Article

A study on Nilpotent graph in genetic code algebra

Birinchi Kumar Boruah
Department of Mathematics, Sibsagar University, Sivasagar, Assam 785665, India
E-mail:birinchikumar49@gmail.com

Received 08 May 2025; *Accepted 28 May 2025*; *Published online 30 May 2025*; *Published 1 December 2025*

Abstract
The genetic code is a set of codons that contains genetic information regarding the creation of protein molecules. We studied nilpotent graphs in genetic code algebra in this work. The vertex set is the set of all non-nilpotent elements of a ring, and two vertices are neighbouring if and only if their sum is nilpotent. Different measurements of centrality have been thoroughly examined in our current paper. We also investigated three network parameters: clustering coefficient, degree of dispersion, and skewness.

Keywords amino acid; centrality measure; correlation coefficient; clustering coefficient; degree of distribution; genetic code; Nilpotent graph.

Network Biology
ISSN 2220-8879
URL: http://www.iaees.org/publications/journals/nb/online-version.asp
RSS: http://www.iaees.org/publications/journals/nb/rss.xml
E-mail: networkbiology@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

The genetic code is the set of principles that govern the translation of information stored in genetic material (DNA or RNA sequences) into proteins. Proteins are the fundamental building blocks and functional components of all living creatures. Proteins are made up of amino acids, and each protein is made up of a linear chain of amino acids. There are now 20 distinct amino acids identified in proteins. Each amino acid is a triplet code (codon) comprising four potential bases: A, C, G, and T in DNA (T (Thymine) is substituted by U (Uracil) in RNA). The chain of amino acids takes on diverse forms to generate different proteins. The process of transmitting information from DNA to protein construction involves two steps: transcription and translation. The sequence of nucleotides is not properly duplicated while reproducing the strand of DNA due to mutation. This alters protein formation. Codons may undergo several types of alterations, including point mutations, deletions, insertions, and inversions. We shall solely look at point mutation in this study. A point mutation is characterised by a single base change in the gene sequence. It substitutes one base nucleotide with another in the genetic material, DNA or RNA. This mutation might occur in a single point, two points, or more. Again, transition mutation refers to a point mutation from purine (A, G) to purine or pyrimidine (C, T) to pyrimidine, whereas transversion mutation refers to a point mutation from purine to pyrimidine or vice versa.

Point mutations typically occur during DNA replication.

The genetic code in DNA (or RNA) is made up of 64 codons, however there are only 20 amino acids. This suggests that there is some overlap, i.e., many codons code for the same amino acid. Synonymous codons are codons that code for the same amino acid. This may be viewed as a function of many to one when it comes to conveying codons to amino acids. As a result, it's worth investigating if the genetic code has any mathematical feature that optimises as the number of codons approaches three times the number of amino acids (Rudin et al., 2011). The mistake (acceptable mutation) frequency discovered in codons indicates the relevance of base position. From the third base to the first base, then to the second base, the number of mistakes falls. The second base is the most physiologically important, whereas the third is the least important codon.

Sanchez et al. discovered an optimum codon arrangement of the genetic code (Bagler and Sinha, 2007). Furthermore, the codons are organised in such a way that some key links between codons and amino acids are highlighted. Different formal mathematical models of the genetic code have been developed for this aim.

Many researchers studied in this topic and attempted to provide an algebraic formulation of the genetic code's structure (Crick et al. 1961; Bonacich, 1972; Wong, 1975; Bertholf and Walls, 1978; Kimura, 1981; Osawa et al., 1992; Watts and Strogatz, 1998; Szathmáry, 1999; Balakrishnan, 2002; Newman, 2002; Wuchty and Stadler, 2003; Kundu, 2005; Sanchez et al., 2005; Aftabuddin and Kundu, 2007; Bagler and Sinha, 2007; Koonin and Novozhilov, 2009; Sengupta and Kundu, 2012; Akhtar and Ali, 2014; Akhtar et al., 2015; Koonin and Frozen, 2017; Ali and Bora, 2021; Boruah and Ali, 2022). Hornos and Hornos were the first to use group theoretical approaches to investigate genetic coding (Kundu, 2005). Sanchez et al. proposed a novel approach for quantifying the link between DNA genomic sequences (Bagler and Sinha, 2007; Bertholf and Walls, 1978). They stated that genetic code developed to reduce the consequences of transcription and translation mistakes. Ali and Borah (2021) and Gohain et al. worked in the same topic and described various algebraic and topological structures of the genetic code (Akhtar et al., 2015). Some interesting correlations between algebraic and biological features have been discovered using these structures. $G(R)$ is the nilpotent graph of a ring R, where two unique vertices x and y are near if $x + y$ is nilpotent, where $Nil(R)$ is the set of nilpotent elements of a ring R. The primary objective for this effort is to investigate mathematical structures, namely graph structures that may naturally emerge in genetic code.

