Kevin Zhou

15-112 Fundamentals of Programming and Computer Science

Section S

4/15/15

Term Project Proposal

For my 112 Term Project, I want to create a data visualization program that handles data that consists of multiple individual data points where each data point has a “weight” value associated with it, and some of the data points are connected to each other for whatever reason is specified in the data set (there can potentially be multiple connections between any two given points). The program will display each data point as a vertex/node on a graph, and each connection between two vertices as an edge between them, and will use a physics engine which simulates vertices pushing and pulling on each other (depending on their weight and number of connections) in order to create a dynamic graph that allows the user to see which nodes are closely related to each other and which one are farther away. The user can also interact with the graph by clicking and dragging nodes and allowing the physics engine to run in order to re-shape the graph and look for more patterns in the arrangement of the data. The program will then be able to handle arbitrary data sets, as long as each data set contains a weight value for each node and the number of connections between any two data points. The goal is to create a visualization that allows the user to see the relationships between and potential groupings of the data points.

One example of data that would be applicable would be Twitter hashtags. One way of representing data about a given set of hashtags would be to look at the trends in posts over a given time period. Each hashtag would be one point in our data set, and every time a Tweet contains that hashtag, we would add one to the weight of that hashtag. We would then connect two hashtags every time someone makes a Tweet containing both hashtags. Once we run the simulation on this dataset, we would be able to see different groupings of the hashtags and learn more about how certain topics are related (or unrelated) to each other.

Relating to creating a visualization from Twitter hashtags, I would like to implement a web scraping component that allows users to easily take data from the Internet. They could input a set a hashtags that they want to analyze, and the program would pull all the information necessary from Twitter so that the user can easily generate the graph visualization. Another possibility could be to look at news article topics from a news website (e.g. NYTimes). If the user wanted to manually input data instead of pulling it from the web, I could also provide a way for them to do so.

Another potential feature I could add would be the ability to take data from a given time period and see how it changes and evolves. The program could create an animation that shows how the weights and connections change over time, and how that causes the produced graph to change over time.

In this project, I plan on using TKinter for graphics, and otherwise I do not anticipate needing any other modules or technologies.