

Big Data Processing: Assignment 2

Marks: 10

Deadline: 31.03.23, 11.55 pm IST

Implement Karger's mincut algorithm on unweighted undirected graph in python to find two communities in the graph. Your program must produce the following two things:

1. The value of the probable mincut.
2. The two communities that we get after removing the mincut edges. It should be in the following format: two column list, where the first column is the node id and the second column is the community id (1 or 2) and the columns are separated by space.

The input graph data will be a two column (space separated) plain text file, where each line represents an edge. The following gives an example file:

```
1 6
2 7
1 7
```

where the first line means there is an edge between the node 1 and 6. Similarly other lines represent other edges.

You can create a dataset on your own to test your code. Alternatively, you can find a number of graph datasets here <http://snap.stanford.edu/data/index.html>, which you want to make use of.

We will evaluate your program on a linux system from command line with the arguments as follows:

```
python <your-code.py> <path to data two column graph file>
```

The above format is very important for evaluation. Thus, your program arguments must follow the sequence.

Submission guidelines:

You need to submit the program as a single python file in moodle. The file name must follow the format: **assignment-2-roll.py** (where the roll denotes your roll number that must match exactly with your IITKGP roll number). Please note that if you fail to follow the format, your program may not be evaluated at all.

Important notes:

1. No credit will be given if your program does not run and produces wrong output.
2. No submission will be accepted after deadline.
3. It is your responsibility to check that the file has been submitted successfully.
4. Plagiarism from friend or from web will invite negative (-10) marks.