapid development. The Model-View-Controller framework, which splits the application into three separate, but interconnected parts, was chosen as the new architectural paradigm for GRNsight. This architectural pattern vastly improved GRNsight’s organization and reduced state conflicts which lead to the numerous bugs and deficiencies.

Introduction

The central model of molecular biology describes how the flow of information in a cell during gene expression goes from DNA to RNA to proteins. Transcription factors control gene expression by binding to regulatory DNA sequences. Activators increase gene expression. Repressors decrease gene expression. Transcription factors are themselves proteins encoded by genes. A gene regulatory network (GRN) consists of genes, transcription factors, and the regulatory connections between them, which govern the level of expression of mRNA and proteins from those genes. These gene regulatory networks can be represented in a graph format, with nodes and edges. Each node in a graph represents a regulatory transcription factor, with each edge representing a regulatory relationship, either an activation or repression relationship from the transcription factor to its target gene.

The Dahlquist lab has been investigating the early transcriptional response to cold shock in budding yeast. The lab has measured gene expression using DNA microarrays for the wild type and transcription factor deletion strains. This has led to the identification of gene regulatory networks that have been hypothesized to be important to the regulation of this type of environmental stress response. The Dahlquist lab uses the microarray data as inputs to a mathematical model that models the dynamics of gene expression over time. This model is named GRNmap and is a sister project of GRNsight. GRNmap is a differential equations model of the changes in gene expression over time for a gene regulatory network (Dahlquist et al. 2015). Each gene (node) in the network has an equation. The parameters in the model are estimated from laboratory data. The weight parameter, *w*, gives the direction (activation or repression) and magnitude of the regulatory relationship. GRNmap produces an Excel spreadsheet with an adjacency matrix representing the network. 0 represents no relationship between a regulator and its target gene. Initially, 1 represents a regulatory relationship where the gene specified by the column controls the gene specified by the rows. After the parameters have been estimated, a positive weight value indicates activation and a negative weight value indicates repression. The magnitude of the weight defines the strength of the relationship. However, GRNmap does not generate a visual representation of GRNs. Thus, a need for a visualization tool arises.

GRNsight was developed after the Dahlquist lab observed that commonly used graphing software for similar adjacency matrices was most optimized for large data sets. Software like yEd (Wiese et al., 2004), Cytoscape (Shannon et al., 2003) and Gephi (Bastian et al., 2009) are common software packages that allow for the visualization of networks. However, these tools are most effective in organizing and displaying large data sets (larger than 100 nodes). This means that the emphasis of the visualization is on important or central nodes. In the Dahlquist lab, the primary interest lies in the relationships between the nodes (or genes) rather than the nodes themselves. All three of these tools are also stand-alone applications, meaning that they must be installed on the machine that the tool would be run on. While this provides additional options and features not able to be provided with a web-based program, it also steepens the learning curve for a visualization software and limits the ability of users to access visualizations from anywhere. As a result, of the increased options and complexity, the user interfaces are extremely convoluted and difficult to use, creating an additional barrier to the ability of the Dahlquist lab to be able to integrate them into our data processing pipeline. Furthermore, yEd is not an open source project; this means that there is no opportunity for outside contributions to the project and the software cannot be built upon by others. Additionally, Cytoscape does not take in adjacency matrices as a data input, limiting the ability of compatibility with the data types that the Dahlquist lab works with. Gephi, on the other hand, is not targeted towards biological applications and thus lack visualization features and ease of use customized for a dynamic systems biology laboratory like the Dahlquist lab. However, some functions of Gephi are currently utilized by the lab for graph statistics, a feature currently unavailable in GRNsight.

To develop a successful visualization tool, the Dahlquist lab determined that it was essential that GRNsight contain features common to any successful visualization tool for biological pathways. These features are best described by Pavlopoulos et al. (2015) and Saraiya et al. (2005). One of the first requirements is that pathways must be automatically constructed; for example, this would mean that a pathway constructed in Adobe Illustrator or Microsoft PowerPoint is not a good representation for visualizing a pathway. The visualization should also not be represented statically, there must be some way to dynamically modulate the graph. Other requirements included being able to gain context for the pathway within the application. Additionally, they stated that the visualizations should be able to provide information through their visual representations (Saraiya et al., 2005). Building on this, Gostner et al. (2015) state that the aesthetic appearance of the model must promote the readability and maintainability of the model. Finally, many of the reviews of current systems biology visualization tools state that the visualization tool must promote the sharing of data and results throughout the scientific community. This includes features like data interoperability, visual sharing, and the ability for collaboration (Saraiya et al., 2005 & Pavlopoulos et al., 2015).

GRNsight is optimized for visualizing the relationships in small- to medium-scale gene regulatory networks (Dahlquist et al., 2015). Version 1.18.1 is shown in Figure 1, and the latest version, 3.1.15 is shown in Figure 2. GRNsight uses the Data-Driven Documents (D3) (Bostock, 2011) JavaScript library to generate a graph derived from input network data. D3 dynamically manipulates HTML and Scalable Vector Graphics (SVG) to form the elements of the graph. GRNsight implements D3’s force layout algorithm which applies a physics-based simulation to the graph. Users are able to modify the physics of this simulation. GRNsight then allows users to easily move nodes or lay them out automatically in an alphabetized grid. Nodes can also display time course data sets as a heat map to allow users to better understand the relationship of the edge to the expression of the gene. GRNsight also allows users to import networks in the commonly used formats SIF and GraphML. Features like these, among others, have enabled GRNsight to become an essential utility in the data processing pipeline utilized by the Dahlquist lab.

GRNsight has been built and revised to its current version over the course of five years. The initial development began in 2014, and since then has had 9 different developers contribute towards its codebase. GRNsight developers have tried to follow the Test-Driven-Development best practice consistently that encourages developers to write tests for any potential changes before actually making those changes (Bissi, 2016). However, even given this aspirational best practice, the codebase has gotten quite disorganized having been quickly developed by successive generations of undergraduate students. This has led to detrimental effects, such as a rapidly increasing number of bugs in the software, which correlated with the increasing number of features. The codebase is not as user-friendly as it could be for new developers on the project.

We determined that a central state register was the best way to resolve the core issues GRNsight was having: an increasing number of bugs and vastly disorganized code. We believe that the Model-View-Controller (MVC) software architectural pattern was the best implementation of this system. MVC was first described to the general public by Krasner & Pope as a pattern for implementation of user interfaces in 1988. After the rapid rise of web applications in the dawn of the internet, MVC quickly became the software architecture of choice for most use-cases. Leff and Rayfield (2001) described the implementation of a basic web application using this paradigm. The architectural paradigm calls for the separation of the central state of the application (the Model) from the implementation of the code that was the basis of what the user saw (the View), from the code that handled any particular interactions the user could have with the web application (the Controller).

I will describe the major revisions to GRNsight since its last major release of version 3.0.0 in April of 2018. This includes the addition of new features to make GRNsight more accessible to users, allow for better sharing of graphs, and increase visualization options. Furthermore, I will also describe the implementation of the new MVC architecture for GRNsight and its significance to how GRNsight operates today. I will show that GRNsight is a powerful tool that accomplishes its task of visualizing small- to medium-scale gene regulatory networks in a lightweight, easy-to-use, and feature-rich manner, all while implementing industry-standard best-practice software development practices.

