A Graph Theory Based Proposal for the Analysis of the RNA Processing Pathway in COVID-19 Disease Pathology

Anusha Senapati

June, 2021

Abstract

Graph theory has been used for years as a tool to quickly solve mathematical problems. Its importance has grown as people are starting to realize its significant impact when applied to different areas including biological sciences. This paper will explore graph theory in general and specifically apply graph theory to study the interactions between COVID-19 and human proteins, giving further insight into the disease. This paper attempts to delineate how incidence matrices can be used to efficiently describe PPI (Protein-Protein Interaction) networks.

1 Introduction

Graph Theory is the study of graphs, and is used to help manipulate graphs in order to portray relationships of different data sets. The study of Graph Theory was first created by Leonard Euler (1735), while working on the Seven Bridges of Konigsberg problem. Later, this led to the discovery of the eulerian graph. Today, biologists around the world use graph theory, specifically for the subject areas of molecular biology and genomics, to analyze lab data. One particular use of Graph Theory in biology is to model and analyze proteins in networks called protein-protein interaction (PPI) networks. This paper explores graph theory initially and then examines a specific connection Graph Theory has with biological networks. This paper includes a proposal to develop an incidence matrix to study PPI networks specific to the RNA (ribonucleic acid) pathway in COVID-19 pathology.

2 Conceptual Background to Graph Theory

Some important graph theory concepts rely on a few fundamental definitions.

Definition 1. Graph: a graph is the mathematical arrangement which is used to show and analyze the relationships between different objects.

Graphs contain edges (or a "line") and vertices (also known as "nodes").

Definition 2. Vertex: A vertex is a point where multiple lines meet. It is also called a node (mentioned above). A vertex set is the set of all vertices of a graph. This can be written as V(G) for any graph G. The size of a graph is the number of vertices.

Definition 3. An edge is one of the connections between the vertices of the network (graph). Edges can be directed, meaning they point from one node to the next. Many edges can be formed from a single vertex. Without a vertex, an edge cannot be formed. There must be a starting vertex and an ending vertex for an edge. In addition, the edge set of a graph (the set of edges in a specific graph) for a graph G can be written as E(G). The number of edges in a graph is known as the order of the graph. Each edge can also be assigned a certain "weight", or in other words, the measure of how connected an edge is to its attached vertex.

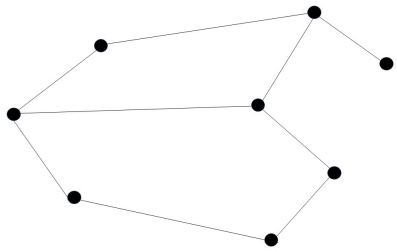

Figure 1: Graph G contains a vertex set V(G) = 7 and E(G) = 9.

We can practice using these simple concepts by solving the following problem:

For a graph G with a rule that i + j ∈ S and —i - j— ∈ S, where i and j belong to vertex set S, find and draw the graph G.

In order to solve, we first need to write out the vertex set S.

Since it is not specified as to which numbers should be used, we create our own. For example, $S = \{2, 4, 5, 8, 9, 11\}$.

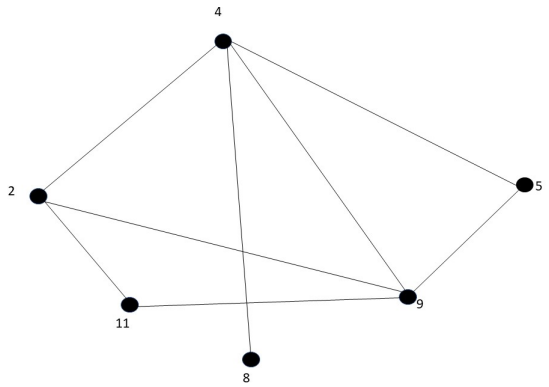

Figure 2: Graph G with a vertex set $S = \{2,4,5,8,9,11\}$

In this case, each number can preform the given functions of being added to produce a number which also belongs to the set, or, subtracted from/subtracted by a number from the set and then finding its distance from zero to get a number which belongs to set S as well.

