Article

Analysis of Z_{64} genetic code network

Adil Akhtar[1], **Nasrin Irshad Hussain**[2]
[1]Department of Mathematics, Golaghat Engineering College, Golaghat, Assam-785621, India;
[2]Department of Computer Application, Sibsagar University, Assam-785665, India
E-mail: adil.akhtar19@gmail.com, hussainnasrin531@gmail.com

Received 27 November 2024; *Accepted 31 December 2024*; *Published online 15 January 2025*; *Published 1 September 2025*

Abstract
The genetic code is the set of rules defining how the sequence of nucleotides in DNA or RNA determines the specific amino acid sequence in the synthesis of protein. Proteins are the basic functional elements of living organisms and amino acids are the building blocks of proteins. Each protein is formed by a linear chain of amino acids. The DNA consists of two complementary long chains of nucleotides, viz. Adenine (A), Cytosine (C), Guanine (G) or Thymine (T) (Uracil (U) in case of RNA). Three consecutive DNA nucleotides form a codon. Each codon specifies a particular amino acid. The set of 64 codons can be equipped with a ring structure isomorphic to the ring of integers modulo 64 (Z_{64}). Different graph structures can be generated from a ring. In this paper we have discussed total graph in this ring. Total graph of a ring R is an undirected graph where vertex set is the set of all elements of the ring and for distinct $x, y \in R$, the vertices x and y are adjacent if $x+y \in Z(R)$, where $Z(R)$ be the set of zero divisor of R. In this manuscript using signed graph we have try to explain that the graph of the genetic code is an unbalanced graph.

Keywords amino acid; genetic code; signed graph; total 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

Proteins are prior macromolecules in living organisms, virtually exist in all biological activities (Xin and Zhang, 2021; Yang and Zhang, 2022; Sun and Zhang, 2023). The genetic codes of an organism are stored in DNA and are transcribed into messenger RNA (mRNA) during protein synthesis. Sánchez et al. (2005) arranged DNA bases according to their physico-chemical properties. As a result, they got two orders {A, C, G, U} and {U, G, C, A} in these base set arrays which made it possible to define a sum operation in several ways. From these, they obtained two cyclic groups isomorphic to Z_4 (group Z_4 of integer's module 4). Recently Ali and Phukan (2012) discussed algebraic and topological structures of genetic code. They have represented the genetic code as $Z_4 \times Z_4 \times Z_4$ and established some relationship between algebraic and biological aspect of genetic

code. In this paper we attempt to study graph theoretic aspect of the genetic code with respect to the algebraic structure discussed by Sanchez.

Different types of graph structures can be introduced corresponding to a given algebraic structure. Cameron and Ghosh (2011) introduced the power graph of a finite group. The power graph of a group is the graph whose vertex set is the group, and two elements being adjacent if one is a power of the other. Dennts Bertholf et al. (1976) introduced graphs of finite abelian groups whose vertices are in one-to-one correspondence with the non-identity subgroups of *G*, and two vertices are joined by an edge iff the corresponding subgroups intersect. Anderson and Badawi (2008) introduced the total graph of a commutative ring and is denoted by $T(\Gamma(R))$, where *R* is the commutative ring. It is the (undirected) graph with all elements of *R* as vertices, and for distinct $x, y \in R$, the vertices *x* and *y* are adjacent iff $x + y \in Z(R)$, where *Z*(*R*) is the set of zero-divisors of *R*. They also discussed the three (induced) subgraphs namely, $Nil(\Gamma(R))$, $Z(\Gamma(R))$, and $Reg(\Gamma(R))$ of $T(\Gamma(R))$, with vertices *Nil*(*R*), *Z*(*R*) and *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*. Beck(1988) introduced the concept of the graph of the zero divisors of *R*, where he was mainly interested in colorings. In his work all elements of the ring were vertices of the graph. The investigation of colorings of a commutative ring was then continued by Anderson and Naseer in 1993. In 1999, D. F. Anderson and Livingston associate a graph, Γ(*R*), to *R* with vertices *Z*(*R*) \ {0}, where *Z*(R) is the set of zero-divisors of *R*, and for distinct $x, y \in Z(R) \setminus \{0\}$, the vertices *x* and *y* are adjacent if and only if $xy = 0$. In 2009 Akbari et al. proved that the total graph is a Hamiltonian graph if this graph is connected. We have discussed total graph in the genetic code algebra. Then we have transformed this total graph to a signed graph by defining positive and negative edge.

