Article

Graphical representation of genetic code algebra

Birinchi Kumar Boruah, **Tazid Ali**
Department of Mathematics, Dibrugarh University, Assam 786004, India
E-mail: birinchikumar49@gmail.com, tazid@dibru.ac.in

Received 26 January 2023; *Accepted 26 February 2023*; *Published online 6 March 2023*; *Published 1 September 2023*

Abstract
During the translation process, the genetic code is the nucleotide sequence that determines the amino acid sequence of protein molecules. A codon is a universal triplet of nucleotides that codes for an amino acid. A group G's identity graph is a graph in which the vertex set is the set of all group elements and two vertices in G are adjacent if $a.b = e$, where e is the group's identity element. In this study, we looked at the identity graph in the genetic code algebra. In our current study, we have thoroughly discussed various measures of centrality. In addition to this analysis, the correlation coefficients between various centrality measures, clustering coefficient, degree of distribution, and skewness are all examined.

Keywords amino acid; centrality measure; correlation coefficient; clustering coefficient; degree of distribution; genetic code; identity 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

Cells are the building blocks of all living things. Each cell contains a set of chromosomes that serve as the blueprint for the entire living organism. A chromosome is composed of genes (DNA sequences), each of which encodes a unique protein. A protein is a linear sequence of amino acids, which are the fundamental building blocks and functional components of living organisms. Twenty amino acids have been discovered to exist in proteins to date. Each of the three sequencing bases is a codon, which is a unit that specifies an amino acid. There are 64 codons with four bases. As a result, there must be some similarity, which means that more than one codon codes for the same amino acid. Synonymous codons are those that encode for the same amino acid. This is a one-to-many mapping of codons to amino acids. Furthermore, the UAA, UAG, and UGA triplets are known as stop codons and serve to terminate the translation process.

The processes that allow information to flow from DNA to protein are transcription and translation (Shu, 2017). Because of mutation, the sequencing bases are not duplicated exactly when the DNA strand is replicated. This influences protein formation. Mutations in genetics are classified as deletions, insertions, inversions, point mutations, and frame shifts. A point mutation is a one-base substitution in a genetic

sequence. Transition is a point mutation in genetics and molecular biology that converts the purine A and G to a purine or the pyrimidine C and U to a pyrimidine. A point mutation that converts a purine to a pyrimidine or vice versa is referred to as transversion.

Various researchers have made a diverse range of contributions to the field of biological networks (Bagler and Sinha, 2007; Khansari et al., 2016; Zhang, 2016). Kundu (2005) investigated how hydrophobic and hydrophilic networks contribute to proteins' "small-world properties". He also discovered that the average degree of nodes in the hydrophobic network is higher than in the hydrophilic network. Jiao et al. (2007) used contact energy to investigate the weighted amino acids network and demonstrated that it satisfies the small-world property. Akhtar and Ali (2014) discovered an amino acid network that is dependent on codon mutations. According to their findings, the amino acids Arginine (high hydrophilic) and Serine (low hydrophilic) have the highest centrality values regardless of centrality measurement. Wuchty and Stadler (2003) investigated various biological network centrality measures. They concluded that the degree of vertex centrality alone is insufficient to distinguish between lethal and viable proteins. Newman (2002) investigated assortative mixing characteristics in protein association networks, neural networks, and food networks. Furthermore, he discovered that an assortative network can transmit information more efficiently than a disassortative network. The centralities of two biological networks: the transcriptional network and the PPI network, were investigated by Koschutzki and Schreiber (2004). Their findings suggested that when studying biological networks, different centrality measures should be considered. Ali and Akhtar (2016) created an amino acid network that depicts the amino acid's evolutionary pattern. They discussed different centrality measures for that network and discovered that when considering centrality measures such as degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality, the hydrophobic amino acid Tyrosine (Y) has the highest centrality values. They also investigated the correlation coefficients between various measures of centrality. Zhang (2016) identified node attributes that have a significant impact on network node centrality. Identity graph of a group G is a graph where vertex set is the set of all elements of the group and two vertices say a and b in G are adjacent if $a.b = e$, where e is the identity element of G.