Results

*Codebase*

Since its inception, GRNsight’s client and server code have been written in JavaScript. GRNsight uses Node.js (Node.js Foundation, 2009) as a server-side runtime environment, with ExpressJS (Holowaychuck & Wilson, 2010) as a framework; it delivers key information to the front-end client code that runs the network visualization. Since 2018, GRNsight has used Webpack (Koppers et al., 2015) in order to attain faster delivery of the service, as well as increased modularity of code. Webpack allows developers to bundle assets, such as images and web elements, and create bundles within the code itself, thereby reducing the amount of front-end code that a browser must process in order to deliver an intended service. The introduction of Webpack to bundle assets necessitated a move of the codebase from the fifth ECMAScript specification of JavaScript, which was released in 2009, to the sixth ECMAScript specification released in 2015, commonly known as ECMAScript 2015 or ES6 (Ecma International, 2015). This allowed for the use of advanced features; this includes concise function declarations, as well as exports and imports which allow developers to centralize functions, constants, and variables used throughout an application’s codebase. Furthermore, the addition of classes gives the GRNsight development team the option of introducing further modularity. However, the conversion to ES6 also necessitated the addition of Babel (Tschinder et al., 2014), a compiler, to convert ES6 and newer specification code into a backwards compatible version of JavaScript used in current and older browsers or environments. This ensures the ES6 code in the GRNsight codebase is compatible with browsers across the entire spectrum of our potential users. Additionally, the addition of Babel prepares the GRNsight codebase for the porting to the popular ReactJS (Facebook, 2015) framework, as it compiles JSX syntax, a portmanteau of HTML and JavaScript used heavily in ReactJS. Other important packages include D3.js (Bostock, 2011), the main visualization library, ESLint (Zakas et al., 2015) for code formatting, and Chai (Luer et al., 2014) and Mocha (Hiller et al., 2014) for testing. D3.js is one of the most commonly used JavaScript libraries for the creation of visualizations.

*Current features*

Currently, GRNsight contains a plethora of features that are used to help provide better visualization options. Since the graph is rendered with physics enabled on nodes and edges, there is an option to automatically lay out the graph in an alphabetized graph (*Fig. 5*); there are also sliders provided to modify the charge of nodes and increase or decrease the link distance of edges if the graph remains in a force graph layout, as well as lock those sliders. If timecourse expression data is provided in an uploaded worksheet, options are provided to choose which expression data set to display on the top or bottom half of the node; users are also provided options to average the replicate values of the timecourse data and to modify the log fold change max value. One can also change the size of the viewport to three predetermined options or to automatically fit the screen, as well as choose to restrict the graph to the viewport. Inside the viewport, arrows and sliders are provided to move the viewport around, or to zoom in and out of the graph. The user may also use their mouse for those controls. Finally, users may modulate the value against which edge weight magnitudes are normalized against, choose a threshold at which to display an edge weight as gray, or show gray edges as dashed.

*Architecture of GRNsight v. 1.0.0 to GRNsight v. 3.1.3*

The GRNsight project was originally written in the spring of 2014 with the first release coming online on March 20, 2014. The original codebase was less than a third of the size than it currently is. GRNsight version 1.0.0 features were limited to simply importing a Microsoft Excel workbook and displaying a graph with unweighted or weighted edges (*Fig. 2*); thus, very few files were contained in GRNsight’s codebase. At the end of 2014, the folder that contained the JavaScript code that runs the web client contained only five files. Only two of those files contained code written by GRNsight’s original developers. The other three files were necessary imports, such as the jQuery (jQuery Foundation, 2006) library and a package called imagesLoaded (DeSandro, 2012) to detect when all images have been loaded. As additions of new front-end features accelerated, a similar structure was retained to the initial file structure of GRNsight, with each new feature getting a new file unless absolutely unnecessary. Immediately before the beginning of the current refactor work that started in May of 2018, nine files existed in the folder containing the web client code, of which only two files were for external library imports.

In the original architecture, all three aspects of the application were contained in singular files (*Fig. 1*). These three aspects are what the user sees, the controls for how the user can modulate the application, and what the state of a feature is. I will illustrate this pattern using the force graph layout function. Upon uploading an appropriate file, the user sees the force graph layout with default physics enabled. The sidebar displays sliders that can be used to alter the current values of link distance and charge. The link distance slider controls the length of the edges between nodes and the charge determines the repulsive force between nodes. The state of the application consists of the current value of the link distance and charge sliders. Under the previous architecture, all aspects related to these sliders co-existed in the same file, and often all code for the sliders were contained in the same function. This file structure of GRNsight had the potential to be very problematic due to conflicts between different aspects and features of GRNsight was great. These clashes could lead to bugs when a user selected a certain combination of actions. The number of bugs that arose led to the decision to convert the entire front-end codebase to an architecture following the commonly used Model-View-Controller paradigm.

*New architecture for GRNsight v4.0.0*

GRNsight’s application of the Model-View-Controller paradigm (*Fig. 4*) has a similar web application interface design proposed by Lee and Rayfield (2001). This proposal for architecture proposes information displayed to the user be controlled by three separate segments: a Controller portion of the application which processes the user's interaction, a Model portion of the application that contains both the information represented by the View portion. While following this basic paradigm, the state of the application interface is separated from the logic that controls changes to that application state.

The majority of GRNsight’s user interface code now resides in six files: setup-handlers.js, grnstate.js, update-app.js, setup-load-and-import-handlers.js, constants.js, and grnsight.js (*s*). The grnsight.js file’s functions to import and compile the various modules necessary for proper web application function into one complete file, a minified version of which is then served to a browser. The file constants.js acts as a module for the Webpack software to parse and populate multiple repeated variables across most of these files. This improves code comprehensibility and allows for descriptive naming of these variables across the main components of the new architecture. The remaining four files contain the bulk of the application’s functional codebase.

The central store for the application state resides in an object called GRNstate, which is in the eponymous grnstate.js. This file serves as the Model component of the MVC architecture (*Fig. 4*). The file contains one central object with 30 properties, such as the default link distance and charge slider values, and two functions for getting and setting the network parsed by the server-side code. The state of every property that can be modulated, except for the size of the container of the graph, is stored inside this central model. It additionally contains a helper method to annotate the value of each of the edges in the network. Update-app.js is the View component of the new GRNsight architecture (*Fig. 4*). The update-app is triggered by changes detected by the Controller component, called setup-handlers in GRNsight. Along with displaying the user interface of GRNsight, update-app.js contains code to handle any potential actions allowed to be taken by the user. Given the various options for changing the graph, the code has been refactored in such a way so as to allow for the least number of conflicts. No functions for the handling of different actions interact with each other, and none of the helper functions are called by functions handling different states of the application. The last central file to the new GRNsight architecture is named setup-handlers.js, which serves as the View component (*Fig. 4*). This file contains a function named setup-handlers which receives changes from the user and updates the central store of the app state. A secondary file imported into setup-handlers.js is named setup-load-and-import-handlers.js and contains the code necessary for the loading of the application container and the import of adjacency matrices into GRNsight. Setup-handlers.js contains a separate handler function for every action available to the user; each function changes a property of the GRNstate object and triggers the update-app function in order to execute the appropriate action. These two files would constitute the Controller portion of the MVC architecture. The remaining files contain the basic code using the D3.js library necessary to create a graph with physics.

The order of movement of code during the refactor was especially important in transitioning the entire codebase. Each distinct feature was relocated to the new set of files separately to ensure the least disruption of the feature set. Additionally, each feature was tested completely to ensure that it worked fully before the work to transition the next feature to the new architecture. The first feature to be fully moved was displaying gray edges as dashed lines. This feature was moved first because it consisted simply of a checkbox that changed a simple appearance of graph edges. After accomplishing this, the order of the move followed a relatively simple logical progression, moving progressively larger and larger features. After establishing a process to follow when moving features, the set of functions that allowed users to change the values that edge widths were normalized against were moved into update-app.js, grnstate.js, and setup-handlers.js. Next, the slider that controls the edge weight threshold to display edges as gray instead of colored was moved. The next relatively easy feature to move over was the three radio buttons which allow the user to decide when to display edge weights. However, after these features were moved, the largest portion of the work followed, with the movement of the sliders that control the graph layout physics, as well as the recently added button which allows a user to automatically layout nodes in a grid. These features were heavily inter-mixed into the code which controls the main visualization of the graph. Thus, movement of these features often broke other features, requiring their move to be completed together. Finally, the last feature that was moved was the node coloring feature, which existed solely in one file, making its movement straightforward, but arduous, because of the tight integration of all the functions and variables contained within the feature’s code file.