2 Genetic Code Algebra

All living organisms consist of cells. In each cell there is the same set of chromosomes. Chromosomes are strings of DNA and serves as a model for the whole organism. A chromosome consists of genes, block of DNA. Each gene encodes a particular protein. Proteins are the basic constructional blocks and functional elements of living organisms. Amino acids are the building blocks of proteins. Each protein formed by a linear chain of amino acid. There are 20 different amino acids being found till now that occurs in proteins. Each amino acid is a triplet code of four possible bases. The chain of amino acids takes on different shapes to form different proteins.

Deoxyribonucleic acid (DNA) stores genetic information about how to construct or synthesize proteins. It is made up of unit called nucleotides. Each nucleotide contains one sugar group (deoxyribose), one phosphate group and one nitrogenous base (Adenine (A), Cytosine(C), Guanine (G) or Thymine (T)). The sugar and phosphate group are responsible for the helical backbone of DNA. The purines (Adenine and Guanine) and the pyrimidines (Cytosine and Thymine) form the double helix. The bases are paired and joined together by hydrogen bonds. Two hydrogen bonds attach Adenine (A) to Thymine (T) and three hydrogen bonds attach Cytosine (C) to Guanine (G). Therefore, DNA consists of two complementary long chains of nucleotides. According to Watson-Crick, a purine of one chain is always paired with a pyrimidine of the other. The sequence of one side is enough to deduce the other. Mathematically, DNA can be considered as a sequence of four letters: A, G, C, and T. A sequence of three bases forms a unit called codon. A codon specifies one amino acid. The genetic code is a series of codons that specify which amino acids are required to make up specific protein. As there are four bases, this gives us 64 codons. Out of these 64, the three triplets UAA, UAG and

UGA are known as stop codons or nonsense codons and their role is to stop the biosynthesis. The codon AUG codes for the initiation of the translation process and is therefore also known as start codon.

Sánchez et al. (2005) arranged DNA bases according to their physico-chemical properties. As a result, they got two orders {A, C, G, U} and {U, G, C, A} in these base set arrays made it possible to define a sum operation in several ways. From these, they obtained two cyclic groups isomorphic to Z_4 (group Z_4 of integer's module 4). In Table 1 A and B we show the sum tables of bases obtained from the two possible orders. Therefore, there are two cyclic groups: the primal and the dual group, corresponding to the ordered sets {A, C, G, U} and {U, G, C, A}. They are also able to establish the genetic code tables starting with the codon having the fewest hydrogen bonds. So, two possible codon set arrangements are induced. In the tree-entry tables, the order (0 to 63) is read in the following direction: the second base corresponding to 4 essential columns, the first base to 4 essential rows and the third base to sub-rows (Table 2).

Table 1 Sum Operation tables defined in the set of four bases of the DNA.

A: Primal algebra.

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

B: Dual algebra

+	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 The primal genetic code table induced by the primal order {A, C, G, T}. The bijection between the primal genetic code Abelian group with Z_{64} is also shown in the table.

	A			**C**			**G**			**U**			
	No	(1)	(2)	No	(1)	(2)	No	(1)	(2)	No	(1)	(2)	
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**

[(1)]The base triplets (codons). [(2)]The one letter symbol of amino acids; "-"corresponds to stop codons.

3 Signed Graph

A graph G is defined usually by a finite set of vertices (or nodes) V and the set E of edges denoted by $G = (V, E)$. In general, the nodes represent objects or entities while the edges represent relationship between the nodes. A signed graph is simply a graph where each edge between the nodes is labeled as either positive (+) or negative (–). It is useful for representing a symmetric binary relation between any two quantities. The concept of signed graph is used if the relationship between each pair of nodes is symmetric. With the help of signed graph we may represent the interpersonal relationships between groups of individuals. The simplest approach to study such a group of individuals is to draw a graph in which the individuals are the nodes and in which there is an edge for 'x' to node 'y' if x is in some relation to y. This relationship may be like or dislike, associate with or avoids and so on. We can include two different relationship in a graph by using two different signs i.e. positive (+) and negative (–) to distinguish them. Then the (+) sign of an edge represent that there is a relationship between the nodes and the indication of a positive relation such as like, agree, etc. and the (–) sign represents the other relation such as dislike, disagree, hate etc. which lead to the existence of signed graph.