In this paper, we tried to investigate the connection between identity graphs and genetic codes. The paper's outline is as follows. Section 2 gives a brief overview of the various centrality measures and introduces some basic graph theory concepts that serve as the foundation for our analysis. The identity graph derived from the codons and various centrality measures are discussed in Section 3. The bivariate correlation in centrality measures is also covered. We will examine some of the most crucial network parameters in section 4. The paper's conclusion can be found in Section 5.

2 Fundamental Ideas for Graphs

A graph G is an ordered triple $(V(G), E(G), \psi_G)$ consisting of a nonempty set $V(G)$ of vertices, a set $E(G)$, disjoint from $V(G)$, of edges, and an incidence function ψ_G that associates with each edge of G an unordered pair of (not necessarily distinct) vertices of G. If e is an edge and u and v are vertices such that $\psi_G(e) = uv$, then e is said to join u and v; the vertices u and v are called the ends of e. The ends of an edge are said to be incident with the edge, and vice versa. Two vertices which are incident with a common edge are adjacent, as are two edges which are incident with a common vertex. A walk in G is a finite non-null sequence $W = v_0e_1v_1e_2v_2 \ldots e_qv_q$, whose terms are alternately vertices and edges, such that, for $1 \leq i \leq q$, the ends of e_i are v_{i-1} and v_i. We say that W is a walk from v_0 to v_q, or a (v_0, v_q)-walk. If the edges $e_1, e_2, \ldots, e_q$ of a walk W are distinct, W is called a trail. If the vertices $v_1, v_2, \ldots, v_q$ are distinct then W is called a path. Two vertices u and v of G are said to be connected if there is a (u, v)-path in G. If not, it is disconnected. Assume that G is 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 an edge exists between vertex v_i to vertex v_j and $a_{ij} = 0$ otherwise.

2.1 The centrality of the graph

In graph theory, centrality measure of a vertex represents its relative importance within the graph G . It is a real valued function $f: V \to R$, where V is the vertex set of the graph G.

2.1.1 Degree centrality

Degree centrality of a node u, denoted by $C_d(u)$, is the number of nodes to which u is directly connected.

2.1.2 Closeness centrality

Closeness centrality is defined as follows (Zhang, 2016)

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

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

2.1.3 Betweenness centrality

Betweenness centrality (Watts and 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 and Strogatz, 1998).

2.1.4 Eigenvector centrality

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

2.2 Network parameters

Biological networks employ a variety of network parameters. We have gone over three fundamental network parameters: clustering coefficient, degree of distribution, and skewness.

2.2.1 Clustering coefficient

Assume there is a node v with degree k in the undirected graph G, and there are e edges between v's k neighbours in G. Then, in G, the clustering coefficient of v is defined as

$$C_v = \frac{2e}{k(k-1)}$$

Here, C_v calculates the ratio between the edge numbers among the neighbours of v and the total possible edge numbers: $k(k-1)/2$, where $0 \leq C_v \leq 1$.

2.2.2 Degree distribution

In an undirected graph, the degree of a vertex is the number of links or edges the vertex has to the other vertices (Zhang, 2016, 2018). The degree distribution, $P(k), k = 0, 1, \ldots,$ is then used to calculate the ratio of vertices

in the network with degree k. Mathematically, $P_k = \frac{n_k}{n}$, where n is the network size and n_k is the total number of vertices of degree k in the network.

2.2.3 Skewness

Karl Pearson proposed measuring skewness for the first time in 1895. When the mean, median, and curve mode are not the same, the situation of skewness, which implies the absence of symmetry, exists in a curve. Positive skewness and negative skewness emerge in the distribution depending on the vertices and relative location of the mode, mean, and median. In our analysis, we take into account Karl Pearson's coefficient of skewness, denoted by S_k and defined by the formula $S_k = \frac{3(Mean-Median)}{Standard\ Deviation}$. The skewness value is within the range of -3 to $+3$.

