K L HYDERABAD

FRESHMAN ENGINEERING DEPARTMENT

A Project-Based Lab Report

On

Isomorphism in biological networks

SUBMITTED BY:

I.D NUMBER	NAME
2010030046	E.Pravallika
2010030168	Tahseen Begum
2010030344	N. Sowgna
2010030445	Keerthana Pulugam

UNDER THE ESTEEMED GUIDANCE OF

Dr. P. Sree Lakshmi

<DESIGNATION>

KONERU LAKSHMAIAH EDUCATION FOUNDATION

(Deemed to be University)
Moinabad Road, Aziz Nagar, Hyderabad - 500075

DEPARTMENT OF BASIC ENGINEERING SCIENCES

CERTIFICATE

This is to certify that the project-based laboratory report entitled "Isomorphism in biological networks" submitted by Mr./Ms**. Tahseen Begum, E.Pravallika, N.Sowgna, Keerthana Pulugam** bearing Regd. No. 2010030168, 2010030046, 2010030344, 2010030445 to the **Department of Basic Engineering Sciences, KL University** in partial fulfillment of the requirements for the completion of a project in the "DAA Project" course in II B Tech IV Semester, is a Bonafede record of the work carried out by him/her under my supervision during the academic year 2021-22.

Signature of the Supervisor

Name and
Designation

Signature of the HOD

Signature of the External Examiner

ACKNOWLEDGEMENTS

It is great pleasure for me to express my gratitude to our honorable President **Sri. Koneru Satyanarayana**, for giving the opportunity and platform with facilities in accomplishing the project-based laboratory report.

I express my sincere gratitude to our Principal **Dr. L. Koteswara Rao** for his administration of our academic growth.

I express sincere gratitude to our Coordinator for her leadership and constant motivation provided in the successful completion of our academic semester. I record it as my privilege to deeply thank you for providing us with the efficient faculty and facilities to make our ideas into reality.

I express my sincere thanks to our project **supervisor Dr. P. Sree Lakshmi** for her novel association of ideas, encouragement, appreciation, and intellectual zeal which motivated us to venture into this project successfully.

Finally, it is pleased to acknowledge the indebtedness to all those who devoted themselves directly or indirectly to making this project report success.

Name: Tahseen Begum

Regd. No: 2010030168

Name: E.Pravallika

Regd. No: 2010030046

Name: N.Sowgna

Regd. No: 2010030334

Name: Keerthana Pulugam

Regd. No: 2010030445

ABSTRACT

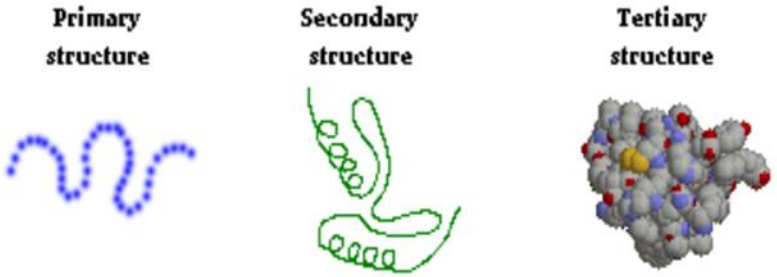
Number of real-world problems is represented by graph. Graph isomorphism is the area of pattern matching and widely used in various applications such as image processing, protein structure, computer and information system, chemical bond structure, Social Networks. This project surveys both various applications of graph isomorphism and their importance in the society.

INDEX

Problem Statement

We Use Protein Structure, Nodes Represents Protein And Edges Represents Their Interactions Between Nodes.

- We Have Three Levels Of Protein Structure I.E., Primary Structure, Secondary Structure, Tertiary Structure.

