Home assignment
Machine Learning position

1 Algorithmic task

1.1 Problem definition

Let G be a connected undirected graph without cycles (a tree graph), where each leaf is colored either using white, blue or red color. The task is to find the maximum number of different pairs of leaves, where one leaf is red and the other blue such that all pairs can connected by mutually disjunctive paths — these paths connects leaves of each pair and these path has no common node.

1.2 Task

Your goal is to write a program in Python 3, that takes the input describing the graph from the standard input/file and returns a single number on the standard output with the answer to the problem. The program has to be efficient and quick (e.g., brute-forcing is inadmissible). **The solution should follow all coding standards common in python** (including docstrings for building documentation), comments, etc.

The solution should also be reproducible with as least work as possible. Ideally, you prepare a docker image which can be used for running the program; if you are unfamiliar with the docker, usage of the Anaconda virtual environment is recommended (provide also the file defining the environment). You are given some tests inputs, please incorporate them to your unit tests but do not limit the unit tests to them, also test other parts you think should be tested using unit tests. Your program should contain a basic CLI. The CLI can be used , for example, for setting the verbosity, for setting whether the input is taken from the standard input or a provided file, etc.

You are also supposed to write a brief report with illustrations describing your approach and algorithms you have used to tacking the problem. The report should clearly contain the information about the asymptotic complexity of your solution with explanation. More details about the report requirements are in the section 5.

1.3 Input

- The first line contains 3 numbers (integers); denoted M, R, and B, where
  - M — the number of edges,
  - R — number of red leaves,
  - B — number of blue leaves.
- Then there are M rows, where each row represents an edge. It contains two numbers n_1 and n_2 — indices of nodes that are connected by the edge. The nodes are indexed from 1.
- Then you get one row of R numbers — indices of leaves colored red.
- Then you get one row of B numbers — indices of leaves colored blue.

You can expect $3 \leq N \leq 10^6$ and that there is at least one blue leaf and at least one red leaf.

2.3.1 *Example input*

```
6 2 2
1 2
2 3
4 2
5 7
6 5
4 5
1 3
7 6
```

The graph listed above is also shown in fig. 1.Another example (Case 3) is shown in fig. 2.

2.3.2 *Expected output*

```
1
```

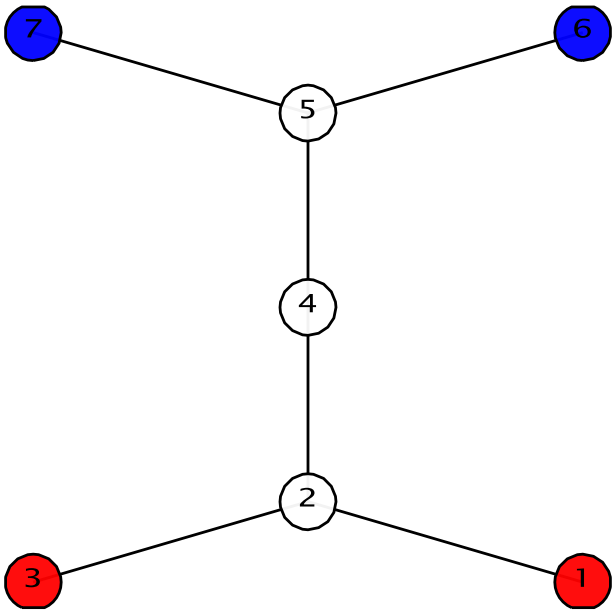

Figure 1. The input graph for the Case 1

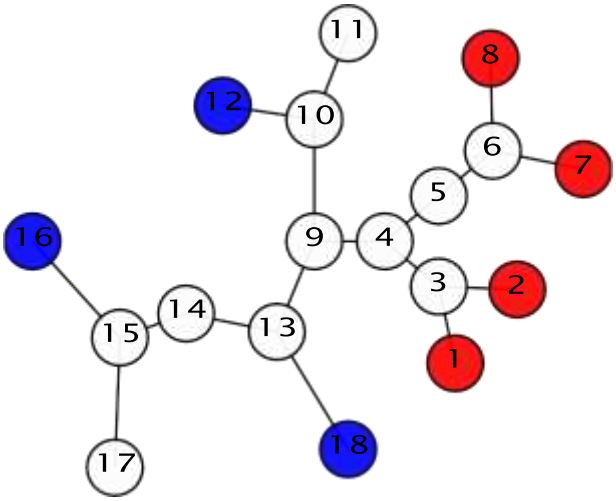

Figure 2. The input graph for the Case 3