International Journal of Basic and Applied Sciences, 14 (4) (2025) 194-198

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/01r2zb80
Research paper

Graph Theoretical Analysis of SARS-COV-2 Spike Protein

Akhil C. K. [1], Athul T. B. [2]*, Roy John [3], Manju N. V.[4], Shubha A. B. [5]

[1] Department of Mathematics, University College, Thiruvananthapuram, Kerala, India
[2] Department of Mathematics, Sree Narayana College, Punalur, Kollam, Kerala, India
[3] Department of Mathematics, St.Stephen's College, Pathanapuram, Kollam, Kerala, India
[4] Department of Mathematics, University of Kerala, Thiruvananthapuram, Kerala, India
[5] Department of General Science, Government Engineering College,
Bartonhill, Thiruvananthapuram, Kerala, India.
**Corresponding author E-mail: athultbacl@gmail.com*

Received: June 23, 2025, Accepted: August 3, 2025, Published: August 8, 2025

Abstract

The spike protein of the severe acute respiratory syndrome coronavirus-2 (SARS-CoV-2) is the main surface antigen of the coronavirus. The global pandemic that affected most countries and territories in 2019 was caused by the emergence of the SARS-CoV-2 spike protein. Graph theoretical analysis plays an important role in biological networks. First, we construct the Pt-graph of SARS-CoV-2 based on the physicochemical properties and adjacency of amino acids in the corresponding peptide/protein in SARS-CoV-2. By analyzing the of Pt-graph, we receive some observations about the relations among the amino acids, physiochemical properties of amino acids and peptide/protein, and it will help in the development of and drug design. In this study, we analyze the Pt-graph of the SARS-CoV-2 spike protein to examine its structural properties.

Keywords: *Amino acid, SARS-CoV-2 spike protein, Graph, Pt-graph, biological networks.*

1. Introduction

The coronavirus, later identified as SARS-CoV-2, was initially detected in Wuhan. December 2019, China. According to investigations, the infection could have started in bats and spread to people via a possible intermediary host, a wild animal that was sold at a Wuhan seafood market. Zoonotic diseases in which pathogens are prevalent spread from animals to people and cause infectious diseases. Likewise, outbreaks have been zoonotic. For example, the H1N1 influenza virus, the Middle East Respiratory Syndrome (MERS) coronavirus, and the Ebola virus. These viruses can adapt and spread by going through genetic reassortment or mutations. Researching the causes of these diseases is essential to avoid and control them. Testing for the origin of such diseases is essential for prevention and control. Ongoing international research aims to understand the facts that contribute to the initial transfer of SARS-CoV-2 to humans.

Proteins [4] are the most prevalent biological macromolecules, found in every cell and all cellular components. The fundamental units of proteins are known as amino acids, which are organic compounds characterized by the presence of at least one amino group (-NH2) and one carboxyl group (-COOH). The substances possess various physicochemical properties [1], including hydrophobicity, polarity, non-polarity, aliphaticity, hydrophilicity, aromaticity, and positive and negative charge. The prearrangement of amino acids in a protein is characteristic of that protein and is referred to as its key structure. Peptides and proteins are composed of amino acids, where the carboxyl group of one amino acid bonds with the amino group of another.

The precise types of peptides, called Neuropeptides, are synthesized and released by neurons. Typically found in axon terminals at synapses, they are categorized as potential neurotransmitters, although certain neuropeptides also function as hormones.

The SARS-CoV-2 spike protein, commonly referred to as the S protein, is a surface protein found on the SARS-CoV-2 virus responsible for COVID-19. This protein serves as a key target for neutralizing antibodies. Many vaccines are designed to elicit immune responses specifically against the spike protein. Changes in the spike protein can influence the virus's transmission, infectivity, and ability to evade the immune system. Variants of concern (VOCs) frequently exhibit mutations in the spike protein, such as those seen in the Alpha, Beta, Delta, and Omicron variants.