2 Initial Graph Concepts

A graph is an ordered triple $G = (V(G), E(G), I_G)$, where $V(G)$ is a nonempty set, $E(G)$ is a set disjoint from $V(G)$, and I_G is an "incidence" relation that associates with each element of $E(G)$ an unordered pair of elements (same or distinct) of $V(G)$ (Zhang, 2016, 2018). Elements of $V(G)$ are called the vertices (or nodes or points) of G, and elements of $E(G)$ are called the edges (or lines) of G. $V(G)$ and $E(G)$ are the vertex set and edge set of G, respectively. If for the edge e of G, $I_G(e) = \{u, v\}$, we write $I_G(e) = uv$.

If $I_G(e) = \{u, v\}$, then the vertices u and v are called the end vertices or ends of the edge e. Each edge is said to join its ends & we say that e is incident with each one of its ends. Also, the vertices $u\&v$ are then incident with e. A vertex u is a neighbor of v in G if uv is an edge of G and u is not equal to v. A walk in a graph G is an alternating sequence $W: v_0\, e_1\, v_1\, e_2\, v_2\, \ldots\ldots\ldots e_p v_p$ of vertices and edges beginning and ending with vertices in which v_{i-1} and v_i are the ends of e_i. A walk is called a trail if all the edges appearing in the walk are distinct. It is called a path if all the vertices are distinct. Two vertices u and v of G are said to be connected if there is a $u - v$ path in G, otherwise it is disconnected. Let G be a graph of order n with vertex set $V = \{v_1, \ldots\ldots\ldots, v_n\}$. The adjacency matrix of G is the $n \times n$ matrix $A = (a_{ij})$, where $a_{ij} = 1$ if there is an edge from vertex v_i to vertex v_j and $a_{ij} = 0$ otherwise.

The centrality measure of a vertex in graph theory shows its relative significance inside the graph G

(Zhang, 2016, 2018). It is a real valued function $f: V \to R$, where V is the vertex set of the graph G.

Degree centrality

The number of nodes to which a node u is directly linked, indicated by $C_d(u)$, is the degree centrality of that node (Zhang, 2016, 2018).

Eigenvector centrality

The eigenvector centrality of the associated graph is the eigenvector of the largest eigenvalue of the adjacency matrix (Bonacich, 1972).

Betweenness centrality

Betweenness centrality (Watts DJ, Strogatz, 1998) of a node v is defined as-

$$C_{btw}(v) = \sum_{m \neq v \in V} \sum_{n \neq v \in V} \frac{\sigma_{mn}(v)}{\sigma_{mn}}$$

Where, σ_{mn} and $\sigma_{mn}(v)$ are the number of shortest paths from vertex m to n and the number of shortest paths from m to n that pass through v (Watts DJ, Strogatz, 1998; Zhang, 2016, 2018).

Closeness centrality

Closeness centrality is defined as follows-

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

Where, n and $d(u,v)$are the total number of nodes of the network and shortest path distance between u and v (Zhang, 2016, 2018).

3 Graph on Genetic Code

Sanchez et al. (2005) discovered that the four RNA (or DNA) bases may be organised or ordered based on their codon-anticodon relationships. The hydrogen bond number and chemical type (purine and pyrimidine) of bases are crucial factors in this. Two ordering of the basis sets are obtained: $\{A, C, G, U\}$ and $\{U, G, C, A\}$. They studied the sum operation of the bases produced from the aforementioned two alternative ordering (A, C, G, U and U, G, C, A), which makes the two sets isomorphic to Z_4. Table 1 shows the sum operation of the bases produced from the two-alternative ordering. They also established the sum and product operations on the set of codons. It was observed that the group obtained on the set of codons is isomorphic to the group of integer module 64, $(Z_{64}, +)$. These two sum & product operations on the set of whole codons represents a ring structure isomorphic to the ring of $(Z_{64}, +, .)$. In 2015 Akhtar et al. discussed the total graph of this ring structure Z_{64}. By taking the ordered base set $\{A, C, G, U\}$ isomorphic Z_4 from Sanchez et al., 2005, Boruah et al. (2022) observe that $Z_4 \times Z_4 \times Z_4$ forms a ring structure isomorphic to the ring of $(Z_4 \times Z_4 \times Z_4, +, .)$. In this paper we have discussed nilpotent graph by taking base set $\{A, C, G, U\}$ and ring $(Z_4 \times Z_4 \times Z_4, +, .)$ and $(Z_{64}, +, .)$.