A graph is said to be balanced or stable (Roberts, 1978) if all the cycle of the signed graph are positive (+). Similarly, a graph is called unbalanced if there exists at least one cycle which is negative (-). Here it means that for a negative cycle, there is an odd number of negative edges in the cycle while a positive cycle has an even number of negative edges in the cycle.

In the following Fig. 1 gives a balanced graph and Fig. 2 gives an unbalanced graph.

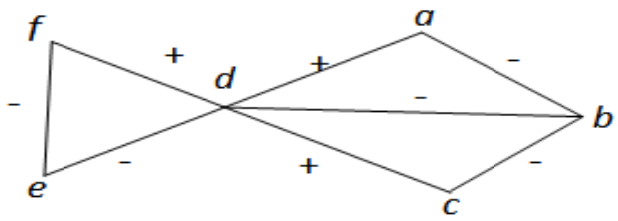

Fig. 1 A balanced graph.

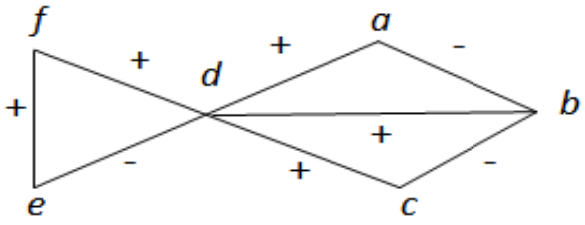

Fig. 2 An unbalanced graph.

We state the following theorem that will be required in the sequel.

Theorem 1 (Harary, 1954): A signed graph G, is called balanced if and only if its vertex set V can be separated into two disjoint subsets namely V_1 and V_2, in such a way that each positive line of G joins two points of the same subset and each negative line joins two points of different subsets.

4 Graph in Genetic Code Algebra

In Z_{64}, $Z(Z_{64}) = \{2,4,8,16,32\}$ is the set of zero divisors. To consider total graph in Z_{64} we define two codons to be adjacent iff their sum is zero divisor. Using Table 2 we obtain the total graph of Z_{64} as in Fig. 3.

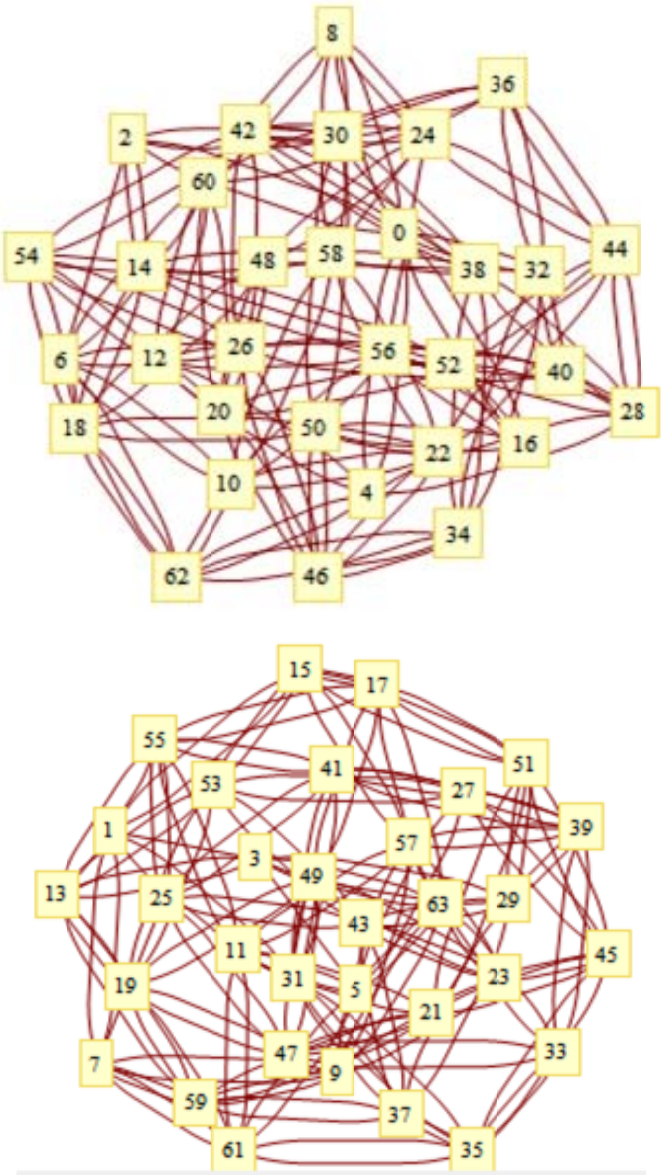

Fig. 3 Total graph of Z_{64}.