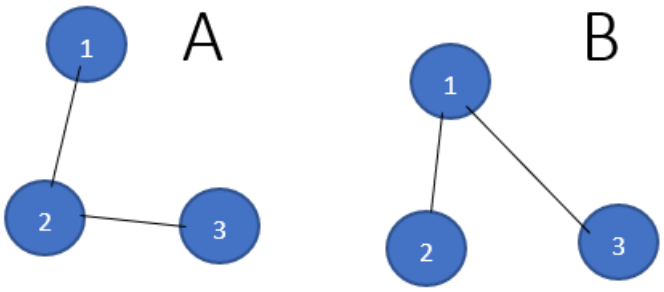


- For Example Collection Of Food Is Available And Every Food Is Having their Structure That Is Graph Structure. For Providing The Food And That Food Contains Some Proteins, Then Find Protein Graph Structure First And Check Where It Is Available In The Food Structure Or Not.

Solution strategy- Algorithm used

We have Graph A and B, now we have to check the two graphs are isomorphic are not using the VF2 Algorithm

Our vertex are 1,2,3 in graphs A and B



Step 1 :

- I match empty A with empty B it always works.
- We can match 1A with 1B,2B,3B
- Now we take 1A with 1B it always works

Step 2 :

- I can match 2A with 2B or 3B
- I match 2A with 2B always works because 1A,2A, and 1B,2B are connected.

Step 3 :

- I can match 3A with any node in Graph B we cannot connect because there is no edge between 2B and 3B in graph B, so we will go back to again step2.

Step 4:

- I can match 2A with 3B we cannot connect because there is no edge between 2B and 3B in graph B, so we will go back to again step2, But in step2 we didn't have a solution, so we will go to step1.

Step 5:

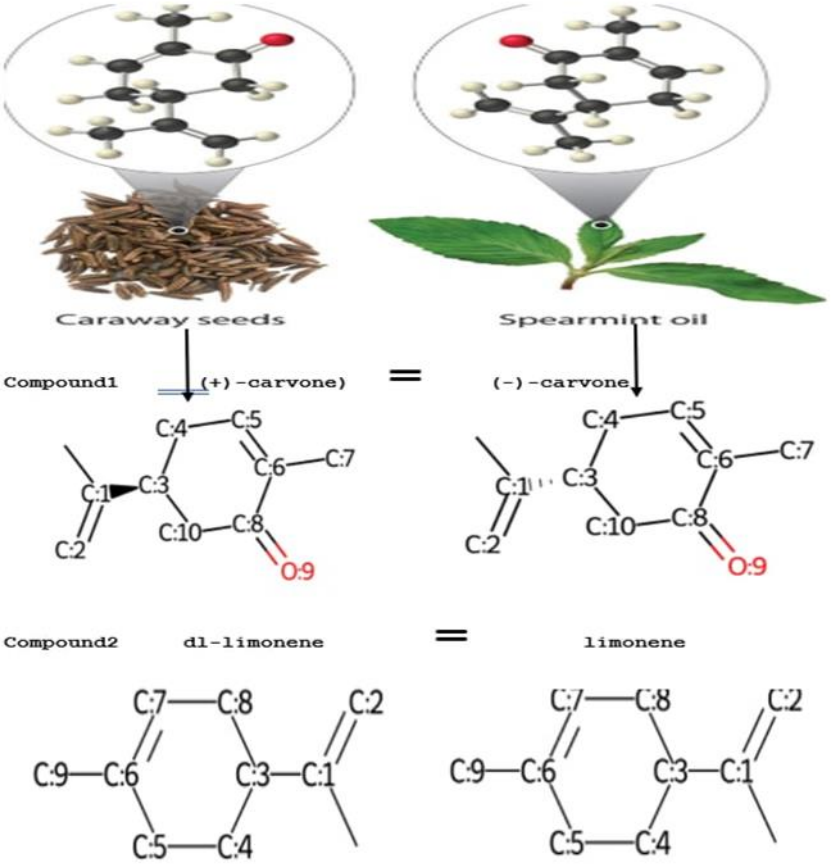
- I match 1A with 2B and 2A with 1B and 3A with 3B
- The graphs are isomorphic

Numeric example of the algorithm (Sample input - expected output)

