Project Report
on

Isomorphism in biological networks

Design and Analysis of Algorithms

By

N. Sowgna (2010030344)

Tahseen Begum (2010030168)

E. Pravallika (2010030046)

Keerthana (2010030445)

under the supervision of

Ms. P. Sree Lakshmi

Assistant Professor

Department of Computer Science and Engineering

K L University Hyderabad,

Aziz Nagar, Moinabad Road, Hyderabad – 500 075, Telangana, India.

March, 2022

Declaration

The Project Report entitled "Isomorphism in biological networks" is a record of bonafide work of 2010030344(N.Sowgna) ,2010030046 (E.Pravallika),2010030168(Tahseen Begum) , 2010030445(Keerthana Pulugam), submitted as a requirement for the completion of the course **Design and Analysis of Algorithms** in the Department of Computer Science and Engineering to the K L University, Hyderabad. The results embodied in this report have not been copied from any other Departments/University/Institute.

Signature of the Students

N. Sowgna (2010030344)
Tahseen Begum (2010030168)
E. Pravallika (2010030046)
Keerthana (2010030445)

Certificate

This is to certify that the Project Report entitled "Isomorphism in biological networks" is being submitted by 2010030344(N.Sowgna), 2010030046(E.Pravallika), 2010030168 (Tahseen Begum) , 2010030445 (Keerthana Pulugam) , as a requirement for the completion of the course **Design and Analysis of Algorithms** in the Department of Computer Science and Engineering, K L University, Hyderabad is a record of bonafide work carried out under our guidance and supervision.

The results embodied in this report have not been copied from any other departments/ University/Institute.

Signature of the Supervisor

Ms. P. Sree Lakshmi

Assistant Professor

Signature of the HOD

Signature of the Examiner

ACKNOWLEDGEMENT

First and foremost, we thank the lord almighty for all his grace & mercy showered upon us, for completing this project successfully.

We take grateful opportunity to thank our beloved Founder and Chairman who has given constant encouragement during our course and motivated us to do this project. We are grateful to our Principal **Dr. L. Koteswara Rao** who has been constantly bearing the torch for all the curricular activities undertaken by us.

We pay our grateful acknowledgement & sincere thanks to our Head of the Department **Dr. Chiranjeevi Manike** for his exemplary guidance, monitoring and constant encouragement throughout the course of the project. We thank **Ms. P. Sree Lakshmi** of our department who has supported throughout this project holding a position of supervisor.

We whole heartedly thank all the teaching and non-teaching staff of our department without whom we won't have made this project a reality. We would like to extend our sincere thanks especially to our parents, our family members and friends who have supported us to make this project a grand success.

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. Graphs are used to represent an image which will be very useful in the area of Image processing where graph shows structural description of image. In such representation vertices represent region of image and relation of regions are represented by links called as edges. Subgraph Isomorphism algorithm for chemical structures is Maximum Common Subgraph which is used for matching 2D and 3D Structures.

TABLE OF CONTENTS

LIST OF FIGURES

1.INTRODUCTION

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.[7]

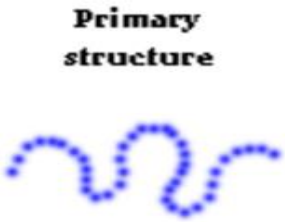


Figure 1.1: primary Structure

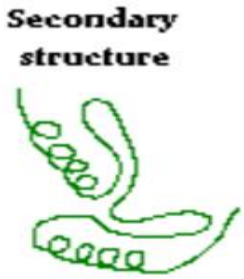


Figure 1.2: Secondary Structure

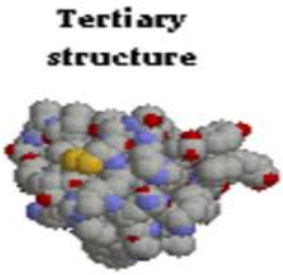


Figure 1.3: Tertiary Structure

For Example Collection Of Food Is Available And Every Food Is Having their Structure That Is Graph Structure.[7] 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.