3 Graph on Genetic Code

According to Sanchez et al. (2005), the four RNA (or DNA) bases can be arranged or ordered by taking into account their interactions with one another's codons and anticodons. The number of hydrogen bonds and the chemical type of bases (purine and pyrimidine) are important factors in this. From this, two orders of the base sets are obtained: $\{A, C, G, U\}$ and $\{U, G, C, A\}$. They explained sum operation of the bases obtained from the above two possible orders ($\{A, C, G, U\}$ and $\{U, G, C, A\}$) which makes the two sets isomorphic to Z_4. The sum operation of the bases obtained from the two possible orders is represented in Table 1. They also defined the sum and product operations on the codon set. The group derived from the set of codons was found to be isomorphic to the group of integer module 64, $(Z_{64}, +)$. These two sum and product operations represent a ring structure that is isomorphic to the ring of $(Z_{64}, +, .)$ on the set of whole codons. The total graph of this ring structure Z_{64} was described by Akhtar et al. in 2015. Boruah et al. (2022) observe that $Z_4 \times Z_4 \times Z_4$ forms a ring structure isomorphic to the ring of $(Z_4 \times Z_4 \times Z_4, +, .)$ by using the ordered base set $\{A, C, G, U\}$ isomorphic Z_4 from Sanchez et al., 2005. Here we have discussed identity graph by taking base set $\{A, C, G, U\}$ and group $(Z_4 \times Z_4 \times Z_4, +)$.

Codons 000–303 (codons XAZ) code for the most hydrophilic amino acids, while codons 030–333 (codons XUZ) code for the most hydrophobic amino acids (Table 3).

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	AminoAcid	No	Codon	AminoAcid	No	Codon	AminoAcid	No	Codon	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

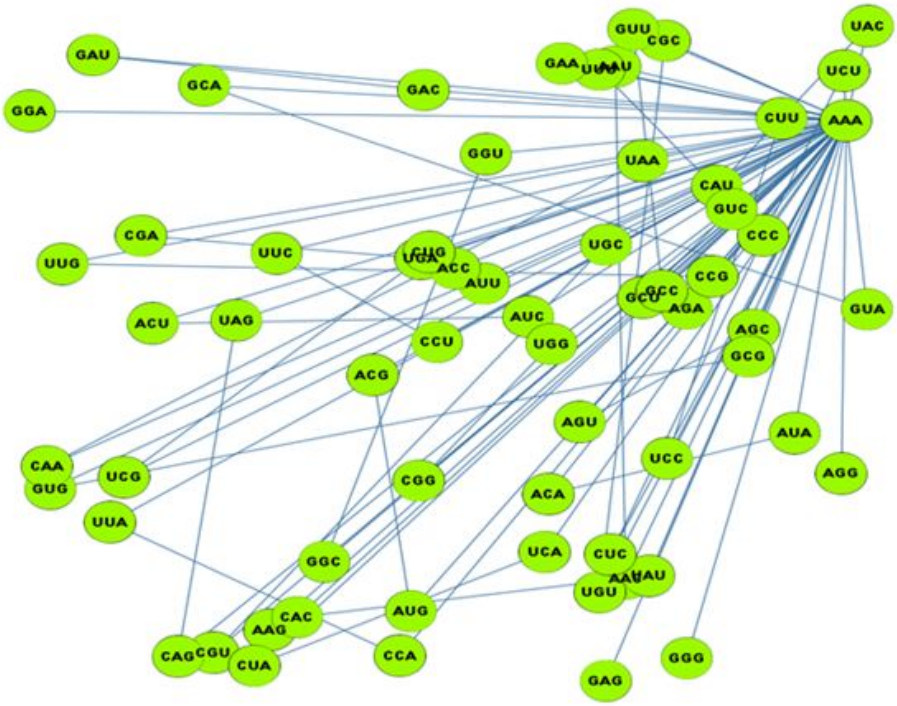

Fig. 1 Identity graph of the 64 codons.

3.1 The identity graph's centralities

In this section, we computed various centrality measures to analyse the identity graph (Fig. 1) and displayed all results in Table 4.

Table 4 Different codon centrality measures.

