CAS CS 562: Advanced Database
Programming Assignment#2

Yanshan Song U93089867
Juntao Wang U18801732

Hadoop naive algorithm
Job1:
Map: Input v_1 v_2, emit $< v_1, v_2 >$
Reduce: Input $< v, \Gamma(v) >$
For each $(u, w)(u\_id < w\_id)$ in $\Gamma(v)$, produce $< (u, w), v >$
The result of Job1:

```
1     2,43    1
2     2,60    1
3     2,128   1
4     2,161   1
5     2,185   1
6     2,199   1
7     2,200   1
8     2,4954  1
9     2,5673  1
10    2,6020  1
```

Job2:
Map: Input $< (u, w), v >$
Produce three items: $< (u, w), v >$, $< (u, v), \$ >$, $< (w, v), \$ >$
Reduce: Input $< key, values >$
If \$ in *values*, For each v in *values*, $triangleCount$ += 1/3
Finally, output *triangleCount*
The result of Job2:

```
1     TriangleCount   1597979
```

Hadoop improved algorithm
Job1:
Map: Input $v_1 \quad v_2$, emit $< v_1, v_2 >$
Reduce: Input $< v, \Gamma(v) >$
For each u in $\Gamma(v)$,
if $v\_id < u\_id, produce < (v, u), (" smallerIdDegree", |\Gamma(v)|) >$
else produce $< (u, v), (" largerIdDegree", |\Gamma(v)|) >$
The result of Job1:

```
1,2 smallerIdDegree,51
1,43    smallerIdDegree,51
1,60    smallerIdDegree,51
1,128   smallerIdDegree,51
1,161   smallerIdDegree,51
1,185   smallerIdDegree,51
1,199   smallerIdDegree,51
1,200   smallerIdDegree,51
1,4954  smallerIdDegree,51
1,5673  smallerIdDegree,51
```

Job2:
Map: Input $< key, value >$, produce $< key, value >$
Reduce: Input $< (v, u), [(" smallerIdDegree", smallerIdDegree), (" largerIdDegree", largerIdDegree)] >$
If $smallerIdDegree \leq largerIdDegree$, produce $< v, u >$, else produce $< u, v >$
The result of Job2:

```
1   10127
103827  1
103917  1
104012  1
104013  1
104452  1
104453  1
104454  1
105008  1
105058  1
```

Job3:
Map: Input $< v, u >$, produce $< v, u >$
Reduce: Input $< v, \Gamma(v) >$
For each u in $\Gamma(v)$, if $v\_id \leq u\_id$, produce $< v, (u, null) >$ else produce $< u, (v, null) >$

For each $(u, w)(u\_id < w\_id)$ in $\Gamma(v)$, produce $(v, (u, w))$

The result of Job3:

```
1   1   10127,null
2   1   43,null
3   1   43,10127
4   1   2,null
5   1   2,10127
6   1   2,43
7   9   10,null
8   100 201,null
9   62  100,null
10  100 62,201
```

Job4:

Map: Input $< key, values >$

if $values[1] = null$, produce $< values, key >$ else produce $< (key, values[0]), \$) >$

Reduce: Input $< key, values >$

If \$ in $values$, For each v in $values$, $triangleCount$ += 1

Finally, output $triangleCount$

The result of Job4:

```
1   TriangleCount   1597979
```

Pig