Anindita Varshneya

BIOL 598-04: Advanced Systems Biology Research

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

Annotated Bibliography

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. *ICWSM*, *8*, 361-362.

Gephi is a popular visualization software that competes with Cytoscape and GRNsight. Like most other biological network visualization systems, Gephi is focused on large-scale networks. It is worth mentioning Gephi in my paper because it further highlights how there are several different options for large-scale network analysis, but none of these options cater to small- and medium-scale networks.

Breitkreutz, B. J., Stark, C., & Tyers, M. (2003). Osprey: a network visualization system. *Genome biology*, *4*(3), R22. doi: 10.1186/gb-2002-3-12-preprint0012

Osprey is another software platform that creates visualizations of complex large-scale gene interaction networks. The program provides several options for user-defined interactions and customizations including node relaxation (similar to gravity in GRNsight), exports, and mouse-over functionality for more information. Osprey also takes advantage of other databases, such as Gene Ontology, to integrate additional information about the genes in the networks. Because I worked on hiding and showing weights, and because Osprey is another competitor of GRNsight, this paper is worth studying and mentioning.

DeRisi, J. L., Iyer, V. R., & Brown, P. O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science*, *278*(5338), 680-686. doi: 10.1126/science.278.5338.680

This paper analyzes the effectiveness and importance of using DNA microarrays when studying gene expression patterns. The researchers who wrote this paper also studied *Saccharomyces cerevisiae*, and wrote about the processes they used to understand gene expression during the metabolic shift from fermentation to respiration. Because this paper finished on the importance of good analysis tools when interpreting microarray data, I plan on using this paper to set up the usefulness of visualization tools in general.

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*, btv557. doi: 10.1093/bioinformatics/btv557

Cytoscape.js is an open-source graph library that we used within the GRNsight project. It provides algorithms for several different types of graph statistics and can also be used to create interactive graphs within web-browsers. Because one of my projects on the GRNsight project was beginning our graph statistics table, it will be important to mention and cite the library I used to create those statistics. As a side note, this paper was published with “visualisation” misspelled.

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

This paper created a list of important functionalities of graphical modeling tools for systems biology and analyzed several different visualization tools on how well they incorporated those functionalities. This paper will help build background information about the important features that graphical visualization systems should have, what other programs are lacking, and what GRNsight provides.

Kitano, H. (2002). Systems biology: a brief overview. *Science*, *295*(5560), 1662-1664. doi: 10.1126/science.1069492

This paper defines systems biology and explains the importance of approaching a research question with system-level understanding. Because the LMU Bioinfomatics and Biomathematics group functions as a system in our work overall, I felt this paper would provide background as to how and why this research group is split the way it is, and how the work that I have been doing for the past couple years fits into the bigger picture.

Longabaugh, W. J., Davidson, E. H., & Bolouri, H. (2009). Visualization, documentation, analysis, and communication of large-scale gene regulatory networks. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, *1789*(4), 363-374. doi: 10.1016/j.bbagrm.2008.07.014

This paper actually focuses on presenting another visualization software called BioTapestry. While BioTapestry is not as well known as the other visualization software I have cited in this bibliography, this paper is especially interesting because it outlines why GRN visualizations are so interesting and how these visualizations could be simplified so they are as understandable as possible. This paper can serve as further evidence and validation of some design decisions made on GRNsight.

MacNeil, L. T., & Walhout, A. J. (2011). Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome research*, *21*(5), 645-657. doi: 10.1101/gr.097378.109

While most other papers that I read have focused on visualization software and GRNsight’s competing software, this paper actually defines gene regulatory networks (GRNs). It describes all of the different types of research questions a GRN can answer as well as several different methods through which researchers can analyze and visualize GRNs. This will be important background information as the entire project is based on GRNs.

Mugridge, R. (2003, June). Test driven development and the scientific method. In *Agile Development Conference, 2003. ADC 2003. Proceedings of the* (pp. 47-52). IEEE. doi: 10.1109/ADC.2003.1231452

This paper defines test-driven development through the lens of the scientific method. The content of this article is highly applicable to GRNsight and my own paper because it identifies how TDD can be incorporated into the scientific method, similar to how GRNsight attempted to incorporate TDD into our own research practices. This paper is essential to bring light to one of my major contributions to the project.

Pastrello, C., Otasek, D., Fortney, K., Agapito, G., Cannataro, M., Shirdel, E., & Jurisica, I. (2013). Visual data mining of biological networks: one size does not fit all. *PLoS Comput Biol*, *9*(1), e1002833. doi: 10.1371/journal.pcbi.1002833

This article is especially interesting because it describes the usefulness of programs that build visualizations of large-scale networks vs. small-scale networks. The author details the importance of several different biological visualization tools and how they should be integrated to provide the most comprehensive analysis of biological processes. By providing a workflow through which the authors analyzed anti-cancer networks, they provide an outline of tools they find most important for this type of research. Comparing and contrasting the tools the authors mentioned with the tools provided by GRNsight could validate some of our design decisions while also providing some potential next-steps for the project.

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

This paper summarizes the pros and cons of several different network visualization tools for biological applications. While some of the software mentioned in this survey is redundant to other papers I have cited, this paper provides a good overview of all of the characteristics that make up a good visualization system. Comparing the strengths of the software assessed with the features on GRNsight could provide additional insight regarding features and functionality that GRNsight should consider in the future.

Rafique, Y., & Mišić, V. B. (2013). The effects of test-driven development on external quality and productivity: A meta-analysis. *IEEE Transactions on Software Engineering*, *39*(6), 835-856. doi: 10.1109/TSE.2012.28

Test-driven development (TDD) was one of my biggest contributions to the GRNsight project, so explaining the importance of TDD and what it entails is essential. This journal article is a study of the impacts of TDD on academic projects and projects in industry. The study found that there was substantial improvement in quality among academic research teams that utilized TDD. While academic groups were less likely to commit to TDD and lacked the efficiency that the industry groups had, the positive impacts of applying TDD were noticeable. This idea, as well as several others mentioned in this paper, is important to address in my paper to bring importance to the work that I did.

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

Cytoscape is one of the most popular software for visualizing large-scale biomolecular data, so it is essential to study and analyze this software in comparison to GRNsight. While Cytoscape has had several papers published since the start of the project, this paper covers more detail regarding the purpose and applications of the project, as it is one of the first papers published by the group.