G_1= {CCC, CCU, CUC, CUU, UCC, UCU, UUC, UUU}

G_2= {AAC, AAU, CAC, CAU, GAC, GAU, UAC, UAU, ACC, ACU, CCC, CCU, GCC, GCU, UCC, UCU, AGC, AGU, CGC, CGU, GGC, GGU, UGC, UGU, AUC, AUU, CUC, CUU, GUC, GUU, UUC, UUU}

Table 1 Sum operation on $\{A, C, G, U\}$&$\{U, G, C, A\}$.

+	A	C	G	U
A	A	C	G	U
C	C	G	U	A
G	G	U	A	C
U	U	A	C	G

+	U	G	C	A
U	U	G	C	A
G	G	C	A	U
C	C	A	U	G
A	A	U	G	C

Table 2 Product operation on $\{A, C, G, U\}$.

.	A	C	G	U
A	A	A	A	A
C	A	C	G	U
G	A	G	A	G
U	A	U	G	C

Table 3 The genetic code table induced by the order {A, C, G, U}.

A			C			G			U			
No		Codon	No		Codon	No		Codon	No	Codon	AminoAcid	
AminoAcid			AminoAcid			AminoAcid						
A 000	AAA	K	010	ACA	T	020	AGA	R	030	AUA	I	A
001	AAC	N	011	ACC	T	021	AGC	S	031	AUC	I	C
002	AAG	K	012	ACG	T	022	AGG	R	032	AUG	M	G
003	AAU	N	013	ACU	T	023	AGU	S	033	AUU	I	U
C 100	CAA	Q	110	CCA	P	120	CGA	R	130	CUA	L	A
101	CAC	H	111	CCC	P	121	CGC	R	131	CUC	L	C
102	CAG	Q	112	CCG	P	122	CGG	R	132	CUG	L	G
103	CAU	H	113	CCU	P	123	CGU	R	133	CUU	L	U
G 200	GAA	E	210	GCA	A	220	GGA	G	230	GUA	V	A
201	GAC	D	211	GCC	A	221	GGC	G	231	GUC	V	C
202	GAG	E	212	GCG	A	222	GGG	G	232	GUG	V	G
203	GAU	D	213	GCU	A	223	GGU	G	233	GUU	V	U
U 300	UAA	-	310	UCA	S	320	UGA	-	330	UUA	L	A
301	UAC	Y	311	UCC	S	321	UGC	C	331	UUC	F	C
302	UAG	-	312	UCG	S	322	UGG	W	332	UUG	L	G
303	UAU	Y	313	UCU	S	323	UGU	C	333	UUU	F	U

Table 4 The genetic code table induced by the order {A, C, G, U}.

A			C			G			U			
No	Codon	AminoAcid	No	Codon	AminoAcid	No	Codon	AminoAcid	No	Codon	AminoAcid	
A0	AAA	K	16	ACA	T	32	AGA	R	48	AUA	I	A
1	AAC	N	17	ACC	T	33	AGC	S	49	AUC	I	C
2	AAG	K	18	ACG	T	34	AGG	R	50	AUG	M	G
3	AAU	N	19	ACU	T	35	AGU	S	51	AUU	I	U
C4	CAA	Q	20	CCA	P	36	CGA	R	52	CUA	L	A
5	CAC	H	21	CCC	P	37	CGC	R	53	CUC	L	C
6	CAG	Q	22	CCG	P	38	CGG	R	54	CUG	L	G
7	CAU	H	23	CCU	P	39	CGU	R	55	CUU	L	U
G	8	GAA E	24	GCA	A	40	GGA	G	56	GUA	V	A
9	GAC	D	25	GCC	A	41	GGC	G	57	GUC	V	C
10	GAG	E	26	GCG	A	42	GGG	G	58	GUG	V	G
11	GAU	D	27	GCU	A	43	GGU	G	59	GUU	V	U
U	12	UAA -	28	UCA	S	44	UGA	-	60	UUA	L	A
13	UAC	Y	29	UCC	S	45	UGC	C	61	UUC	F	C
14	UAG	-	30	UCG	S	46	UGG	W	62	UUG	L	G
15	UAU	Y	31	UCU	S	47	UGU	C	63	UUU	F	U

