Research Article

A Study on the Identity Graph in Genetic Code

Birinchi Kumar Boruah*, Tazid Ali, Shyamalima Saikia

Department of Mathematics, Dibrugarh University, Dibrugarh, Assam, INDIA.

Submission Date: 28-02-2022; Revision Date: 13-03-2022; Accepted Date: 10-04-2022.

ABSTRACT

The genetic code is the sequence of nucleotides that determines the amino acid sequence of protein molecules during the process of translation. The term codon is used to denote the universal nucleotide triplet that codes for an amino acid. 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 the present study, we have analyzed the Identity graph in the genetic code algebra. In our current study different measures of centrality have been comprehensively discussed. Apart from this investigation, a study on the correlation coefficients between different measures of centrality is carried out in addition to clustering coefficient, degree of distribution as well as skewness.

Keywords: Amino acid, Centrality measure, Correlation coefficient, Clustering coefficient, Degree of distribution, Genetic code, Identity graph.

Correspondence:
Mr. Birinchi Kumar Boruah,
Department of Mathematics, Dibrugarh University, Dibrugarh-786004, Assam, INDIA.

Email: birinchikumar49@gmail.com

INTRODUCTION

Networks can now be found almost anywhere in natural and artificial systems. Amino acids are essential for cellular metabolism and must be synthesized by organisms. The process of synthesis of protein molecule is entirely based on the genetic codes; the universal process of conversion of codons to amino acid molecules which in turn transformed to functional proteins through a number of cellular modification processes. The genetic codes are the basis of life; which contains instructions for the formation of a protein molecule. It is the set of rules that govern the translation of information encoded in genetic material (DNA or RNA sequences) into proteins. The transfer RNAs are the code adaptor molecules that bear anticodon complementary to the codon and direct the process of amino acid or protein synthesis. Transcription and translation are the processes by which information is transferred from DNA to protein. As of now, a total of 22 amino acids are present, each of which is specified by the triplet code of nitrogenous bases, known as genetic code. Adenine (A), Guanine (G), Cytosine (C), and Uracil (U) code for these amino acids in 64 different triplet combination, however out of these 64 genetic codes UAA, UAG, and UGA doesn't code for any amino acid, known as stop or termination codon.[1-2] Thus, a genetic code can be defined as the sequence of three bases that codes for a specific amino acid in triplet form. The term codon is used to denote nucleotide triplet that codes for an amino acid. The genetic codes are unambiguous and degenerative, i.e. each triplet code for specific amino acid and specific amino acid can be encoded by more than one triplet base respectively, and they are comma less, non-overlapping as well as universal in nature (Crick *et al.*, 1961;[1] E. V. Koonin and Novozhilov, 2009;[3] Rodin *et al.*, 2011;[4] Szathmary, 1999;[5] Wong, 1975). Marshall Nirenberg and Heinrich J. Matthaei (1961) first elucidated genetic code in *E. coli* emphasizing 64 codons for 20 amino acids.

Several theories have been postulated to answer the question of origin and evolution of genetic codes and they are believed to be evolved through chemical constrain as well as under the influence of evolutionary forces. According to the stereochemical theory of evolution of genetic codes, stereochemical properties of amino acids are directly associated with its triplet

