

Project 3 Final Submission

1. Explain all design decisions

- Optimizations
 - Although I didn't optimize my milestone too much I did make many needed changes for performance for the final submission.
 - I cut down the correlation coefficient calculations from 7 minutes on GCC to 3 minutes by removing an array dimension, cutting down on loop runs and adding undirected edges at the same time for both nodes
 - I cut down the Floyd-Warshall calculations from 4 minutes per graph to 2:45 minutes by adding in a check for infinite distances and reducing the overall graph size
- Graph Representation
 - No major changes from the milestone state
- Worst-Case Bounds
 - The time bounds have changed from $O(n^6)$ to $O(n^4)$, and the time per step has been significantly reduced by the optimizations from $O(n^4)$ to something more akin to $O(n^3)$ functionally.
 - The space bounds have increased as a result of including the Floyd-Warshall algorithm which requires creating a 3969 x 3969 graph to perform on top of the adjacency list. This brings up the total memory usage from 170mb to closer to 260mb.

2. Problems Encountered in Implementation

- The largest problems in this section of the project were checking to make sure the values I received were correct. Asking on Piazza helped a lot in determining this as well as consistently checking each function as I created it to check for robustness during changes and optimization.
- Floyd-Warshall and the clustering coefficient were very easy to implement overall.

3. Known Bugs or Incomplete Implementation

- At the present I'm not seeing any bugs in the program and any problems loading files are handled before operations occur.

4. Significant Interaction with Peers

- Tingyu Luo
 - We compared our results during most of the final submission implementation as she and I were working ahead consistently.
 - I helped her a bit with implementing the random graphs and Floyd-Warshall and she assisted me with ideas for creating the clustering coefficient
- Piazza was a major source of help for this section and I also attempted to share what I had gathered with the rest of the classes

5. Citations to External Books or References

- Introduction to Algorithms textbook – Used it to assist with implementation of Floyd-Warshall and creating the adjacency matrices
- Cplusplus.com – Used primarily for specific cases where I got stuck on implementation and needed exact functional definitions

Project 3 Milestone

6. Explain all design decisions

- Graph Representation
 - I went for an adjacency list similar in style to the project 2 Storm structs. I created a single array of vertices 3969 in size and then attached a linked list of connected edges to each of them in order to simulate the graph.
- Optimizations
 - I didn't make too many algorithmic optimizations since doing so would require even more complicated optimizations in later coding. My primary goal was to make the code work first.
 - I did optimize for data gathering though. For each layer of calculation, I did store the necessary fields, namely in the correlation coefficient I stored the mean, S_{xx}/S_{yy} , S_{xy} and R values after calculating them all. This saved a ton of time and simplified calculating the thresholds later.
- Worst-Case Bound
 - I believe the worst case bound of my program is in calculating the S_{xy} value as part of the correlation coefficient. It's a $O(n^6)$ function, although functionally for our data a $O(n^4)$ which requires iterating through every potential piece of weekly data to calculate the final outcome of S_{xy} . This singular function dominates the entire run time of the program. The others would only take a few seconds otherwise.
 - Space-wise the program eats up about 170mb of memory while the program is running.

7. Problems Encountered in Implementation

- The largest problem I had with implementation was with the correlation coefficient. It took me a few days just to understand how to code it for all the necessary fields. In the end I also had to ask for some help in figuring out how to work it for the huge data set we were given.
- I also had a problem fully understanding what the project was asking us to do and show. The directions were a bit unclear about what we needed to present.

8. Known Bugs or Incomplete Implementation

- At the present there are no known bugs or incomplete sections in the project

9. Significant Interaction with Peers

- Tingyu Luo
 - Helped her with implementing the DFS since she was having segmentation fault issues
 - She helped me with part of the correlation coefficient algorithm since I was having problems conceptualizing parts of it
- Asked a couple of questions, shared my results and gave feedback on Piazza

10. Citations to External Books or References

- GeeksForGeeks.com – Used it to help build a basic component iterator and then re-tooled it to also count component sizes
 - Kishore, Brij Raj. "Connected Components in an undirected graph." *GeeksforGeeks*, 11 June 2017, <https://www.geeksforgeeks.org/connected-components-in-an-undirected-graph/>
- StackOverflow – Helped me with figuring out how to open binary files in a simple way
 - Retired Ninja. "reading floating numbers from bin file continuously and outputting in console window." *Stack Overflow* 27 October 2013 5:20, <https://stackoverflow.com/questions/19614581/reading-floating-numbers-from-bin-file-continuosly-and-outputting-in-console-win>
- Cplusplus.com – Used primarily for specific cases where I got stuck on implementation and needed exact functional definitions
- Introduction to Algorithms textbook – Used it to assist with implementation of the DFS and the adjacency list