From the total graph of Z_{64} we observe that there are two disjoint graphs: one containing vertices representing even numbers (V_E) and another containing vertices representing odd numbers (V_O). So the total graph is disconnected. However the two disjoint graphs are individually connected.

Next we consider signed graph to the genetic code algebra.

For that we assign positive sign to each edge of the above considered total graph, and for every pair of vertices that were not connected under the total graph, we assign a negative edge. Then we have the signed graph of the genetic code as:

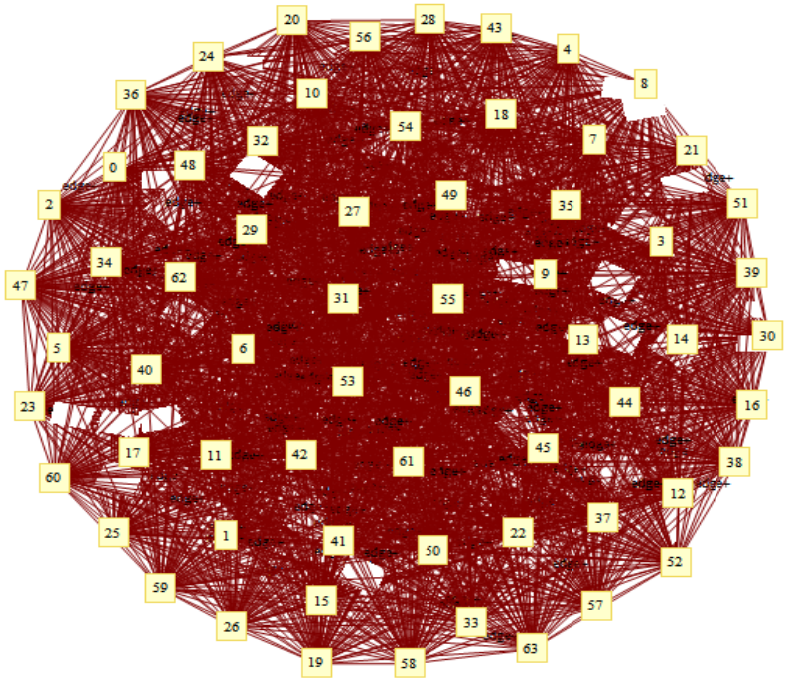

Fig. 4 Signed graph of Z_{64}. Edge (+)ve means (+)ve edge.

Following pairs of vertices represent (+)ve edges.

V={{0,2}{0,4},{0,8},{0,16},{0,32},{1,3},{1,7},{1,15},{1,31},{2,6},{2,14},{2,30},
{3,5},{3,13},{3,29},{3,63},{4,12},{4,28},{4,62},{5,11},{5,27},{5,61},{5,63},{6,10},
{6,26},{6,60},{6,62},{7,9},{7,59},{7,61},{8,24},{8,58},{8,60},{9,23},{9,57},{9,59},
{9,63},{10,22},{10,56},{10,58},{10,62},{11,21},{11,55},{11,57},{11,61},{12,20},
{12,54},{12,56},{12,60},{13,19},{13,53},{13,55},{13,59},{14,18},{14,52},{14,54},
{14,58},{15,17},{15,51},{15,53},{15,57},{16,50},{16,52},{16,56},{17,49},{17,51},
{17,55},{17,63},{18,48},{18,50},{18,54},{18,62},{19,47},{19,49},{19,53},{19,61},
{20,46},{20,48},{20,52},{20,60},{21,45},{21,47},{21,51},{21,59},{22,44},{22,46},
{22,50},{22,58},{23,43},{23,45},{23,49},{23,57},{24,42},{24,44},{24,48},{24,58},
{25,41},{25,43},{25,47},{25,55},{26,40},{26,42},{26,46},{26,54},{27,39},{27,41},
{27,45},{27,53},{28,38},{28,40},{28,44},{28,52},{29,37},{29,39},{29,43},{29,51},
{30,36},{30,38},{30,42},{30,50},{31,35},{31,37},{31,41},{31,49},{32,34},{32,36},
{32,40},{33,35},{33,39},{33,47},{33,63},{34,38},{34,46},{34,62},{35,37},{35,45},
{35,61},{36,44},{36,60},{37,43},{37,59},{38,42},{38,58},{39,41},{39,57},{40,56},
{41,55},{42,54},{43,53}, {44,52},{45,51},{46,50},{47,49}}.