SCAN QR CODE TO VIEW ONLINE

www.ajbls.com

DOI: 10.5530/ajbls.2022.11.24

codons whereas frozen accident theory explains it as an event that occurred purely by chance (Crick *et al.*, 1961;[1] E. Koonin, 2017;[2] E. V. Koonin and Novozhilov, 2009;[3] Osawa *et al.*, 1992).[6] On the other hand, according to the theory of co-evolution genetic codes, the codon system coevolved with the pathways of biosynthesis of amino acids along with precursor product relationships involved in these pathways (Kimura, 1981;[7] E. Koonin, 2017;[2] Wong, 1975). Point mutation, translational frameshifts, and translational mislead are prime reasons for the evolution of codons. Related amino acids have related codons, for example, hydrophobic amino acids have base U in the second position and they are similar in properties. Codons that differ only in a single nucleotide code for the same or closely related amino acid that share similar physicochemical properties (Crick *et al.*, 1961;[1] E. Koonin, 2017;[2] E. V. Koonin and Novozhilov, 2009;[3] Osawa *et al.*, 1992;[6] Rodin *et al.*, 2011;[4] Szathmary, 1999;[5] Wong, 1975). Mutation is one of the important factors due to which the sequence of bases cannot be exactly copied in replicating the strand of DNA. The consequences of mutation affect the formation of protein. Codons can be muted in a variety of ways, including deletion, insertion, inversion, point, and frame-shift mutation. Point mutation can occur at a single or multiple points. A point mutation from a purine(pyrimidine) to a purine(pyrimidine) is called transition and a point mutation from a pyrimidine to purine or vice versa is called transversion. The genetic code is made up of 64 codons that code for the 20 amino acids. That is, different codons may code the same amino acid. From Mathematical point of view, it is a many to one function from codons to amino acids. From the study carried out by Balakrishnan (2002),[8] it is found that the total number of codons are almost three times the total number of amino acids, so some mathematical structures on genetic code seems to be surely present. The error frequency (accepted mutations) of codons influences the importance of different base positions in codons. From the third base to the first base, and then to the second base, the frequency of errors decreases.

Researchers have been making sincere effort and notable contributions in the area connected to biological network. Kundu[9] discovered that the hydrophobic network had a significantly higher average node degree than the hydrophilic network. To demonstrate this further, he demonstrated that long-range network interactions have a scale-free distribution. The lecture focused on hydrophobic, hydrophilic, and charged protein networks, with Aftabuddin and Kundu[10] discussing the different types of networks found within proteins. Akhtar and Ali[11] created an amino acid network based on codon mutations. As a result, regardless of the centrality measures used, the amino acids Serine(S) and Arginine(R) have the greatest centrality values. Further the degree of distribution follows Weibull distribution pattern. Wuchty and Stadler[12] discussed different types of centrality measures in biological network. Only the degree of vertex centrality was found to distinguish deadly proteins from those that could lead to survival. Bagler and Sinha[13] discovered that the average long-range clustering coefficient has a negative correlation with protein folding rate.

One of the most prominent non-numerical branches of mathematics is Graph Theory which is closely connected to algebra and matrix theory though it's considered as part of Topology. Graph theory is more reliable to study Biological Networking mainly structures of microscopic organism. Biological network is applicable for all kinds of protein-protein interaction, neuronal network and many researchers did study on it. For any particular algebraic structure, its corresponding graphical structure can be introduced. Dennts Bertholf *et al.*[14] found the graph structures of finite abelian group. And in this kind of group, vertices of the graph are in one-to-one mapping with non-identity subgroup of G, and if the subgroups cross with each other, then we join the two vertices with an edge. Anderson and Badawi[15] gave the idea of Total graph of commutative ring. 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.

Preliminary Concepts of Graph

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 $V(G)$. 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 and 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 $W : v_0\, e_1\, v_1\, e_2\, v_2 \cdots\cdots\cdots e_p v_p$ 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 $V=\{v_1,\ldots\ldots,v_n\}$. The adjacency matrix of G is the $n\times n$ matrix , where if there is an edge from vertex v_i to vertex v_j and $a_{ij}=0$ otherwise.

Centrality in graph

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

Degree centrality

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

Eigenvector centrality

The eigenvector of the greatest eigenvalue of the adjacency matrix of the corresponding graph is the eigenvector centrality (Bonacich, 1972).[16]

Betweenness centrality

Betweenness centrality (Watts and Strogatz, 1998)[17] 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.

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.

Graph on Genetic Code

Two different orderings of the RNA bases were introduced by Sanchez *et al.*[18] They obtained two ordering of base sets $\{A, C, G, U\}$ and $\{U, G, C, A\}$. A sum operation (Table 1) is defined on these two base sets such that the two sets are isomorphic to the cyclic group Z_4 (group Z_4 of integer module 4). Ali and Phukan,[19,20] defined a product operation (Table 2) on the base set $Y=\{A, C, G, U\}$, such that Y form a commutative ring structure. In the ring $(Y, +,.)$; A, C represents additive identity and multiplicative identity respectively. Sanchez also defined an addition operation on the set of 64 codons. These addition operations together with the set of codons form a group which is isomorphic to the group $(Z_{64}, +)$ as shown in Table 3. Akhtar *et al.*[21] discussed total graph on the group of codons. We have discussed identity graph on the same group structure of codons.

