Jessica Lee, jal2262

QMSS G4063 Data Visualization

Assignment 4

14 April 2014

**Network Visualization**

In exploring the dataset, it took me quite a while to figure out how best to work with the formatting so that I could convert it in a meaningful way. Even though I think Gephi offers more attractive visualizations that are more easily customizable in terms of the endproduct of what I wanted the visualization to look like, I don’t find Gephi to be that intuitive. In the end, I used R Studio to construct an appropriate .gefx file.

And, to be totally honest, I consciously partitioned the data so that I could have an easier time managing it. For instance, I had initially wanted to build a visualization with innate communities by tournament year, but what kept happening was that the year itself would become this massive node, with all the other players clinging to it. I don’t find such visualizations helpful or informative. So, I stuck with the 2012 tournament.

In the network structure, players were the nodes and edges were the relationships between them (i.e. when they played one another). Here, the degree centrality indicates how far the tennis player came along in the US Open 2012 tournament (i.e. Murray and Djokovic both have the highest degree of 7, since they are the final players in the tournament). The algorithm calculated node/edge attributes based on node properties and similarities, so that the more a node or player appeared, the more his node is weighted, both in size and in color.

Once the .gefx file was generated via R Studio, I used the Fan layout—I think it offered the best look at the actual matching of the tournament. The interactive website (very convenient for Gephi users!) offers an even better look at the tournament line-up, and with fun mouseover tricks.