We will to show that whether the graph is balanced or unbalanced. If possible let the vertex set V be separated into two disjoint subset say V_1 and V_2, satisfying the condition of Theorem 1. Suppose $0 \in V_1$, then we must have $1 \in V_2$. Now by Theorem 1, $\{2,4,8,16,32\} \subseteq V_1$ and $\{3,7,15,31\} \subseteq V_2$. From here we see that relation between any element of V_1 with any element of V_2 is negative. But edges between the elements 3 and $7 \in V_2$ are negatively connected. Hence the pair (V_1, V_2) does not satisfy Theorem 1. Since (V_1, V_2) is an arbitrary pair,

we conclude that no pair of disjoint vertices will satisfy Theorem 1. Consequently signed graph of Z_{64} is an unbalanced graph.

5 Conclusion

The paper proposes the graphical representation of the genetic code. From the total graph of Z_{64} we observe that there are two disjoint graphs where one containing vertices representing even numbers and another containing vertices representing odd numbers. So the total graph is disconnected. However the two disjoint graphs are individually connected. Then by introducing signed graph we have established that the genetic code algebra is an unbalanced graph. As we know that for a balanced graph sharing the genetic information is efficient and easier in comparison to an unbalanced graph. From the above discussion it may conclude that the total graph of genetic code is unbalanced and the efficiency of sharing genetic information is less.

References

Akbari S, Kiani, D, Mohammadi F, Moradi S. 2009. The total graph and regular graph of a commutative ring. Journal of Pure and Applied Algebra, 213(12): 2224-2228

Anderson DD, Naseer M. 1993. Beck′ s coloring of a commutative ring. Journal of Algebra, 159(2): 500-514

Anderson DF, Badawi A. 2008. The total graph of a commutative ring. Journal of Algebra, 320(7): 2706-2719

Anderson DF. 2011. Zero-divisor Graphs In Commutative Rings. Commutative Algebra, Noetherian and Non-Noetherian Perspectives. Springer-Verlag

Banu PS. 2024. Applications of nondeterministic zerodivisor graph. Communications on Applied Nonlinear Analysis, 31(2s): 173

Beck I. 1988. Coloring of commutative rings. Journal of Algebra, 116(1): 208-226

Bertholf D, Walls G. 1978. Graphs of finite Abelian groups. Czechoslovak Mathematical Journal, 28(3): 365-368

Boruah BK, Ali T, Saikia S. 2022. A study on the identity graph in genetic code. Asian Journal of Biological and Life Sciences, 11(1): 173

Cameron PJ, Ghosh S. 2011. The power graph of a finite group. Discrete Mathematics, 311(13): 1220-1222

Harary F. 1953. On the notion of balance of a signed graph. Michigan Mathematical Journal, 2(2): 143-146

Hussain NI, Boruah K. 2024. Analysis of amino acids network based on graph mining. Network Biology, 14(3): 242

Hussain NI, Boruah K, Akhtar A. 2025. Predicting evolutionary importance of amino acids through mutation of codons using *k*-means clustering. Journal of Electronics, Electromedical Engineering, and Medical Informatics, 7(1): 13-26

Riaz A, Kousar S, Kausar N, Pamucar D, Addis GM. 2022. Research Article An Analysis of Algebraic Codes over Lattice Valued Intuitionistic Fuzzy Type-3 R-Submodules.

Roberts FS. 1978. Graph theory and its applications to problems of society. Society for Industrial and Applied Mathematics

Rogers M. 2022. The Pandoras Box Congress: 140 Scientists Ask: Now that We Can Rewrite the Genetic Code, What Are We Going To Say? In: The Ethics of Biotechnology. 215-224, Routledge, USA

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

Shahid Z, Simpson B, Miao KH, Singh G. 2023. Genetics, histone code. In: StatPearls [Internet]. StatPearls

Publishing

Sun JN, Zhang WJ. 2023. Construction and analysis of the protein-protein interaction network of visual system in Drosophila. Network Biology, 13(4): 155-185

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

Yang S, Zhang WJ. 2022. Systematic analysis of olfactory protein-protein interactions network of fruitfly, *Drosophila melanogaster*. Archives of Insect Biochemistry and Physiology, 110(2): e21882