Article

Special identity subgraph in genetic code

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

Received 1 March 2022; *Accepted 12 March 2022*; *Published 1 June 2022*

Abstract
The genetic code is a series of codons that stores genetic information about protein molecule formation. The identity graph of a group G is a graph in which the vertex set is the set of all elements of the group and two vertices in G are adjacent if $a.b = e$, where e is the group's identity element. Let H be a subgroup of G then the identity graph drawn for the subgroup H is known as the identity special subgraph of G (special identity subgraph of G). In this study, we looked at the special identity graph in the genetic code algebra. Different measures of centrality have been thoroughly discussed in our current study. Aside from this investigation, research is being conducted on the correlation coefficients between different measures of centrality, as well as the clustering coefficient, degree of distribution, and skewness.

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

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

Biological science is one of the important areas in which mathematics have been successfully applied for long time. This area may be referred as mathematical biology or biomathematics. Various mathematical formulae and method of working are used in biology. Genetics is one of the important branches of biology. Every organism is constituted by the cells. Chromosomes can be thought of as being made up of strings of genes. The DNA is found in the cells of all living organisms and is situated in the nucleus, organized into chromosomes. Other than the nucleus, a very small amount of DNA is also found in the mitochondria. The genetic code is the set of rules that guide the translation of DNA into 20 amino acids, which constitute the fundamental units of protein in living cells (Fig. 1). A codon is a sequence of three bases of DNA from the 4 bases: Adenine (A), Cytosine (C), Guanine (G), and Thymine (T) which specifies one amino acid. As there are four bases so we have total 64 codons. The codons UAA, UGA, UAG signal the end of the polypeptide chain during translation which are known as stop codons. The codon AUG initiates the translation process which is known as start codon. Mathematically, for a sequence of DNA, four letters: A,

G, C, and T are taken. Any change to the gene sequence may change the information that is encoded by it. As a result, this changes the protein that is produced. These phenomena are referred to as mutations. Various types of mutation exist: point mutation, frameshift mutation, deletion, insertion, inversion.

Genetic codes can be understood as a many one function between the sets: the 64 potential codons and the 20 amino acids along with the stop codon (Boruah and Ali, 2022). Clearly this mapping is degenerate. Due to degeneracy more than one codon is coded for some amino acids. For example, the codons GUU, GUC, GUA, GUG codes amino acid valine. Over the years, the biological implications of degeneracy have been rigorously studied.