*Visualization improvements*

A couple of visualization features were added over the course of my work on the web application. The first was the ability of a user to control the edge weight threshold that determined whether to display edges as gray. Previously, edges were displayed as gray if their weight values were below five percent of the normalized maximum edge weight value. Implementation of this feature was simple; a user-input slider that displays a percentage was coded. If the normalized edge weight values are below the user-set percentage of the maximum edge weight value, the edges are displayed as a gray edge (*Fig. 5*). This feature is important because it allows the user to be able to focus only on edges with certain distinguished weights. Users control this feature via a slider that corresponded to a percentage of the largest magnitude edge weight. By default, the normalization factor is set such that edges with an edge weight magnitude of less than 5% of the largest magnitude edge weight are visualized as gray. Movement of the slider increases the percentage by one percent increments, up to a maximum of one hundred percent and a minimum of one percent. Furthermore, a checkbox was implemented to show gray edges as dashed so color blind users can better distinguish a gray edge from a colored edge. This was achieved by adding a boolean variable that updates to true upon the change of the checkbox’s status. If the edge weight is below the threshold and the boolean is true, then the edge takes on a dashed style (*Fig. 8*). Finally, another feature which I worked on allows users to normalize edge weight visualization against a user-inputted value, instead of the default largest magnitude edge weight value. A text box was created along with set and reset buttons. The reset button was set to default to the largest magnitude edge weight value. Inputting a value and clicking the set button runs a function which normalizes each of the edge weights against the user input and refreshes the graph (*Fig. 6*). In order to work within d3.js’s visualization limitations, edges are grouped into twelve “bins”, corresponding to edge thickness. The edge with the largest weight magnitude is placed in the bin that corresponds to the largest possible edge thickness. Edges are then assigned to subsequent bins based on what percentage of the largest weight their weight magnitudes are. All three of these new features were created before the implementation of the current architecture and were thus moved into the new architecture.

Discussion

*Central store for application state*

The addition of a central state for GRNsight in the form of an object creates a cross-reference that is easily accessible by both the code itself and the developers working on the code (Hansen & Fossum, 2005). This central state, in the form of an object in GRNsight, stores most of the information regarding the application state, save for the position and physics of the actual nodes and edges displayed in the simulation. There are many advantages to having this central store of information. First, having a central state store prevents unintended state conflicts. Conflicts between application sub-states often create bugs that, especially in front-end code, may affect a user’s experience. According to Ceaparu et al. (2004), errors in an application are a leading cause for end-user frustration. A few bugs that were present under GRNsight’s previous architecture were immediately removed with the conversion of the architecture to MVC. Some of these bugs occurred when a user changed the normalization factor of a graph’s edges while a graph was in a grid layout configuration. Under the previous architecture, the graph would revert back to a force graph layout. This was resolved after the transition to the MVC architecture, as the separate controllers for these features referenced different parameters within the central state store. However, when bugs do occur under the MVC architecture, having a central state store makes the diagnosis of the bugs straightforward. After the transition of the application to the new architecture, a bug arose where, even upon clicking on the “Lock Force Graph Parameters” check box, the Force Graph Parameter sliders would not be locked. Diagnosing and creating a solution involved simply looking for all functions that relied on that parameter of the central state store, finding the conflict, and removing the conflict.

Finally, having a central state store allows ease of use for future development (Leff & Rayfield, 2001). In addition to allowing for minimization of potential state conflicts that could arise from new features, the central state allows the developer to access the code more easily and add new modules. Under the previous architecture, blocks of code that might have controlled the same feature were interspersed throughout various files or in different parts of the same file. Under MVC, this code is now guaranteed to be contained between three files, and functions are organized according to the specific feature that they control or interact with. Functions that are shared between multiple functions are located as close to the top of the file as possible. Furthermore, the transition of GRNsight to ECMAScript 2015 (ES6) allows for the importation of code from different locations, allowing code to be much more organized and increasing potential modularity. Components of the MVC architecture are separated into modules located in separate files because ES6 allows each file to nonetheless interact with each other. Webpack additionally allows for this modularity to continue to work across all platforms, including browsers that may not yet support the new ECMAScript featureset.

The newly centralized architecture of GRNsight allows developers to add new features in a highly intuitive and modular fashion. The way that the Model-View-Controller is implemented has a profound impact on how easy the code is to develop, maintain and reuse (Hansen & Fossum, 2005). Thus, the way that MVC was implemented in GRNsight was made such that harmful interactions were minimized, and that minimal code was situated in locations outside the MVC architecture. Most code that faces the client-side is situated in three files each of which constitutes a component of MVC. Thus, adding new features simply requires the addition of the HTML into the static template files, creation of a handler for the feature into setup-handlers.js, addition of a new application state to grnstate.js, and addition of code to execute the user’s input and update the view to update-app.js. Furthermore, the entire separation of Model, View, and Controllers enable a developer to create pluggable views/controllers; in other words, the complete separation of all components allows a developer to change the functionality or appearance of a different component of a feature without the necessity to completely refactor all code associated with a feature because the code resides completely in one function. However, in the same vein, some modularity can also be hindered by the strict MVC architecture. It is significantly harder to reuse view and controller code because both are tied to the specific model. Thus, a developer must take extra and often superfluous steps to ensure full code compatibility if there are any changes to the model, or in order to ensure reusability of model or controller code (Veit & Hermann, 2003).

*New features*

Three new visualization features were added which allow users to set a threshold at which to visualize edges as gray or colored, view those gray edges as dashed lines, and to allow users to normalize edge weight visualization against a user inputted value. Allowing users to set a threshold at which to visualize gray edges allows users the opportunity to determine for themselves edges of importance, rather than having GRNsight doing it for them. The default of five percent was chosen as these relationships tend to be fairly weak in a small- or medium-scale gene regulatory network. However, this feature was added because it allows users to be able to decide for themselves which edges are important and share those with peers. Since edges are colored blue and red to correspond to activation or repression relationships, setting an edge weight visualization percentage allows viewers to pay attention to those colored edges rather than the thin, weak relationships.

The addition of the feature that allows users to control the visualization of gray edge weights as dashed was motivated primarily because a member of the research group was color-blind. The color-blind member of the research group pointed out that it was difficult for him to distinguish between thin magenta/red edges and gray edges. Red-green color blindness affects approximately 8% of Caucasian males, 5% of Asian males, and 4% of African males worldwide (Birch, 2012); in order to make GRNsight available to a larger portion of the scientific community, solutions were considered to allow for more accessible viewing options. Two solutions were ultimately chosen; the first to change the edge weight coloring from cyan and magenta to blue and red to make edges easier to distinguish, and the second to display gray edges as dotted lines, so that even without color users would be easily able to distinguish thin colored edges from gray edges (Okabe & Ito, 2002).

Finally, the last feature added in to improve GRNsight’s functionality was the option to set an edge weight normalization factor. Testing within the Dahlquist lab revealed that biologists are often trying to compare graphs of multiple gene regulatory networks in which the strongest magnitude relationship may not always be the same value. Thus, the original design of GRNsight did not facilitate accurate comparisons of graphs. Allowing the user to set edge thickness normalization factor to a certain value facilitates the comparison of graphs with differing maximum edge weight magnitudes. Thus, research that relies on the comparison of different gene regulatory networks can be easily shared.