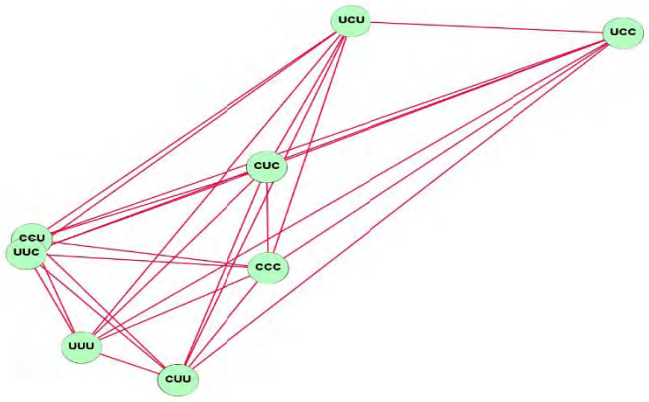

Fig. 1 Nilpotent graph of $Z_4 \times Z_4 \times Z_4$ (G_1).

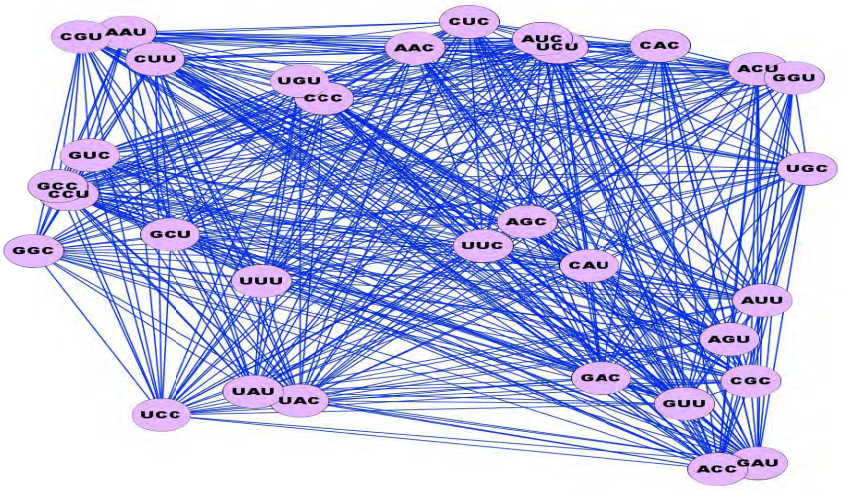

Fig. 2 Nilpotent graph of Z_{64} (G_2).

We can define a function u from the set of vertices of G_1 to the set of vertices to G_2 given by $u: G_1 \to G_2$ such that

$$u(LMN) = \begin{cases} LMC \ if \ N = U \\ LMU \ if \ N = C \end{cases} \quad \forall \ LMN \in G_1$$

Here it is clear that the function u is one-one mapping. From a biological point of view, it is recognized that this function represents the transition mutation of the third base of a codon. Therefore, we can conclude that the transition of the third base of codons can be represented in terms of nilpotent graph of the genetic code. Also, from our defined function it is clear that the transition of the third base of all codons gives a one-one map. Also, from the function $u: G_1 \to G_2$, we observed that for any codon whose third base is pyrimidine, then under the function it changes to pyrimidine.

4 Centralities in Nilpotent graph

Different measures of centrality have been calculated to analyze Nilpotent graphs (Fig. 1 & Fig. 2). Table 5 and Table 6, gives the different centrality values for the codons.

Table 5 Centrality values of codons for G_1.