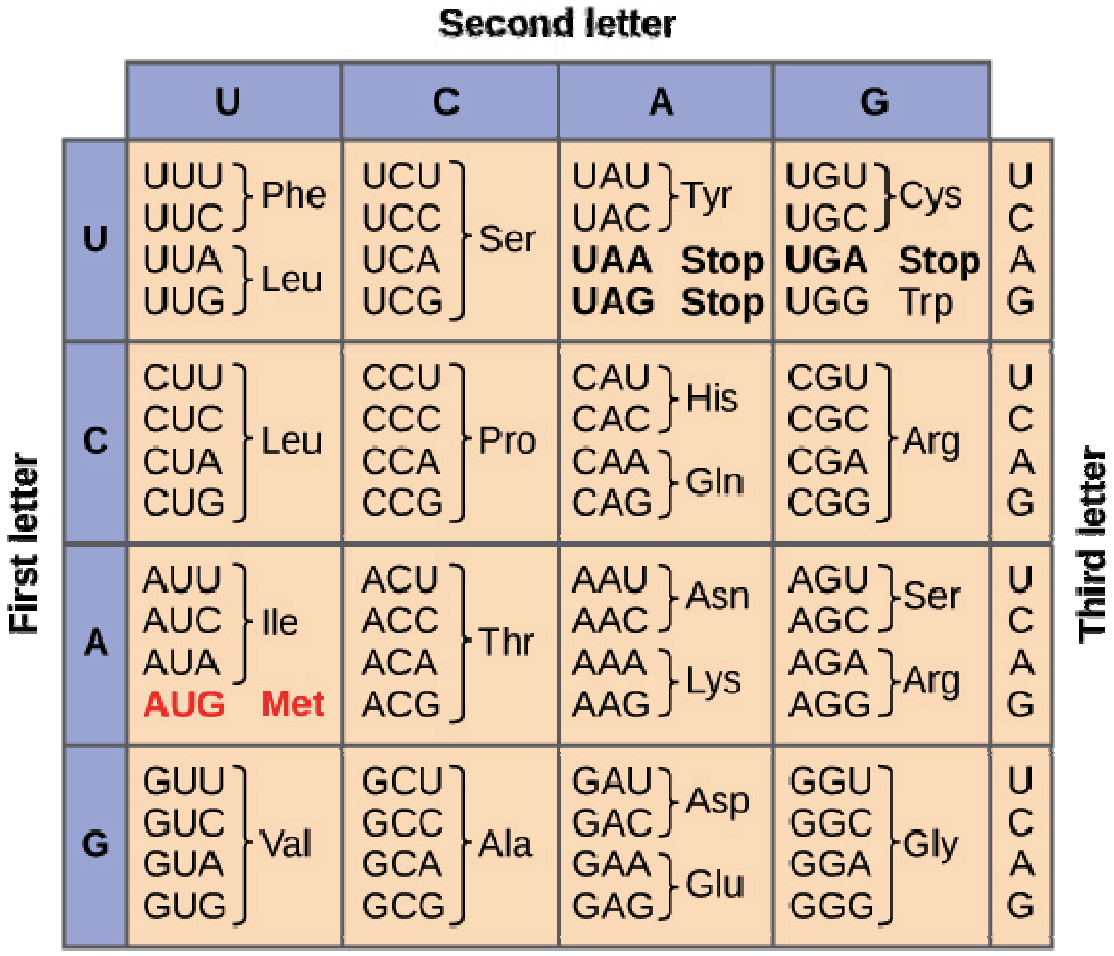


Fig. 1 Genetic codes.

We have observed that the series of codons are strongly linked to the physicochemical properties of the bases A, C, G and U such as number of hydrogen bonds ($A = U$ and $G \equiv C$) and chemical types (pyrimidine $\{A, G\}$ and purine $\{C, U\}$). Three bases form a codon, and the significance of the base varies depending on its position on the codon. In a codon the most biologically significant base is second base. The hydrophobicity as well as hydrophilicity of an amino acid is also associated with the most important base, i.e., the second codon base. The hydrophobic amino acids are associated with the codons having Uracil (U) as second codon base and in case of hydrophilic amino acids the codons have Adenine (A) as second base. In recent times study of Genetic Code with some Mathematical Structures has become one of the popular research topics. At present the structure of the genetic code is quite familiar. For the development of the

genetic code the algebraic pattern is first proposed by Hornos & Hornos (In 1993). In the survey of the genetic code the group theoretical pattern is established by them. Due to symmetry breaking process they could interpret the degeneracy of amino acids. In 2007, Al-Zaharani et al. explained by using analytical way of a group the degeneracy of series of codons in better way. Osawa et al. (1992) discussed that the basic knowledge of the source and evolution of the genetic code should completely rely on the information about the relation between the molecular biochemistry of amino acid synthesis, and repossession from the nucleic acid of the proteins stored design information. But on the other hand, Lacey and Mullinsin (1983) stated that though nature of an evolutionary biochemical structure should reflect its portions, properties from the information of prebiotic origins, but it's not relevant to contemporary systems. To know the correlation of synonymous codon usage and protein structure in *Homo sapiens* and *E. Coli*, Gu et al. had tried too. Jungck (1978) discussed that chemistry between amino acids and their corresponding anti-codon dinucleotides acted strongly on the measure of hydrophobicity/ hydrophilicity or of molecular volume/polarity. The application of graph theory in genetics, as analysis of pedigrees, determination of inbreeding coefficients systematization of the properties of the genetic code (Bertman and Jungck, 1979; Jungck, 1978); protein and sequencing of nucleic acid (Jungck et al., 1982); Also, Gilmore and Hoffman (1964) explained the solution to the Benzer problem which was again explained by another two texts on graph theory namely (Roberts, 1976; Busacker and Saaty, 1965). Recently Ali and Phukan (2012) discussed algebraic and topological structures of molecular biology. They have established some relationship between algebraic and biological aspect of genetic code, e.g., all hydrophilic amino acids characterize zero divisors and inverse of it gives hydrophobic/ stop codons and stop codons form a basis of the whole set of codons which is coded by amino acids and only two codons are either both zero divisor or both non-zero divisors. Different authors have discussed ways of generating graphs from algebraic structures. In a finite group, the power graph is introduced by Cameron and Ghosh (2009), where every vertex are the elements of a group and two vertices are joined if one is a power of the other. Bertholf et al. (1976) identified graphs of finite abelian groups whose vertices are injective with the non- identity subgroups of *G*, and an edge is connected by two vertices iff the corresponding subgroups intersect. In a commutative ring the total graph and a few sub-graphs were established and explored by Anderson and Badawi in 2008. Also, examined subgraphs (induced) specially $Nil(\Gamma(R))$, $Z(\Gamma(R))$, & $Reg(\Gamma(R))$ of $T(\Gamma(R))$, with vertices $Nil(R)$, $Z(R)$ & $Reg(R)$, respectively, where $Nil(R)$ is the ideal of nilpotent elements of R, and $Reg(R)$ is the set of regular elements of R. The lattice structure of the genetic code with some relations & physicochemical properties of amino acids is discussed by Gohain, Ali and Akhtar (2015). Akhtar and Gohain (2015) studied properties of the amino acids and established amino acid network. Also, discussed different centrality measures, correlation coefficient, clustering coefficient and degree of distribution for networks. Akhtar et al. (2015) discussed graph structure in gene algebra. Sanchez et al. (2005) have shown that the set of 64 codons can be equipped with a ring structure isomorphic to the ring of integers modulo 64, Z_{64}. From this algebraic structure they have constructed total graph in genetic code. They have shown that some specific types of mutations in the codon set. The identity graph of a group G is a graph in which the vertex set is the set of all elements of the group and two vertices in G are adjacent if $a.b = e$, where e is the group's identity element. Let H be a subgroup of G then the identity graph drawn for the subgroup H is known as the identity special subgraph of G (special identity subgraph of G).