*Fulfillment of initial goals*

GRNsight fulfills many of the goals set forth at the inception of the project. The new features, in combination with the existing features, fulfill the guidelines of a successful visualization tool ascribed by Saraiya et al. (2005), Pavlopoulos et al. (2015) and Gostner et al., (2015) discussed earlier in this paper. The pathways are automatically constructed upon the import of a compatible file type and dynamically displayed in a manner that allows users maximum control over the placement of nodes. Dynamic display of the graph includes options like displaying the node in a grid and control over the charge of nodes and link distances between them. The visual language of GRNsight is standardized in both edge relationships between the nodes and the display of timecourse gene expression data within a node. All controls are standardized in two locations, the sidebar, and the top menu bar to provide maximum accessibility to graph controls. The new features added to GRNsight since version 3.0.0 also maximize accessibility for the greatest number of users and promote the sharing of data within the scientific community by allowing uncomplicated controls over visual sharing options. Finally, users may visualize pathways in a larger context when they click on a node. This opens a webpage which compiles data on a gene in *Saccharomyces cerevisiae* from five different prominent databases. Because GRNsight is hosted on a webpage, it is theoretically accessible from anywhere at any time because the only requirement for GRNsight operation is a relatively modern web browser. Furthermore, it is an extremely lightweight application. Internal testing reveals that the GRNsight webpage loads within approximately 1200 milliseconds. Demo graphs are loaded within approximately 500 milliseconds. Finally, being open source allows GRNsight to be completely transparent and flexible for the scientific community.

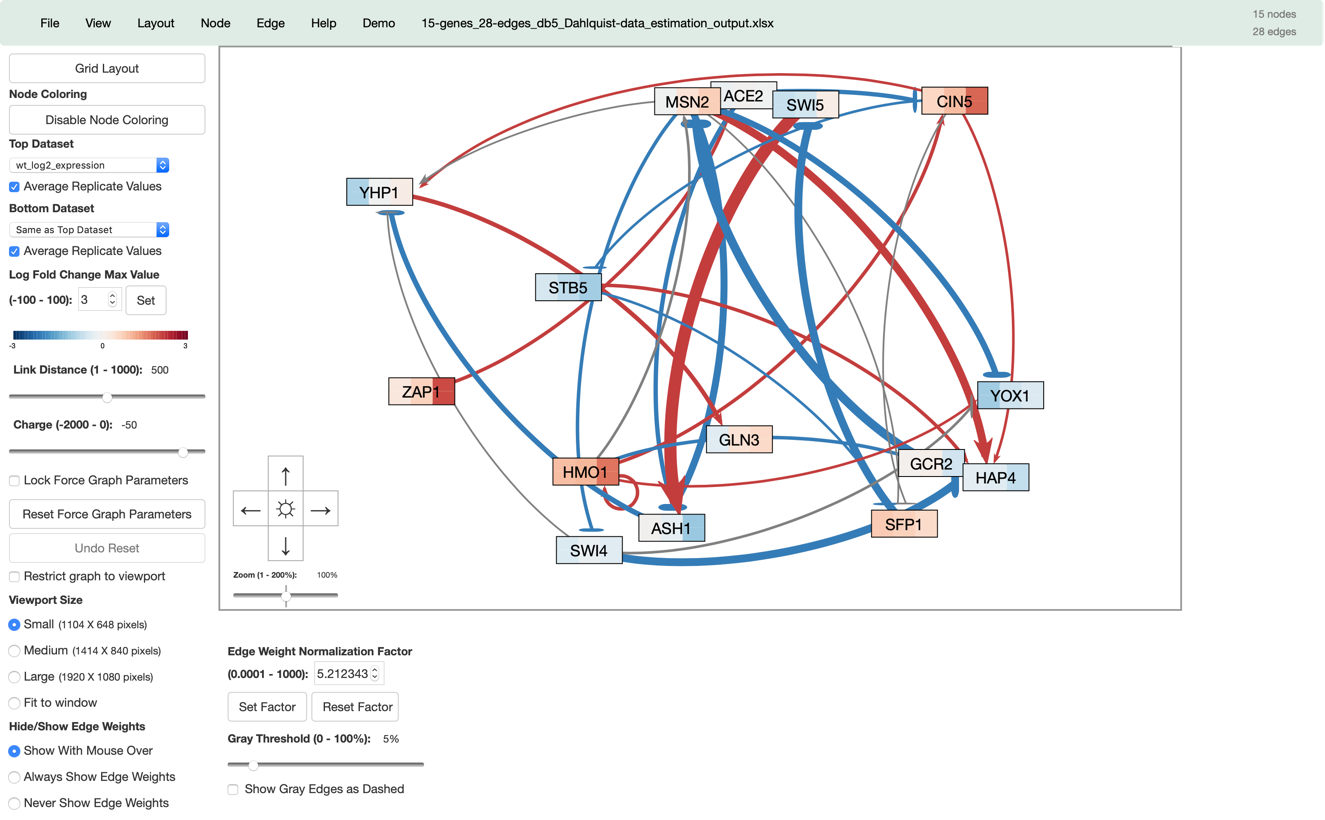
*Future directions*

Now that GRNsight is a much more modular, developer-friendly, and user-friendly software following the refactor to MVC and addition of new features, there are many possible directions that the GRNsight team could take with our software. One feature currently in development allows users to export GRNsight data to Excel workbooks. This would hopefully allow for the future automation of the creation of GRNsight compatible workbooks. JavaScript libraries like Cytoscape.js library (Franz et al., 2015) allows the potential addition of useful graph statistics. These statistics include global clustering coefficient, betweenness centrality, and degree distribution. Statistics like these allow scientists to visualize important or insignificant relationships between genes (nodes) and to determine important properties regarding the significance of certain nodes or graphs. Systems biology analysis tools such as InnateDB (Breuer et al., 2012) and ConsensusPathDB (Kamburov et al., 2012) are two examples of software that have already successfully implemented the functions that use the graph statistics in Cytoscape.js. However, Cytoscape.js does not take into account directed edges so significant modification would have to be done to the application in order to achieve the desired results within GRNsight. Additionally, use of a new testing architecture in conjunction with the existing client-side testing document to better follow the new web-client architecture would allow GRNsight to retain a more intuitive testing process and would help to better attain bug-free software. This transition would also go hand in hand with the addition of tests specific to the new architecture in order to help GRNsight return to the test-driven development (Beck, 2003) philosophy, which it has strayed from in recent history. Additionally, in order to better increase GRNsight’s share-ability, a feature could be added to allow export of a graph visualization to an easily shareable format such as Portable Network Graphics (PNG) or Scalable Vector Graphics (SVG). Finally, in the near future, a move of the front-end architecture to ReactJS is planned. ReactJS is a JavaScript library developed by Facebook designed especially for single-page applications (Vipul & Sonpakti, 2016). The advantages of using ReactJS include increased modularity and the creation of a “virtual” Document Object Model (DOM), or tree structure for encoding every web page. ReactJS creates an in-memory cache of the DOM, and, upon user-changes, only renders the sub-components that have been changed, allowing a developer to code the project with extreme modularity. ReactJS is an important step for GRNsight to take as it allows developers to future proof the addition of new features.

