

**Team Members**

Jerome Freudenberg

Jake Goldman

**Period**

6

**Team Name****Project Title**

Interactive Protein Network Visualization

**I. Project Idea**

The journal Nature published an article on May 17 titled *Architecture of the human interactome defines protein communities and disease networks*.

(<http://www.nature.com/nature/journal/v545/n7655/full/nature22366.html>) The paper examines protein interaction and includes a dataset of around 56000 proteins interactions with each other. Protein interactions involve proteins binding together and activating new functions such as gene regulation or enzymatic reactions. One example of a protein interaction would be hemoglobin A and hemoglobin B. The two

proteins are useless without each other, but together transport oxygen through our blood.

We will use the dataset from the article to create a visualization of the network of these interactions. To accomplish this we will look at network algorithms and attempt to implement one ourselves. This may involve the use of a library or API if writing the algorithm from scratch proves too difficult. Our overall goal is to make the visualization interactive, allowing users to highlight proteins, zoom in, change layout algorithms, etc with the aid of widgets. Ultimately, we will create a unique, visual interpretation of the data that makes finding connections between proteins easy.

## **II. Critical Features**

- 1) Working visualization for at least some subset of the data
  - We will attempt to write this from scratch, but we will use a library/API if necessary
- 2) Feature to highlight gene/protein name by click
- 3) Feature allowing zoom in
- 4) Search option based on protein/gene name
  - Search based from file input: the user can create/input a file with several protein names listed, and the program will highlight them
  - Search based from widget: the user can input a protein name in a text box and it will be highlighted/zoomed in on

### III. To Be Added Later

- 5) Allow users to try different network visualization algorithms
- 6) Show important nodes (e.g. most connected/longest path)

### IV. Developmental Stages

*Both/Peer Programing*

**Jerome**

**Jake**

- 1) 6/2/17: *Research and decide on a network visualization algorithm*
  - a) *Determine whether writing our own algorithm from scratch will be feasible (especially given the need for fast computation due to the large dataset)*
  - b) <http://profs.etsmtl.ca/mmcguffin/research/2012-mcguffin-simpleNetVis/mcguffin-2012-simpleNetVis.pdf>
- 2) 6/3/17: *Implement basic class skeleton*
- 3) 6/(4-5)/17: *Implement algorithm for layout of visualization*
  - a) *Possibly create a subset of the data if performance becomes an issue*
- 4) 6/8/17 (Demo Date): *Features 2 to 4*
  - a) **Highlight gene/protein by click**
  - b) **Zoom in**
  - c) **Search option**
- 5) 6/12/17: *Bug fixes*

If running ahead of schedule: work on different visualization algorithms

## V. UML Diagram

In our current stage, we are unsure of specific methods necessary for the project to work, as they depend on the type of algorithm we settle on. However, we have constructed a basic UML diagram, to be updated over the course of our project.