Vertex	Degree Centrality (C_d)	Closeness Centrality (C_{cl})	Betweenness Centrality (C_{bwt})	Eigenvector Centrality (C_λ)
CCC	7	1	0	1
CCU	7	1	0	1
CUC	7	1	0	1
CUU	7	1	0	1
UCC	7	1	0	1
UCU	7	1	0	1
UUC	7	1	0	1

UUU	7	1	0	1

Table 6 Centrality values of codons for G_2.

Vertex	Degree Centrality (C_d)	Closeness Centrality (C_{cl})	Betweenness Centrality (C_{bwt})	Eigenvector Centrality (C_λ)
AAC	31	1	0	1
AAU	31	1	0	1
CAC	31	1	0	1
CAU	31	1	0	1
GAC	31	1	0	1
GAU	31	1	0	1
UAC	31	1	0	1
UAU	31	1	0	1
ACC	31	1	0	1
ACU	31	1	0	1
CCC	31	1	0	1
CCU	31	1	0	1
GCC	31	1	0	1
GCU	31	1	0	1
UCC	31	1	0	1
UCU	31	1	0	1
AGC	31	1	0	1
AGU	31	1	0	1
CGC	31	1	0	1
CGU	31	1	0	1
GGC	31	1	0	1
GGU	31	1	0	1
UGC	31	1	0	1
UGU	31	1	0	1
AUC	31	1	0	1
AUU	31	1	0	1
CUC	31	1	0	1
CUU	31	1	0	1
GUC	31	1	0	1
GUU	31	1	0	1
UUC	31	1	0	1
UUU	31	1	0	1

In a nilpotent network, centrality parameters emphasize the degree of similarity or closeness of one node to another or to its neighbors. The average number of connections a node has with other nodes in the network

is referred to as degree centrality. The greater the value of degree centrality, the greater the number of connections. After examining all of the codons (Table 5 and Table 6), it is discovered that all codons have the highest value in all centrality measures. All codons are linked to every other codon. As a result, all codons have the minimum cumulative shortest path distance and a high closeness centrality value. All of the observations show that the degree of the all codons are high, and thus the betweenness centralities are also high. Furthermore, the sum of the direct and indirect links of all the codons is found to be the greatest. As a result, eigenvector centrality is greatest in all codons. The shorter the distance between two nodes, the stronger the connection, and the higher the value of betweenness centrality, the greater the influence in the network. As a result, these codons have a greater evolutionary contribution (Chakrabarty and Parekh, 2014). The codons with the highest eigenvector centrality are said to play an important role in the evolutionary process.

5 Network Parameters

To analyse biological networks, we employ various network parameters. In the following sections, we will look at a few of them in order to interpret the network's communication pattern.

5.1 Codon clustering coefficients

Clustering coefficient is defined as the capacity of a graph to be divided into clusters. Clusters are a subset of the set that includes edges that connect vertices to vertices. The clustering coefficient C_i of a specific node $'i'$ is defined as the ratio of the total number of links e_i of neighbours to its nearest neighbours. The average clustering coefficient for the entire network is $C_i(C_i = \frac{2e_i}{K_i(K_i-1)})$, where $K_i$is the degree of node $'i'$ and $0 \leq C_i \leq 1$). The relationships between neighbouring nodes become stronger as the value of the clustering coefficient increases. As a result, it slows the spread of information (Sengupta and Kundu, 2012). Clustering coefficients for G_1 and G_2 are given in Table 7 and 8, respectively.

Table 7 Clustering coefficient of the codons for G_1.

CCC	**1**
$\boldsymbol{CCU}$	1
$\boldsymbol{CUC}$	1
$\boldsymbol{CUU}$	1
$\boldsymbol{UCC}$	1
$\boldsymbol{UCU}$	1
$\boldsymbol{UUC}$	1
UUU	1

Table 8 Clustering coefficient of the codons for G_2.

AAC	**1**
$\boldsymbol{AAU}$	1
$\boldsymbol{CAC}$	1
$\boldsymbol{CAU}$	1
$\boldsymbol{GAC}$	1
$\boldsymbol{GAU}$	1
$\boldsymbol{UAC}$	1

UAU	1
ACC	1
ACU	1
CCC	1
CCU	1
GCC	1
GCU	1
UCC	1
UCU	1
AGC	1
AGU	1
CGC	1
CGU	1
GGC	1
GGU	1
UGC	1
UGU	1
AUC	1
AUU	1
CUC	1
CUU	1
GUC	1
GUU	1
UUC	1
UUU	1

The clustering coefficient of the amino acid depends on the degree of amino acid as well as the number of direct interactions between the neighbouring amino acids. For the network G_1 and G_2, we observe that all the codons have a high clustering coefficient value of 1. The whole networks have a clustering coefficient value of 1, which is same as all the codons. The clustering coefficient is getting higher with the higher number of links between neighbours. So, the higher clustering coefficient values of the network slow down the flow of evolutionary messages. From the clustering coefficient of the whole networks and the clustering coefficients of the codons, we can say that the evolutionary mechanism is comparatively slow in the neighbourhood of all codons in comparison to the whole network.