Vertex	Degree Centrality (C_d)	Closeness Centrality (C_{cl})	Betweenness Centrality (C_{bwt})	Eigenvector Centrality (C_λ)
AAA	63	1	1925	1
AUA	2	0.508065	0	0.081907
ACA	2	0.508065	0	0.081907
AGA	1	0.504	0	0.06594
UUG	2	0.508065	0	0.081907
CCG	2	0.508065	0	0.081907
AGC	2	0.508065	0	0.081907
AGU	2	0.508065	0	0.081907
CCU	2	0.508065	0	0.081907
UUC	2	0.508065	0	0.081907
AUC	2	0.508065	0	0.081907
ACU	2	0.508065	0	0.081907
AAC	2	0.508065	0	0.081907
AAU	2	0.508065	0	0.081907
AAG	1	0.504	0	0.06594
CGG	2	0.508065	0	0.081907
UGG	2	0.508065	0	0.081907
AUG	2	0.508065	0	0.081907
ACG	2	0.508065	0	0.081907
AGG	1	0.504	0	0.06594
CGC	2	0.508065	0	0.081907

UGU	2	0.508065	0	0.081907
CGA	2	0.508065	0	0.081907
UGA	2	0.508065	0	0.081907
AUU	2	0.508065	0	0.081907
ACC	2	0.508065	0	0.081907
CGU	2	0.508065	0	0.081907
UGC	2	0.508065	0	0.081907
GGG	1	0.504	0	0.06594
GGA	1	0.504	0	0.06594
UCG	2	0.508065	0	0.081907
CUG	2	0.508065	0	0.081907
GGU	2	0.508065	0	0.081907
GGC	2	0.508065	0	0.081907
CAU	2	0.508065	0	0.081907
UAC	2	0.508065	0	0.081907
UUA	2	0.508065	0	0.081907
CCA	2	0.508065	0	0.081907
GUG	2	0.508065	0	0.081907
GCG	2	0.508065	0	0.081907
UCA	2	0.508065	0	0.081907
CUA	2	0.508065	0	0.081907
UUU	2	0.508065	0	0.081907
CCC	2	0.508065	0	0.081907
GAA	1	0.504	0	0.06594
GAG	1	0.504	0	0.06594
UAA	2	0.508065	0	0.081907
CAA	2	0.508065	0	0.081907

UAU	2	0.508065	0	0.081907
CAC	2	0.508065	0	0.081907
UCU	2	0.508065	0	0.081907
CUC	2	0.508065	0	0.081907
UCC	2	0.508065	0	0.081907
CUU	2	0.508065	0	0.081907
CAG	2	0.508065	0	0.081907
UAG	2	0.508065	0	0.081907
GAC	2	0.508065	0	0.081907
GAU	2	0.508065	0	0.081907
GUC	2	0.508065	0	0.081907
GCU	2	0.508065	0	0.081907
GUU	2	0.508065	0	0.081907
GCC	2	0.508065	0	0.081907
GCA	2	0.508065	0	0.081907
GUA	2	0.508065	0	0.081907

In an identity 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 4), it is discovered that codon AAA has the highest value in all centrality measures. Codon AAA is the only codon that is linked to every other codon. As a result, codon AAA has the minimum cumulative shortest path distance and a high closeness centrality value. All of the observations show that the degree of the codon AAA is high, and thus the betweenness centrality is also high. Furthermore, the sum of the direct and indirect links of the codon AAA is found to be the greatest. As a result, eigenvector centrality is greatest in codon AAA. The codon AAA, which codes for lysine, had the highest degree of centrality and shared connections with all other codons, whereas the node AGA, which codes for arginine, shared only AAA's edge. Because arginine and lysine are both basic hydrophilic amino acids, AGA showed only association with AAA, differing only by a single nucleotide, i.e., in the second position. 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. In both cases, the analysis of these parameters results in a higher value for the AAA node.

3.2 The bivariate correlation of numerous centralities

The bivariate correlation of different centrality measures for the identity graph was examined in this section. Correlation is the most important factor to consider when studying assortative or disassortative networks. The network is said to be assortative when nodes with higher degrees of connectivity appear to communicate with other nodes with higher degrees of connectivity. Higher degree nodes appear to communicate with lower degree nodes in a disassortative network (Newman, 2002). The correlation coefficients for each centrality measure are shown in Table 5.

Table 5 Correlation coefficients for various measures of centrality.