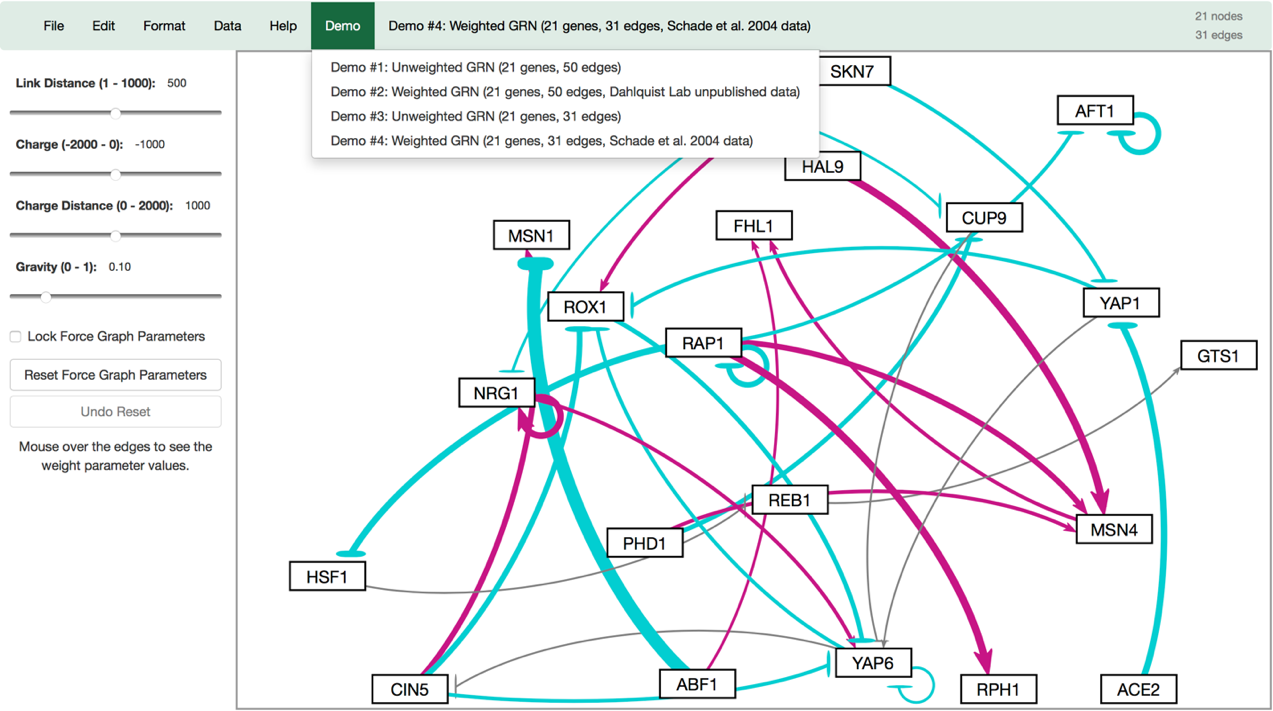
*Conclusion*

GRNsight has undergone many changes since its inception in 2014; however, it has retained the same architecture for the majority of its lifetime. During the course of my work on this project, a major effort was undertaken to convert GRNsight to more sustainable architecture, termed Model-View-Controller. This architecture breaks GRNsight into modules resulting in better developer access, and fewer bugs. Finally, we anticipate the addition of new features that allow users to better modulate the network graphs created by GRNsight allow users better options and more flexibility.

Figures



*Figure 2: A view of what GRNsight v. 3.1.15 looks like with the addition of all the new features. This is in comparison to the features present in GRNsight v 1.18.1, seen in Figure 2.*



*Figure 1: A screenshot of the GRNsight application interface at the release of version 1.18.1. Only a few options exist to allow users to modulate the graph. The gravity and charge distance sliders have since been eliminated.*