We attempted to investigate the relationship between Identity graph and genetic code in this paper. The following is the sequence of the paper: In section 2, we go through the graph's basic concept. Section 3

discusses the identity graph derived from the codons, followed by an examination of several centrality measures. In addition, we discuss the bivariate correlation in centrality measures. We go over a few network parameters in Section 4. In section 5, we've highlighted the most important findings from our current study.

2 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 & 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.

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 Degree centrality

Degree centrality of a node u, denoted by $C_d(u)$, is the number of nodes to which u is directly connected (Shams and Khansari, 2014; Xin and Zhang, 2021; Zhang, 2016).

2.2 Eigenvector centrality

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

2.3 Betweenness centrality

Betweenness centrality (Watts and Strogatz, 1998; Zhang, 2016) 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.

2.4 Closeness centrality

Closeness centrality (Shams and Khansari, 2014; Xin and Zhang, 2021; Zhang, 2016) 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.

3 Graph on Genetic Code

Two different orderings of the RNA bases were introduced by Sanchez et al. (2005). 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 (2013) defined a product operation (Table 2) on the base set $Y = \{A, C, G, U\}$, such that Y forms a commutative ring structure. In the ring $(Y, +, \cdot)$; A, C represents additive identity and multiplicative identity respectively. Sanchez also defined an addition operation on the set of 64 codons. These addition operations on the set of codons form a group which is isomorphic to the group $(Z_{64}, +)$ as shown in Table 3. Akhter et al. (2015) discussed total graph on the group of codons. We have identity graph on the same group structure of codons. From this we have discussed special identity subgraph (identity subgraph) 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

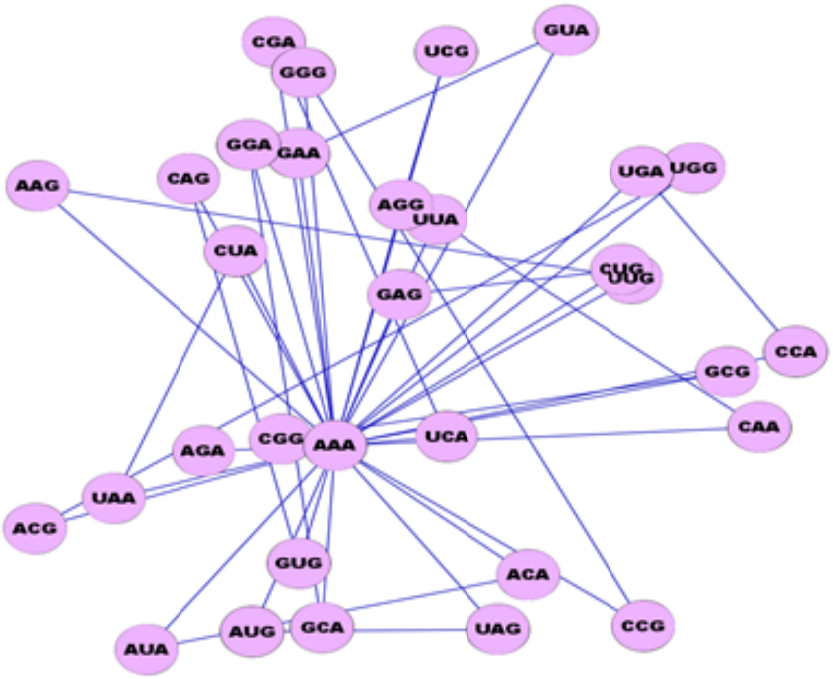
Table 3 The genetic code table of 64 codons.

	A			C			G			U			
	No	Codon	Amino Acid	No	Codon	Amino Acid	No	Codon	Amino Acid	No	Codon	Amino Acid	
A	0	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
C	4	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

