Parallel PageRank Algorithm (Google Search Simulation)

Project Report

Name: Ayush Gangwar
Roll No.: 202211006
Environment: WSL Ubuntu, MPI + OpenMP

Codes:

1) Pagerank.cpp:

```
#include <mpi.h>
#include <omp.h>

#include <vector>
#include <algorithm>
#include <numeric>
#include <fstream>
#include <iostream>
#include <sstream>
#include <string>
#include <cmath>
#include <cstdlib> // getenv

using idx_t = int64_t;
using val_t = double;

struct CSR {
   std::vector<idx_t> row_ptr; // size n+1
   std::vector<idx_t> col;     // incoming neighbors
};

void build_csr_from_edgelist(idx_t n, const std::vector<std::pair<idx_t,idx_t>>& edges,
                             CSR &csr, std::vector<idx_t> &outdeg) {
   outdeg.assign(n, 0);
   for (auto &e : edges) {
      idx_t u = e.first;
      if (u>=0 && u<n) outdeg[u]++;
   }
   std::vector<idx_t> indeg(n,0);
```

```
    for (auto &e : edges) {
        idx_t v = e.second;
        if (v>=0 && v<n) indeg[v]++;
    }
    csr.row_ptr.resize(n+1);
    csr.row_ptr[0]=0;
    for (idx_t i=0;i<n;i++) csr.row_ptr[i+1]=csr.row_ptr[i]+indeg[i];
    csr.col.resize(csr.row_ptr[n]);
    std::vector<idx_t> cur = csr.row_ptr;
    for (auto &e : edges) {
        idx_t u = e.first;
        idx_t v = e.second;
        if (v<0 || v>=n || u<0 || u>=n) continue;
        csr.col[cur[v]++] = u;
    }
}

void read_edge_list(const std::string &fname, std::vector<std::pair<idx_t,idx_t>>& edges, idx_t
&maxnode) {
    std::ifstream in(fname);
    if (!in.is_open()) {
        std::cerr<<"Cannot open "<<fname<<"\n";
        exit(1);
    }
    std::string line;
    edges.clear();
    maxnode = -1;
    while (std::getline(in,line)) {
        if (line.size()==0) continue;
        if (line[0]=='#') continue;
        std::istringstream iss(line);
        idx_t u,v;
        if (!(iss>>u>>v)) continue;
        if (u<0 || v<0) continue;
        edges.emplace_back(u,v);
        if (u>maxnode) maxnode=u;
        if (v>maxnode) maxnode=v;
    }
    in.close();
}

int main(int argc, char** argv) {
    MPI_Init(&argc,&argv);
    int rank, nprocs;
    MPI_Comm_rank(MPI_COMM_WORLD,&rank);
    MPI_Comm_size(MPI_COMM_WORLD,&nprocs);
```

```
if (argc < 6) {
    if (rank==0) {
        std::cout<<"Usage: "<<argv[0]<<" <edge_list.txt> <N_nodes> <max_iter> <tol> <damping>\n";
        std::cout<<"If N_nodes==0, it reads max node index from file and uses max+1\n";
    }
    MPI_Finalize();
    return 1;
}

std::string edgefile = argv[1];
idx_t N = atoll(argv[2]);
int max_iter = atoi(argv[3]);
val_t tol = atof(argv[4]);
val_t damping = atof(argv[5]);

// Read edge list (each rank reads full file)
std::vector<std::pair<idx_t,idx_t>> edges;
idx_t maxnode;
read_edge_list(edgefile, edges, maxnode);
if (N==0) N = maxnode+1;
if (rank==0) {
    std::cout<<"Read "<<edges.size()<<" edges, N="<<N<<", procs="<<nprocs<<"\n";
}

CSR csr;
std::vector<idx_t> outdeg;
build_csr_from_edgelist(N, edges, csr, outdeg);

// Partition rows
idx_t rows_per = (N + nprocs - 1) / nprocs;
idx_t row_start = rank * rows_per;
idx_t row_end = std::min(N, row_start + rows_per);
idx_t local_n = std::max((idx_t)0, row_end - row_start);

// Prepare Allgatherv params
std::vector<int> counts(nprocs), displs(nprocs);
for (int p=0;p<nprocs;p++) {
    idx_t rs = p * rows_per;
    idx_t re = std::min(N, rs + rows_per);
    counts[p] = (int)(re - rs);
    displs[p] = (int)(rs);
}

std::vector<val_t> pr_local(local_n, 1.0 / (val_t)N);
std::vector<val_t> pr_all(N, 0.0);
MPI_Allgatherv(pr_local.data(), (int)local_n, MPI_DOUBLE,
               pr_all.data(), counts.data(), displs.data(), MPI_DOUBLE,
```

```
            MPI_COMM_WORLD);

val_t base = (1.0 - damping) / (val_t)N;
bool debug = (getenv("DEBUG_PAGERANK") != nullptr);

if (debug) {
    val_t s=0.0;
    for (idx_t i=0;i<N;i++) s += pr_all[i];
    std::cout<<"RANK "<<rank<<" initial pr_all sum="<<s<<" sample[0..4]=";
    for (int k=0;k<5 && k<N; ++k) std::cout<<pr_all[k]<<" ";
    std::cout<<"\n";
}

double total_start = MPI_Wtime();
double comp_time = 0.0, comm_time = 0.0;
val_t global_diff = 0.0;
int iter;

for (iter=0; iter<max_iter; ++iter) {
    double iter_start = MPI_Wtime();

    val_t local_dang = 0.0;
    for (idx_t i = 0; i < local_n; ++i) {
        idx_t node = row_start + i;
        if (node < N && outdeg[node] == 0) local_dang += pr_local[i];
    }
    val_t global_dang = 0.0;
    MPI_Allreduce(&local_dang, &global_dang, 1, MPI_DOUBLE, MPI_SUM, MPI_COMM_WORLD);

    std::vector<val_t> pr_new(local_n, 0.0);
    double comp_s = MPI_Wtime();
    #pragma omp parallel for schedule(dynamic,64)
    for (idx_t i = row_start; i < row_end; ++i) {
        val_t sum = 0.0;
        idx_t r0 = csr.row_ptr[i];
        idx_t r1 = csr.row_ptr[i+1];
        for (idx_t p = r0; p < r1; ++p) {
            idx_t src = csr.col[p];
            if (src >= 0 && src < N) {
                if (outdeg[src] > 0) sum += pr_all[src] / (val_t)outdeg[src];
            }
        }
        pr_new[i - row_start] = base + damping * (sum + global_dang / (val_t)N);
    }
    comp_time += MPI_Wtime() - comp_s;

    val_t local_diff = 0.0;
```

```
    for (idx_t i=0;i<local_n;i++) local_diff += std::abs(pr_new[i] - pr_local[i]);
    pr_local.swap(pr_new);

    double comm_s = MPI_Wtime();
    MPI_Allgatherv(pr_local.data(), (int)local_n, MPI_DOUBLE,
                   pr_all.data(), counts.data(), displs.data(), MPI_DOUBLE,
                   MPI_COMM_WORLD);
    comm_time += MPI_Wtime() - comm_s;

    val_t local_sum = 0.0;
    for (idx_t i = 0; i < (idx_t)local_n; ++i) local_sum += pr_local[i];
    val_t global_sum = 0.0;
    MPI_Allreduce(&local_sum, &global_sum, 1, MPI_DOUBLE, MPI_SUM, MPI_COMM_WORLD);
    if (global_sum > 0.0 && std::abs(global_sum - 1.0) > 1e-15) {
        for (idx_t i=0;i<N;i++) pr_all[i] /= global_sum;
        for (idx_t i = row_start; i < row_end; ++i)
            pr_local[i - row_start] = pr_all[i];
    }

    MPI_Allreduce(&local_diff, &global_diff, 1, MPI_DOUBLE, MPI_SUM, MPI_COMM_WORLD);

    if (debug && iter < 5) {
        val_t s = 0.0;
        for (idx_t i=0;i<N;i++) s += pr_all[i];
        std::cout<<"RANK "<<rank<<" iter="<<iter<<" pr_all sum="<<s<<" sample[0..4]=";
        for (int k=0;k<5 && k<N; ++k) std::cout<<pr_all[k]<<" ";
        std::cout<<"\n";
    }

    if (rank==0) {
        double iter_time = MPI_Wtime() - iter_start;
        std::cout<<"Iter "<<iter<<" diff="<<global_diff<<", time(s)="<<iter_time<<"\n";
    }
    if (global_diff < tol) break;
}

double total_end = MPI_Wtime();
double total_time = total_end - total_start;

if (rank==0) {
    std::vector<std::pair<val_t, idx_t>> pairs;
    pairs.reserve(N);
    for (idx_t i=0;i<N;i++) pairs.emplace_back(pr_all[i], i);
    std::sort(pairs.begin(), pairs.end(), std::greater<>());
    std::cout<<"Top 10 PageRank (val, node):\n";
    for (int k=0;k<10 && k<(int)pairs.size();k++) {
        std::cout<<k<<": "<<pairs[k].first<<" , "<<pairs[k].second<<"\n";
```

```
    }
  }

  double max_total_time, max_comp_time, max_comm_time;
  MPI_Reduce(&total_time, &max_total_time, 1, MPI_DOUBLE, MPI_MAX, 0, MPI_COMM_WORLD);
  MPI_Reduce(&comp_time, &max_comp_time, 1, MPI_DOUBLE, MPI_MAX, 0,
MPI_COMM_WORLD);
  MPI_Reduce(&comm_time, &max_comm_time, 1, MPI_DOUBLE, MPI_MAX, 0,
MPI_COMM_WORLD);

  if (rank==0) {
    std::cout<<"Total time (max over ranks) = "<<max_total_time<<" s\n";
    std::cout<<"Comp time (max) = "<<max_comp_time<<" s, Comm time (max) =
"<<max_comm_time<<" s\n";
    std::cout<<"Iterations = "<<(iter+1)<<"\n";
  }

  MPI_Finalize();
  return 0;
}
```