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

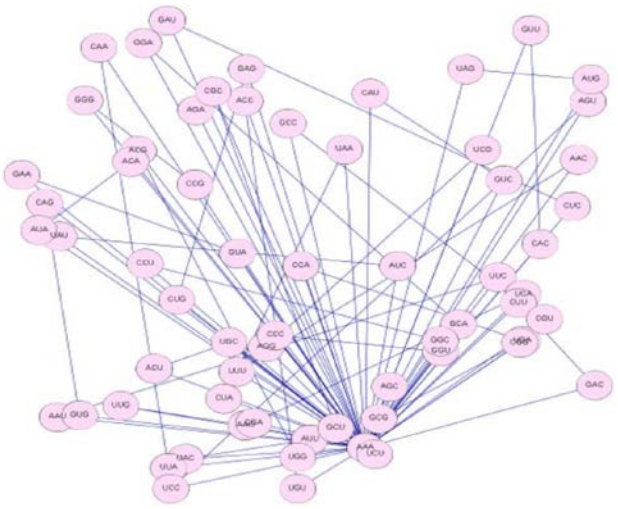

Figure 1: Identity graph of the 64 codons.

Table 3: The genetic code table of 64 codons.

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

Centralities in Identity graph

Different measures of centrality have been calculated to analyze Identity graph (Figure 1). Table 4, give the different centrality values for the codons.

Table 4: Centrality measures for the codons.

Vertex	Degree Centrality (C_d)	Closeness Centrality (C_c)	Betweenness Centrality (C_{bwt})	Eigenvector Centrality (C_λ)
AAA	63	1	1922	1
AUU	2	0.508065	0	0.081524
UAC	2	0.508065	0	0.081524
UCG	2	0.508065	0	0.081524
AGG	2	0.508065	0	0.081524
CAA	2	0.508065	0	0.081524
UUA	2	0.508065	0	0.081524
CCG	2	0.508065	0	0.081524
GGG	2	0.508065	0	0.081524
GCA	2	0.508065	0	0.081524
GGA	2	0.508065	0	0.081524
AAU	2	0.508065	0	0.081524
UUC	2	0.508065	0	0.081524
CCC	2	0.508065	0	0.081524
GGU	2	0.508065	0	0.081524
UCC	2	0.508065	0	0.081524
AGU	2	0.508065	0	0.081524
ACC	2	0.508065	0	0.081524
UGU	2	0.508065	0	0.081524
AAG	2	0.508065	0	0.081524
UUG	2	0.508065	0	0.081524
CCA	2	0.508065	0	0.081524
UGA	2	0.508065	0	0.081524
UCA	2	0.508065	0	0.081524
CGA	2	0.508065	0	0.081524
UAG	2	0.508065	0	0.081524
AUG	2	0.508065	0	0.081524
AAC	2	0.508065	0	0.081524
UUU	2	0.508065	0	0.081524
AGA	1	0.504	0	0.065638
UAU	2	0.508065	0	0.081524
AUC	2	0.508065	0	0.081524
UCU	2	0.508065	0	0.081524
AGC	2	0.508065	0	0.081524
GAC	2	0.508065	0	0.081524
CUU	2	0.508065	0	0.081524
CCU	2	0.508065	0	0.081524
GGC	2	0.508065	0	0.081524
GCC	2	0.508065	0	0.081524
CGU	2	0.508065	0	0.081524
GAG	2	0.508065	0	0.081524
CUG	2	0.508065	0	0.081524
ACG	2	0.508065	0	0.081524
UGG	2	0.508065	0	0.081524
GAU	2	0.508065	0	0.081524

Continued...

Continued...

CUC	2	0.508065	0	0.081524
UAA	2	0.508065	0	0.081524
CUA	2	0.508065	0	0.081524
GCU	2	0.508065	0	0.081524
CGC	2	0.508065	0	0.081524
GAA	2	0.508065	0	0.081524
GUA	2	0.508065	0	0.081524
ACU	2	0.508065	0	0.081524
UGC	2	0.508065	0	0.081524
CAU	2	0.508065	0	0.081524
GUC	2	0.508065	0	0.081524
GUG	2	0.508065	0	0.081524
CAG	2	0.508065	0	0.081524
ACA	2	0.508065	0	0.081524
AUA	2	0.508065	0	0.081524
CAC	2	0.508065	0	0.081524
GUU	2	0.508065	0	0.081524
GCG	2	0.508065	0	0.081524
CGG	2	0.508065	0	0.081524