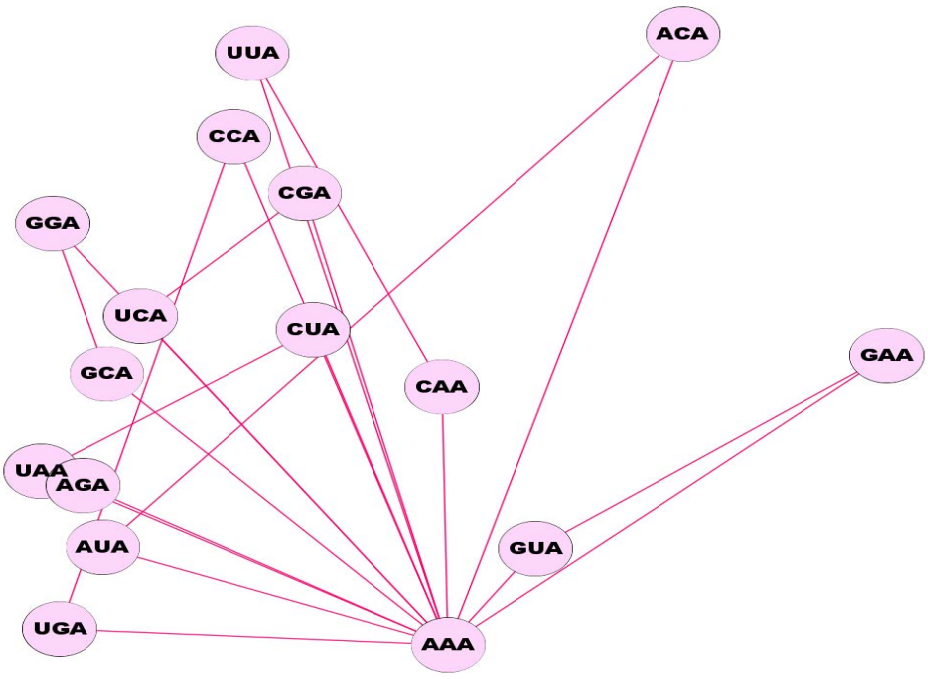
The subgroups of the above codon group are

$$H_1 =< AAG >$$
$$= \{AAG, UUG, CAA, UUA, CAG, GUG, GAA, GUA, AUG, UAG, CUA, UAA, CUG, GAG,$$
$$ACA, AUA, UCA, CGA, UGG, ACG, GGG, CCG, GGA, GCA, GCG, CGG, CCA, UGA, AGG, UCG, AAA, AGA\}$$
$$H_2 =< CAA >$$
$$= \{CAA, UUA, AAA, GAA, GUA, UAA, CUA, ACA, AUA, GCA, GGA, CCA, UGA, CGA, UCA, AGA\}$$
$$H_3 =< GAA >= \{GAA, GUA, AAA, ACA, AUA, GCA, GGA, AGA\}$$
$$H_4 =< ACA >= \{ACA, AUA, AAA, AGA\}$$
$$H_5 =< AGA >= \{AAA, AGA\}$$

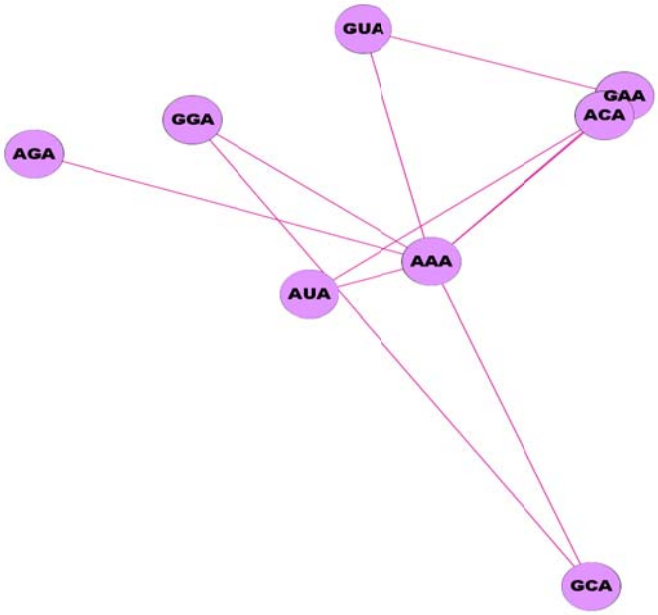
The corresponding special identity subgraphs are shown in Fig. 2.


CGA
GGG
UCG
GUA
GGA
GAA
AAG
CAG
AGG
UUA
UGA
UGG
CUA
CUG
UUG
GAG
CCA
GCG
CAA
AGA
CGG
AAA
UCA
UAA
ACG
GUG
ACA
AUA
AUG
GCA
UAG
CCG