2.LITERATURE SURVEY

2.1 EXISTING SOLUTIONS

A graph G1 = (V1, E1) is isomorphic to a subgraph of a graph G2 = (V2, E2) if there exists a subgraph of G2, say G2a, such that G1 =~ G2a Assume G1 has n nodes. [7]Lets examine each subset of G that has n nodes, check if they have the same labels as nodes in G1,and if yes, check if the edge in G1 exists also in the selected set. we get an exponential algorithm.[7]

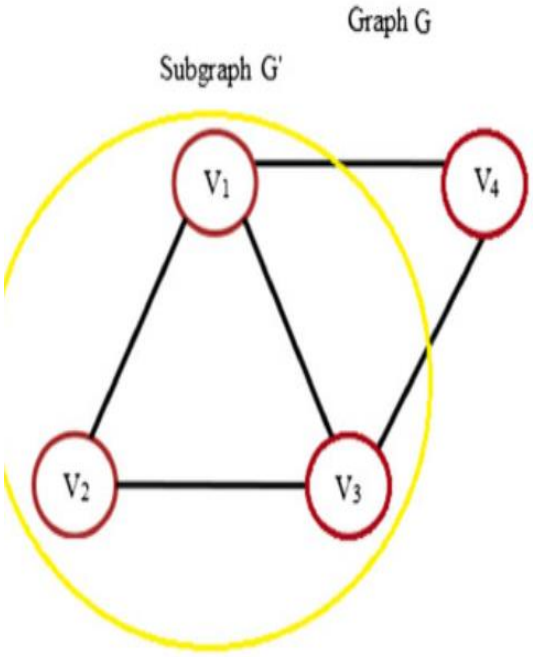


Figure 2.1:Graph G

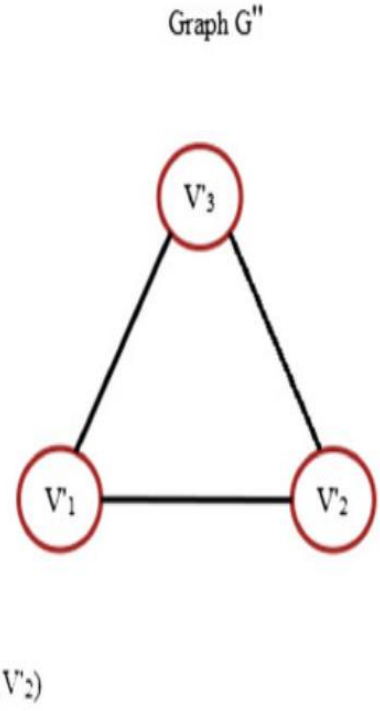


Figure 2.2: Graph G''

3.HARDWARE & SOFTWARE REQUIREMENTS

1.WINDOWS 11 FOR X64-BASED SYSTEM

Ensure the PC you want to install Windows 11: Has a 64-bit CPU: Windows 11 can only run on 64-bit CPUs. To see if your PC has one, go to Settings > System > About, or search "System Information" in Windows and look under "System Type." The Windows 11 ISO is only available for devices with x64 processors.

2.BRACKETS

Brackets is a source code editor with a primary focus on web development. Created by Adobe Inc., it is free and open-source software licensed under the MIT License, and is currently maintained on GitHub by open-source developers. It is written in JavaScript, HTML and CSS. Brackets is cross-platform, available for macOS, Windows, and most Linux distributions. The main purpose of Brackets is its live HTML, CSS and JavaScript editing functionality.

3.JUPITER NOTEBOOK

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations, and narrative text. Its uses include data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

4.FUNCTIONAL & NON-FUNCTIONAL REQUIREMENTS

4.1 DATA STRUCTURES

We Use Graphs For Our Project As A Data Structure.[7]

Graphs Are Powerful Data Structure To Represent Objects And Their Concept.

Objects Are Nothing But Nodes And Edges Describes Relation Among Object

Graphs Are Remains Same If And Only If We Are Not Changing Their Label.

4.2 NUMERIC EXAMPLE OF THE ALGORITHM

An implementation of the VF2 algorithm for graph isomorphism testing. The simplest internet to is to call netnetwork_isomorphic().[6]

```
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)}
```