**activation**

**repression**

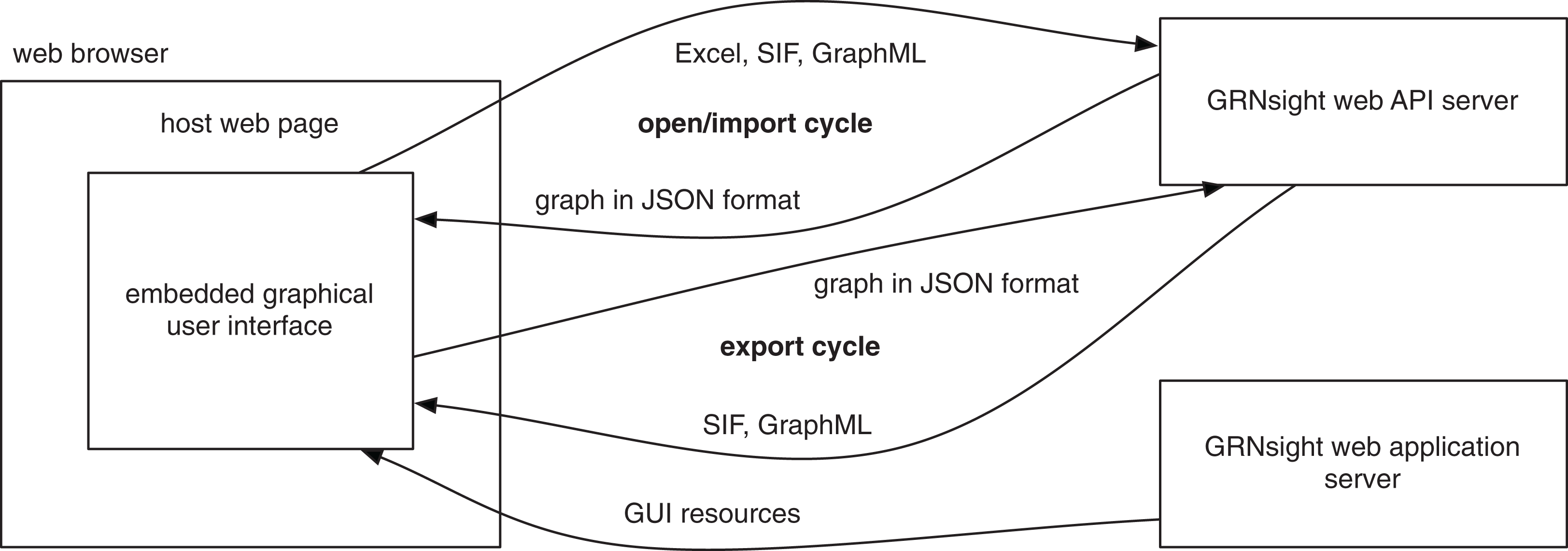
**weak influence**

**Key:**



**gene**

s

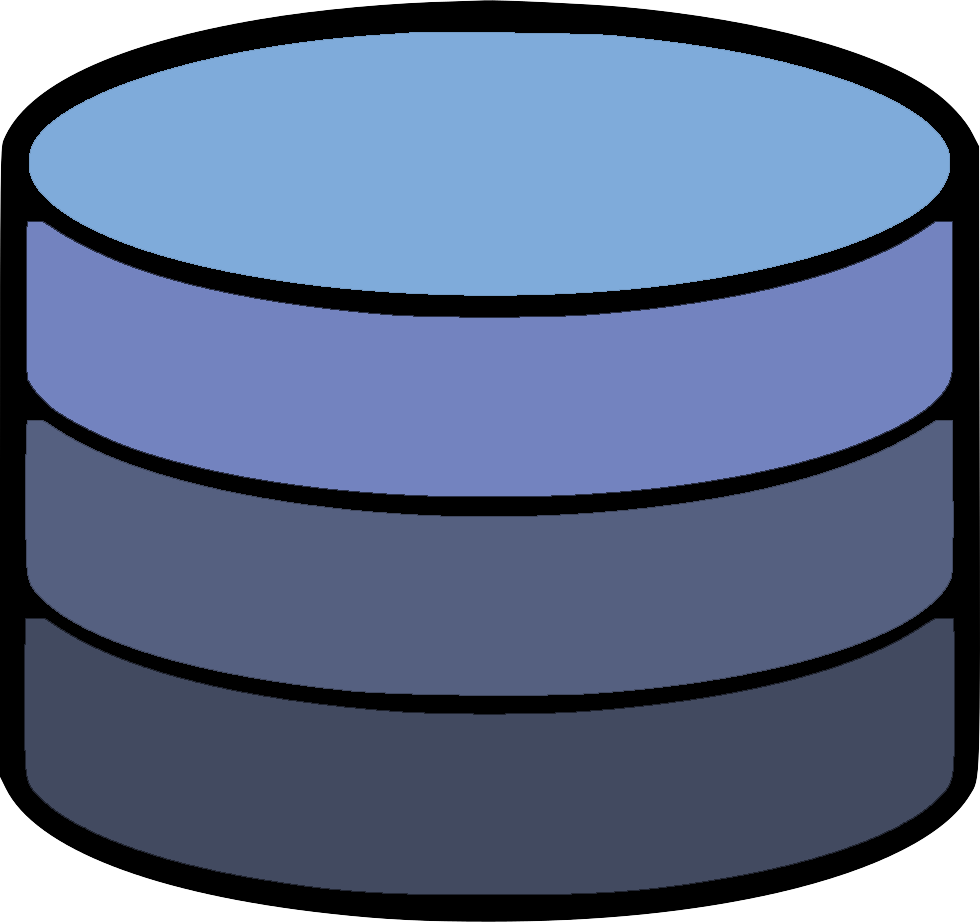


*Figure 3: A representation of GRNsight's version 1.0 architecture. Notably, there is no emphasis on maintaining a Model-View-Controller paradigm to reduce application instability (Dahlquist et al., 2016).*

User

Controller

View



Uses

Manipulates

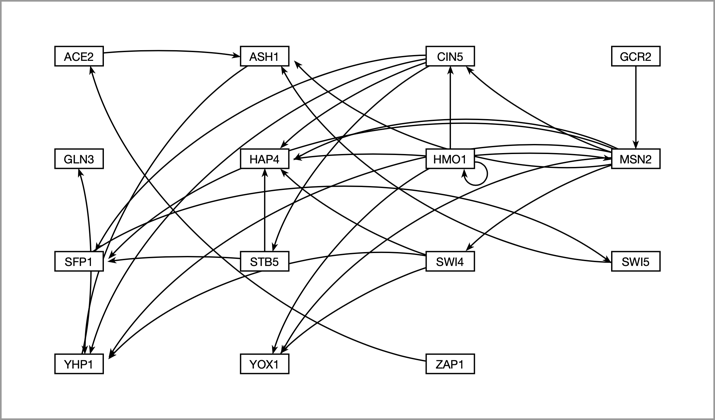
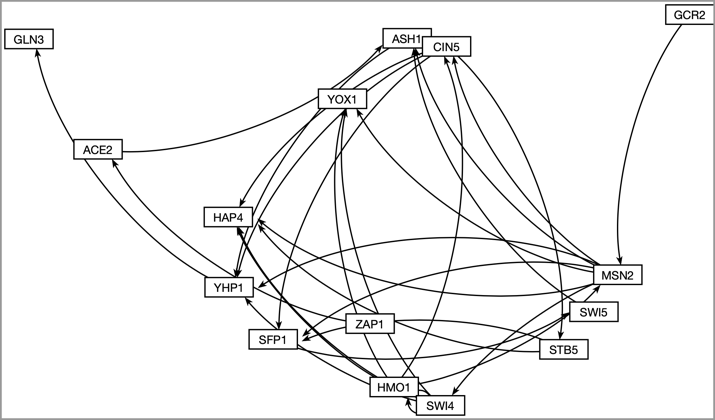
Updates

Seen By

Model

Figure 4: A visual representation of the Model-View-Controller paradigm. This software architectural pattern requires the codebase to be contained in three distinct processes that contain specific controls of the application. The user uses a portion of the application, which is heard by the controller, and relayed to the model. This updates the state of the application, which in turn is heard by the View code, which updates what the user sees.

*Figure 5: An example of a graph without edge weights or node coloring rendered by GRNsight based on worksheet data on the left, and the same graph laid out automatically in an alphabetized grid layout on the right.*



**Force Graph Layout**

**Grid Layout**

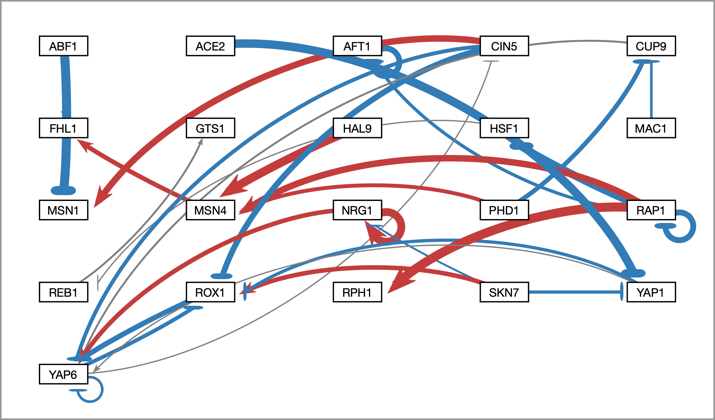
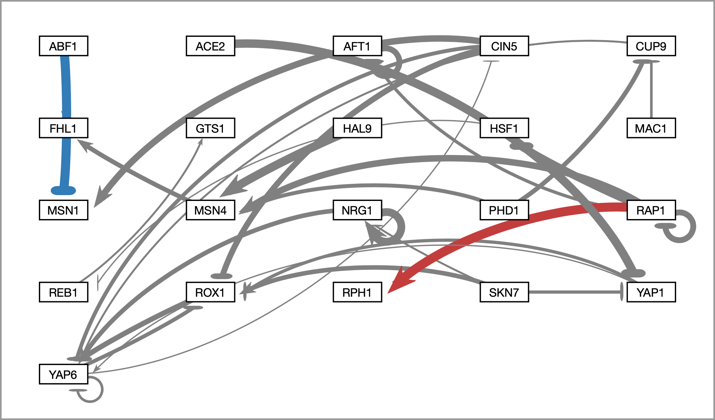


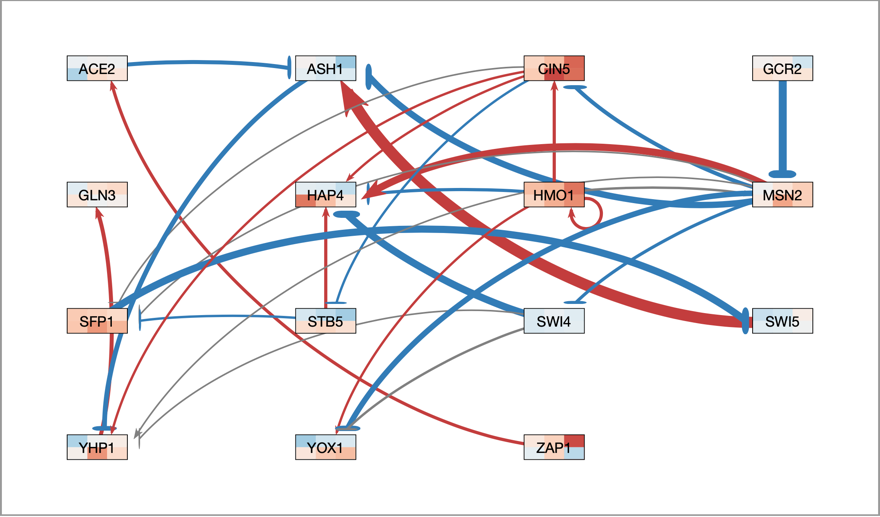
Figure 6: Medium-scale gene regulatory networks automatically produced from GRNsight. Graph A has its edge weight threshold for edges to be shown as gray set at the default 5 percent of the largest edge weight in the graph. Graph B has the threshold set at 80 percent of the largest edge weight. Nodes remain in a similar location when changing the threshold on the slider when in force graph mode. Changing the threshold does not affect the position of the nodes when in grid layout.