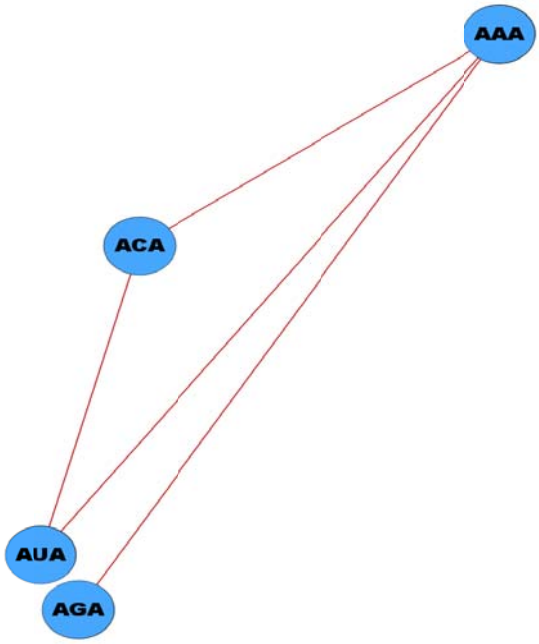
H_1


ACA
UUA
CCA
CGA
GGA
UCA
CUA
GAA
GCA
CAA
UAA
AGA
AUA
GUA
UGA
AAA

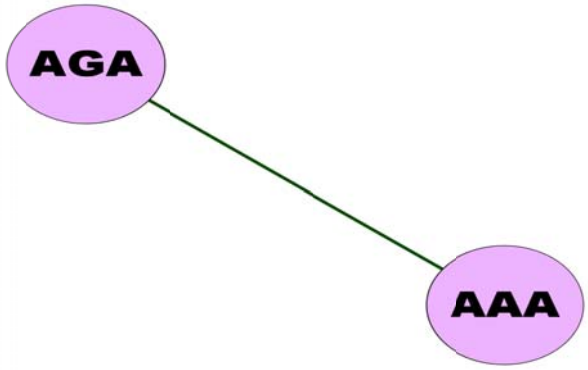
H_2



H_3



H_4



H_5

Fig. 2 Special identity subgraph of Z_{64}.

3.1 Centralities in special identity subgraph

Here, we have computed different centrality measures to analyze the special identity subgraphs H_1, H_2, H_3, H_4 and H_5 (Fig. 2) and displayed all values in Table 4, Table 5, Table 6, Table 7 and Table 8.

Table 4 Centrality values of codons for H_1.

Vertex	Degree Centrality (C_d)	Closeness Centrality (C_{cl})	Betweenness Centrality (C_{bwt})	Eigenvector Centrality (C_λ)
AAG	2	0.516667	0	0.147292
UUG	2	0.516667	0	0.147292
CAA	2	0.516667	0	0.147292
UUA	2	0.516667	0	0.147292
CAG	2	0.516667	0	0.147292
GUG	2	0.516667	0	0.147292
GAA	2	0.516667	0	0.147292
GUA	2	0.516667	0	0.147292
AUG	2	0.516667	0	0.147292
UAG	2	0.516667	0	0.147292
CUA	2	0.516667	0	0.147292
UAA	2	0.516667	0	0.147292
CUG	2	0.516667	0	0.147292
GAG	2	0.516667	0	0.147292
ACA	2	0.516667	0	0.147292
AUA	2	0.516667	0	0.147292
UCA	2	0.516667	0	0.147292

CGA	2	0.516667	0	0.147292
UGG	2	0.516667	0	0.147292
ACG	2	0.516667	0	0.147292
GGG	2	0.516667	0	0.147292
CCG	2	0.516667	0	0.147292
GGA	2	0.516667	0	0.147292
GCA	2	0.516667	0	0.147292
GCG	2	0.516667	0	0.147292
CGG	2	0.516667	0	0.147292
CCA	2	0.516667	0	0.147292
UGA	2	0.516667	0	0.147292
AGG	2	0.516667	0	0.147292
UCG	2	0.516667	0	0.147292
AAA	31	1	450	1
AGA	1	0.508197	0	0.114938

Table 5 Centrality values of codons for H_2.

Vertex	Degree Centrality (C_d)	Closeness Centrality C_{cl}	Betweenness Centrality C_{bwt}	Eigenvector Centrality C_{λ}
CAA	2	0.535714	0	0.256468
UUA	2	0.535714	0	0.256468
AAA	15	1	98	1
GAA	2	0.535714	0	0.256468
GUA	2	0.535714	0	0.256468
UAA	2	0.535714	0	0.256468
CUA	2	0.535714	0	0.256468
ACA	2	0.535714	0	0.256468
AUA	2	0.535714	0	0.256468
GCA	2	0.535714	0	0.256468
GGA	2	0.535714	0	0.256468
CCA	2	0.535714	0	0.256468
UGA	2	0.535714	0	0.256468
CGA	2	0.535714	0	0.256468
UCA	2	0.535714	0	0.256468
AGA	1	0.517241	0	0.189126

Table 6 Centrality values of codons for H_3.

Vertex	Degree Centrality (C_d)	Closeness Centrality (C_{cl})	Betweenness Centrality (C_{bwt})	Eigenvector Centrality (C_λ)
GAA	2	0.583333	0	0.443242
GUA	2	0.583333	0	0.443242
AAA	7	1	18	1
ACA	2	0.583333	0	0.443242
AUA	2	0.583333	0	0.443242
GCA	2	0.583333	0	0.443242
GGA	2	0.583333	0	0.443242
AGA	1	0.538462	0	0.295473

Table 7 Centrality values of codons for H_4.

Vertex	Degree Centrality (C_d)	Closeness Centrality (C_{cl})	Betweenness Centrality (C_{bwt})	Eigenvector Centrality (C_λ)
ACA	2	0.75	0	0.837194
AUA	2	0.75	0	0.837194
AAA	3	1	2	1
AGA	1	0.6	0	0.459951

Table 8 Centrality values of codons for H_5.