Graph theory [7] has numerous applications in the field of biology. The amino acid network [5] within the protein was examined by S. Kundu. A study conducted by Adil Akthar and Nisha Gohan explored the application of graph theory to the analysis of amino acid networks [1]. In 2014, Adil Akhtar and Tazid Ali [2] utilized centralities within amino acid networks. Utilizing the concept of amino acid networks. In [6], G. Suresh Singh and Akhil C. K introduces a novel type of graph known as the Pt-graph for peptides and proteins. In 2024, C. K. Akhil and G. Suresh Singh [3] analyze the pt graph of Endomorphin. In this paper, we conduct an analysis of the Pt-graph associated with the SARS-CoV-2 spike protein. Ultimately, we obtain insights regarding the relationships between amino acids, their physicochemical

Copyright © Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

properties, and peptides/proteins. This understanding may contribute to advancements in the evolution of peptides/proteins and the design of pharmaceuticals in the future.

1.1 Basic Concepts of Graph Theory

A graph G consists of a pair G = (V, E), where V is a non-empty set and E is the set of 2-element subsets of V. The members of V are termed as vertices, while the members of E are called edges. The collection V is termed the vertex set of G, and E is referred to as the edge set of G. Two vertices, u and v, are considered adjacent if there exists an edge connecting them, and two edges are deemed adjacent if they share a common vertex. For basic definitions and notations, we refer to [7].
In the field of graph theory, the centrality measure of a vertex indicates its vital role in the overall structure of the graph. Centrality is demarcated as a function on the vertices of a graph. That is, centrality is a function f assigns a value $f\ (v)\ \in\ \boldsymbol{R}$ to each vertex $v\ \in\ V$ in a graph G the degree of centrality is denoted by $c_d(u)$ and is defined as the number of vertices to which the vertex u is adjacent. The vertices that are connected to a given vertex u are also called the neighbours of u. Degree centrality demonstrates that an important vertex is involved in many communications. The contact gives the instant importance or peril of the vertex in the corresponding network. Mathematically, it is defined as

$$c_d(u)\ =\ deg(u)$$

However, in physical problems, the degree of centrality is not an authentic measurement for finding a significant vertex since it may be connected circuitously with other vertices. Degree centrality is used to identify important vertices in biological networks, such as protein interaction networks or food webs. Nodes with high degree centrality are often critical for maintaining the stability and functionality of the network.
The closeness centrality measures the idea about how a vertex is close to other vertices, not only to the first neighbour but also in a global gauge. Generally, a vertex is central means it is close to all other vertices. If a vertex is close to other vertices, then it can speedily relate to all other vertices. In general, closeness centrality is defined as the reciprocal of the sum of the shortest path distances between each pair of vertices and every other vertex in the network. The closeness centrality of a vertex represents a significant vertex that can easily reach or interconnect with another vertex of the network. Mathematically, it is defined as

$$c_{cl}(u) = \frac{n-1}{\sum_{v\in V} d(u,v)}$$

where n denotes the number of vertices of the network and $d(u,v)$ is the length of the shortest path between the vertices u and v. If a vertex has the minimum accumulative shortest path distance, then that vertex has maximum closeness centrality. And the maximum closeness centrality vertex is very well connected to all other vertices. Closeness centrality is used to identify individuals who are at high risk of being infected with a disease in a network. Individuals with high closeness centrality are often in close contact with many other individuals, making them more likely to spread the disease.
Another well-known centrality measure is the betweenness centrality. Betweenness centrality connections between two nonadjacent vertices depend on the other vertex, generally on those on the paths between the two. The betweenness centrality of a vertex is the number of shortest paths going through it. Mathematically, it is defined as