An implementation of the VF2 algorithm for graph isomorphism testing. The simplest internet to is to call netnetwork_isomorphic().

```
In [31]: import networkx as nx
         from networkx.algorithms import isomorphism
         G1=nx.path_graph([(1,2),(2,3)])
         G2=nx.path_graph([(1,2),(1,3)])

         GM = isomorphism.GraphMatcher(G1, G2)
         GM.is_isomorphic()

Out[31]: True

In [32]: GM.mapping

Out[32]: {(1, 2): (1, 2), (2, 3): (1, 3)}
```

Flow chart


Caraway seeds
Spearmint oil
Compound1
(+)-carvone)
=
(-)-carvone
C:4
C:5
C:6
C:7
C:1
C:3
C:10
C:8
C:2
O:9
Compound2
dl-limonene
=
limonene
C:7
C:8
C:2
C:9
C:6
C:3
C:1
C:5
C:4

Hardware and Software used

1)Windows 11 for x64-based system

2) brackets

3) Jupiter notebook

IMPLEMENTATION

```
def mol_with_atom_index(mol):
  for atom in mol.GetAtoms():
    atom.SetAtomMapNum(atom.GetIdx())
  return mol


from urllib.request import urlopen
from urllib.parse import quote


def CIRconvert(ids):
  try:
    url = 'https://cactus.nci.nih.gov/chemical/structure/' + quote(ids) +
'/smiles'
    ans = urlopen(url).read().decode('utf8')
    return ans
  except:
    return 'Did not work'


from rdkit import Chem
p__c1_name1 = input("Enter primary protien compoud1 of caraway seeds  :
")
s_c1_formula1 = CIRconvert(p__c1_name1)
print(s_c1_formula1)
mol1 = Chem.MolFromSmiles(s_c1_formula1)
mol_with_atom_index(mol1)


from rdkit import Chem
```

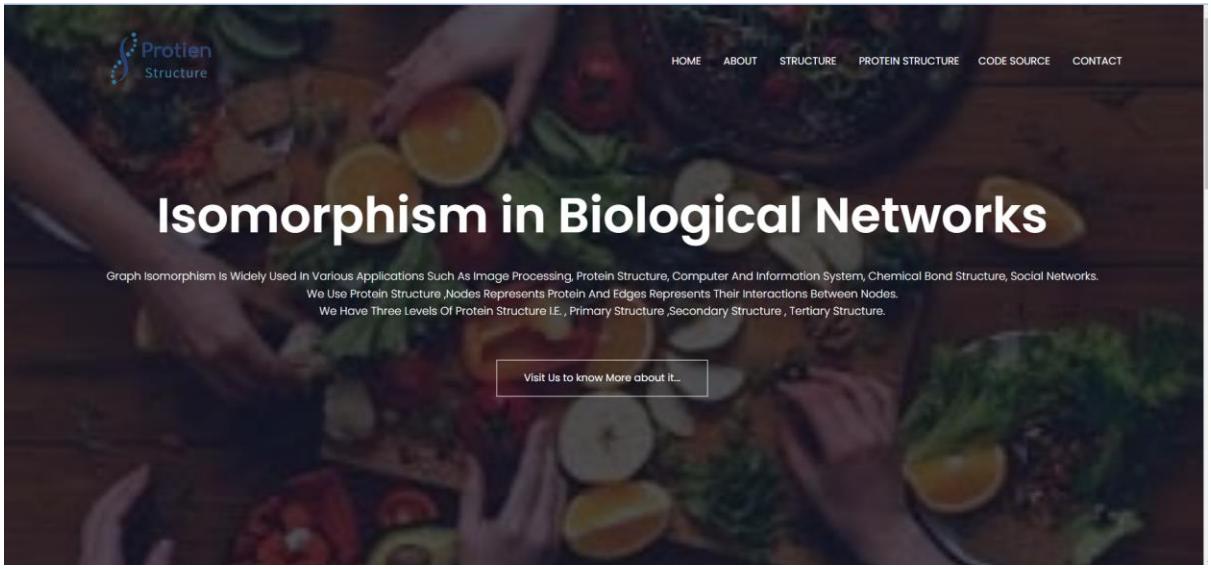
```
p__c1_name2 = input("Enter primary protien compoud1 of spearmint oil :
")

s_c1_formula2 = CIRconvert(p__c1_name2)

print(s_c1_formula2)

mol2 = Chem.MolFromSmiles(s_c1_formula2)

mol_with_atom_index(mol2)


from rdkit.Chem.EnumerateStereoisomers import
EnumerateStereoisomers,StereoEnumerationOptions

isomers = tuple(EnumerateStereoisomers(mol1))

len(isomers)


from rdkit import Chem

for smi in sorted(Chem.MolToSmiles(x,isomericSmiles=True)for x in
isomers):

   print(smi)

   isomer=Chem.MolFromSmiles(smi)

   mol_with_atom_index(isomer)


from pysmiles import read_smiles

import networkx as nx

smiles='CC(=C)[C@H]1CC=C(C)C(=O)C1'

mol_1=read_smiles(s_c1_formula1)

mol_2=read_smiles(s_c1_formula2)


nodes_c1 =mol_1.nodes(data='element')

nodes_c2=mol_2.nodes(data='element')

adj_matrix_c1=nx.to_numpy_matrix(mol_1)
```