Centrality parameters in an identity network emphasize the strength of similarity or closeness of a node to another node or to its neighbors. Degree centrality denotes the average number of connections a node shares with other nodes present in the network. The higher the value of degree centrality larger will be the number of connections. After observing all the codons (From Table 4), it is found that codon *AAA* has highest value in all centrality measures. Codon *AAA* is unique among all the codons that is connected to all other codons. Thus, codon *AAA* attain the minimum cumulative shortest path distance and has high closeness centrality value. From all the observation it is clearly seen that degree of the codon *AAA* is high and so betweenness centrality is also high. Moreover, the sum of direct and indirect link of the codon *AAA* is found to be maximum. Therefore, eigenvector centrality is maximum in the codon *AAA*. The codon AAA code for lysine showed the highest degree centrality share connection with all other codons whereas the node AGA, which code for arginine shared edge only with AAA. Since Arginine and lysine both are the basic hydrophilic amino acid, AGA showed association only with AAA, differing only in a single nucleotide i.e., in the second position. The shorter the distance between two nodes higher will be the connection and the higher the value of betweenness centrality higher will be the influence in the network. Analysis of these parameters results in higher value of AAA node in both cases.

Correlation within different centralities

In this section we have focused primarily on correlation coefficients between various measures of centralities for the Identity graph. Table 5 represents the correlation coefficients between the centrality measures obtained using Pearson's method. When correlation value is positive ($r > 0$) then the network is assortative but if negative ($r < 0$) then the network is disassortative (Newman, 2002).[22] Table 5, informs us that all the centrality measures for the graph is highly correlated. Therefore, these centrality measures representing different centrality features are closely associated in this network. Hence using any measure is as good as another. It is well known fact that assortative network transfers information more easily than disassortative network (Newman, 2002).[22] Also, from the above correlation coefficient we observe that network is assortative type ($r > 0$) and consequently evolutionary information flow will be easy.

Table 5: Correlation coefficients for the centrality measures.

	C_d	C_{cl}	C_{bwt}	C_λ
C_d	1	0.999967	0.999866	1
C_{cl}	0.999967	1	0.999966	0.999959
C_{bwt}	0.999866	0.999966	1	0.999851
C_λ	1	0.999959	0.999851	1

Network Parameters

Various parameters are being used in biological networks. In this research we mainly used three parameters. Firstly, clustering coefficients, secondly degree of distribution and lastly skewness.

Clustering coefficient

Clustering coefficient is defined as the measurement of the capacity of a graph that can be splitted into clusters. Clusters are subset of the set that contains edges connecting vertices with vertices. Clustering coefficient C_i of a particular node '*i*' is defined by the ratio of total number of links e_i of neighbours with its nearest neighbours. Clustering coefficient for whole network is actually the average of C_i $C_i = (C_i = \frac{2e_i}{K_i(K_i - 1)}$, where K_i is the degree of node '*i*'). As the value of clustering coefficient increases the relations of neighboring nodes become strong. As a result, it slows down the information spread (Sengupta and Kundu, 2012).[23]

Table 6, depicts clustering coefficients of all the codons. Clustering coefficient of a codon depends upon two factors viz. degree of the codon and number

of direct connections in between two neighboring codons. It is observed that except for the codons *AAA*&*AGA*, other codons have degree 2. Also, the number of links in between the neighboring codons is 1, so except *AAA*&*AGA* all the codons have high clustering coefficient i.e., 1. We observed that clustering coefficient of the whole graph is 0.9689. Thus, after observing the clustering coefficients of the graph and clustering coefficients of the codons, we found that, except *AAA*&*AGA*, in the neighbourhood of other codons, flow of evolutionary process is comparatively slow as compared to the whole graph.

Table 6: Clustering coefficient of the codons.