Definition 4. Degree: The degree of a vertex is the number of edges attached to it.

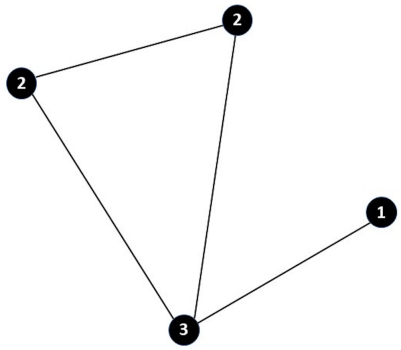

Figure 3: Each vertex labelled with its degree.

The number of degrees a vertex has can quickly help us understand the relationship between any node and edge in a graph.

Definition 5. Incident: A vertex and edge are considered incident to each other if directly connected.

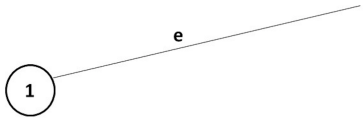

Figure 4: Vertex 1 is incident to edge e.

2.1 Matrices in Graph Theory

Another important part of graph theory is using matrices to describe graphs.

Consider a Graph G of order n and size m.

In such a graph, V(G) = $\{v_1, v_2, v_3, ..., v_n\}$ and E(G) = $\{e_1, e_2, e_3, ..., e_n\}$.

With this information, it is possible to derive two different types of matrices: an adjacency matrix and an incidence matrix.

Concept 1. Adjacency Matrix: an adjacency matrix is an n x n matrix, defined as A = $[a_{ij}]$, if

$$a_{ij} = \begin{cases} 1 & \text{If } V_i V_j \in E(G) \\ 0 & \text{otherwise ;} \end{cases}$$

Concept 2. Incidence Matrix: an incidence matrix is an n x m matrix, defined as B = $[b_{ij}]$, if

$$b_{ij} = \begin{cases} 1 & \text{If } V_i \text{ is incident with } e_j \\ 0 & \text{otherwise.} \end{cases}$$

Now we can apply this knowledge to find the adjacency and incidence matrices of the following graph:

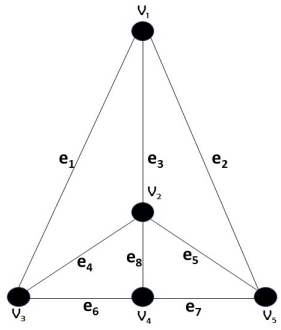

Figure 5: Graph with V(G) = $\{v_1, v_2, v_3, v_4, v_5\}$ and E(G) = $\{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8\}$

If we align the matrix in such a way that the vertices are rows and columns, starting from v_1 to v_5, then we achieve the following empty matrix:

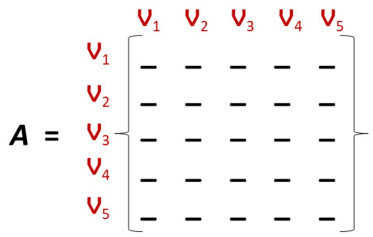

Figure 6: Empty adjacency matrix of graph in Figure 5

Then, we can add the values into the empty spaces on the matrix. So, if two vertices are connected by an edge, the intersection of the row and column will be marked with a 1. On the other hand, if they are not connected by a common edge, then the intersection is marked as a 0. We achieve the following matrix:

$$A = \begin{bmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 1 & 0 \end{bmatrix}$$

Figure 7: Adjacency matrix of graph from Figure 5

In order to find the Incidence Matrix, we need to create an empty matrix with vertices aligned as rows and edges aligned as columns.

