

Assignment # 3: Random Graphs

October 8, 2020

Advice:

1. The file `2016-05-26.tgz` contains binary NetFlow data. You can convert them into `.csv` files on Linux using `flow-tools`. They can be installed on Ubuntu 18.04 LTS by

```
1 sudo apt-get install flow-tools
```

They are not available on Ubuntu 20.04 LTS and newer Linux distros because they require Python 2 to work. If you have them, use

```
1 export-flow -f2 < INPUTNAME > OUTPUTNAME.csv
```

Otherwise, I created a new archive `exported-flows.zip`, you can use it and just skip first step of the assignment.

2. The amount of transported packets is in the column `dpkts` (it's the 4th one - indexing from 0).
3. The source address is in the column `srcaddr` (10th).
4. The destination address is in the column `dstaddr` (11th).
5. It is a netflow with TCP protocol if its 17th column is equal to 6.
6. It is a netflow with UDP protocol if its 17th column is equal to 17.