Network Visualization

Sarah Weintraub, Aidan Pecorale

1 Introduction

Network visualizations are notoriously difficult to parse and extract relevant biological information from multi-omics datasets. By creating a multilayer network will allow researchers to sort not only by annotation, but by clutered trends as well. Currently, this type of search is done with two excel sheets and manually search through clusters of interest and then switching sheets to that cluster and filtering through by annotation of interest. This visualization will provide ease of access to search the entire trasncriptome for annotation as well as creating easier access per cluster.

2 One-sentence description

We aim to build a tool for multilayer network visualization for RNA-Seq transcriptomic data.

3 Project Type

Visualization / Application

4 Audience

Who is the audience for this project? How does it meet their needs? What happens if their needs remain unmet?

Scientists looking at multi-omic datasets is the audience - specifically, this tool can be implemented for anyone looking at transcriptomic data. It can be used to elucidate meaning from individual RNA-seq datasets as well as for comparison across organisms for comparative genomics.

5 Approach

5.1 Details

What is your approach?

We plan to use d3 to implement the multilayer network. The first layer will be the trends for each node/cluster with the connection strength, and the second layer to be the count, FC, and annotation information for each gene in the cluster.

5.2 Evidence for Success

Why do you think it will work?

We have all the data for this to work, as long as we can implement our ideas into d3, it should work. The papers below show similar visualization with less interactivity, and with our technical experience in making interactive charts, we are confident we will be able to accomplish this task.

6 Best-case Impact Statement

In the best-case scenario, what would be the impact statement (conclusion statement) for this project?

We set out to build an interactive multilater network tool for multi-omics analysis.

7 Major Milestones

1. Format data

2. Build skeleton multilayer network

3. Add interactivity

4. Add search tools

5. Add in big datasets

6. Make tool publicly available

8 Obstacles

One obstacle will be the large datasets, but if this is unavailable, we plan to parse down our data to make analysis easier.

Search individual nodes versus the entire network could also prove difficult as we have not done this before.

8.1 Major obstacles

Building a multilayer network.

8.2 Minor obstacles

- Organizing and filtering networks.

- Implementing whole dataset.

- Adding filtering options and ease of access specifications

9 Resources Needed

What additional resources do you need to complete this project?

1. Sketched images to base network layers off of.

2. Correlation matrix of data to build network.

3. Network trend images to add to top layer nodes.

4. Code to generate top network layer.

5. Code to generate bottom network layer.

6. Code to connect top and bottom network layers.

7. Code to build search customization in layers and between.

8. Maybe use Spark or Shiny R app to run the visualizations.

10 5 Related Publications

List 5 major publications that are most relevant to this project, and how they are related.

- Analysis of Network Clustering Algorithms and Cluster Quality Metrics at Scale [1].

- GraphOmics: an interactive platform to explore and integrate multi-omics data [5].

- The State of the Art in Visualizing Multivariate Networks [3].

- The State of the Art in Multilayer Network Visualization [2].

- Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks [4]

11 Define Success

What is the minimum amount of work necessary for this work be publishable?

Building an interactive multilayer network showing trends of each cluster and then more detail when zoomed in is the minimum to be published and used regularly by labs. Even just an interactive network showing trends of clusters could be useful to labs when studying omics data.

References

[1] S. Emmons, S. Kobourov, M. Gallant, and K. Börner. Analysis of network clustering algorithms and cluster quality metrics at scale. *PloS one*, 11(7):e0159161, 2016.

[2] F. Mcgee, M. Ghoniem, G. Melançon, B. Otjacques, and B. Pinaud. The state of the art in multilayer network visualization. In *Computer Graphics Forum*, volume 38, pages 125–149. Wiley Online Library, 2019.

[3] C. Nobre, M. Meyer, M. Streit, and A. Lex. The state of the art in visualizing multivariate networks. In *Computer Graphics Forum*, volume 38, pages 807–832. Wiley Online Library, 2019.

[4] P. Shannon, A. Markiel, O. Ozier, N. S. Baliga, J. T. Wang, D. Ramage, N. Amin, B. Schwikowski, and T. Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*, 13(11):2498–2504, 2003.

[5] J. Wandy and R. Daly. Graphomics: an interactive platform to explore and integrate multi-omics data. *BMC bioinformatics*, 22(1):1–19, 2021.