**Gray Edge Threshold = 5%**

**Gray Edge Threshold = 50%**

**Top Node Coloring Data Set = wild type microarray data**

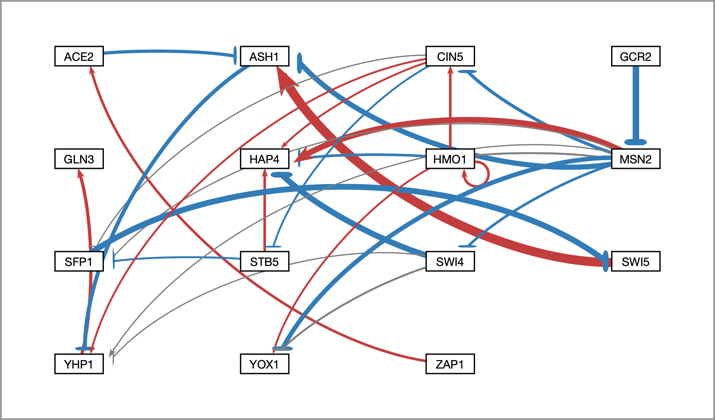
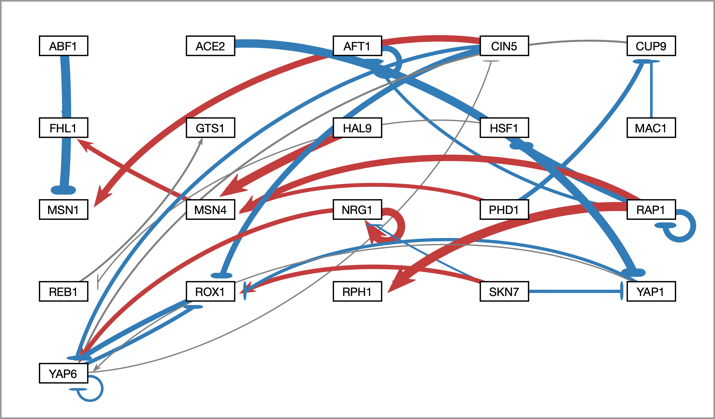
**Bottom Node Coloring Data Set = *gln3* microarray data**



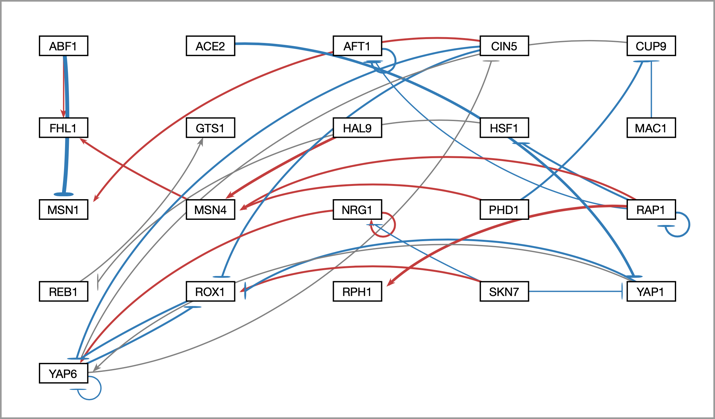
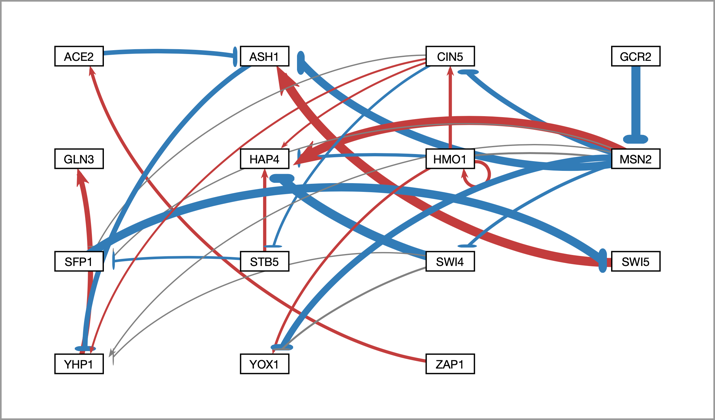
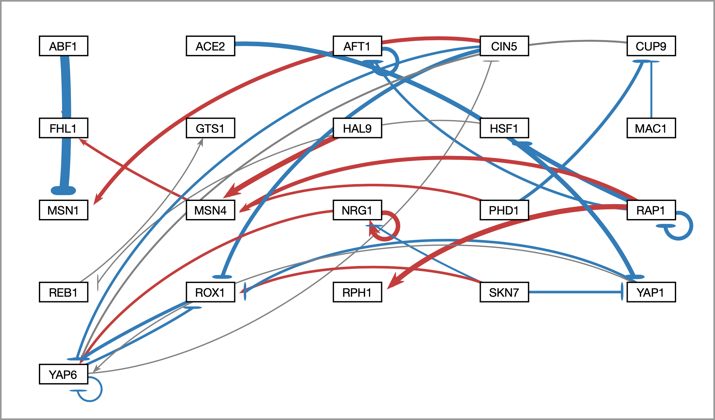
*Figure 7: Node coloring allows users to see how timecourse expression changes affect the overall relationship that the transcription factors may have on the expression of other genes. Users have the option to be able to choose up to two different timecourse gene expression data sets to overlay on each node. In this case one can see the mild overall expression relationship comes of the outlined genes over a period of time in both the top wild type microarray data and the bottom* ∆*gln3 microarray data.*

Figure 7: This shows a small graph rendered by GRNsight, with different edge weights. The top two graphs have been rendered as is, and the graphs below have been paired and set to use two different normalization factors to show the effect of setting various normalization factors. The middle left graph now has an edge thickness rendered by GRNsight that is more accurate to the actual relationship between nodes when compared to the middle right graph.

**Normalization Factor = 1.5**



**Normalization Factor = 5.21**



**Normalization Factor = 3**

**Normalization Factor = 3**

**Normalization Factor = 10**

**Normalization Factor = 10**

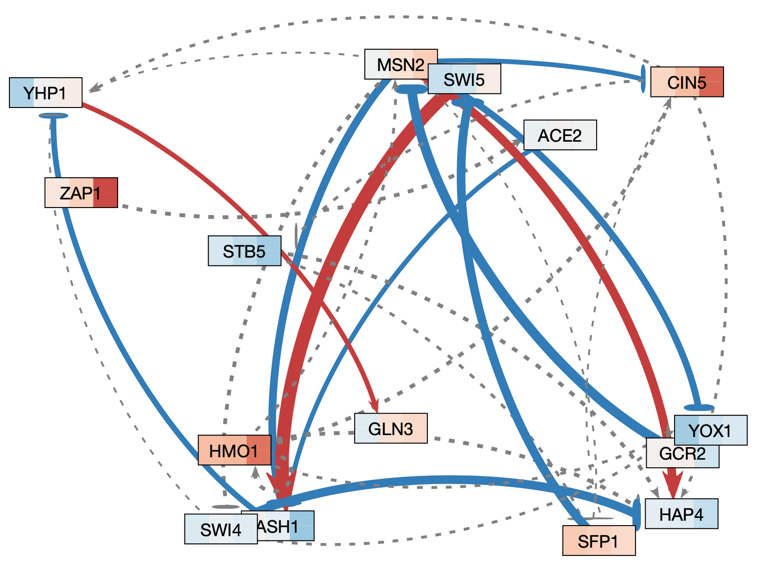
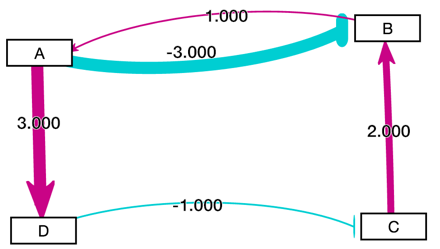


Figure 8: A medium scale gene regulatory network laid out automatically by GRNsight. Gray edges are shown visualized as dashed lines. This feature was added to assist colorblind users with distinguishing between thin edges with an activation or repression relationship, and edges with a negligible effect on target genes.