AAA	0.015873
AUU	1
UAC	1
UCG	1
AGG	1
CAA	1
UUA	1
CCG	1
GGG	1
GCA	1
GGA	1
AAU	1
UUC	1
CCC	1
GGU	1
UCC	1
AGU	1
ACC	1
UGU	1
AAG	1
UUG	1
CCA	1
UGA	1
UCA	1
CGA	1
UAG	1
AUG	1
AAC	1
UUU	1
AGA	0
UAU	1
AUC	1
UCU	1
AGC	1
GAC	1
CUU	1
CCU	1
GGC	1
GCC	1
CGU	1
GAG	1
CUG	1
ACG	1
UGG	1
GAU	1
CUC	1
UAA	1
CUA	1
GCU	1
CGC	1
GAA	1
GUA	1
ACU	1
UGC	1
CAU	1
GUC	1
GUG	1
CAG	1
ACA	1
AUA	1
CAC	1
GUU	1
GCG	1
CGG	1

Degree of Distribution and Skewness

In this section we shall discuss the degree of distribution and 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 ornegative. In our study, we have used the Karl Pearson's coefficient of skewness, defined as

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

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 7, shows the degree of distribution values of different codons. From Table 7, Pearson's coefficient of skewness is found to be −0.53459. The negative value led us conclude that the degree of distribution of the codons are negatively skewed distribution.

Table 7: *Degree distribution of the codons.*

AAA	0.015625
AUU	0.96875
UAC	0.96875
UCG	0.96875
AGG	0.96875
CAA	0.96875
UUA	0.96875
CCG	0.96875
GGG	0.96875
GCA	0.96875
GGA	0.96875
AAU	0.96875
UUC	0.96875
CCC	0.96875
GGU	0.96875
UCC	0.96875
AGU	0.96875
ACC	0.96875
UGU	0.96875
AAG	0.96875
UUG	0.96875
CCA	0.96875
UGA	0.96875
UCA	0.96875
CGA	0.96875
UAG	0.96875
AUG	0.96875
AAC	0.96875
UUU	0.96875
AGA	0.015625
UAU	0.96875
AUC	0.96875
UCU	0.96875
AGC	0.96875
GAC	0.96875
CUU	0.96875
CCU	0.96875
GGC	0.96875
GCC	0.96875
CGU	0.96875
GAG	0.96875
CUG	0.96875
ACG	0.96875
UGG	0.96875
GAU	0.96875
CUC	0.96875
UAA	0.96875
CUA	0.96875
GCU	0.96875
CGC	0.96875
GAA	0.96875
GUA	0.96875
ACU	0.96875
UGC	0.96875
CAU	0.96875
GUC	0.96875
GUG	0.96875
CAG	0.96875
ACA	0.96875
AUA	0.96875
CAC	0.96875
GUU	0.96875
GCG	0.96875
CGG	0.96875

CONCLUSION

We made an effort to analyze the Identity graph structure of the genetic code. To investigate the impact of each codon, different centrality measures were adopted. After discussing several centralities measures it was found that Codon *AAA* has the highest centrality value among all the centrality measures. The *AAA* codon encoding Lysine shared connection with all other codons. *AGA* coding arginine is connected to only the node *AAA* (Lysine). This might give insight to the contiguous property of genetic codes that played a crucial role in the process of co-evolution of the codon system. For example, the amino acid Lysine was first synthesized from Asp (Aspartic acid), several different arrangements of genetic code might have occurred within Lysine in domains and gradually *AAA* and *AAG* version might have favored over others for Lysine. So, the ultimate conclusion we have derived is that codon *AAA* have a significant role in amino acids evolution. Further we have analyzed correlation coefficients of the different centrality measures of codons. It was observed that all centrality measures are highly correlated. Again, observation from correlation coefficient is that network is assortative type and hence evolutionary information flow will be quite easy. While looking at the clustering value of the codons, it was clearly seen that except codons *AAA* and *AGA* all the codons have high clustering coefficients. Therefore except *AAA* and *AGA*, in the neighborhood of other codons, flow of evolutionary process is quite slow as compared to the

Continued...

entire network. Another important observation is that the degree of distribution is negatively skewed which follows Weibull distribution pattern.

