Exercise 4. *In this assignment, the goal is to implement a set of simple Map Reduce tasks. Please include the Python scripts used in the submitted report. You will work with a collection of e-mail data downloadable from: https: // snap. stanford. edu/ data/ email-EuAll. txt. gz The data forms a graph G of e-mails between users, with each line being of the form sender receiver. Compute the following on G:*

a) Number of nodes in the graph

Ans:

265214

Following are the assumption for the below calculations:
1) Duplicate edges are ignored for calculation of indegree and outdegree
2) Self loop – 1 outdegree and 1 indegree for that particular node
3) For the two hops case, self-loops are ignored

b) Average (and median) indegree and out degree

Ans:

	Median	**Average**
InDegree	**0**	**1.584**
OutDegree	**1**	**1.584**

c) Average (and median) number of nodes reachable in two hops

Ans:

Average = 127.118
Median = 63.0,

d) Number of nodes with indegree > 100

Ans:

702