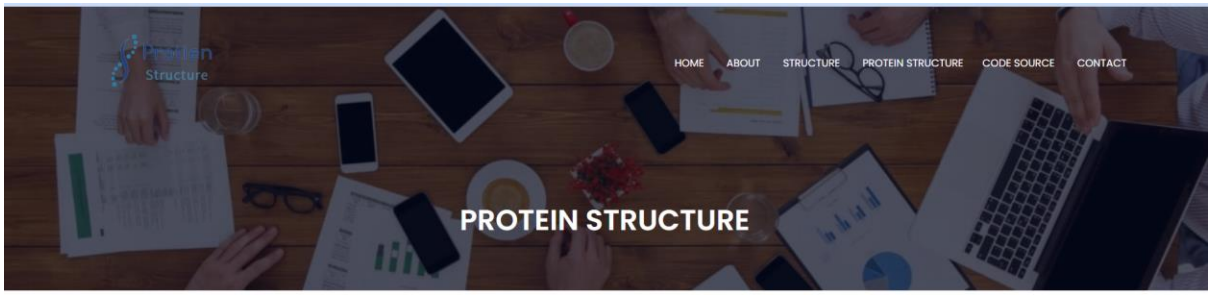
```
adj_matrix_c2=nx.to_numpy_matrix(mol_2)
print(nodes_c1)
print(adj_matrix_c1)

print(nodes_c2)
print(adj_matrix_c2)

from network.algorithms import isomorphism
G1=nx.from_numpy_matrix(adj_matrix_c1)
G2=nx.from_numpy_matrix(adj_matrix_c2)
GM=isomorphism.GraphMatcher(G1,G2)
GM.is_isomorphic()

from networkx.algorithms import isomorphism
G1=nx.from_numpy_matrix(adj_matrix_c1)
G2=nx.from_numpy_matrix(adj_matrix_c2)
GM=isomorphism.GraphMatcher(G1,G2)
GM.is_isomorphic()

from rdkit import Chem
p__c2_name2 = input("Enter primary protien compoud2 of spearmint oil :
")
s_c2_formula2 = CIRconvert(p__c2_name2)
print(s_c2_formula2)
mol2 = Chem.MolFromSmiles(s_c2_formula2)
mol_with_atom_index(mol2)
```

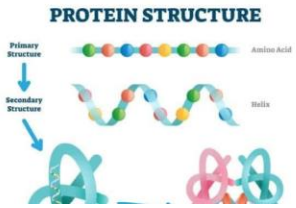
```
from rdkit.Chem.EnumerateStereoisomers import
EnumerateStereoisomers,StereoEnumerationOptions

isomers = tuple(EnumerateStereoisomers(mol1))

len(isomers)


from rdkit import Chem

for smi in sorted(Chem.MolToSmiles(x,isomericSmiles=True)for x in
isomers):
    print(smi)
    isomer=Chem.MolFromSmiles(smi)
    mol_with_atom_index(isomer)


from pysmiles import read_smiles

import networkx as nx

smiles='CC(=C)C1CCC(=CC1)C'

mol_1=read_smiles(s_c2_formula1)

mol_2=read_smiles(s_c2_formula2)


nodes_c1 =mol_1.nodes(data='element')

nodes_c2=mol_2.nodes(data='element')

adj_matrix_c1=nx.to_numpy_matrix(mol_1)


adj_matrix_c2=nx.to_numpy_matrix(mol_2)

print(nodes_c1)

print(adj_matrix_c1)


print(nodes_c2)

print(adj_matrix_c2)
```

```
from networkx.algorithms import isomorphism
G1=nx.from_numpy_matrix(adj_matrix_c1)
G2=nx.from_numpy_matrix(adj_matrix_c2)
GM=isomorphism.GraphMatcher(G1,G2)
GM.is_isomorphic()
```