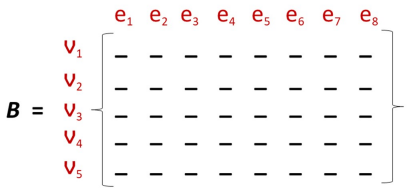

Figure 8: Empty incidence matrix of graph from Figure 5

Then, we can insert 0's and 1's into their respective spots. This means that when an edge and vertex are incident to one another, their intersection on the matrix should be marked by a 1. However, if not, it should be marked by a 0. After this process, we create a matrix like this:

$$B = \begin{Bmatrix} 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \end{Bmatrix}$$

Figure 9: Incidence matrix of graph from Figure 5

3 Applications of Graph Theory to Biological Networks

One of Graph Theory's most common applications in Graph Theory is using graphs to model certain biological networks.

3.1 Protein-Protein Interactions (PPI) Within a Cell

In the cell, molecules known as proteins control most of the biological systems. Some proteins work alone. However, the majority of proteins present in cells interact with one another to facilitate certain biological activities. Major cellular functions like DNA replication, RNA processing, cellular metabolism, and Ion transport are all mediated by various protein-protein interactions known as PPI networks.

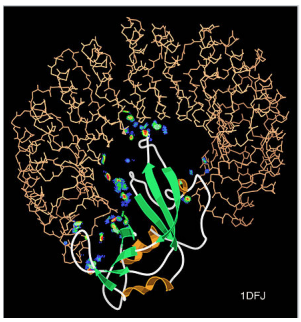

Figure 10: The protein-protein interaction between a ribonuclease inhibitor and a ribonuclease protein. [2]

Therefore, understanding these networks is crucial in understanding the cell activity in normal and diseased states. Understanding the interactions can be beneficial in developing drugs for specific diseases by targeting certain pathways within the cell. Using Graph Theory, the protein-protein interactions can be represented schematically as shown in Figure 11 below. An altered protein-protein interaction schematic can be seen in a diseased state as compared to the homeostatic state. This kind of study can be beneficial in understanding the disease related proteins and their role in the pathological progression of the disease. For example, the node shown in dark red in the figure below can be seen as disease related given its connectivity to the other nodes. A dark red node would be more involved in the disease pathology. However, the white colored nodes which are not directly connected to the dark red nodes are less likely to be involved in the disease pathology.

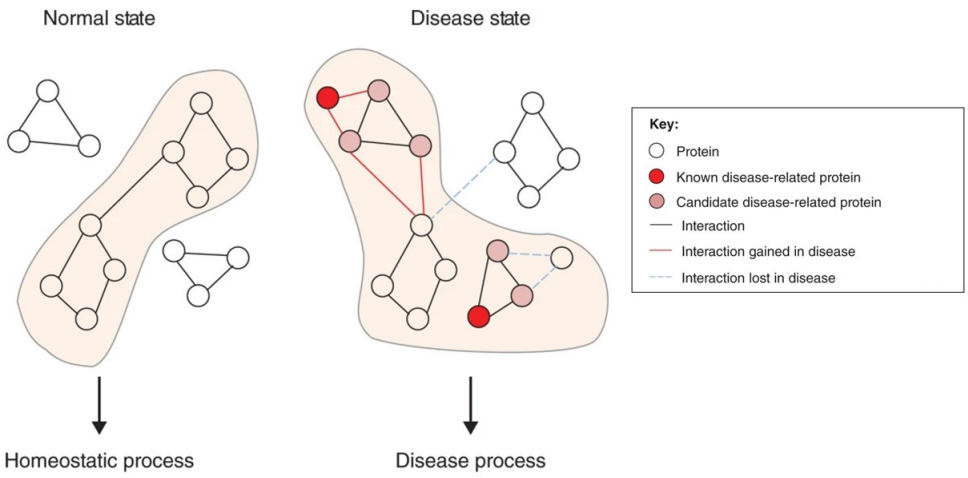

Figure 11: Alteration in the protein-protein interactions under pathological conditions. [3]

3.2 PPI in relation to COVID-19

SARS-CoV-19 (COVID-19) is an acute respiratory disease which is highly contagious and has spread worldwide, since December 2019 leading to an ongoing pandemic. Although vaccines have been developed recently against the disease, majority of the world's population remain unvaccinated. Although work is underway to develop drugs that inhibit the virus, the primary treatment is symptomatic [5]. For a disease like COVID-19, graph theory analysis of the protein-protein interactions could be an alternative approach to understanding the disease pathology in more depth.