Figure 4.1:Example for vf2 algorithm

4.3 FLOW CHART

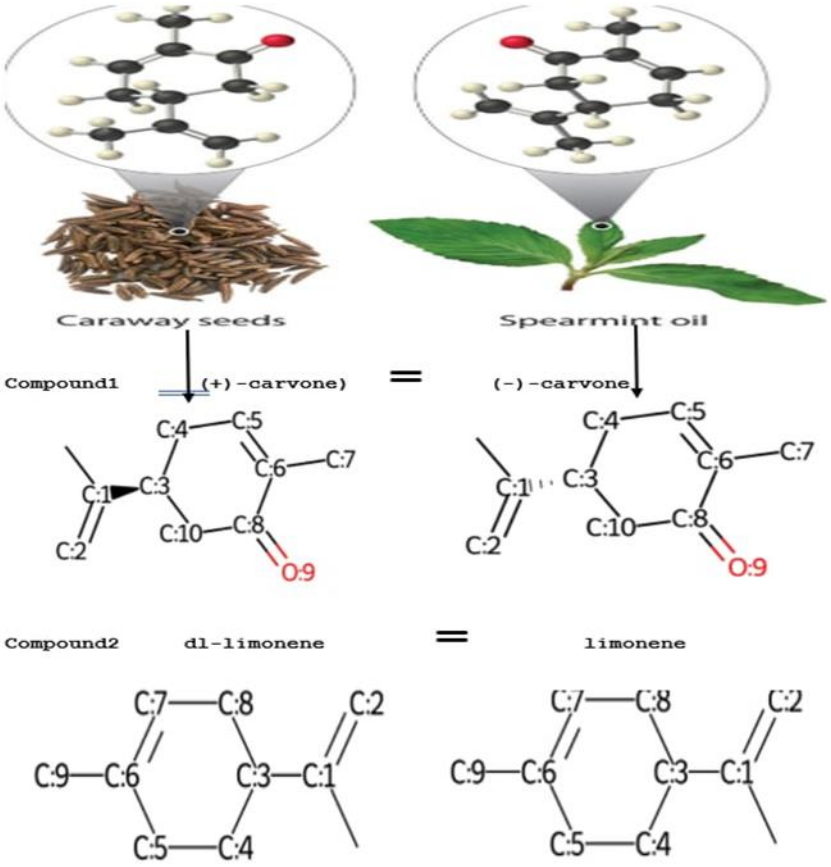
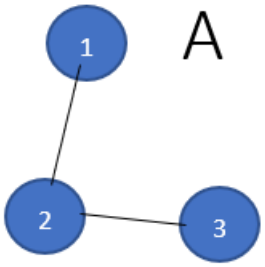


Figure 4.2:Flow chart

5.PROPOSED SYSTEM

5.1 SOLUTION STRATEGY- ALGORITHM

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[5]



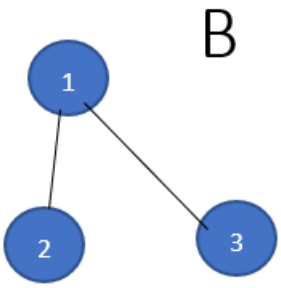


Figure 5.1:Graph A Figure 5.2:Graph B

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

6.IMPLEMENTATION

Nodes represents protein and Edges represents their interactions between nodes. Proteins are represented by networks . To compare two proteins structures number of algorithms are reviewed in ,which plays an important role in biological networks. mol_with_atom_index method is used to give index for the nodes.[4]

RDKit cant convert names to SMILES. Chemical Identifier Resolver can convert names and other identifiers (like CAS No) and has an API so you can convert with a script[4]. We use https://cactus.nci.nih.gov/chemical/structure[3] this link and CIRconvert method to convert given real protein structure inko SMILES.

Next we took two items first one is that caraway seeds and spearmint oil from caraway seeds we took two compounds (+)-carvone, dl-limonene and from spearmint oil we took two compounds (-)-carvone, limonene .[1]

Before comparing first we convert the SMILES[1] structure to matrix using networkx method.we compare (+)-carvone, (-)-carvone primary protein structures if we get Yes we are assuming that caraway seeds and spearmint oil are isomorphic .In the same way we compare more two compounds. We use GraphMatcher[6] to check they are isomorphic .