$$c_{btw}(u) = \sum_{s,t\neq u} \frac{\sigma_{st}(u)}{\sigma_{st}}$$

where σ_{st} is the number of the shortest (s,t) paths and $\sigma_{st}(u)$ is the number of the shortest (s,t) paths that pass through u. Betweenness centrality represents identifying the vertices that make the maximum information flow of the network. A significant vertex will lie on many paths between other vertices in the network. From this vertex, we can regulate the information of the network. Without these vertices, there would be no method for two neighbours to interconnect with each other. In general, the high degree vertex has high betweenness centrality. However, a high betweenness centrality vertex need not always be a high degree vertex. Also, end vertices have zero betweenness centrality. Betweenness centrality is used to identify individuals who are likely to be important for the spread of a disease in a network. Individuals with high betweenness centrality are often able to spread the disease quickly through the network, and targeting these individuals for interventions can be an effective strategy for controlling the spread of the disease. Overall, betweenness centrality is a useful measure that can provide valuable insights into the structure and dynamics of various types of networks, particularly those that involve transportation, communication, or disease spread.
For a square matrix A, is a value λ is called an eigenvalue for A if $det(A\ -\ \lambda I)\ =\ 0$ where I is an $n\times\ n$ identity matrix. Eigenvector centrality is the principal eigenvector of the adjacency matrix of the corresponding graph. In matrix vector notation, we can write

$$A(G)X\ =\ \lambda X$$

where $A(G)$ is the adjacency matrix of the graph is the eigenvalue and X is the corresponding eigenvector. In general, there will be different eigenvalues for $A(G)$. However, the eigenvector associated with the greatest eigenvalue is the eigenvector centrality. Eigenvector centrality gives the direct as well as indirect importance of a vertex in a network. Eigenvector centrality is used to identify proteins that are central to metabolic pathways in a biological network. Proteins with high eigenvector centrality are often critical for maintaining the stability and functionality of the network. Overall, eigenvector centrality is a versatile and widely used measure that can provide valuable insights into the structure and dynamics of various types of networks, particularly those that involve social, information, or financial networks.

2. P t-graph of Peptides/ Proteins

In the amino acid network [1], a vertex set is established as the complete collection of the twenty natural amino acids, while the edge set is characterized by the connections between pairs of amino acids that share at least one (or in some instances, two) common properties.

Proteins are polymers of amino acids, with each amino acid residue joined to its neighbour by a specific type of covalent bond. Twenty different types of natural amino acids are commonly found in peptides/proteins. The twenty natural amino acids and their abbreviations are shown in Table 1. The sequence of amino acids in a protein is characteristic of that protein and is called its primary structure. Peptides/proteins are the compounds of amino acids in which a carboxyl group of one is united with an amino group of another

Table 1: Twenty natural amino acids and their physico-chemical properties

Amino Acids Properties	G	A	V	M	W	L	I	F	P	Y	S	T	E	C	N	Q	D	K	H	R
Hydrophobic	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
Hydrophilic	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
Polar	0	0	0	0	1	0	0	0	0	1	1	1	1	1	1	1	1	1	1	1
Non-polar	1	1	1	1	0	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0
Aliphatic	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
Aromatic	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0
N(A)	1	1	0	1	0	0	0	0	1	0	1	1	1	1	1	1	1	1	0	1
(+)ve	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
(-)ve	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0
N(B)	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	0	0	0

Definition 3.1. [6] A Pt-graph is characterized as a graph G = (V, E) representing a peptide or protein, where the vertex set V comprises all distinct amino acids found within the peptide or protein. The weight of a vertex in G corresponds to the frequency of its occurrence in the peptide or protein sequence. Two vertices are considered adjacent in G if they are consecutive elements in the sequence and share at least one common physicochemical property.

Remark 3.2. [6] The weight assigned to a vertex reflects the frequency with which a specific amino acid appears within a sequence. A higher weight for a vertex in a Pt-graph indicates that more significant characteristics can be associated with that amino acid about the peptide or protein.