RESULTS

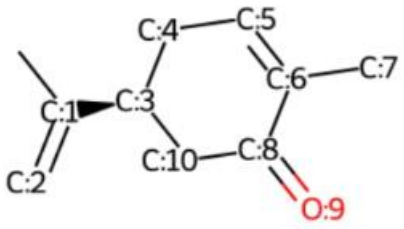

Protien
Structure
HOME
ABOUT
STRUCTURE
PROTEIN STRUCTURE
CODE SOURCE
CONTACT
Isomorphism in Biological Networks
Graph Isomorphism Is Widely Used In Various Applications Such As Image Processing, Protein Structure, Computer And Information System, Chemical Bond Structure, Social Networks.
We Use Protein Structure ,Nodes Represents Protein And Edges Represents Their Interactions Between Nodes.
We Have Three Levels Of Protein Structure I.E. , Primary Structure ,Secondary Structure , Tertiary Structure.
Visit Us to know More about it...


Protien
Structure
HOME
ABOUT
STRUCTURE
PROTEIN STRUCTURE
CODE SOURCE
CONTACT
PROTEIN STRUCTURE

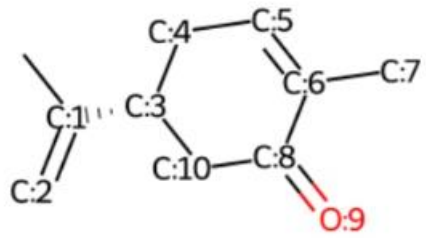

PROTEIN STRUCTURE
Primary Structure
Amino Acid
Secondary Structure
Helix

PROTEIN STRUCTURE

```
Enter primary protien compoud1 of caraway seeds  : (+)-carvone
CC(=C)[C@H]1CC=C(C)C(=O)C1
```

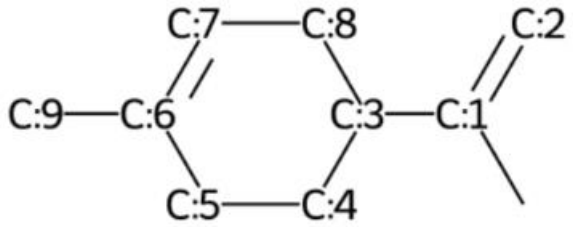



```
Enter primary protien compoud1 of spearmint oil : (-)-carvone
CC(=C)[C@@H]1CC=C(C)C(=O)C1
```

