Mathematical Forum
Vol.28(1), 2020
ISSN: 0972-9852

SIGNED GRAPH APPROACH TO STUDY AMINO ACIDS NETWORKS

Arun Kumar Baruah[a], Pranjal Kumar Bora[*b]

[a]Department of Mathematics, Dibrugarh University, Dibrugarh- 786004, India

[*b]Centre for Computer Science and Applications, Dibrugarh University, Dibrugarh-786004, India
Email: arun@dibru.ac.in , pranjaly2k@gmail.com

Abstract

In the process of protein evolution, if two amino acids have more similar common properties the chance of substitution or mutation occurs between them will increase with a decrease in physico-chemical distance. Again, if two amino acids have more common properties, they will be structurally more similar and hence they will have always a tendency to evolve into another in the process of evolution. There are 20 amino acids found till now in protein and each of them has different physico-chemical properties due to the variation in the structure of the R group. Here in this manuscript, we have applied the signed graph approach to study the evolutionary aspects of amino acids and hence the protein structure. We have considered mainly four physico-chemical properties to study the structural similarity of amino acids. Finally, from the balance property point of a signed graph, we have discussed the relationship between the evolutions with structural similarities among amino acids.

Keywords: Mutation, Physico-chemical properties, Balance theory

010 AMS Classification: 92B10, 05C90

1. Introduction

The basic building blocks of proteins are amino acids and each protein is formed by a long linear chain of amino acids.20 different amino acids are being found till now which take part in proteins formation. Each amino acid is a triplet code of four possible bases which are Adenine (A), Guanine (G), Cytosine (C), and Thymine (T) and a sequence of three bases form a unit called a codon. All the 20 amino acids have different physico-chemical properties which are mainly due to the change in R group attached to the α-carbon. This property is important for structural analysis of each amino acid which plays an important role in protein formation [17].

Different structures of amino acids are due to variance in physico-chemical and physical properties like hydropathy, charge, solubility, volume, etc. If two amino acids have more common properties than they are structurally more similar than other two amino acids with less common properties. From the evolution concept, it is well known that the percentage of evolution from one amino acid to another will increase when they are structurally more similar [18], [19]. Since chemically similar amino acids are known to be more interchangeable than dissimilar ones, due to the structure of the genetic code, they can substitute easily in the protein tertiary structure, while the replacement is increasingly troublesome between amino acids that lean toward various spatial neighbors[7]. It is interpreted like this, if two amino acids have more common properties than their tendency to evolve into other amino acids will increase [7].

Amino acids replacement is mainly due to the change of one amino acid to another in corresponding DNA and RNA because of point mutation, although in all amino acids substitution is not equally contributing towards better characterization of protein. Also, it is important to know what types of amino acid substitutions are more likely to be under selection and what types are mostly neutral to study the evolution of the protein. This is due to some facts which include positioning of amino acids in sequence and how similar and dissimilar amino acids are with whom substitution takes place. There are two main types of amino acid substitution or mutation namely conservative substitution and radical substitution [7]. Conservative substitution take place between two amino acids with similar properties and radical substitution take place if there is no such similarity exists which is mainly in abnormal human hemoglobin [12].For example, in humans, sickle cell anemia is formed due to the mutation in beta-globin where at position 6, negatively charged glutamic acid is exchanged with not charged valine. Another way we may say that conservative mutation take place between amino acids with similar structure and vice versa. As conservative replacements can be less important phenotypic changes, they are more common than radical replacements and less conformation of tertiary structure of protein [12].

From this discussion, it is clear that there is an important relationship between properties of amino acids with structural similarity and it leads to the substitution of amino acids and ultimately to the evolution of protein. To study this interrelationship, we have applied a mathematical tool called Signed Graph. We have used Heider's balance theory [9] for the analysis purpose of a graph which is stable or balanced and unstable or unbalanced. In our proposed methods to find out the evolutionary importance with respect to structural similarity. Here we have considered only four Physico-chemical properties as shown in Table1[1],[11] beloware taken for observation and also try to find out dependency of individual property in protein evolution.