Remark 3.3. [6] The centrality measures of a Pt-graph assist in determining the number of associated amino acids, either directly or indirectly, with neighbouring amino acids within the sequence of the relevant peptide or protein that possesses at least one shared physicochemical property

3. Graph Theoretical analysis of Pt-graph of SARS-CoV-2

The sequence of the SARS-CoV-2 spike protein is

VYYHKNNKSWMESEFRVYSSANNCTFEVSQPFLMDLEGK QGNFKNLREFVFKNIDG YFK IYSKH TPI NV R DLPQGFSA
LEPLVDLPIGINITRFQTLLALHRSYLTPGDSSSGWTAGAAAYYVGYLQPRFLLKYNENGTITDAVDCALDPLSETKCTLKSF
T VEK GIYQTSNFRVQPTESIVRFPNITNLCPFGEVFNATRFASVYAWNRKRISNCVADYSVL YNSASFSTFKCYGV SPTKLN
DLCFTNVYADSFVIRGDEVRQIAPGQTGKIADYNYKLPDDFTGCVIAWNSNNLDSKVGGNYNYLYRLFRKSNLKPFERDIST
EIYQAGSTPCNGVEGFNCYFPLQSYGFQPTNGVGYQPYRVVVLSFELLHAPATVCGPKK STNLVKNKCVNF NFNGLTGTG
VLTESNKKFLPFQQFGR DIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNT SNQVAVL YQDVNCTEVPVAIHADQLTPTW
RVYSTGSNVFQTRAGCLIGAEHVNNSYECDIPIGAGICASYQTQTNSPRRARSVASQSIIAYTMSLGAENSVAYSNNSIAIPTNF
TISVTTEILPVSMTKTSVDCTMYICGDSTECSNLLLQYGSFCTQLNRALTGIAVEQDKNTQEVFAQVKQIYKTPPIKDFGGFNFS
QILPDPSKPSKRSFIEDLLFNKVTLADAGFIKQYGDCLGDIAARDLICAQKFNGLTVLPPLLTDEMIAQYTSALLAGTITSGWTF
GAGAALQIPFAMQMAYRFNGIGVTQNVLYENQKLIANQFNSAIGKIQDSLSSTASALGKLQDVVNQNAQALNTLVKQLS SNF
GAISSVLNDILSRLDKVEAEVQIDRLITGRLQSLQTYVTQQLIRAAEIRASANLAATKSECVLGQSKRVDFCGKGYHLMSFPQS
APHGVVFLHVTYVPAQEKNFTTAPAICHDGKAHFPREGVFVSNGTHWFVTQRNFYEPQIITTDNTFVSGNCDVVIGIVNNTVY
DPLQPELDSFKEELDKYFKNHTSPDVDLGDISGINASVVNIQKEIDRLNEVAKNLNESLIDLQELGKYEQYIKWPWYIWLGFI
AGLIAIVMVTIMLCCMTSCCSCLKGCCSCGSCCKFDEDDSEPVLKGVK LHYT

For the construction of Pt-graph [4] G of SARS-CoV-2 spike protein, the vertex set is taken as the collection of different amino acids presented in the sequence, and the weight of a vertex in G is the number of times it appears in the sequence. Also, two vertices are adjacent in G if they are consecutive elements in the sequence (in other words they have a peptide bond between them) and have at least one common physicochemical property. Here Alanine (A), Arginine (R), Aspargine (N), Aspartic Acid (D), Cystein (C), Glutamic Acid (E), Glutamine (Q), Glycine (G), Histidine (H), Isoleucine (I), Leucine (L), Lysine (K), Methionine (M), Phenylalanine (F), Proline (P), Serine (S), Thronine (T), Tryptophan (W), Tyrosine (Y), and Valine (V) are the 20 amino acids presented in the sequence of SARS-CoV-2 spike protein. Then the vertex set for the Pt-graph will be

$$V = \{M_{14}, F_{77}, V_{97}, L_{108}, P_{57}, S_{99}, Q_{62}, C_{40}, N_{88}, T_{97}, R_{42}, A_{79}, Y_{54}, G_{82}, D_{62}, K_{61}, H_{17}, W_{12}, I_{76}, E_{48}\}$$

The Pt-graph of SARS-CoV-2 spike protein is shown in Figure 1

Next, we calculate the degree centrality, closeness centrality, betweenness centrality and eigen- vector centrality) for the vertices of Pt-graph and is shown in Table 2. These measures are calculated using the open softer ware SageMath. From the analysis of Pt-graph of SARS-CoV-2 spike protein, we have some simple observations and are given below.