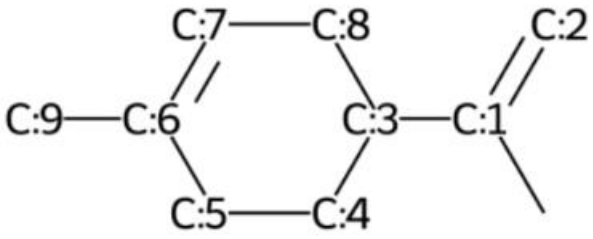



```
[(0, 'C'), (1, 'C'), (2, 'C'), (3, 'C'), (4, 'C'), (5, 'C'), (6, 'C'), (7, 'C'), (8, 'C'), (9, 'O'), (10, 'C')]
[[0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [1. 0. 1. 1. 0. 0. 0. 0. 0. 0. 0.]
 [0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 1. 0. 0. 1. 0. 0. 0. 0. 0. 1.]
 [0. 0. 0. 1. 0. 1. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 1. 0. 1. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 1. 0. 1. 1. 0. 0.]
 [0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 1. 0. 1. 1.]
 [0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0.]
 [0. 0. 0. 1. 0. 0. 0. 0. 1. 0. 0.]]
[(0, 'C'), (1, 'C'), (2, 'C'), (3, 'C'), (4, 'C'), (5, 'C'), (6, 'C'), (7, 'C'), (8, 'C'), (9, 'O'), (10, 'C')]
[[0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [1. 0. 1. 1. 0. 0. 0. 0. 0. 0. 0.]
 [0. 1. 0. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 1. 0. 0. 1. 0. 0. 0. 0. 0. 1.]
 [0. 0. 0. 1. 0. 1. 0. 0. 0. 0. 0.]
 [0. 0. 0. 0. 1. 0. 1. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 1. 0. 1. 1. 0. 0.]
 [0. 0. 0. 0. 0. 0. 1. 0. 0. 0. 0.]
 [0. 0. 0. 0. 0. 0. 0. 1. 0. 1. 1.]
 [0. 0. 0. 0. 0. 0. 0. 0. 1. 0. 0.]
 [0. 0. 0. 1. 0. 0. 0. 0. 1. 0. 0.]]
```

```
Enter primary protien compoud2 of caraway seeds  : dl-limonene
CC(=C)C1CCC(=CC1)C
```




```
Enter primary protien compoud2 of spearmint oil : limonene
CC(=C)C1CCC(=CC1)C
```




```
[(0, 'C'), (1, 'C'), (2, 'C'), (3, 'C'), (4, 'C'), (5, 'C'), (6, 'C'), (7, 'C'), (8, 'C'), (9, 'C')]
[[0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
 [1. 0. 1. 1. 0. 0. 0. 0. 0. 0.]
 [0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 1. 0. 0. 1. 0. 0. 0. 1. 0.]
 [0. 0. 0. 1. 0. 1. 0. 0. 0. 0.]
 [0. 0. 0. 0. 1. 0. 1. 0. 0. 0.]
 [0. 0. 0. 0. 0. 1. 0. 1. 0. 1.]
 [0. 0. 0. 0. 0. 0. 1. 0. 1. 0.]
 [0. 0. 0. 1. 0. 0. 0. 1. 0. 0.]
 [0. 0. 0. 0. 0. 0. 1. 0. 0. 0.]]
[(0, 'C'), (1, 'C'), (2, 'C'), (3, 'C'), (4, 'C'), (5, 'C'), (6, 'C'), (7, 'C'), (8, 'C'), (9, 'C')]
[[0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
 [1. 0. 1. 1. 0. 0. 0. 0. 0. 0.]
 [0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 1. 0. 0. 1. 0. 0. 0. 1. 0.]
 [0. 0. 0. 1. 0. 1. 0. 0. 0. 0.]
 [0. 0. 0. 0. 1. 0. 1. 0. 0. 0.]
 [0. 0. 0. 0. 0. 1. 0. 1. 0. 1.]
 [0. 0. 0. 0. 0. 0. 1. 0. 1. 0.]
 [0. 0. 0. 1. 0. 0. 0. 1. 0. 0.]
 [0. 0. 0. 0. 0. 0. 1. 0. 0. 0.]]
```

Future scope of improvement

This project can be further enhanced to provide a 3D shape generator so that we can show the diagram in a 3D shape. Now it shows a normal diagram in the future we make that diagram into a 3D shape by doing the changes.

REFERENCES

https://github.com/K-L-U-H/Isomorphism-in-biological-networks

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3152790/

https://cactus.nci.nih.gov/chemical/structure/

https://stackoverflow.com/questions/54930121/converting-molecule-name-to-smiles

https://stackoverflow.com/questions/19883567/vf2-algorithm-implementation

https://networkx.org/documentation/stable/reference/algorithms/isomorphism.vf2.html

https://www.ijcaonline.org/archives/volume162/number7/somkunwar-2017-ijca-913414.pdf