 YRF 2023-24-Mid Term ReportFrame 24

| Name of the YRF Fellow | Shaun Mathew |
| --- | --- |
| Project Guide Name | Prof. M Hamsa Priya |
| Co - Guide Name | Aniket Malakar |
| Project Title | Exploring Protein Conformational Changes and Protein-Solvent Interactions through Graph Network Analysis |

1. Brief technical summary of the project progress and current status at the moment of the mid-term report. (up to 1000 words)

* Understanding of the current graph network creation and analysis methods applied on the biomolecular networks through reading multiple research papers.
  + Went through some codes on bio-molecular network analysis.
  + Explored methods like eigen-analysis, spectral-analysis, etc., done on the molecular networks and explored various testing methods to ensure relevant results matching real-life observations.
* Understanding the codebases for modules like GROMACS, NetworkX, MDSimulations, iGraphs, etc.
  + Went through the entire documentation and familiarised myself with the various features for network analysis in these modules.
  + Incorporated the relevant optimisation techniques into the code and documented the observations.
* Incorporating appropriate methods that help in faster analysis of the networks.
  + Unweighted analysis of Centrality and Shortest Path by using weight = None in the NetworkX functions.
  + Tested Multithreading for parallel analysis of the centrality functions.
  + Tested Multiprocessing for optimised analysis of the centrality functions.
  + Tested and used KDTree for optimised construction of the graph network from the position of the molecules.
* Using GROMACS features like frame-wise analysis using gmx commands to speed up the code and ensure optimal and low time complexity in graph network construction and analysis.
* Applying and coding additional methods of Graph Network Analysis like Community Analysis, Subgraph-Centrality Analysis, Eigen-Centrality Analysis and Hydrogen Bond Properties Analysis.
  + Community Analysis: Helps in understanding of the common patterns in the graph network.
  + Subgraph Centrality Analysis: Division of the graph network into disjoint components and parallel analysis of their centralities for faster code execution for big networks.
  + Eigen-Centrality Analysis: Measures the influence of a node in a network based on the concept that a node's importance is proportional to the importance of its neighbours.
  + Hydrogen-Bond Properties Analysis: Calculate the likelihood and strength of the bonds by analysing the frames in which the hydrogen bonds occur.
* Testing the code and its analysis parameters on water solutions.
* Adding flags to set the analysis code as per the analysis methods.
  + Weighted vs Unweighted Graph Analysis.
  + KD-Tree Loop Optimization vs Normal Brute Force Method for Graph Construction.
  + Disjoint Subgraph Centrality Parallelization.
  + Multithreading in Centrality Calculations.
  + Multiprocessing in Centrality Calculations.
  + Shortest Path Modes: Unweighted, All-Pair Dijkstras, Floyd Warshall, etc.
* Codes and the changes made in it, along with the observations: [[LINK](https://docs.google.com/document/d/1BFuI11TN2lKxndEhUN0SYFc5VViQIZ0IkVnzXjXuBds/edit?usp=sharing)]
* Github Coding Repository: [[LINK](https://github.com/SMath0510/YRF-Project/tree/master/Codes)]
* Summary of Research Papers Read and the Learning from them: [[LINK](https://github.com/SMath0510/YRF-Project/tree/master/Research%20Papers)]

2. Next step plans for remaining project work?

* Installation of g\_permute library and integration into the protein analysis methods.
* Extending the code analysis to binary solutions and then protein-osmolyte solutions.
* Reading additional papers and getting more useful and relevant graph analysis methods.
* Further, optimise the code blocks with the help of advanced processing methods and data structures.
* Simulate complex biomolecular mixtures and conduct network-based analysis on them.
* Explore libraries in C++ like Boost that can carry out similar tasks but much faster (due to the fast nature of languages like C++).

3. Expected project results

* A code for analysis of the protein osmolyte solution.
* Getting the expected results through analysis of any solution with the help of graph networks.
* Getting and reporting the results of analysis done on various protein conformation and osmolyte solutions.

4. Any foreseeable risks in the near future that may hinder achieving the final project objectives?

* Lack of codebases in the biomolecular network analysis field.
* Complexity in implementing the analysis algorithms from scratch in faster languages like C++.
* Lack of support for trajectory/simulation output analysis tools in C++.
* Version compatibility-related issues in installing g\_permute and its integration in the graph network analysis code.

5. Do you think the results from the YRF project will be ready by April 2023?

Yes NoRectangle: Rounded Corners 3Rectangle: Rounded Corners 6

6. If No, are you ready to contribute during summer for YRF Project?

Part time Full timeRectangle: Rounded Corners 3Rectangle: Rounded Corners 9

I expect the project to be over by April 2023, but in case of some unexpected hurdles, if it doesn't end by that time, I am willing to work on the project in the upcoming summer in part-time mode.

| Signature of Guide |  |
| --- | --- |
| Signature of Co-guide |  |

Flowchart: Manual Input 25