5.2 Degree of distribution & skewness

In this section we shall discuss the degree of distribution & Pearson's skewness of the codons. The degree distribution $P(k)$ is actually the fraction of nodes with degree k. If we have n nodes with n_k number of nodes having degree k, then $P(k) = \frac{n_k}{n}$. In general, the degree distribution represents the probability that a chosen node will have accurately k links.

Another important statistical parameter is skewness. Skewness is defined with the measure of symmetry or asymmetry of the distribution. Skewness idea was first introduced by Karl Pearson in 1895. It's denoted as S_k. Depending upon mean and median, skewness may be positive or negative. In our study, we have used the Karl Pearson's coefficient of skewness, defined as

$$S_k = \frac{3(\text{Mean} - \text{Median})}{\text{Standard deviation}}, -3 \leq S_k \leq 3$$

For symmetrical (i.e., normal) distribution $S_k = 0$. If $S_k > 0$, then it's positively skewed. If $S_k < 0$, then we consider negatively skewed.

Table 9 and Table 10, shows the degree of distribution values of different codons for G_1 and G_2. From Table 9 and Table 10, Pearson's coefficient of skewness is found to be 0. The zero-value led us conclude that the degree of distribution of the codons (for G_1 and G_2) are symmetrical distribution.

Table 9 Degree distribution of the codons for G_1.

CCC	**1**
$\boldsymbol{CCU}$	1
$\boldsymbol{CUC}$	1
$\boldsymbol{CUU}$	1
$\boldsymbol{UCC}$	1
$\boldsymbol{UCU}$	1
$\boldsymbol{UUC}$	1
UUU	1

Table 10 Degree distribution of the codons for G_2.

AAC	**1**
$\boldsymbol{AAU}$	1
$\boldsymbol{CAC}$	1
$\boldsymbol{CAU}$	1
$\boldsymbol{GAC}$	1
$\boldsymbol{GAU}$	1
$\boldsymbol{UAC}$	1
$\boldsymbol{UAU}$	1
$\boldsymbol{ACC}$	1
$\boldsymbol{ACU}$	1
$\boldsymbol{CCC}$	1
$\boldsymbol{CCU}$	1
$\boldsymbol{GCC}$	1
$\boldsymbol{GCU}$	1
$\boldsymbol{UCC}$	1
$\boldsymbol{UCU}$	1
$\boldsymbol{AGC}$	1
$\boldsymbol{AGU}$	1
$\boldsymbol{CGC}$	1
$\boldsymbol{CGU}$	1
$\boldsymbol{GGC}$	1
$\boldsymbol{GGU}$	1

UGC	1
UGU	1
AUC	1
AUU	1
CUC	1
CUU	1
GUC	1
GUU	1
UUC	1
UUU	1

6 Conclusion

As shown by Sanchez et al., 2005 an algebraic structure, viz., ring $(Z_{64}, +, .)$ naturally occurs in the genetic code. Again, weknow that given an algebraic structure, different graph structures can be obtained from it. In this paper we have made an attempt to investigate the nilpotent graph structure of the genetic code. A one-one mapping between each pair of graphs have been obtained. Next, different centrality measures are applied here as a graph theoretic tool to study the influence of each codon (Both G_1 and G_2). From these centrality measures we can summarize that all codons (Both G_1 and G_2) have equal importance in the evolutionary process. We have also observed that all codons (Table 7 and Table 8) have high clustering coefficient values. So, the rate of the evolutionary process is comparatively slow in the vicinity of all codons (Both G_1 and G_2). Lastly, we have observed that the degree of distribution is symmetrical.

