FAST National University of

Computer and Emerging Sciences

Parallel Distributed Computing:

Project Report

Group Members:

Muneel Haider	21i0640
Abdullah Zahoor	21i2481
Muhammad Abdullah	21i0643

Challenges:

We faced a lot of problems in the pre-processing phase, which included understanding the given datasets and how to process them into graphs.

Pre-processing:

- **Data Formatting:**
  - Converting the provided datasets into a graph the program understands.
  - A lot of our problems were simplified, thanks to the geeksforgeeks provided code.
  - Transformation of edge lists into adjacency lists was also a great challenge.
- **Data Distribution:**
  - Determining the best way to partition the graph for distribution across MPI processes was a complex task.

Implementation:

- **Memory Management:**
  - Efficiently managing memory in a parallel environment, ensuring that each MPI process has access to data without redundancy.
- **Load balancing:**
  - Ensuring that all processes and threads have an equal amount of work to prevent race conditions or overloading of work to some processes whereas some sit idle.

Testing:

- **Scalability Issues:**
  - Due to limited computing power, we are not able to process graphs with more than 2000 nodes.
  - The program may not run as expected when increasing the number of nodes and edges, which will result in insufficient memory errors.

Optimizations:

- **Parallel Design:**
  - Ensuring that the algorithm minimizes inter-process communication;
- **Communication Balancing:**
  - Minimizing the data transfer between processes.

Results:

1. Email-Enroll.txt:

- Sequential Performance:
  - With printing all read edges and their weights:
    - More than 60 seconds.
  - Without printing all read edges and their weights:
    - 50-60 seconds.
- Parallel Performance:
  - With printing all read edges and their weights:
    - 50-60 seconds.
  - Without printing all read edges and their weights:
    - 30-40 seconds.
- Speed-Up:
  - 20 seconds.

2. Email-EuAll.txt:

- Sequential Performance:
  - With printing all read edges and their weights:
    - More than 1 minute 20 seconds.
  - Without printing all read edges and their weights:
    - More than 60 seconds.
- Parallel Performance:
  - With printing all read edges and their weights:
    - More than 60 seconds.

    - Without printing all read edges and their weights:
        - 50-60 seconds.
- Speed-Up:
    - 15 seconds.

DoctorWho:

- Sequential Performance:
    - With printing all read edges and their weights:
        - More than 1 minute 40 seconds.
    - Without printing all read edges and their weights:
        - More than 1 minute 10 seconds.
- Parallel Performance:
    - With printing all read edges and their weights:
        - More than 1 minute 10 seconds.
    - Without printing all read edges and their weights:
        - Around 60 seconds, varies each time.
- Speed-Up:
    - 20 seconds.