Vertex	Degree Centrality (C_d)	Closeness Centrality (C_{cl})	Betweenness Centrality (C_{bwt})	Eigenvector Centrality (C_λ)
AAA	1	1	0	1
AGA	1	1	0	1

All other codons can easily interact with a codon with a high closeness centrality value. As a result, the evolutionary process uses it to communicate quickly with the remaining codons. The closeness centrality value of the codon AAA is 1, according to Tables 4, 5, 6, and 7. As a result, we can assume that the flow of evolutionary information continues at a similar rate through AAA. The betweenness centrality assesses the codon's contribution to expressing the evolutionary mechanism. A codon with a high betweenness centrality value represents the identification of codons responsible for the majority of the network's information flow. For example, the betweenness centralities for the codon AAA are 450, 98, 18, and 2, whereas the betweenness centralities for other codons are 0 (Table 4, Table 5, Table 6 and Table 7). Thus, AAA is related to more pairs of codons through the evolutionary mechanism than other codons, i.e., codon AAA appears as an intermediate between more pairs of codons than others. In a network, eigenvector centrality appears to be more active and prominent than degree centrality. A node is considered large if it has a large number of neighbours and/or important neighbours. Because the sum of the codon AAA's direct and indirect links is the

greatest, the codon AAA has the greatest eigenvector centrality (Table 4, Table 5, Table 6 and Table 7). With the exception of the codon's AAA and AGA, the eigenvector centrality of the other codons is equal in magnitude because these codons share the same neighbours (Table 4, Table 5, Table 6 and Table 7). 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.

3.2 Correlation of different centralities

In this section, we looked at the correlation coefficients between different measures of centrality in special identity subgraph networks. Correlation analysis is possible when there is at least one relationship between two variables. The term correlation refers to the relationship between two variables in which when the values of one variable change, the values of the other variable change as well. The Karl Pearson coefficient of correlation is defined as $r = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{n\sigma_x\sigma_y}$ where σ_x and σ_y are the standard deviations of the X and Y series, respectively. The value of r ranges from $+1$ and -1. Correlation is the most important feature to investigate in assortative or disassortative networks. If the correlation value is greater than zero $(r > 0)$, the network is assortative; if the correlation value is less than zero $(r < 0)$, the network is disassortative (Newman, 2002). Tables 9, 10, 11, 12, and 13 show the correlation coefficients between the centrality measures of the special identity subgraphs H_1, H_2, H_3, H_4, and H_5. Pearson's method is used to compute all correlation coefficients.

Table 9 Correlation coefficients of the centrality measures for H_1.

	C_d	C_{cl}	C_{bwt}	C_λ
C_d	1	0.999857	0.999408	0.999994
C_{cl}	0.999857	1	0.999847	0.999793
C_{bwt}	0.999408	0.999847	1	0.999283
C_λ	0.999994	0.999793	0.999283	1

Table 10 Correlation coefficients of the centrality measures for H_2.

	C_d	C_{cl}	C_{bwt}	C_λ
C_d	1	0.999329	0.997097	0.999911
C_{cl}	0.999329	1	0.999217	0.99875
C_{bwt}	0.997097236	0.999217061	1	0.995989847
C_λ	0.999910574	0.998749643	0.995989847	1

Table 11 Correlation coefficients of the centrality measures for H_3.

	C_d	C_{cl}	C_{bwt}	C_λ
C_d	1	0.996344	0.981981	0.998332
C_{cl}	0.996344	1	0.994536	0.98975
C_{bwt}	0.981981	0.994536	1	0.969433
C_λ	0.998332	0.98975	0.969433	1

Table 12 Correlation coefficients of the centrality measures for H_4.

	C_d	C_{cl}	C_{bwt}	C_λ
C_d	1	0.984732	0.816497	0.962771
C_{cl}	0.984732	1	0.904534	0.901015
C_{bwt}	0.816497	0.999966	1	0.630031
C_λ	0.962771	0.901015	0.630031	1

Table 13 Correlation coefficients of the centrality measures for H_5.

	C_d	C_{cl}	C_{bwt}	C_λ
C_d	0	0	0	0
C_{cl}	0	0	0	0
C_{bwt}	0	0	0	0
C_λ	0	0	0	0

Tables 9, 10, 11, and 12 show that all of the centrality measures for all of the special identity subgraph networks are highly correlated. As a result, these centrality measures, which represent various centrality features, are closely related in these networks. As a result, using any measure is equivalent to using any other. It is well known that information can be transferred more easily through an assortative network than through a disassortative network (Newman, 2002). Furthermore, we can see from the above correlation coefficient that all three networks are assortative types $(r > 0)$, implying that evolutionary information flow will be simple.

4 Network Parameters

In biological networks, various network parameters are used. We've talked about the clustering coefficient, the degree of distribution, and Pearson's skewness in this section.

