3D Biological Network Visualization

Ke Deng, Kalina Borkiewicz, Charles Blatti, Donna Cox, Colleen Bushell
Department of Computer Science, College of Engineering, University of Illinois at Urbana-Champaign

INTRODUCTION

Main Purpose
The goal of this project is to visualize complex biological network data in an immersive 3D environment. This tool can rank and provide information about genes that are highly correlated to some input phenotype. We aim to present knowledge-guided gene prioritization and gene set characterization information in a clear way.

Target Users
This tool is designed mainly for biological researchers who study some phenotype and related genes, so we researched on the best way to present the information that users are most interested in.

Background Information
I. Knowledge-Guided Gene Prioritization
We create a rank list that ranks the genes by their correlation to some phenotype of interest. Since we are in knowledge-guided, the ranking is decided by two scores:
- Baseline score: correlation between a gene and the user specified phenotype
- Network score: correlation between the network of a gene (a group of genes) and the user specified phenotype

Input: Gene by Samples Expression Matrix, Sample Phenotype Labels, Gene Knowledge Network

Output: Gene Rankings
Gene information to present:
- Node labels
- Network score
- Baseline score
- Score difference
- Node degree
- Node neighbors

II. Gene Set Characterization
This is a test of ranking biological pathways based on their overlap with the top 200 genes.

Input: Pathway Gene Sets, Top200 Gene List

Output: Pathways Rankings
Pathway information to present:
- Pathway name
- Pathway score / Intersection
- Top genes in pathway
- Similarity between pathways
- Network edges between genes

RESULTS

Overview
After inputting files, this is the home page, an overview of the ranking of top 200 significant genes.
- Nodes in the circle are sorted by visualization score.
- Baseline score is represented by gene color, which can be interpreted by the dynamic legend on top right.
- Node degree is indicated by node size - larger node means bigger degree (more neighbors).

Features
Buttons and toggles on top left provide more features.
- Rectangle Zoom In: Draw a rectangle and jump (zoom in) to the rectangle covered field.
- Node Label: Hover over a node shows its label.
- NetworkOnlyTop200: The top 200 genes are splitted into two rings.
  - Inner loop: genes with high absolute baseline score. They are significant since themselves are highly correlated with the specified phenotype.
  - Outer loop: genes with high network score. They are significant because their network (their neighbors) are highly correlated with the specified phenotype.
- Show Neighbors:
  - Use the toggle: Neighbors of top200 genes are shown on the stack.
  - Clicking on one node can also highlight its neighbors in yellow.
- Slider: Show other amounts of nodes in the circle, with a maximum of the number of genes loaded.
- Show Pathway:
  - Ten rings are shown.
    - The top 9 rings are the top 9 pathways, ranked by their overlap with the top 200 genes.
    - Last ring contains genes that are in top 200 genes but not belong to any top 9 pathways.
    - Genes in every ring stay in their original position - the same position as in the main circle sorted by the visualization score.
    - A gene may appear in more than 1 pathway.
    - The label of the current pathway, which is the nearest ring to the camera, is shown on the right side.

DISCUSSION

- The main problems in 2D visualization are that it usually requires multiple manual steps to select a node’s neighbors and the nodes are easily overlapped with each other. These can be easily solved by 3D visualization. The third dimension gives more space to expand the visualization, so it can show immediate neighbors and resolve some overlapping issues. In addition, it can add more user-friendly features, such as changing viewpoints, layout and adding the depth perception.
- For showing pathway, traditional 2D network usually have many edges overlapped and can only convey information in limited dimension. Our 3D pathway use the depth to show all top pathways, and we plan to convey multidimensional information at once.
- We also plan to add more motion to genes to indicate useful information. For example, in pathway mode, we use “wiggling” to show the similarities between pathways (a gene wiggle more if it appears less in the top pathways).

CONCLUSIONS

Since this app is a visualization tool of biological network information, we still need to do more research, try different approaches and get feedback to find the best way to present the information.

While we are making efforts in adding more features, we are still working on optimization the running time to improve our 3D model. We hope to discover more potential of 3D visualization in the future.

To make a suggestion, feel free to contact: kedeng3@illinois.edu

ACKNOWLEDGEMENTS

NCSA | National Center for Supercomputing Applications