References

Bastian, M., Heymann, S., & Jacomy, M. (2009, March). Gephi: an open source software for exploring and manipulating networks. *In Third international AAAI conference on weblogs and social media*. Retrieved 1 May 2019, from https://www.aaai.org/ocs/index.php/ICWSM/09/paper/view/154/1009

Ashbacher, C. (2003). " Test-Driven Development: By Example" by Kent Beck. *Journal of Object Technology*, 2(2), 203-204. doi:10.5381/jot.2003.2.2.r1.

Birch, J. (2012). Worldwide prevalence of red-green color deficiency. *Journal of the Optical Society of America A*, 29(3), 313-320. doi: 10.1364/JOSAA.29.000313

Bissi, W., Neto, A. G. S. S., & Emer, M. C. F. P. (2016). The effects of test driven development on internal quality, external quality and productivity: A systematic review. *Information and Software Technology*, 74, 45-54. doi:10.1016/j.infsof.2016.02.004

Bostock, M. (2011). D3: Data-Driven Documents. Retrieved 1 May 2019, from *https://github.com/d3/d3.*

Dahlquist, K. D., Fitzpatrick, B. G., Camacho, E. T., Entzminger, S. D., & Wanner, N. C. (2015). Parameter estimation for gene regulatory networks from microarray data: cold shock response in Saccharomyces cerevisiae. *Bulletin of mathematical biology*, *77*(8), 1457-1492. doi:10.1007/s11538-015-0092-6

Dahlquist, K. D., Dionisio, J. D. N., Fitzpatrick, B. G., Anguiano, N. A., Varshneya, A., Southwick, B. J., & Samdarshi, M. (2016). GRNsight: a web application and service for visualizing models of small-to medium-scale gene regulatory networks. PeerJ Computer Science, 2, e85. doi:10.7717/peerj-cs.85

DeSandro, D. (2012). imagesLoaded: JavaScript is all like "Your images done yet or what?". Retrieved 1 May 2019, from *https://github.com/desandro/imagesloaded.*

Ecma International (2015). ECMA-262 6th Edition, The ECMAScript 2015 Language Specification. Geneva. Retrieved 1 May 2019, from http://www.ecma-international.org/ecma-262/6.0/ECMA-262.pdf/.

Facebook (2015). React: A declarative, efficient, and flexible JavaScript library for building user interfaces. Retrieved 1 May 2019, fromhttps://github.com/facebook/react/.

Franz, M., Lopes, C. T., Huck, G., Dong, Y., Sumer, O., & Bader, G. D. (2015). Cytoscape. js: a graph theory library for visualisation and analysis. *Bioinformatics*, 32(2), 309-311. doi:10.1093/bioinformatics/btv557

Gostner, R., Baldacci, B., Morine, M. J., & Priami, C. (2015). Graphical modeling tools for systems biology. *ACM Computing Surveys (CSUR)*, *47*(2), 16. doi:10.1145/2633461

Hansen, S., & Fossum, T. V. (2005). Refactoring model-view-controller. *Journal of Computing Sciences in Colleges*, *21*(1), 120-129.

Hiller, C., Holowaychuck, T.J., Roebuck, P., Müller, P., Juerg, B., Singh, G., Cataldo, K., & Bamieh, A. (2014). Mocha: a lovely testing framework for JavaScript. Retrieved 1 May 2019, from *https://github.com/mochajs/mocha.*

Holowaychuck, T.J. & Wilson, D.C. (2010). ExpressJS. Retrieved 1 May 2019, fromhttps://github.com/expressjs/express/.

jQuery Foundation (2006). jQuery – New Wave JavaScript. Retrieved 1 May 2019, fromhttps://jquery.com/.

Kamburov, A., Stelzl, U., Lehrach, H., & Herwig, R. (2012). The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Research*, 41(D1), D793-D800. doi:10.1093/nar/gks1055

Koppers, T., Ewald, J., Larkin, S., & Kluskens, K. (2015). Webpack: Module bundler. Retrieved 1 May 2019, fromhttps://github.com/webpack/webpack.

Krasner, G. E., & Pope, S. T. (1988). A description of the model-view-controller user interface paradigm in the smalltalk-80 system. *Journal of object oriented programming*, *1*(3), 26-49.

Leff, A., & Rayfield, J. T. (2001). Web-application development using the model/view/controller design pattern. In *Proceedings fifth ieee international enterprise distributed object computing conference* (pp. 118-127). IEEE.

Luer, J., Todorov, V., Cirkel, K., Fernandes da Costa, L., & Snodgrass, G. (2014). Chai: BDD / TDD assertion framework for node.js and the browser that can be paired with any testing framework. Retrieved 1 May 2019, fromhttps://github.com/chaijs/chai*.*

Node.js Foundation (2009). Node.js: Node.js JavaScript Runtime. Retrieved 1 May 2019, fromhttps://nodejs.org/en/.

Okabe, M., & Ito, K. (2002, November 20). *How to make figures and presentations that are friendly to color blind people*(Rep.). Retrieved May 1, 2019, from University of Tokyo website: http://jfly.iam.u-tokyo.ac.jp/html/color\_blind/

Pavlopoulos, G. A., Wegener, A. L., & Schneider, R. (2008). A survey of visualization tools for biological network analysis. *Biodata mining*, 1(1), 12. doi:10.1186/1756-0381-1-12

Saraiya, P., North, C., & Duca, K. (2005). Visualizing biological pathways: requirements analysis, systems evaluation and research agenda. *Information Visualization*, *4*(3), 191-205. doi:10.1057/palgrave.ivs.9500102

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., ... & Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, *13*(11), 2498-2504. doi:10.1101/gr.1239303

Tschinder, D., Ng, B., Smyth, L., Ribaudo, N., Sauleau, S., & Zhu, H. (2014). Babel: The compiler for writing next generation JavaScript. Retrieved 1 May, 2019, from *https://github.com/babel/babel.*

Veit, M., & Herrmann, S. (2003, March). Model-view-controller and object teams: A perfect match of paradigms. In *Proceedings of the 2nd international conference on Aspect-oriented software development* (pp. 140-149). ACM. doi:10.1145/643603.643618

Vipul, A. M., & Sonpatki, P. (2016). *ReactJS by Example - Building Modern Web Applications with React.* Packt Publishing Ltd.

Wiese, R., Eiglsperger, M., & Kaufmann, M. (2004). yFiles — Visualization and Automatic Layout of Graphs. *Mathematics and Visualization*, 173–191. doi:10.1007/978-3-642-18638-7\_8

Zakas, N.C., Partington, K., Volodin, I., Mills, B., Nagashima, T., Singh, G., Cataldo, K., & Katz, T. (2015). ESLint: A fully pluggable tool for identifying and reporting on patterns in JavaScript. Retrieved 1 May 2019, from *https://github.com/eslint/eslint.*