7.RESULTS DISCUSSION

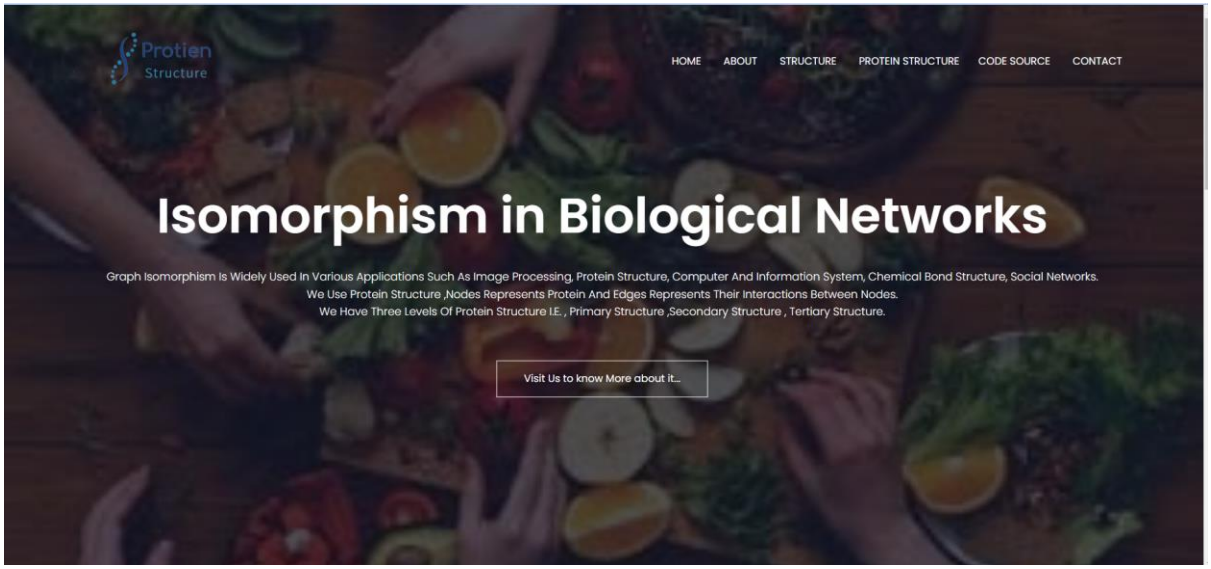
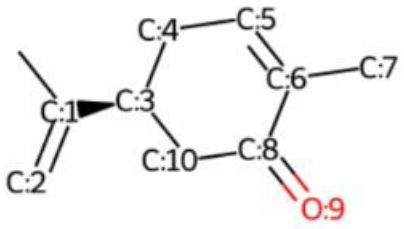


Figure 7.1:Home Page

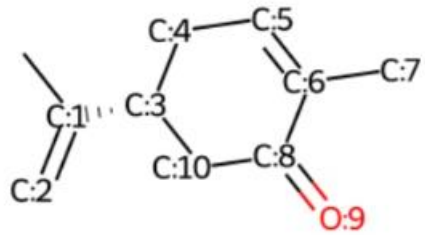
Figure 7.2:About page

PROTEIN STRUCTURE

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



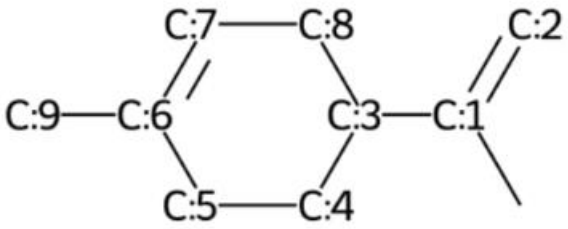

```
Enter primary protien compoud1 of spearmint oil : (-)-carvone
CC(=C)[C@@H]1CC=C(C)C(=O)C1
```



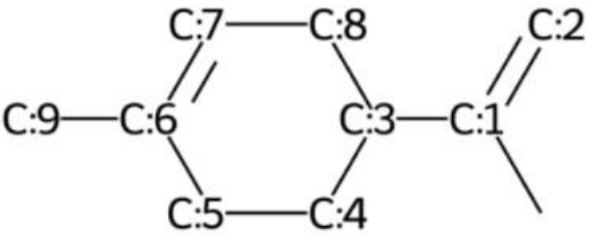

```
[(0, 'C'), (1, 'C'), (2, 'C'), (3, 'C'), (4, 'C'), (5, 'C'), (6, 'C'), (7, 'C'), (8, 'C'), (9, 'O'), (10, 'C')]
[[0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
 [1. 0. 1. 1. 0. 0. 0. 0. 0. 0.]
 [0. 1. 0. 0. 0. 0. 0. 0. 0. 0.]
 [0. 1. 0. 0. 1. 0. 0. 0. 0. 1.]
 [0. 0. 0. 1. 0. 1. 0. 0. 0. 0.]
 [0. 0. 0. 0. 1. 0.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

Figure 7.3:Protein structure of carvone

```
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.]]
```

Figure 7.3:Protein structure of limonene

8.CONCLUSION

Two graphs are ‘isomorphic’ if there exists a one-to-one mapping between their nodes such that each edge in one graph can be mapped to an edge in the other graph. Hierarchical graphs provide more intuitive formal representations of proteins and other structured molecules with multiple functional components than do the regular graphs of current languages for specifying rule-based models, such as the BioNetGen language (BNGL). Thus, the proposed use of hierarchical graphs should promote clarity and better understanding of rule-based models. Various applications of graph isomorphism and their algorithm are studied thoroughly. These applications play an important role in the society. In this paper graph isomorphism applications and their problems are also discussed. It can be concluded that these applications are useful in pattern matching.

9.FUTURE WORK

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.

10.REFERENCES

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

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

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

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

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

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

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