2) Makefile:

```
CXX=mpicxx
CXXFLAGS=-O3 -fopenmp -std=c++17
TARGET=pagerank

all: $(TARGET)

$(TARGET): pagerank.cpp
$(CXX) $(CXXFLAGS) pagerank.cpp -o $(TARGET)

clean:
rm -f $(TARGET)
```

3) Scaling csv File code:

```
#!/usr/bin/env bash
set -euo pipefail

mkdir -p data
OUTCSV=data/scaling_results.csv
echo "procs,omp_threads,total_time_s,iterations,exit_code,logfile" > "$OUTCSV"

PROCS_LIST="1 2 4"
OMP_T=2
```

```
EDGEFILE="data/web-Google-100k.txt"
BIN="./pagerank"

for p in $PROCS_LIST; do
  export OMP_NUM_THREADS=$OMP_T
  LOG="data/pagerank_p${p}.log"
  TIMEF="data/pagerank_time_p${p}.txt"
  echo "=== Running: mpirun -np $p (OMP=$OMP_NUM_THREADS) ==="
  timeout 600s /usr/bin/time -f "%e" -o "$TIMEF" mpirun --allow-run-as-root --oversubscribe --bind-to none
-np $p $BIN $EDGEFILE 0 500 1e-6 0.85 > "$LOG" 2>&1 || true
  RC=$?
  ELAPSED=$(cat "$TIMEF" 2>/dev/null || echo "")
  if [ -z "$ELAPSED" ]; then ELAPSED="error"; fi
  ITERS=$(grep -oE "Iterations[[:space:]]*[:=]?[[:space:]]*[0-9]+" "$LOG" | head -n1 | grep -oE "[0-9]+" ||
true)
  if [ -z "$ITERS" ]; then
    ITERS=$(grep -oE "Iter[[:space:]]*[0-9]+" "$LOG" | tail -n1 | grep -oE "[0-9]+" || true)
  fi
  if [ -z "$ITERS" ]; then ITERS="error"; fi

  echo "$p,$OMP_T,$ELAPSED,$ITERS,$RC,$LOG" >> "$OUTCSV"
  echo "Recorded -> p=$p time=$ELAPSED iters=$ITERS rc=$RC log=$LOG"
done

echo
echo "Wrote $OUTCSV"
cat "$OUTCSV"
echo
echo "Tails of logs:"
for f in data/pagerank_p*.log; do
  echo "---- $f ----"
  tail -n 50 "$f" || true
  echo
done
```