ACKNOWLEDGEMENT

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

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

1. Crick FH, Barnett L, Brenner S, Watts-Tobin RJ. General nature of the genetic code for proteins. Nature. 1961;192:1227-32. doi: 10.1038/1921227a0, PMID 13882203.
2. Koonin EV. Frozen accident pushing 50: Stereochemistry, expansion, and chance in the evolution of the genetic code. Life (Basel). 2017;7(2):22. doi: 10.3390/life7020022, PMID 28545255.
3. Koonin EV, Novozhilov AS. Origin and evolution of the genetic code: The universal enigma. IUBMB Life. 2009;61(2):99-111. doi: 10.1002/iub.146, PMID 19117371.
4. Rodin AS, Szathmáry E, Rodin SN. On origin of genetic code and tRNA before translation. Biol Direct. 2011;6(1):14. doi: 10.1186/1745-6150-6-14, PMID 21342520.
5. Szathmáry E. The origin of the genetic code: Amino acids as cofactors in an RNA world. Trends Genet. 1999;15(6):223-9. doi: 10.1016/s0168-9525(99)01730-8, PMID 10354582.
6. Osawa S, Jukes TH, Watanabe K, Muto A. Recent evidence for evolution of the genetic code. Microbiol Rev. 1992;56(1):229-64. doi: 10.1128/mr.56.1.229-264.1992, PMID 1579111.
7. Kimura M. Possibility of extensive neutral evolution under stabilizing selection with special reference to non-random usage of synonymous codons. Proc Natl Acad Sci USA. 1981;78(9):5773-7. doi: 10.1073/pnas.78.9.5773, PMID 6946514.
8. Balakrishnan J. Symmetry scheme for amino acid codons. Phys Rev E Stat Nonlin Soft Matter Phys. 2002;65(2 Pt 1):021912. doi: 10.1103/PhysRevE.65.021912.
9. Kundu S. Amino acid network within protein. Phys A. 2005;346(1-2):104-9. doi: 10.1016/j.physa.2004.08.055.
10. Aftabuddin M, Kundu S. Hydrophobic, hydrophilic, and charged amino acid networks within protein. Biophys J. 2007;93(1):225-31. doi: 10.1529/biophysj.106.098004, PMID 17172302.
11. Akhtar A, Ali T. Analysis of unweighted amino acids network. Int Sch Res Notices. 2014;2014:350276. doi: 10.1155/2014/350276, PMID 27355050.
12. Wuchty S, Stadler PF. Centers of complex networks. J Theor Biol. 2003;223(1):45-53. doi: 10.1016/s0022-5193(03)00071-7, PMID 12782116.
13. Bagler G, Sinha S. Assortative mixing in protein contact networks and protein folding kinetics. Bioinformatics. 2007;23(14):1760-7. doi: 10.1093/bioinformatics/btm257, PMID 17519248.
14. Bertholf D, Walls G. Graphs of finite abelian groups. Czech Math J. 1978;28(3):365-8. doi: 10.21136/CMJ.1978.101541.
15. Anderson DF, Badawi A. The total graph of a commutative ring. J Algebra. 2008;320(7):2706-19. doi: 10.1016/j.jalgebra.2008.06.028.
16. Bonacich P. Factoring and weighting approaches to status scores and clique identification. J Math Sociol. 1972;2(1):113-20. doi: 10.1080/0022250X.1972.9989806.
17. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. Nature. 1998;393(6684):440-2. doi: 10.1038/30918, PMID 9623998.
18. Sanchez R, Morgado E, Grau R. Gene algebra from a genetic code algebraic structure. J Math Biol. 2005;51(4):431-57. doi: 10.1007/s00285-005-0332-8, PMID 16012800.
19. Ali T, Phukan CK. Topology in genetic code algebra. Math Sci Int Res [journal:2013];2(2):179-82.
20. Ali T, Borah C. Analysis of amino acids network based on mutation and base positions. Gene Rep. 2021;24. doi: 10.1016/j.genrep.2021.101291, PMID 101291.
21. Akhtar A, Ali T, Gohain N. Graph in genetic code algebra. J Bioinform Intelli Control. 2015;4(1):53-6. doi: 10.1166/jbic.2015.1112.
22. Newman MEJ. Assortative mixing in networks. Phys Rev Lett. 2002;89(20):208701. doi: 10.1103/PhysRevLett.89.208701, PMID 12443515.
23. Sengupta D, Kundu S. Role of long- and short-range hydrophobic, hydrophilic and charged residues contact network in protein's structural organization. BMC Bioinformatics. 2012;13:142. doi: 10.1186/1471-2105-13-142, PMID 22720789.

Cite this article: Boruah BK, Ali T, Saikia S. A Study on the Identity Graph in Genetic Code. Asian J Biol Life Sci. 2022;11(1):173-80.