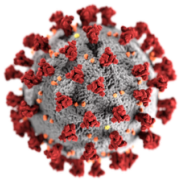

Figure 12: SARS-CoV-19 Virus. [5]

As with every virus, the disease pathology for COVID-19 is influenced by multiple protein-protein interaction pathways [4]. One such pathway is shown in the figure below [Figure 13]. This schematic represents the various PPI's within the RNA processing pathway that are involved in the COVID-19 disease pathology. One of the statistical ways to understand these interactions is by viewing at the MiST (mass spectrometry interaction statistics) score. The MiST score is a combination of abundance, reproducibility, and specificity with regards to protein-protein interactions. In other words, the higher the MiST score the stronger the relation between the proteins are.

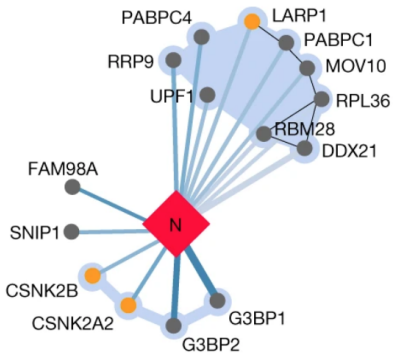

Figure 13: PPI Networks of the RNA Processing of a COVID-19 Protein

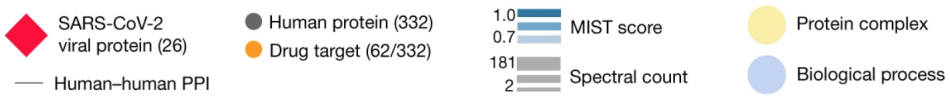

Figure 14: Key to Figure 13

4 Incidence Matrix Proposal for PPI Networks in COVID-19

With the above example [Figure 13], the interactions can also be represented in terms of graph theory. Here, the proteins can be represented by nodes and the interactions between them as edges. In addition this, the MiSt score can represent the weight of each edge. Knowing this, we can determine and analyze the incidence matrix of this graph. Creating an incidence graph would require us to label the edges of the graph as seen in the figure below.

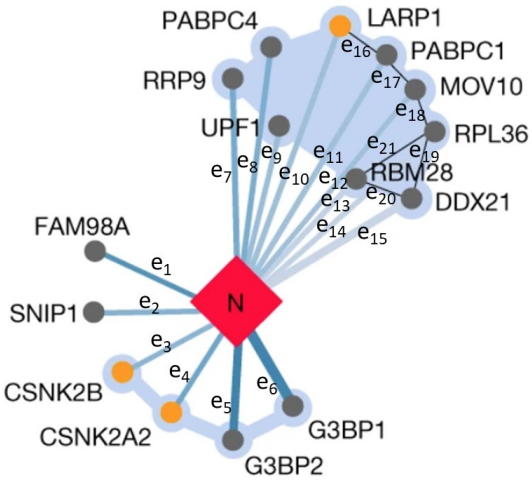

Figure 15: Labelled edges using Figure 13

In order to figure out which proteins have a significant and stronger relationship with one another, we must assign certain values to mean certain things. In this case, we can assume that a MiST score

of 0.7 and below represents a weak interaction. Therefore, we can eliminate those interactions from consideration. However, those with a MiST score above 0.7 can be considered a stronger interaction. From here, we can create an empty adjacency matrix where rows and columns represent the vertices and edges. If a vertex and edge are incident (in this case, have a MiST score above 0.7), then a 1 will be marked in their intersection in the matrix. On the other hand, if the MiST score suggests otherwise, then a 0 will be marked in the intersection within the matrix. An empty matrix with these conditions is seen in Figure 16.