4) Plot_scaling.py:

```
#!/usr/bin/env python3
import pandas as pd
import matplotlib
matplotlib.use("Agg")
import matplotlib.pyplot as plt
import os

csv = "data/scaling_results.csv"
if not os.path.exists(csv):
    print("No CSV found:", csv)
```

```
    raise SystemExit(1)

df = pd.read_csv(csv)
df['total_time_s'] = pd.to_numeric(df['total_time_s'], errors='coerce')
df_sorted = df.sort_values('procs').dropna(subset=['total_time_s']).reset_index(drop=True)
if df_sorted.empty:
    print("No numeric timing rows found in", csv)
    print(df)
    raise SystemExit(1)

t1 = df_sorted.loc[0,'total_time_s']
df_sorted['speedup'] = t1 / df_sorted['total_time_s']
df_sorted['efficiency'] = df_sorted['speedup'] / df_sorted['procs']

print(df_sorted)

plt.figure(figsize=(6,4))
plt.plot(df_sorted['procs'], df_sorted['total_time_s'], marker='o')
plt.xlabel('MPI processes')
plt.ylabel('Total time (s)')
plt.title('Strong scaling: total time')
plt.grid(True)
plt.savefig('data/total_time_plot.png', dpi=200)
plt.close()

plt.figure(figsize=(6,4))
plt.plot(df_sorted['procs'], df_sorted['speedup'], marker='o', label='Measured')
plt.plot(df_sorted['procs'], df_sorted['procs'], '--', label='Ideal')
plt.xlabel('MPI processes')
plt.ylabel('Speedup')
plt.title('Strong scaling: speedup')
plt.legend()
plt.grid(True)
plt.savefig('data/speedup_plot.png', dpi=200)
plt.close()

plt.figure(figsize=(6,4))
plt.plot(df_sorted['procs'], df_sorted['efficiency']*100, marker='o')
plt.xlabel('MPI processes')
plt.ylabel('Efficiency (%)')
plt.title('Strong scaling: efficiency')
plt.grid(True)
plt.savefig('data/efficiency_plot.png', dpi=200)
plt.close()

print("Saved plots to data/*.png")
```

5) Building the project:

```
make clean
make -j
ls -l pagerank
```

6) Test Run code on small subset:

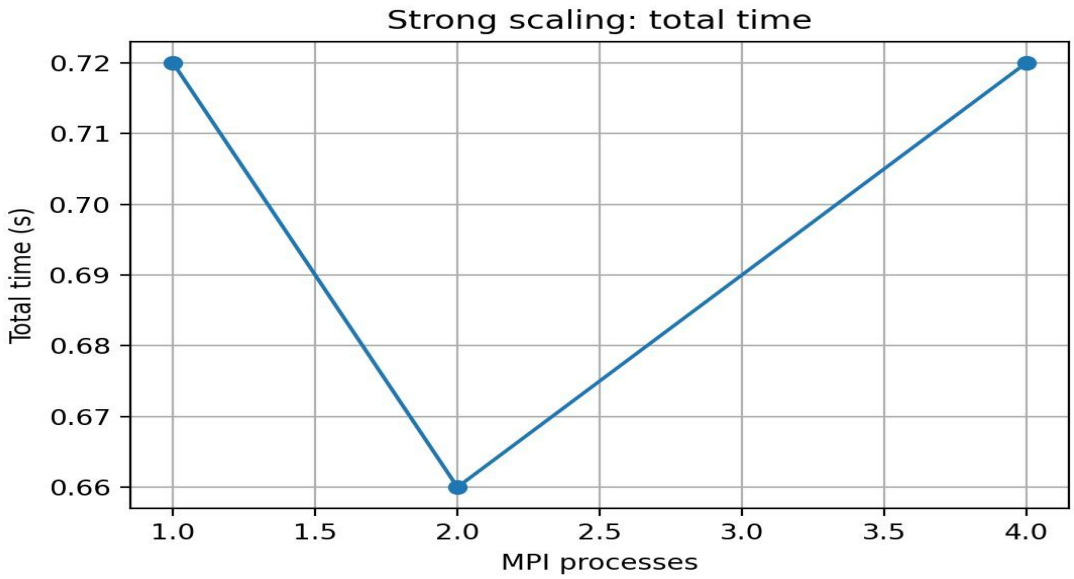
```
export OMP_NUM_THREADS=4
unset DEBUG_PAGERANK
mpirun --allow-run-as-root --oversubscribe -np 1 ./pagerank data/web-Google-100k.txt 0 100 1e-6 0.85 |
tee data/pagerank_run_p1.log
tail -n 40 data/pagerank_run_p1.log
```

7) Plot summary:

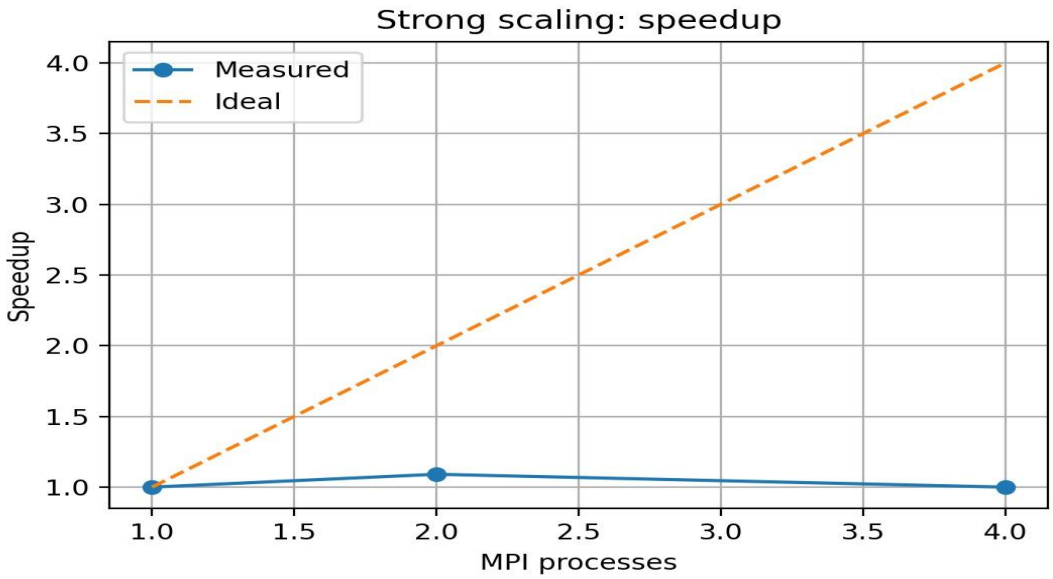
```
python3 plot_scaling.py
```

5. Result Output Images

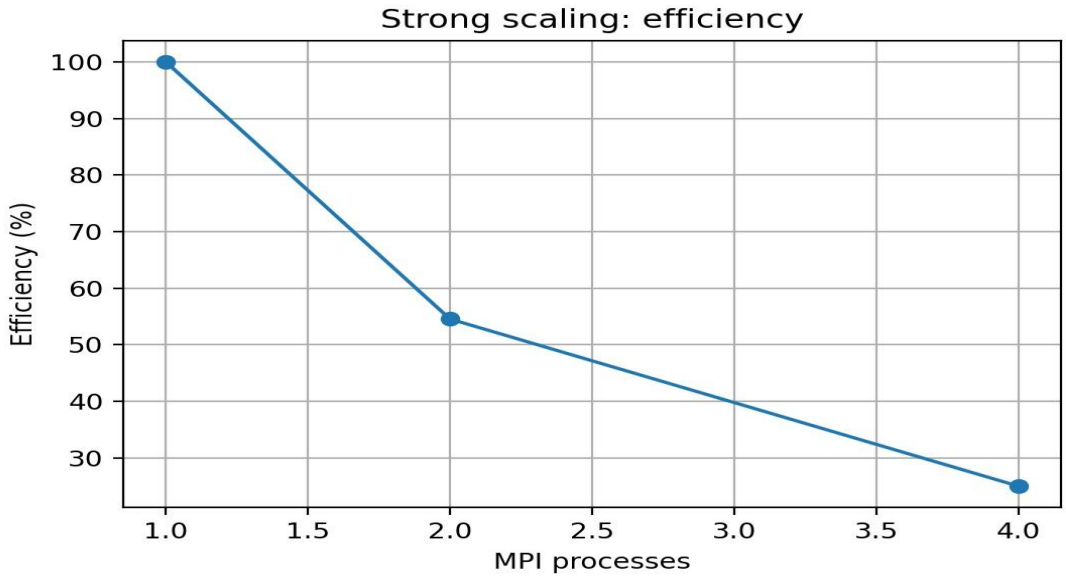
Total Time (strong scaling)



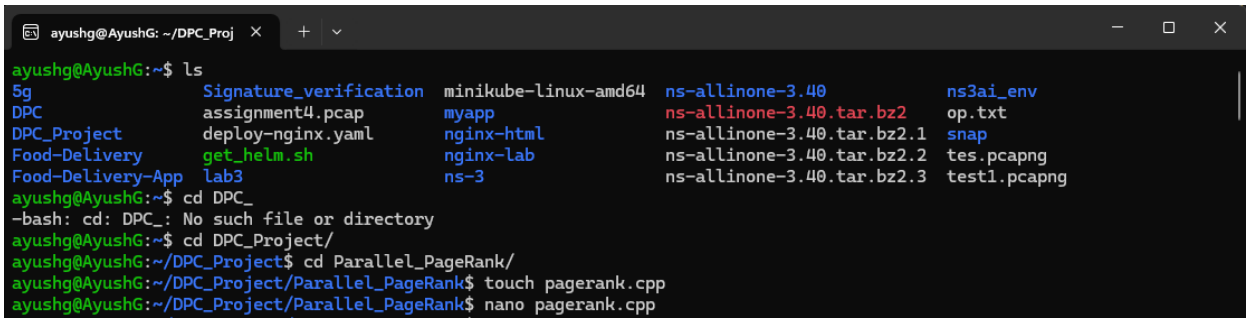
Speedup (measured vs ideal)

