Results

As an overview these are the goals we selected for this project. The data set we decided to use for our project was Amazon product co-purchasing network which was from a study done in Carnegie Mellon called the Dynamics of Viral Marketing. For our traversal, we used DFS which is an algorithm that parses through our data. Our algorithm traverses to its adjacent vertex (say x) that has not been visited before and mark as "visited" (true) and goes on with the adjacent vertex. We used DFS for generating our Strongly Connected Components Graph, which accordingly uses an Adjacency List. The complex/ uncovered algorithm we decided to use was a strongly connected graph and graph coloring. We used the strongly connected graph to shows items or nodes that are most commonly bought together. The results of our strongly connected components are displayed in a CSV file called stronglyconnected.csv. We ensure the accuracy of the algorithm through starting with a smaller data set and debugging until it could handel the entire dataset. We used graph coloring and its unique color rules and algorithm to visually demonstrate the different clusters of items. We also organized the results of the data for the graph coloring into a CSV file called graphcolor.csv. The results show the frequency of colors which helps us determine how many clusters or nodes are related to each other. We discovered that the a lot of the clusters are not related to each other as they have the same color. Since nodes or clusters next to each other cant have the same color, we came to the conclusion that those clusters are not related to each other. The result of our graph coloring was 6 max colors for 24766 nodes. We expected the Graph Coloring and Strongly Connected Components to produce more clusters, since we expected that there would be more purchased products that were bought together. However, our output seemed to contradict this since we got less clusters, and fewer connected nodes in our Graph Coloring output. Overall, this project was a great ending to the

semester. It was both a challenge and a reinforcement to what we have learned. We utilized past concepts and projects to help set our foundation, which helped us smoothly incorporate new ideas, such as strongly connected component graphs and algorithms, such as strongly connected components and DFS.

Visual Results

Strongly Connected Components: screenshot of stronglyconnected.csv

Sample shows a couple strongly connected components

```
22696   3087,
22697   3086,
22698   2573,
22699   2489,
22700   2488,
22701   5330,
22702   5974,
22703   5973,
22704   5972,
22705   7553,
22706   10754,
22707   2619,
22708   2618,
22709   2572,
22710   2571,
22711   2570,
22712   3102,
22713   3101,
22714   7568,
22715   5981,
22716   5980,
22717   5979,
22718   3146,
```

Graph Coloring: screenshot of graphcolor.csv

This cluster of nodes are connected to each other since they go up to the max color code.

```
Vertex 713 ---> Color 2
Vertex 714 ---> Color 3
Vertex 715 ---> Color 1
Vertex 716 ---> Color 2
Vertex 717 ---> Color 2
Vertex 718 ---> Color 2
Vertex 719 ---> Color 0
Vertex 720 ---> Color 3
Vertex 721 ---> Color 4
Vertex 722 ---> Color 5
Vertex 723 ---> Color 6
Vertex 724 ---> Color 3
Vertex 725 ---> Color 1
Vertex 726 ---> Color 1
Vertex 727 ---> Color 2
Vertex 728 ---> Color 2
Vertex 729 ---> Color 3
Vertex 730 ---> Color 3
Vertex 731 ---> Color 2
Vertex 732 ---> Color 4
```