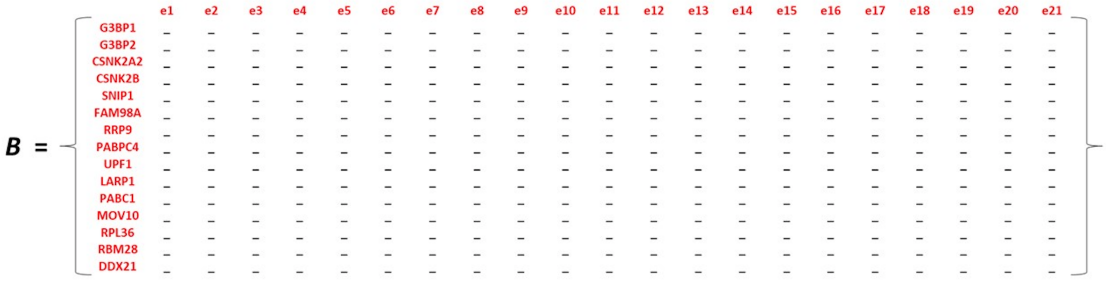

Figure 16: Empty Incidence Matrix for Figure 13

We can then input the values of 0 and 1 into the matrix using the rule in Concept 2. From here, we achieve a matrix shown in Figure 17.

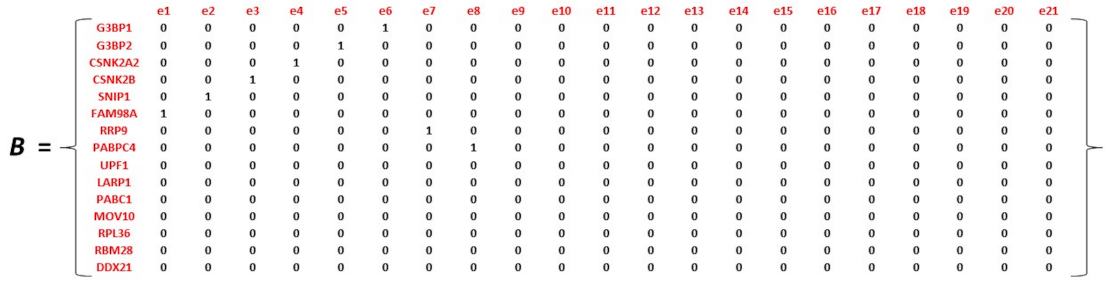

Figure 17: Incidence Matrix for the Figure 13

Here, we have organized the data in such a way that it is easy to visualize which proteins have significant effect on the disease pathology. For example, using the Graph Theory analysis, proteins G3BP1, G3BP2, CSNK2A2, CSNK2B, SNIP1, FAM98A, RRP9, and PABPC4 seem to have a significant impact on the disease pathology as compared to the other proteins analyzed. This can provide an insight on the drug targeting this RNA processing pathway for the treatment of the COVID-19 disease. This can also be expanded to study the other pathways mentioned in literature [4].

5 Conclusion

Graph Theory can be a powerful tool to analyze biological data to study the interaction of proteins to understand pathology of diseases such as COVID-19. The study presented in this paper can be further expanded to analyze other biological pathways involved in disease research. It can also be a useful tool to understand the disease mechanisms of other multi-factorial diseases like cancer.

6 References

[1] Chartrand, Gary and Zhang, Ping. *A First Course in Graph Theory.* Dover Publication, 2012.

[2] Wikipedia: Protein-protein Interactions. https://en.wikipedia.org/wiki/Protein

[3] Kuzmanov, U., Emili, A. Protein-protein interaction networks: probing disease mechanisms using model systems. Genome Med 5, 37 (2013). https://doi.org/10.1186/gm441

[4] Gordon, D.E., Jang, G.M., Bouhaddou, M. et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature 583, 459–468 (2020). https://doi.org/10.1038/s41586-020-2286-9

[5] Wikipedia: COVID-19. https://en.wikipedia.org/wiki/COVID-19

ACKNOWLEDGMENTS:

I would like to thank and acknowledge my mentor, Bex Nelson for her dedication, enthusiasm, and inspiration which helped me understand the concepts of Graph Theory and its applications. Additionally, I would like to thank my parents for their constant support and guidance.