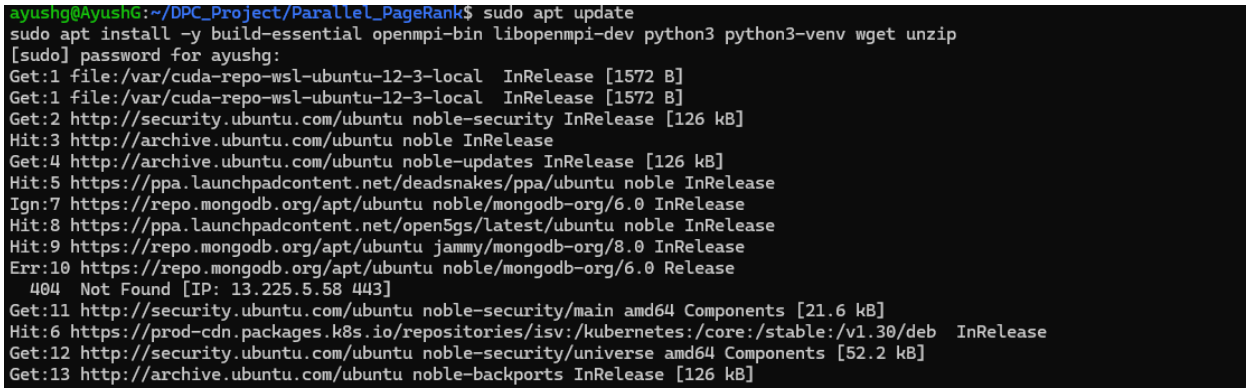


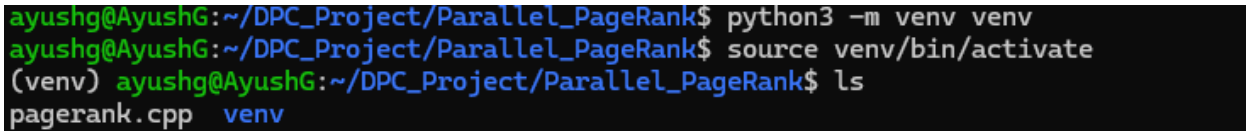
Efficiency (%)

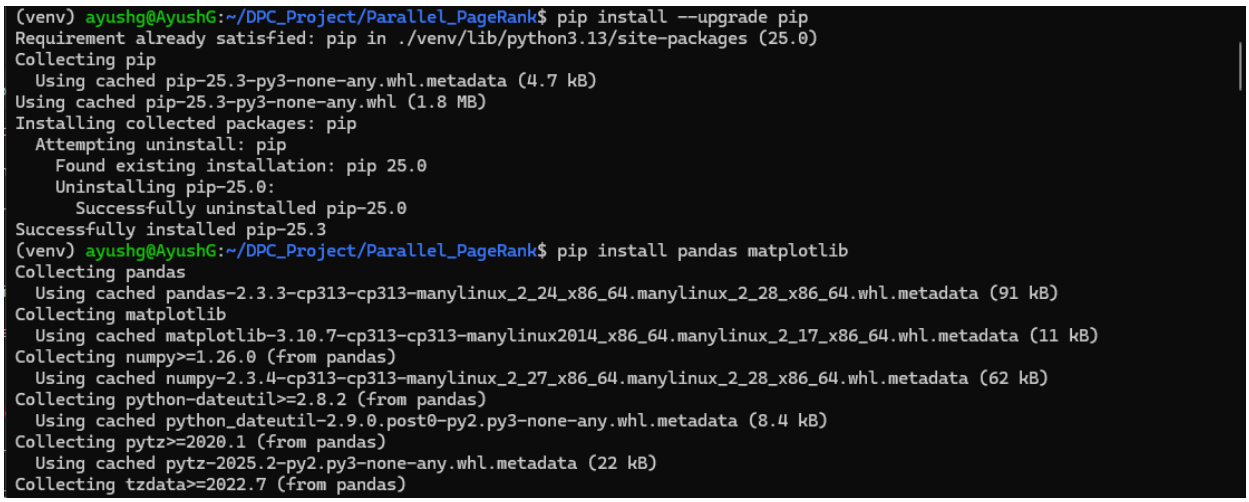


5. WSL Output Images:


ayushg@AyushG: ~/DPC_Proj
ayushg@AyushG:~$ ls
5g Signature_verification minikube-linux-amd64 ns-allinone-3.40 ns3ai_env
DPC assignment4.pcap myapp ns-allinone-3.40.tar.bz2 op.txt
DPC_Project deploy-nginx.yaml nginx-html ns-allinone-3.40.tar.bz2.1 snap
Food-Delivery get_helm.sh nginx-lab ns-allinone-3.40.tar.bz2.2 tes.pcapng
Food-Delivery-App lab3 ns-3 ns-allinone-3.40.tar.bz2.3 test1.pcapng
ayushg@AyushG:~$ cd DPC_
-bash: cd: DPC_: No such file or directory
ayushg@AyushG:~$ cd DPC_Project/
ayushg@AyushG:~/DPC_Project$ cd Parallel_PageRank/
ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ touch pagerank.cpp
ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ nano pagerank.cpp


ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ sudo apt update
sudo apt install -y build-essential openmpi-bin libopenmpi-dev python3 python3-venv wget unzip
[sudo] password for ayushg:
Get:1 file:/var/cuda-repo-wsl-ubuntu-12-3-local InRelease [1572 B]
Get:1 file:/var/cuda-repo-wsl-ubuntu-12-3-local InRelease [1572 B]
Get:2 http://security.ubuntu.com/ubuntu noble-security InRelease [126 kB]
Hit:3 http://archive.ubuntu.com/ubuntu noble InRelease
Get:4 http://archive.ubuntu.com/ubuntu noble-updates InRelease [126 kB]
Hit:5 https://ppa.launchpadcontent.net/deadsnakes/ppa/ubuntu noble InRelease
Ign:7 https://repo.mongodb.org/apt/ubuntu noble/mongodb-org/6.0 InRelease
Hit:8 https://ppa.launchpadcontent.net/open5gs/latest/ubuntu noble InRelease
Hit:9 https://repo.mongodb.org/apt/ubuntu jammy/mongodb-org/8.0 InRelease
Err:10 https://repo.mongodb.org/apt/ubuntu noble/mongodb-org/6.0 Release
404 Not Found [IP: 13.225.5.58 443]
Get:11 http://security.ubuntu.com/ubuntu noble-security/main amd64 Components [21.6 kB]
Hit:6 https://prod-cdn.packages.k8s.io/repositories/isv:/kubernetes:/core:/stable:/v1.30/deb InRelease
Get:12 http://security.ubuntu.com/ubuntu noble-security/universe amd64 Components [52.2 kB]
Get:13 http://archive.ubuntu.com/ubuntu noble-backports InRelease [126 kB]


ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ python3 -m venv venv
ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ source venv/bin/activate
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ ls
pagerank.cpp venv


(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ pip install --upgrade pip
Requirement already satisfied: pip in ./venv/lib/python3.13/site-packages (25.0)
Collecting pip
Using cached pip-25.3-py3-none-any.whl.metadata (4.7 kB)
Using cached pip-25.3-py3-none-any.whl (1.8 MB)
Installing collected packages: pip
Attempting uninstall: pip
Found existing installation: pip 25.0
Uninstalling pip-25.0:
Successfully uninstalled pip-25.0
Successfully installed pip-25.3
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ pip install pandas matplotlib
Collecting pandas
Using cached pandas-2.3.3-cp313-cp313-manylinux_2_24_x86_64.manylinux_2_28_x86_64.whl.metadata (91 kB)
Collecting matplotlib
Using cached matplotlib-3.10.7-cp313-cp313-manylinux2014_x86_64.manylinux_2_17_x86_64.whl.metadata (11 kB)
Collecting numpy>=1.26.0 (from pandas)
Using cached numpy-2.3.4-cp313-cp313-manylinux_2_27_x86_64.manylinux_2_28_x86_64.whl.metadata (62 kB)
Collecting python-dateutil>=2.8.2 (from pandas)
Using cached python_dateutil-2.9.0.post0-py2.py3-none-any.whl.metadata (8.4 kB)
Collecting pytz>=2020.1 (from pandas)
Using cached pytz-2025.2-py2.py3-none-any.whl.metadata (22 kB)
Collecting tzdata>=2022.7 (from pandas)

```
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ touch Makefile
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ nano Makefile
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ ls
Makefile  pagerank.cpp  venv
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ mkdir data
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ ls
Makefile  data  pagerank.cpp  venv
```

```
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ cd data/
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank/data$ wget https://snap.stanford.edu/data/web-Google.txt.gz
gunzip -f web-Google.txt.gz
--2025-11-11 08:28:57--  https://snap.stanford.edu/data/web-Google.txt.gz
Resolving snap.stanford.edu (snap.stanford.edu)... 171.64.75.80
Connecting to snap.stanford.edu (snap.stanford.edu)|171.64.75.80|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 21168784 (20M) [application/x-gzip]
Saving to: ‘web-Google.txt.gz’

web-Google.txt.gz             100%[================================================>]  20.19M  1.49MB/s    in 52s

2025-11-11 08:29:52 (399 KB/s) - ‘web-Google.txt.gz’ saved [21168784/21168784]

(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank/data$ head -n 100000 web-Google.txt > web-Google-100k.txt
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank/data$ cd ..
```

```
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ make -j
mpicxx -O3 -fopenmp -std=c++17 pagerank.cpp -o pagerank
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ ls
Makefile  data  pagerank  pagerank.cpp  venv
```

```
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ export OMP_NUM_THREADS=4
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ unset DEBUG_PAGERANK
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ mpirun --allow-run-as-root --oversubscribe -np 1 ./pagerank data/w
eb-Google-100k.txt 0 100 1e-6 0.85 | tee data/pagerank_run_p1.log
Read 99996 edges, N=916396, procs=1
Iter 0 diff=0.0158512, time(s)=0.00815441
Iter 1 diff=0.00872863, time(s)=0.00665003
Iter 2 diff=0.00454943, time(s)=0.00657968
Iter 3 diff=0.00269276, time(s)=0.00634367
Iter 4 diff=0.00172763, time(s)=0.00617059
Iter 5 diff=0.00116863, time(s)=0.00636433
Iter 6 diff=0.000823789, time(s)=0.00646665
Iter 7 diff=0.000598627, time(s)=0.00648175
Iter 8 diff=0.000446334, time(s)=0.0064897
Iter 9 diff=0.000339216, time(s)=0.00627329
Iter 10 diff=0.00026211, time(s)=0.00700005
Iter 11 diff=0.000205041, time(s)=0.00674053
Iter 12 diff=0.000162196, time(s)=0.00647074
Iter 13 diff=0.00012937, time(s)=0.00686211
Iter 14 diff=0.000103995, time(s)=0.00657544
Iter 15 diff=8.40758e-05, time(s)=0.00616668
Iter 16 diff=6.83545e-05, time(s)=0.0065111
Iter 17 diff=5.57974e-05, time(s)=0.00608677
Iter 18 diff=4.57381e-05, time(s)=0.00645724
Iter 19 diff=3.7602e-05, time(s)=0.00606711
Iter 20 diff=3.10102e-05, time(s)=0.00617386
Iter 21 diff=2.5629e-05, time(s)=0.00641825
Iter 22 diff=2.12342e-05, time(s)=0.00606223
Iter 23 diff=1.76219e-05, time(s)=0.00617297