4.1 Clustering coefficient

Clustering coefficient is defined as the capacity of a graph to be divided into clusters (Zhang, 2018). 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 $'i'$). 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 H_1, H_2, H_3, H_4, and H_5 are given in Tables 14, 15, 16, 17 and 18, respectively.

Table 14 Clustering coefficient of the codons for H_1.

AAG	1
UUG	1
CAA	1
UUA	1
CAG	1
GUG	1
GAA	1
GUA	1
AUG	1
UAG	1
CUA	1
UAA	1
CUG	1
GAG	1
ACA	1
AUA	1
UCA	1
CGA	1
UGG	1
ACG	1
GGG	1
CCG	1
GGA	1
GCA	1
GCG	1
CGG	1
CCA	1
UGA	1
AGG	1
UCG	1
AAA	0.032258
AGA	0

Table 15 Clustering coefficient of the codons for H_2.

CAA	1
UUA	1
AAA	0.066667
GAA	1
GUA	1
UAA	1
CUA	1
ACA	1
AUA	1
GCA	1
GGA	1
CCA	1
UGA	1
CGA	1
UCA	1
AGA	0

Table 16 Clustering coefficient of the codons for H_3.

GAA	1
GUA	1
AAA	0.142857
ACA	1
AUA	1
GCA	1
GGA	1
AGA	0

Table 17 Clustering coefficient of the codons for H_4.

ACA	1
AUA	1
AAA	0.333333
AGA	0

Table 18 Clustering coefficient of the codons for H_5.

AAA	0
AGA	0

An amino acid's clustering coefficient is determined by the degree of the amino acids as well as the number of direct connections between two neighbouring amino acids. Tables 14, 15, 16 and 17 show that, with the

exception of the codon's AAA and AGA, all other codons have degree 2. Furthermore, the number of links between neighbouring codons is one, implying that, with the exception of AAA and AGA, all codons have a high clustering coefficient, i.e., 1. The clustering coefficients for the entire special identity subgraphs H_1, H_2, H_3 and H_4 are 0.938508125, 0.879166875, 0.767857125 and 0.58333325, respectively. Thus, after examining the clustering coefficients of the special identity subgraphs and the clustering coefficients of the codons, we discovered that, with the exception of AAA and AGA, the flow of evolutionary process in the neighbourhood of other codons is comparatively slow when compared to the entire special identity subgraph.

4.2 Degree of distribution and skewness

The degree of distribution and Pearson's skewness of the codons will be discussed in this section. 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. Skewness is another crucial statistical characteristic. The measure of the distribution's symmetry or asymmetry is used to determine skewness. Karl Pearson first proposed the skewness concept in 1895. It's abbreviated as S_k. Skewness can be positive or negative, depending on the mean and median. We employed the Karl Pearson's skewness coefficient, defined as

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

in our research.

In the case of symmetrical (i.e., normal) distribution $S_k = 0$. If $S_k > 0$, the distribution is positively skewed. If $S_k < 0$, we consider the distribution to be negatively skewed.

We have shown the degree of distribution values of all the codons for special identity subgraphs H_1, H_2, H_3, H_4, and H_5 in Tables 19, 20, 21, 22, and 23.

Table 19 Degree of distribution of the codons for H_1.

AAG	0.9375
UUG	0.9375
CAA	0.9375
UUA	0.9375
CAG	0.9375
GUG	0.9375
GAA	0.9375
GUA	0.9375
AUG	0.9375
UAG	0.9375
CUA	0.9375
UAA	0.9375
CUG	0.9375
GAG	0.9375
ACA	0.9375
AUA	0.9375
UCA	0.9375
CGA	0.9375

UGG	0.9375
ACG	0.9375
GGG	0.9375
CCG	0.9375
GGA	0.9375
GCA	0.9375
GCG	0.9375
CGG	0.9375
CCA	0.9375
UGA	0.9375
AGG	0.9375
UCG	0.9375
AAA	0.03125
AGA	0.03125

Table 20 Degree of distribution of the codons for H_2.

CAA	0.875
UUA	0.875
AAA	0.0625
GAA	0.875
GUA	0.875
UAA	0.875
CUA	0.875
ACA	0.875
AUA	0.875
GCA	0.875
GGA	0.875
CCA	0.875
UGA	0.875
CGA	0.875
UCA	0.875
AGA	0.0625

Table 21 Degree of distribution of the codons for H_3.

GAA	0.75
GUA	0.75
AAA	0.125
ACA	0.75
AUA	0.75
GCA	0.75
GGA	0.75
AGA	0.125

Table 22 Degree of distribution of the codons for H_4.

ACA	0.5
AUA	0.5
AAA	0.25
AGA	0.25

Table 23 Degree of distribution of the codons for H_5.