	C_d	C_{cl}	C_{bwt}	C_λ
C_d	1	0.999793	0.999155	0.999997
C_{cl}	0.999793	1	0.999785	0.999739
C_{bwt}	0.999155	0.999785	1	0.999049
C_λ	0.999997	0.999739	0.999049	1

All correlation coefficients (r) are computed using Pearson's method. The value of r can range between $+1$ and -1. $r > 0$ for an assortative network and $r < 0$ for a disassortative network. Table 5 demonstrates that all of the centrality measures in the graph are highly correlated. Consequently, in this network, these centrality measures representing different centrality features are closely related. It is common knowledge that assortative networks transfer information more quickly than disassortative networks (Newman, 2002). Furthermore, the above correlation coefficient indicates that the network is of the assortative type $(r > 0)$, implying that evolutionary information flow will be simple.

4 Network Parameters

To analyse biological networks, we employ different network parameters. We will look at a few of them in the following sections to figure out how the network communicates.

4.1 Codon clustering coefficients

The clustering coefficient is a metric that depicts the likelihood of dividing a graph into clusters. A cluster is a collection of nodes connected by multiple links. A node with a high clustering coefficient is closely related to neighbouring nodes. A node's clustering coefficient appears to have an effect on its neighbouring node, thereby stabilising the flow of information (Sengupta and Kundu, 2012).

The clustering coefficients for all 64 codons in the Identity graph are shown in Table 6.

Table 6 Clustering coefficient of the codons.

AAA	$\mathbf{0.014337}$
AUA	1
ACA	1

AGA	0
UUG	1
CCG	1
AGC	1
AGU	1
CCU	1
UUC	1
AUC	1
ACU	1
AAC	1
AAU	1
AAG	0
CGG	1
UGG	1
AUG	1
ACG	1
AGG	0
CGC	1
UGU	1
CGA	1
UGA	1
AUU	1
ACC	1
CGU	1
UGC	1
GGG	0
GGA	0

UCG	1
CUG	1
GGU	1
GGC	1
CAU	1
UAC	1
UUA	1
CCA	1
GUG	1
GCG	1
UCA	1
CUA	1
UUU	1
CCC	1
GAA	0
GAG	0
UAA	1
CAA	1
UAU	1
CAC	1
UCU	1
CUC	1
UCC	1
CUU	1
CAG	1
UAG	1
GAC	1

$\boldsymbol{GAU}$	1
$\boldsymbol{GUC}$	1
$\boldsymbol{GCU}$	1
$\boldsymbol{GUU}$	1
$\boldsymbol{GCC}$	1
$\boldsymbol{GCA}$	1
$\boldsymbol{GUA}$	1

A codon's clustering coefficient is determined by two factors: the degree of the codon and the number of direct connections between two neighbouring codons. Except for the codons AAA, AGA, AAG, AGG, GGG, GGA, GAA, and GAG, all other codons have degree 2. Furthermore, the number of links between neighbouring codons is one, so with the exception of AAA, AGA, AAG, AGG, GGG, GGA, GAA, and GAG, all codons have a high clustering coefficient, i.e., 1. We discovered that the clustering coefficient for the entire graph is 0.875224. Thus, after examining the graph's clustering coefficient and the clustering coefficients of the codons, we discovered that, with the exception of AAA, AGA, AAG, AGG, GGG, GGA, GAA, and GAG, the flow of evolutionary process is comparatively slow in the neighbourhood of other codons when compared to the entire graph.

4.2 The degree of distribution

We compute the degree of distribution of the nodes (codons) in the Identity graph of the 64 codons in this section. A node's degree in a network is defined as the number of links it has to other nodes. If a network has n nodes and n_k of them have degree k, we have the degree distribution $P(k) = n_k/n$.

4.3 Skewness

Skewness, as previously stated, is a measure of the asymmetry of a variable distribution (Zhang, 2018). The following formula yields the Karl Pearson's coefficient of skewness (S_k)

$$S_k = \frac{3(Mean - Median)}{Standard\ Deviation}$$

The distribution is symmetrical if $S_k = 0$. If $S_k > 0$, the distribution is positively skewed; otherwise, it is negatively skewed.