Iter 24 diff=1.46527e-05, time(s)=0.00613207
```

```
Top 10 PageRank (val, node):
0: 2.72748e-05 , 32163
1: 2.6021e-05 , 614831
2: 2.52229e-05 , 943
3: 2.26225e-05 , 504140
4: 2.12157e-05 , 1560
5: 2.10375e-05 , 161158
6: 2.10375e-05 , 17754
7: 1.9882e-05 , 846221
8: 1.95812e-05 , 1337
9: 1.93766e-05 , 163075
Total time (max over ranks) = 0.27771 s
Comp time (max) = 0.0735438 s, Comm time (max) = 0.0219386 s
Iterations = 41
```

```
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ tail -n 40 data/pagerank_run_p1.log
Iter 15 diff=8.40758e-05, time(s)=0.00616668
Iter 16 diff=6.83545e-05, time(s)=0.0065111
Iter 17 diff=5.57974e-05, time(s)=0.00608677
Iter 18 diff=4.57381e-05, time(s)=0.00645724
Iter 19 diff=3.7602e-05, time(s)=0.00606711
Iter 20 diff=3.10102e-05, time(s)=0.00617386
Iter 21 diff=2.5629e-05, time(s)=0.00641825
Iter 22 diff=2.12342e-05, time(s)=0.00606223
Iter 23 diff=1.76219e-05, time(s)=0.00617297
Iter 24 diff=1.46527e-05, time(s)=0.00613207
Iter 25 diff=1.21998e-05, time(s)=0.00587537
Iter 26 diff=1.01741e-05, time(s)=0.00693517
Iter 27 diff=8.49405e-06, time(s)=0.00704683
Iter 28 diff=7.10056e-06, time(s)=0.00657495
Iter 29 diff=5.94067e-06, time(s)=0.00688452
Iter 30 diff=4.97582e-06, time(s)=0.00678931
Iter 31 diff=4.17053e-06, time(s)=0.00739137
Iter 32 diff=3.49889e-06, time(s)=0.00684534
Iter 33 diff=2.93713e-06, time(s)=0.00688628
Iter 34 diff=2.46753e-06, time(s)=0.00681442
Iter 35 diff=2.07412e-06, time(s)=0.00663854
Iter 36 diff=1.74453e-06, time(s)=0.0068606
Iter 37 diff=1.46796e-06, time(s)=0.0107477
Iter 38 diff=1.23596e-06, time(s)=0.00726984
Iter 39 diff=1.04104e-06, time(s)=0.00748875
Iter 40 diff=8.77275e-07, time(s)=0.00692482
Top 10 PageRank (val, node):
0: 2.72748e-05 , 32163
1: 2.6021e-05 , 614831
2: 2.52229e-05 , 943
3: 2.26225e-05 , 504140
4: 2.12157e-05 , 1560
5: 2.10375e-05 , 161158
6: 2.10375e-05 , 17754
7: 1.9882e-05 , 846221
8: 1.95812e-05 , 1337
9: 1.93766e-05 , 163075
Total time (max over ranks) = 0.27771 s
Comp time (max) = 0.0735438 s, Comm time (max) = 0.0219386 s
Iterations = 41
```

```
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ cd data
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank/data$ ls
pagerank_run_p1.log  web-Google-100k.txt  web-Google.txt
```

```
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank/data$ touch plot_scaling.py
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank/data$ nano plot_scaling.py
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank/data$ ls
pagerank_run_p1.log  plot_scaling.py  web-Google-100k.txt  web-Google.txt
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank/data$ rm import pandas as pd
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank/data$ ls
                                                           ls
pagerank_run_p1.log  plot_scaling.py  web-Google-100k.txt  web-Google.txt
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank/data$ rm plot_scaling.py
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank/data$ ls
pagerank_run_p1.log  web-Google-100k.txt  web-Google.txt
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank/data$ cd ..
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ ls
Makefile  data  pagerank  pagerank.cpp  venv
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ touch plot_scaling.py
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ nano plot_scaling.py
```

```
=== Running: mpirun -np 1 (OMP=2) ===
Recorded -> p=1 time=0.72 iters=41 rc=0 log=data/pagerank_p1.log
=== Running: mpirun -np 2 (OMP=2) ===
Recorded -> p=2 time=0.66 iters=41 rc=0 log=data/pagerank_p2.log
=== Running: mpirun -np 4 (OMP=2) ===
Recorded -> p=4 time=0.72 iters=41 rc=0 log=data/pagerank_p4.log

Wrote data/scaling_results.csv
procs,omp_threads,total_time_s,iterations,exit_code,logfile
1,2,0.72,41,0,data/pagerank_p1.log
2,2,0.66,41,0,data/pagerank_p2.log
4,2,0.72,41,0,data/pagerank_p4.log

Tails of logs:
---- data/pagerank_p1.log ----
Iter 5 diff=0.00116863, time(s)=0.00636741
Iter 6 diff=0.000823789, time(s)=0.00673184
Iter 7 diff=0.000598627, time(s)=0.00572261
Iter 8 diff=0.000446334, time(s)=0.00615635
Iter 9 diff=0.000339216, time(s)=0.0060078
Iter 10 diff=0.00026211, time(s)=0.00560772
Iter 11 diff=0.000205041, time(s)=0.00603624
Iter 12 diff=0.000162196, time(s)=0.00548438
Iter 13 diff=0.00012937, time(s)=0.00590841
Iter 14 diff=0.000103995, time(s)=0.00584179
Iter 15 diff=8.40758e-05, time(s)=0.00561921
