

Read me

1. Insert the Python script in the same Folder of the file with the data of Pokec
2. Run the script and answers the following questions
 - a. "Select operator
[1]: Neighbor Match
[2]: Common Neighbor Match"
Always Insert 1 for the neighbor match
 - b. "Insert number of edges to import. For example 5000."
 - c. "Insert depth of the walk.
For example:
depth=1 a->b
depth=2 a->b->c
depth=3 a->b->c->d

Minimum depth=1
Maximum depth=11"
 - d. "Insert the starting node.
Nodes are between 1 and [maximum number], according to the selected edges"
 - e. "Select the method to use for the queries:
[1]: HASH Table
[2]: CSR
[3]: Both"
3. After the operator has answered at all the question the program starts to calculate. At the end of the process in the same folder of the script there will be 2 or 3 files depending by the situation:
 - Queries Archive: an excel file that reports the main quantities calculated
 - Hash.txt: a text file containing the neighbors calculated with the hash method (if the operator selects the hash method)
 - Csr.txt: a text file containing the neighbors calculated with the csr method (if the operator selects the csr method)
4. At the end of the calculation the code asks the operator if he wants to perform another query with the same subset of data and the same starting node, changing only length of the chain for the neighbors. If another query is done the txt files are overwritten while the excel file is updated maintaining in memory the past queries. If the operator wants to take into consideration another subset of data, he has to restart the code.