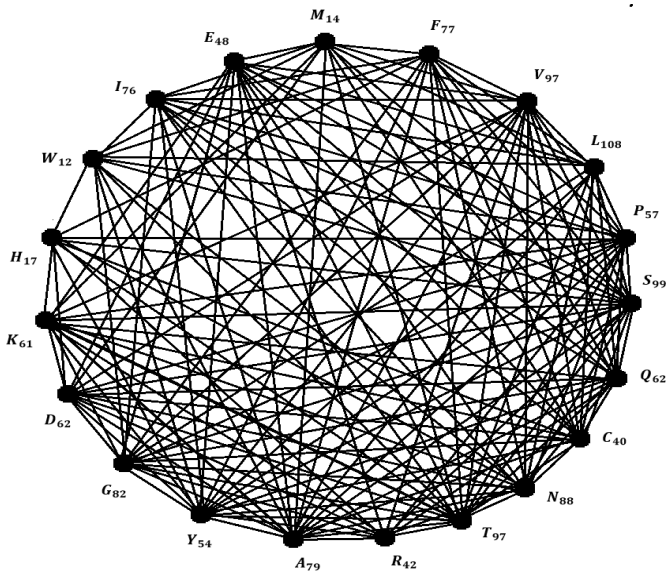

Fig. 1: The Pt-graph of SARS-CoV-2 spike protein

Table 2: The centrality measures for each vertex of the Pt-graph of SARS-CoV-2.

Vertex	Degree	Closeness Centrality	Eigenvector Centrality	Betweenness Centrality
M_{14}	13	0.760000	0.181460	0.0074871522
F_{77}	15	0.826087	0.210592	0.0078802289
L_{108}	15	0.826087	0.214170	0.0049743953
V_{97}	17	0.904762	0.240164	0.0087868109
S_{99}	19	1.000000	0.259071	0.0181977542
Q_{62}	17	0.904762	0.240164	0.0087868109
C_{40}	17	0.904762	0.238222	0.0109666487
N_{88}	18	0.950000	0.248464	0.0148319679
T_{97}	19	1.000000	0.259071	0.0181977542
P_{57}	18	0.950000	0.248464	0.0148319679
A_{79}	18	0.950000	0.250019	0.0130177165
Y_{54}	18	0.950000	0.248464	0.0148319679
R_{42}	13	0.760000	0.188074	0.0028826274
G_{82}	17	0.904762	0.239412	0.0104873522
D62	15	0.826087	0.212606	0.0066982563
K61	15	0.826087	0.212606	0.0066982563
H17	11	0.703704	0.154848	0.0045271655
W12	12	0.730769	0.168590	0.0051600446
I_{76}	14	0.791667	0.201540	0.0033499639
E_{48}	15	0.826087	0.216074	0.0045396611

Observation 4.1. The Pt-graph of SARS-CoV-2 is a connected graph.
Observation 4.2. The average degree of the Pt-graph of SARS-CoV-2 is 15.8.
Observation 4.3. Hydrophilic polar amino acids, Serine (S) and Threonine (T), receive the highest value of all centrality measures.
Observation 4.4. The hydrophilic, polar, aromatic, and positively charged amino acid, Histidine (H), receives the lowest value of all centrality measures.

4. Biological implications

The interconnectedness of the Pt-graph about the spike protein of SARS-CoV-2 demonstrates a robust relationship among the amino acids, suggesting that each amino acid is either directly or indirectly linked within the neuropeptide. The hydrophilic polar amino acids, Serine (S) and Threonine (T), receive the highest value of all centrality measures, and they are the most prominent amino acids in the neuropeptide. The hydrophilic, polar, aromatic, and positively charged amino acid, Histidine (H), receives the lowest value of all centrality measures, and they are not directly connected within the graph.