References

Aftabuddin Md, Kundu S. 2007. Hydrophobic, Hydrophilic, and Charged Amino Acid Networks within Protein. Biophysical Journal, 93(1): 225–231. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0006349507712754

Ali T, Borah C. 2021. Analysis of amino acids network based on mutation and base positions. Gene Reports, 24: 101291

Akhtar A, Ali T. 2014. Analysis of unweighted amino acids network. International Scholarly Research Notices, 1–6. Available from: https://www.hindawi.com/journals/isrn/2014/350276/

Akhtar A, Ali T, Gohain N. 2015. Graph in genetic code algebra. J bioinform Intelli Control, 4(1): 53–56. Available from: http://openurl.ingenta.com/content/xref?genre=article&issn=2326-7496&volume=4&issue=1&spage=53

Bagler G, Sinha S. 2007. Assortative mixing in Protein Contact Networks and protein folding kinetics. Bioinformatics, 23(14): 1760–1767. Available from: https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btm257

Balakrishnan J. 2002. Symmetry scheme for amino acid codons. Phys Rev E, 65(2): 021912. Available from: https://link.aps.org/doi/10.1103/PhysRevE.65.021912

Bertholf D, Walls G. 1978. Graphs of finite abelian groups. Czech Math J, 28(3): 365–2368. Available from:

https://dml.cz/handle/10338.dmlcz/101541

Bonacich P. 1972. Factoring and weighting approaches to status scores and clique identification. The Journal of Mathematical Sociology, 2(1):113–120. Available from: http://www.tandfonline.com/doi/abs/10.1080/0022250X.1972.9989806

Boruah BK, Ali T. 2022. A study of the total graph in genetic code algebra. Network Biology, 12(1): 1-10

Crick FH, Barnett L, Brenner S, Watts-Tobin RJ. 1961. General nature of the genetic code for proteins. Nature, 192: 1227–32

Kimura M. 1981. Possibility of extensive neutral evolution under stabilizing selection with special reference to nonrandom usage of synonymous codons. Proceedings of the National Academy of Sciences, 78(9): 5773–577. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.78.9.5773

Koonin EV. 2017. Frozen Accident Pushing 50: Stereochemistry, Expansion, and Chance in the Evolution of the Genetic Code. Life, 7(2): 22. Available from: http://www.mdpi.com/2075-1729/7/2/22

Koonin EV, Novozhilov AS. 2009. Origin and evolution of the genetic code: the universal enigma. IUBMB Life, 61(2): 99–111

Kundu S. 2005. Amino acid network within protein. Physica A: Statistical Mechanics and its Applications, 346(1–2): 104–109. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0378437104011677

Newman MEJ. 2002. Assortative mixing in networks. Phys Rev Lett, 89(20): 208701. Available from: https://link.aps.org/doi/10.1103/PhysRevLett.89.208701

Osawa S, Jukes TH, Watanabe K, Muto A. 1992. Recent evidence for evolution of the genetic code. Microbiol Rev, 56: 36

Rodin AS, Szathmáry E, Rodin SN. 2011. On origin of genetic code and tRNA before translation. Biol Direct, 6(1): 14. Available from: http://biologydirect.biomedcentral.com/articles/10.1186/1745-6150-6-14

Sanchez R, Morgado E, Grau R. 2005. Gene algebra from a genetic code algebraic structure. J Math Biol, 51(4): 431–457. Available from: http://link.springer.com/10.1007/s00285-005-0332-8

Sengupta D, Kundu S. 2012. Role of long- and short-range hydrophobic, hydrophilic and charged residues contact network in protein's structural organization. BMC Bioinformatics, 13(1): 142. Available from: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-13-142

Szathmáry E. 1999. The origin of the genetic code: amino acids as cofactors in an RNA world. Trends in Genetics, 15(6): 223–229. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0168952599017308

Watts DJ, Strogatz SH. 1998. Collective dynamics of 'small-world' networks. Nature, 393: 3

Wong JTF. 1975. A Co-Evolution Theory of the Genetic Code. Proceedings of the National Academy of Sciences, 72(5): 1909–1912. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.72.5.1909

Wuchty S, Stadler PF. 2003. Centers of complex networks. Journal of Theoretical Biology, 223(1): 45–53. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0022519303000717

Zhang WJ. 2016. Selforganizology: The Science of Self-Organization. World Scientific, Singapore

Zhang WJ. 2018. Fundamentals of Network Biology. World Scientific Europe, London, UK