AAA	1
AGA	1

Pearson's coefficients of skewness for H_1, H_2, H_3, H_4 and H_5 are -0.7624, -1.09789, -1.62019, 0 and 0 respectively, based on the above degree of distributions. We can conclude that the codon networks in H_1, H_2, and H_3 have a negatively skewed distribution, whereas the codon networks in H_4 and H_5 have a symmetrical distribution.

5 Conclusion

We attempted to decipher the genetic code's special identity subgraph structure. To investigate the impact of each codon, various centrality measures were used as a graph theoretic tool to delve deep into the subject. Following a discussion of several centrality measures, it is discovered that Codon AAA has the highest centrality value of all centrality measures (Table 4, Table 5, Table 6 and Table 7). 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 (Table 9, Table 10, Table 11 and Table 12). When the clustering value of the codons is examined, it is clear that, with the exception of codons AAA and AGA, all of the codons have a high clustering coefficient. As a result, with the exception of AAA and AGA in the vicinity of other codons, the evolutionary process is quite slow when compared to the entire network (Table 14, Table 15, Table 16 and Table 17).

References

Anderson DF, Badawi A. 2008. The total graph of a commutative ring. Journal of Algebra, 320: 2706-2719

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

Akhtar A, Ali T, Gohain N. 2015. Graph in genetic code algebra. Journal of Bioinformatics and Intelligent Control, 4(1): 1-4

Antoneli F, Braggion L, Forger M, Hornos JEM. 2003. Extending the search for symmetries in the genetic code. International Journal of Modern Physics B, 17: 3135-3204

Balakrishnan J. 2002. Symmetry scheme for amino acid codons. Physical Review E, 65: 021912-5

Bashford JD, Jarvis PD. 2000. The genetic code as a periodic table: Algebraic aspects. Biosystems, 57: 147-161

Bashford JD, Tsohantjis I, Jarvis PD. 1998. A supersymmetric model for the evolution of the genetic code. Proceedings of the National Academy of Sciences USA, 95: 987-992

Beland P, Allen T. 1994. The origin and evolution of the genetic code. Journal of Theoretical Biology, 170: 359-365

Bertholf DS, Walls GH. 1976. Graph of finite abelian groups. Czechoslovak Mathematical Journal, 28: 103

Bertman MO, Jungck JR. 1979. Group graph of the genetic code. Journal of Heredity, 70: 379-384

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

Cameron JP, Ghosh S. 2011. The power graph of a finite group. Discrete Mathematcs, 311: 1220-1222

Gohain N, Ali T, Akhtar A. 2015. Lattice structure and distance matrix of genetic code. Journal of Biological Systems, 23: 485-504

Hornos JEM, Hornos YMM. 1993. Algebraic model for the evolution of the genetic code. Physical Review Letters, 71(26): 4401-4404

Jiménez-Montaño MA. 1996. The hypercube structure of the genetic code explains conservative and non-conservative amino acid substitutions in vivo and in vitro. Biosystems, 39: 117-125

Jungck JR. 1978. The genetic code as a periodic table. Journal of Molecular Evolution, 11: 211-224

Jungck JR, Dick G, Dick AG. 1982. Computer-assisted sequencing, interval graphs, and molecular evolution. BioSystems, 15: 259-273

Karasev VA, Stefanov VE. 2001. Topological Nature of the Genetic Code. Journal of Theoretical Biology, 209: 303-317

Osawa S, Jukes TH, Watanable K, Myto A. 1992. Recent evidence for evolution of the genetic code. Microbial Rev, 56: 229-264

Phukan CK. 2012. A study on algebraic and topological structures in some aspects of molecular biology. Ph.D thesis. Dibrugarh University, India

Purcell EM. 1977. Life at low Reynolds number. American Journal of Physics, 45(1): 3-11

Roberts FS. 1976. Discrete Mathematical Models with Applications to Social, Biological, and Environmental Problems. 111-140, Prentice-Hall, Engelwood Cliffs, NJ, USA

Sanchez R, Morgado E, Grau R. 2005. Gene Algebra from a genetic Code algebra structure. Journal of Mathematical Biology, 51: 431-457

Sánchez R, Morgado E, Grau R. 2004a. The Genetic Code Boolean Lattice. MATCH Communications in Mathematical and in Computer Chemistry, 52: 29-46

Sánchez R, Morgado E, Grau R. 2004b. A Genetic Code Boolean Structure. I. The Meaning of Boolean Deductions. Bulletin of Mathematical Biology, 67(1): 1-14

Shams B, Khansari M. 2014. Using network properties to evaluate targeted immunization algorithms. Network Biology, 4(3): 74-94

Siemion IZ, Siemion PJ, Krajewski K. 1995. Chou-Fasman conformational amino acid parameters and the genetic code. Biosystems, 36: 231-238

Stadler PF, Stephan K, Cupal J. 1999. RNA shape space in topology. Artificial Life, 3-23

Xin SH, Zhang WJ. 2021. Construction and analysis of the protein-protein interaction network for the detoxification enzymes of the silkworm, *Bombyx mori*. Archives of Insect Biochemistry and Physiology, 108(4): e21850

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