High-centrality amino acids such as serine and threonine play biologically significant roles in the spike protein due to their structural and functional importance. Their centrality in graph-theoretic models—where amino acids are treated as nodes and their interactions as edges—indicates that they act as crucial connectors or hubs within the protein structure. Serine and threonine possess hydroxyl groups, making them common sites for post-translational modifications like phosphorylation and O-linked glycosylation. These modifications are particularly important in the spike protein, where glycosylation can shield the virus from the host immune system and assist in host cell recognition. High-centrality residues are often found in or near functionally critical regions such as the receptor binding domain (RBD), where they influence conformational changes necessary for binding to the ACE2 receptor. Moreover, due to their central roles in maintaining protein structure and function, these residues are typically conserved across viral strains, and mutations at these positions can lead to significant changes in viral infectivity, stability, and immune evasion capabilities.

5. Conclusion

In this article, we discussed the Pt graph and centrality measurements of the SARS-Cov-2 spike protein. This establishes a relationship between amino acids, indicating that all amino acids are connected directly or indirectly within the neuropeptide. Centrality measurements of Pt-graph help to identify the ratio of amino acids directly or indirectly to adjacent amino acids in a suitable peptide/protein sequence with at least one general physicochemical property. The extent of the separation vertices of the Pt-graph results in the direct meaning of the vertex and the number of adjacent amino acids within a sequence with at least one general physicochemical property. The proximity of Pt-graph gives the idea that relationships between amino acids are close to other amino acids within the peptide/protein. The betweenness centrality in the Pt graph refers to the number of shortest ways to pass through the vertices, with the highest value contributing to the importance of peptide/protein amino acids. The eigen vector centrality of the Pt-graph indicates the direct and indirect importance of amino acids in the peptide/protein.

Acknowledgement

We would like to express our profound gratitude to J. Jayakumar, Associate Professor of Zoology (Rted), Govt. College for Women, Thiruvananthapuram, for his support. Also, we express our gratitude towards Dr. G. Suresh Singh, Senior Professor (Retired), University of Kerala, Thiruvananthapuram, Kerala, for his guidance and support.

References

[1] Akhtar, A., & Gohain, N., "Graph theoretic approach to analyze amino acid network", International Journal of Advances in Applied Mathematics and Mechanics, vol. 2, no. 3, p.31-37, 2015.
[2] Akhtar, A., and Ali, T., "Analysis of Unweighted Amino Acids Network", International scholarly research notices, vol. 16, p.350276, 2014,. https://doi.org/10.1155/2014/350276
[3] Akhil, C.K. & Singh, G.S., "Graph Theoretical Analysis of Endomorphin", In: Balasubramaniam, P., Raveendran, P., Mahadevan, G., Ratnavelu, K. (Eds) Discrete Mathematics and Mathematical Modelling in the Digital Era. ICDMMMDE 2023. Springer Proceedings in Mathematics & Statistics, vol.458, 2014. Springer. https://doi.org/10.1007/978-981-97-2640-0_13
[4] David L. Nelson & Michael M. Cox, Lehninger Principles of Biochemistry, 2008.
[5] Kundu, S, Amino acid network within protein. Physica A: Statistical Mechanics and its Applications", vol. 346, p.104-109, 2005 https://doi.org/10.1016/j.physa.2004.08.055
[6] Singh, G.S., & Akhil, C.K., "Analysing Amino Acids in Galanin- Graph Theoretic Approach", International Journal of Recent and Innovation Trends in Computing and Communication, vol. 5, no. 5, 2017. https://doi.org/10.17762/ijritcc.v5i6.774
[7] Singh, G.S., Graph Theory. PHI Learning Private Limited, 2010
[8] Wang, Y., Wang, M., Yin, S., Jang, R., Wang, J., Xue, Z & Xu, T, NeuroPep: a comprehensive resource of neuropeptides. Database: the journal of biological databases and curation, 2015, bav038. https://doi.org/10.1093/database/bav038