Considering degree centrality from Table 4, we have found the Pearson's coefficient of skewness is 0.331. We can see that the skewness coefficient is positive. As a result, the degrees of distribution of the codons are skewed positively.

5 Conclusion

The Identity graph structure of the genetic code was investigated. Numerous centrality measures were employed as a graph theoretic tool to investigate the effects of each codon. Several centrality measures are discussed, and it is found that Codon AAA has the highest centrality value of all of them. The AAA codon, which encodes Lysine, shared a connection with all other codons, whereas the AGA code for Arginine shared only one node

with AAA (Lysine). This could shed light on the contiguous property of genetic codes, which played a critical role in the codon system's co-evolution. For example, when the amino acid Lysine was first synthesized from Asp (Aspartic acid), several different arrangements of genetic code may have occurred within Lysine in domains, and gradually AAA and AAG versions may have favored over others for Lysine, possibly due to the benefits of the presence of an Arginine domain adjacent to Lysine. As a result, we have concluded that codon AAA play an important role in the evolution of amino acids. Furthermore, we examined the correlation coefficients of various codon centrality measures. All centrality measures were found to be highly correlated. Again, the correlation coefficient reveals that the network is assortative, implying that evolutionary information flow will be simple. When the clustering value of the codons is examined, it is clear that, with the exception of codons AAA, AGA, AAG, AGG, GGG, GGA, GAA, and GAG, all of the codons have a high clustering coefficient. As a result, with the exception of AAA, AGA, AAG, AGG, GGG, GGA, GAA, and GAG in the vicinity of other codons, the evolutionary process is quite slow when compared to the entire network. Another noteworthy finding is that the degree of distribution is positively skewed.

Acknowledgment

The authors would like to thank the Department of Mathematics, Dibrugarh University, Dibrugarh, Assam for providing the infrastructure to carry out the research work.

References

Akhtar A, Ali T. 2014. Analysis of unweighted amino acids network. International Scholarly Research Notices, Article ID: 350276

Ali T, Phukan CK. 2013. Topology in genetic code algebra. Mathematical Sciences Research Journal, 2(2): 179-182

Akhtar A, Gohain N. 2015. Graph theoretic approach to analyze amino acid network. International Journal of Advances in Applied Mathematics, 2(3): 31-37

Ali T, Akhtar A, Gohain N. 2016. Analysis of amino acid network based on distance matrix. Physica A, 452: 69-78

Bagler G, Sinha S. 2007. Assortative mixing in protein contact networks and protein folding kinetics. Bioinformatics, 23(14): 1760-1767

Bonacich P. 1972. Factoring and weighting approaches to status scores and clique identification. Journal of Mathematical Sociology, 2: 113-120

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

Freeman L. 1978. Centrality in social networks conceptual classification. Socical Networks, 1: 215-239

Jiao X, Chang S, Li C, Chen W, Wang C. 2007. Construction and application of the weighted amino acid network based on energy. Physical Review E, Article ID: 051903

Khansari M, Kaveh A, Heshmati Z. 2016. Centrality measures for immunization of weighted networks. Network Biology, 6(1): 12-27

Koschutzki D, Schreiber F. 2004. Comparison of centralities for biological networks. In: Proceeding of the German Conference of Bioinformatics (GCB'04). Lecture Notes in Informatics, LNI P-53: 199-206

Kundu S. 2005. Amino acid network within protein. Physica A, 346: 104-109

Newman MEJ. 2002. Assortative mixing in networks. Physiccal Review Letters, 89: 2087011-2087014

Shu JJ. 2017. A new integrated symmetrical table for genetic codes. BioSystems, 151: 21-26

Sanchez R, Morgado E, Grau R, 2005. Gene algebra from a genetic code algebraic structure. Journal of Mathematical Biology, 51(4): 431-457

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: 142

Watts D, Strogatz J, Steven H, 1998. Collective dynamics of small-world networks. Nature, 393: 440-442

Wuchty S, Stadler PF. 2003. Centers of complex networks. Journal of Theoretical Biology, 223: 45–53

Zhang WJ. 2016. Screening node attributes that significantly influence node centrality in the network. Selforganizology, 3(3): 75-86

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