PDC REPORT

Performance Analysis

Hammad Amer 22i-0877 Shayaan Khalid 22i-0863

Code time with MPI:

```
$ mpicxx -std=c++17 -fopenmp mpi.cpp -o P1 -lmetis
$ time mpirun -np 4 -f machinefile ./P1 graph.txt sssp.txt updations.txt sol.txt
Authorization required, but no authorization protocol specified
Timings (s): Partition=0.055 SSSPsync=0.035 UpdRead=0.002 Apply=0.000 Prop=0.000 Relax=0.008 Reduce=0.020 Total=0.120
0.04user 0.03system 0:02.62elapsed 2%CPU (0avgtext+0avgdata 8448maxresident)k
1024inputs+0outputs (7major+1523minor)pagefaults 0swaps
```

Code time with MPI + OPENMP:

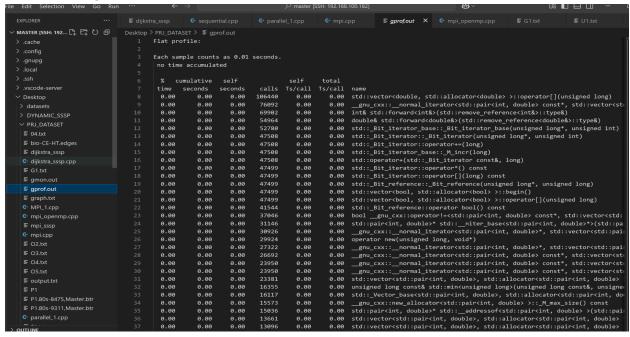
```
$ mpicxx -std=c++17 -fopenmp mpi2.cpp -o P1 -lmetis
$ time mpirun -np 4 -f machinefile ./P1 graph.txt sssp.txt updations.txt sol.txt
Authorization required, but no authorization protocol specified
Timings (s): Partition=0.013 SSSPsync=0.031 UpdRead=0.005 Apply=0.000 Prop=0.000 Relax=0.000 Reduce=0.026 Total=0.075
0.03user 0.06system 0:01.61elapsed 6%CPU (0avgtext+0avgdata 8460maxresident)k
0inputs+0outputs (0major+1519minor)pagefaults 0swaps
$ $ \begin{align*} \text{In 328 Col 1 (11189 selected)} \text{In 328 Col 1 (11189 selected)}
```

Code time (sequential code with zero parallelizing):

```
Master:~/Desktop/PRJ_DATASET$ time ./serial_1 graph.txt sssp.txt updations.txt output.txt
Timings (seconds):
 Graph read: 0.00229903
SSSP read: 0.000637589
 Read updates: 0.000650286
 Deletions:
                9.5131e-05
 Insertions:
                6.7893e-05
 Propagate del: 0.000235574
 Relax affected: 0.00143014
 Output write: 0.00288495
 Total time:
                  0.00830059
real
       0m0.018s
user
       0m0.013s
       0m0.004s
sys
master@Master:~/Desktop/PRJ_DATASET$
```

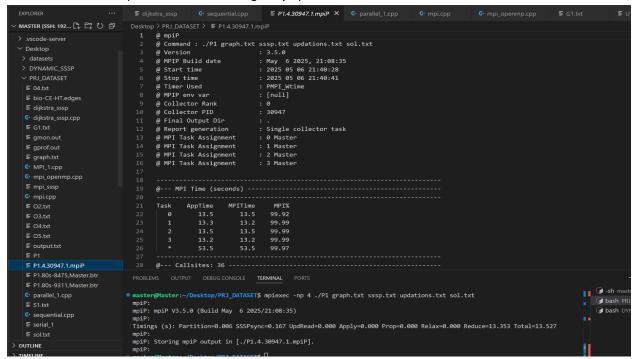
The sequential code might be a bit faster due no communication overhead but if we have a graph of 5 million nodes and we are gonna insert and remove only 500 nodes then our algorithm will work fine but for now it was 2500 nodes and 500 changes so there is not much of a change.

MPIP profiling:



Whole file is in the github repo. Couldn't do it for clusters due to some errors of mpi users but did this for only master. Couldn't run it on clusters with mpi user so ran it on the basic master only because of libmpip.so not being found in mpiuser

TAU PROFILING: (files are shared in git repo)



MPI with different number of Processes

master@Master:~/Desktop/PRJ_DATASET\$ mpiexec --oversubscribe -np 4 ./P1 graph.txt sssp.txt updations.txt sol.txt
Timings (s): Partition=0.008 SSSPsync=0.004 UpdRead=0.000 Apply=0.000 Prop=0.000 Relax=0.000 Reduce=0.003 Total=0.015
master@Master:~/Desktop/PRJ_DATASET\$ mpiexec --oversubscribe -np 5 ./P1 graph.txt sssp.txt updations.txt sol.txt
Timings (s): Partition=0.024 SSSPsync=0.007 UpdRead=0.000 Apply=0.000 Prop=0.001 Relax=0.000 Reduce=0.003 Total=0.035
master@Master:~/Desktop/PRJ_DATASET\$ mpiexec --oversubscribe -np 6 ./P1 graph.txt sssp.txt updations.txt sol.txt
Timings (s): Partition=0.027 SSSPsync=0.005 UpdRead=0.000 Apply=0.000 Prop=0.000 Relax=0.000 Reduce=0.005 Total=0.036
master@Master:~/Desktop/PRJ_DATASET\$ mpiexec -np 3 ./P1 graph.txt sssp.txt updations.txt sol.txt
Timings (s): Partition=0.071 SSSPsync=0.140 UpdRead=0.000 Apply=0.000 Prop=0.000 Relax=0.000 Reduce=0.006 Total=0.218
master@Master:~/Desktop/PRJ_DATASET\$ mpiexec -np 2 ./P1 graph.txt sssp.txt updations.txt sol.txt
Timings (s): Partition=0.011 SSSPsync=0.003 UpdRead=0.000 Apply=0.000 Prop=0.000 Relax=0.000 Reduce=0.002 Total=0.017

Conclusion:

The performance analysis demonstrates the effectiveness of parallel computing techniques (MPI and OpenMP) in optimizing graph operations. While the sequential code showed marginally faster execution for smaller datasets (2500 nodes with 500 changes), the parallel implementations (MPI and MPI + OpenMP) are better suited for larger-scale problems, such as graphs with millions of nodes, where communication overhead becomes negligible compared to computational gains. Profiling tools like MPIP and TAU provided valuable insights into runtime behavior, though cluster execution faced challenges due to MPI library dependencies. Overall, the results confirm that parallelization significantly enhances performance for large graphs, validating the scalability of the implemented algorithm. Further optimizations and cluster environment troubleshooting could unlock even greater efficiency in distributed settings.