Iter 16 diff=6.83545e-05, time(s)=0.00632186
```

```
Iter 37 diff=1.46796e-06, time(s)=0.00612426
Iter 38 diff=1.23596e-06, time(s)=0.00625151
Iter 39 diff=1.04104e-06, time(s)=0.00602259
Iter 40 diff=8.77275e-07, time(s)=0.00657775
Top 10 PageRank (val, node):
0: 2.72748e-05 , 32163
1: 2.6021e-05 , 614831
2: 2.52229e-05 , 943
3: 2.26225e-05 , 504140
4: 2.12157e-05 , 1560
5: 2.10375e-05 , 161158
6: 2.10375e-05 , 17754
7: 1.9882e-05 , 846221
8: 1.95812e-05 , 1337
9: 1.93766e-05 , 163075
Total time (max over ranks) = 0.254319 s
Comp time (max) = 0.0642093 s, Comm time (max) = 0.018007 s
Iterations = 41
```

```
---- data/pagerank_p2.log ----
Iter 5 diff=0.00116863, time(s)=0.004513
Iter 6 diff=0.000823789, time(s)=0.00452312
Iter 7 diff=0.000598627, time(s)=0.00400355
Iter 8 diff=0.000446334, time(s)=0.00451149
Iter 9 diff=0.000339216, time(s)=0.00412528
Iter 10 diff=0.00026211, time(s)=0.00412009
Iter 11 diff=0.000205041, time(s)=0.00397307
Iter 12 diff=0.000162196, time(s)=0.00476689
Iter 13 diff=0.00012937, time(s)=0.00458641
Iter 14 diff=0.000103995, time(s)=0.00434073
Iter 15 diff=8.40758e-05, time(s)=0.00444371
Iter 16 diff=6.83545e-05, time(s)=0.00436083
Iter 17 diff=5.57974e-05, time(s)=0.00440784
Iter 18 diff=4.57381e-05, time(s)=0.00444016
Iter 19 diff=3.7602e-05, time(s)=0.00443679
Iter 20 diff=3.10102e-05, time(s)=0.00442113
Iter 21 diff=2.5629e-05, time(s)=0.00452562
Iter 22 diff=2.12342e-05, time(s)=0.00442811
Iter 23 diff=1.76219e-05, time(s)=0.00433829
Iter 24 diff=1.46527e-05, time(s)=0.00452938
Iter 25 diff=1.21999e-05, time(s)=0.00442939
Iter 26 diff=1.01741e-05, time(s)=0.00439535
Iter 27 diff=8.49404e-06, time(s)=0.00471604
Iter 28 diff=7.10055e-06, time(s)=0.00563574
Iter 29 diff=5.94066e-06, time(s)=0.00550461
Iter 30 diff=4.97581e-06, time(s)=0.00442457
```

```
Top 10 PageRank (val, node):
0: 2.72748e-05 , 32163
1: 2.6021e-05 , 614831
2: 2.52229e-05 , 943
3: 2.26225e-05 , 504140
4: 2.12157e-05 , 1560
5: 2.10375e-05 , 161158
6: 2.10375e-05 , 17754
7: 1.9882e-05 , 846221
8: 1.95812e-05 , 1337
9: 1.93766e-05 , 163075
Total time (max over ranks) = 0.186097 s
Comp time (max) = 0.0399445 s, Comm time (max) = 0.0309879 s
Iterations = 41
```

```
---- data/pagerank_p4.log ----
Iter 5 diff=0.00116863, time(s)=0.00587413
Iter 6 diff=0.000823789, time(s)=0.00642094
Iter 7 diff=0.000598627, time(s)=0.00510527
Iter 8 diff=0.000446334, time(s)=0.00442989
Iter 9 diff=0.000339216, time(s)=0.00584703
Iter 10 diff=0.00026211, time(s)=0.00440266
Iter 11 diff=0.000205041, time(s)=0.00405489
Iter 12 diff=0.000162196, time(s)=0.00473467
Iter 13 diff=0.00012937, time(s)=0.0043773
Iter 14 diff=0.000103995, time(s)=0.00378051
Iter 15 diff=8.40757e-05, time(s)=0.00430355
Iter 16 diff=6.83545e-05, time(s)=0.00508559
Iter 17 diff=5.57974e-05, time(s)=0.00413593
Iter 18 diff=4.57381e-05, time(s)=0.00396492
Iter 19 diff=3.7602e-05, time(s)=0.00621239
Iter 20 diff=3.10102e-05, time(s)=0.00414305
Iter 21 diff=2.5629e-05, time(s)=0.00418422
Iter 22 diff=2.12342e-05, time(s)=0.00554644
Iter 23 diff=1.76219e-05, time(s)=0.00410885
Iter 24 diff=1.46527e-05, time(s)=0.00579041
Iter 25 diff=1.21999e-05, time(s)=0.00578695
Iter 26 diff=1.01741e-05, time(s)=0.00571638
Iter 39 diff=1.04104e-06, time(s)=0.00400196
Iter 40 diff=8.77273e-07, time(s)=0.00375515
Top 10 PageRank (val, node):
0: 2.72748e-05 , 32163
1: 2.6021e-05 , 614831
2: 2.52229e-05 , 943
3: 2.26225e-05 , 504140
4: 2.12157e-05 , 1560
5: 2.10375e-05 , 161158
6: 2.10375e-05 , 17754
7: 1.9882e-05 , 846221
8: 1.95812e-05 , 1337
9: 1.93766e-05 , 163075
Total time (max over ranks) = 0.230056 s
Comp time (max) = 0.040857 s, Comm time (max) = 0.0763997 s
Iterations = 41
```

```
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ cat data/scaling_results.csv
procs,omp_threads,total_time_s,iterations,exit_code,logfile
1,2,0.72,41,0,data/pagerank_p1.log
2,2,0.66,41,0,data/pagerank_p2.log
4,2,0.72,41,0,data/pagerank_p4.log
```

```
(venv) ayushg@AyushG:~/DPC_Project/Parallel_PageRank$ python3 plot_scaling.py
   procs  omp_threads  total_time_s  iterations  exit_code              logfile   speedup  efficiency
0      1           2          0.72          41          0  data/pagerank_p1.log  1.000000    1.000000
1      2           2          0.66          41          0  data/pagerank_p2.log  1.090909    0.545455
2      4           2          0.72          41          0  data/pagerank_p4.log  1.000000    0.250000
Saved plots to data/*.png
```