HW1

Consider the following three molecular structures:

Penicillin A : http://www.chemspider.com/Chemical-Structure.16736211.html
Penicillin G : http://www.chemspider.com/Chemical-Structure.6014.html
Caffeine : http://www.chemspider.com/Chemical-Structure.2424.html

Download the txt files of the structures of the three compounds. In case of water, the structure file looks like this:

```
3 2
1 H
2 O
3 H
1 2 1
2 3 1
```

The first line is the header line, describing the molecule has 3 atoms and 2 bonds. The next 3 lines provide the atom labels, and then the next two lines describe the bonds between atoms. The bond "1 2 1" means there is a bond between atom 1 (H) and atom 2 (O) of type single bond, while "2 3 1" means there is a bond between atom 2 and 3. In this homework, you can ignore the type of bond (single/double/triple) which is the third column in bonds.

(1) Write a script to compute the adjacency matrix E for each of the structures above. For water molecule, the results should look like this:

$$E = \begin{matrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{matrix}$$

(2) Write a script to compute the mapping $\phi(G)$ and test it for the three compounds under the following kernel maps discussed in the lectures:

(A) "Molecular Formula" (in addition to O, C, N, H also consider S)
(B) "Label paired" with "*length=3*"
(C) "Depth First Search" (allow cycle, double traverse, no compression, depth=2). This should be a binary map (either 0 or 1)

(3) Write a script to compute the kernel $k(G_1,G_2)$ and test it on the three compounds for the following kernels discussed in the lectures:
(A) "Molecular formula", and "Molecular-Formula" + MinMax
(B) "Label paired", and "Label paired" + MinMax
(C) "Depth First Search", and "Depth First Search" + Tanimoto

(4) Decide which of the three compounds are more similar according to:
(A) “Molecular formula”, and “Molecular-Formula” + MinMax
(B) “Label paired”, and “Label paired” + MinMax
(C) “Depth First Search”, and “Depth First Search” + Tanimoto

Between “Label paired”, and “Label paired” + MinMax, which one do you think is a better measure for similarity of chemical structure? Make argument for your choice