YRF Work Report

- Shaun Mathew

I. Reading Research Papers

1. Unravelling the Influence of Osmolytes on Water Hydrogen-Bond Network: From Local Structure to Graph Theory Analysis

- DETAILED SUMMARY: [LINK]
- Introduction to the basic methods of graph analysis and formation of graphs in computational biology.
- Learnings:
    - How Graphs are Created?
    - What interactions are mapped?
    - Which measures and properties of Graphs are used in the analysis of the molecular network?
    - How are simulations carried out?
    - What are the different settings that can be used for simulating?

2. Dynamics of Hydrogen Bonds between Water and Intrinsically Disordered and Structured Regions of Proteins

- DETAILED SUMMARY: [LINK]
- Introduces the concept of Intrinsically Disordered Proteins and methods of analyzing them using graphs and networks.
- Learnings:
    - How are ionic interactions taken into consideration?
    - What is the behaviour of the water molecules around the IDR and IDPs?
    - Another setting used for biological simulation.
    - Introduced the concept of Relative Surface Area and its impact on interactions.

3. Spectral Graph Analyses of Water Hydrogen-Bonding Network and Osmolyte Aggregate Structures in Osmolyte–Water Solutions

- DETAILED SUMMARY: [LINK]
- Introduces the eigenanalysis of graph networks in the context of Biological Networks.
- Learnings:
    - What are the methods used for eigenanalysis of a Graph network?
    - What is the relevance of these methods in the context of Biological Networks?
    - How do we interpret the derived results from the analysis?
    - Hyperparameters (cutoff distance, bonding angle, etc.) determination.
    - Calculating the radial distribution functions for different sets of molecular interactions.

4. Revised Centrality Measures Tell a Robust Story of Ion Conduction in Solids

- DETAILED SUMMARY: [LINK]
- Introduces about the conduction in ionic solids and the role of the movement of the ions in the conduction
- Learning:
    - How do we model the ionic movements and calculate the conduction through graph analysis?
    - Earlier Simulation Methods in the domain: KMC (Kinetic Monte Carlo) trajectories. More about this: [LINK].
    - Use of centrality measures to analyse and conclude results.

5. Per|Mut: Spatially Resolved Hydration Entropies from Atomistic Simulations

- DETAILED SUMMARY: [LINK]
- Introduces a new method of calculating spatially resolved hydration entropies with the help of techniques like Permutation Reduction and Mutual Information Expansion. (more about the methods : [LINK])
- Learning:
    - Graph Theoretical, Spatial Distribution Movement and MD Simulation methods of mixture analysis
    - Nearest Neighbour Search using the KNN estimator and the KD Tree method.
    - The essence and importance of entropy in biophysical systems.
    - Estimation methods for Entropy Analysis and Calculations in biophysical systems.

II. Understanding Coding Approaches

- Basic Starter Code provided to carry out further optimizations: [LINK]

Learnings:

- Libraries that are commonly used in graph network formation and analysis.
- Explored various functionalities of NetworkX, MDTraj, and other libraries.
- Understanding the application of Graph analysis methods in the given context (Shortest Paths, Centrality Measurements, Clustering Analysis, etc.)
- Understanding the input and output format obtained from the .xtc and .gro files (obtained from carrying out the simulation of the given mixture under appropriate settings with the help of GROMACS).
- Reading and feeding the input file into the specified library functions.

Improvements:

- Further analysis techniques.
- Optimizations for making the code run faster.

III. Learning Improvement Techniques

Improvement in the speed of the code

- Multi-Threading: Concurrent execution of multiple threads within a process, enabling better parallelization and responsiveness.
- Multi-Processing: Multiprocessing involves the simultaneous execution of multiple processes, each with its own memory space, allowing for parallel computation.
- DataStructures for better Clustering Analysis (KDTree): [LINK]
- Faster Shortest Path Calculation Techniques: Explored various techniques like Multiple_BFS (Unweighted graphs), Floyd Warshall method, Dijkstra's, etc. Checked the best and the fastest method and reported the results.
- Approximated Centrality Calculations: Heuristic measures to calculate approximated centrality values in a much quicker manner.

Learning Further Analysis Techniques

- Learned about eigen and spectral analysis techniques of graph networks and their application in the domain of computational biology.
- Learnt about community and motif analysis of graphs that helps in finding clusters and common patterns in the graphs. More about this: [LINK]

IV. Implementing the Ideas

```
import concurrent.futures  # For implementing threading and multiprocessing
from scipy.spatial import cKDTree  # Using kdtree data structure for
optimization
import networkx as nx  # Using the networkx library for graph-based
calculations
```

- Used these libraries to implement the ideas, and the final working code can be found here: [LINK]
- GitHub Repository with all the codes, results and extra details: [LINK]

--- Thank You ---