In the year1946, Heider first proposed the concept of Structural Balance in the area of social psychology and has been extended for graphs by Harary [5]and Davis[4]. There they suggested that in a social setup, people would prefer forming relationships such that a balanced state emerges in their interpersonal relationships. For example, when there is a group of three nodes or users where there exists an edge or relationship between all pairs of nodes then the relationship between any two nodes in isolation, they could either be friends (positive edge) or enemies (negative edge) [16]. In another study, Guha et al. [8] proposed a model based on social believe where they tried to predict edge signs where they used trust propagation schemes [14]. Recently, Kim et al. [10] proposed a network that is embedding space and it can represent both the sign and the direction of edges and later formulate likelihood functions over both the direct and indirect signed networks.

To learn effective node representations in the signed networks, some recent methods [5] are proposed, where one can take into account both the positive and the negative edges when learning the node representations. Wang et al. [16] have used a neural network as a local structure and proposed an approach to discriminate between the positive edges, the negative edges, the unlabeled edges. Similarly, Wang et al. [15] used amino acids networks and they computed an evolutionary index that was equal to the likelihood of the particular non-synonymous substitution giving rise to the amino acid replacement relative to that of synonymous substitution.

2. Basic Concepts of Graph Theory

2.1 Graph

A Graph G denoted as G=(V,E) consists of a set V={v_1,v_2,v_3,…..} called vertices, also named as nodes, points and set E= {e_1,e_2,e_3,…..} called edges, also called lines or arcs. For vertices u, v of G, if e = uv is an edge then u and v are adjacent means e joins u to v. A graph is connected if there exists a walk between every pair of its vertices.

2.2 Bipartite Graph

It is also called bigraph where vertex-set of a graph G can be portioned into two disjoint sets, V_1 and V_2, such that each edge connects to a vertex of set V_1 to a vertex of set V_2, and there are no edges in G that connect two vertices in V_1 or two vertices in V_2 [16].A graph is bipartite if and only if its cycles are even.

2.3 Signed Graph

In 1946 Heider [9] first proposed the concept of structural balance for undirected graph. He stated that social system is balanced if there is no tension and create a problem if unbalanced and that case change the direction of

balance. Since then, it is a trending area of graph theory which is adopted by number of researchers from different field to study the interpersonal relationship. To check the balancing of a graph, here we use the signed graph approach ofWu and Chen [20]. Vertices define theindividuals in a group or community and edges represent the relationship between them.

A signed networks G = (V,E, σ) is a undirected graph where the set of nodes is denoted by V, with |V| = n and the set of edges is represented by E = {e_1, e_2, e_3 …..e_m}, σ is the sign function σ : E → {−1, +1}that ispartitioned into the set of positive edges E^+ and the set of negative edges E^- adding up toa total of m= m^++ m^- edges[13],[3].The positive degree and negative degree of node i are denoted by d+ and d^- and degree of node i is represented by d_i, equalsthe number of edges connected to node i and calculated based on $d_i = d^+ + d^-$.

A walk of length k in G is a sequence of nodes v_0,$v_{1,.....}$,v_k such that for eachi = 1, 2, 3 …. k there is an edge from v_{i-1} to v_i. A closed walk of length k is defined by $v_0 = v_k$and closed walk if distinct except for the endpoints. The closed walk is called cycle and sign of a cycle is the product of thesigns of its edges. A balanced cycle is measured with a positive sign and an unbalanced one with a negative sign. A graph is balanced if there is no negative cycle.

A signed graph is a simple undirected graph in which each edge is labeled by a sign either +1 or +ve and -1 or -ve where +1 indicate positive relationship between two nodes may be like,agree, associates and -1 indicates dislike, disagree, avoid and so on. For example, here in this graph fig [1], A likes B so +ve edgedo not likesD so –ve sign and C likes A so +ve sign.Using Signed graph one can easily study the evolutionary expect in different time frame period.

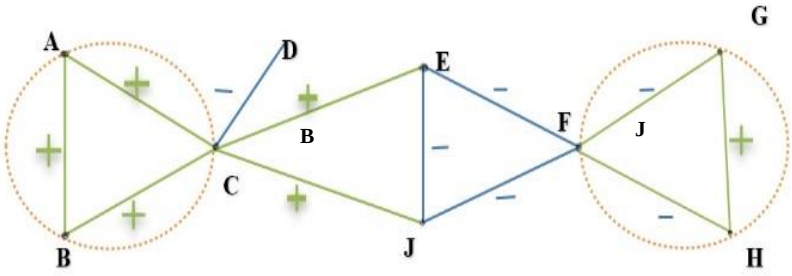

Fig. 1: Example of Signed Graph

In a signed graph network, structural balance property is mostly used and studied theory. There are numbers of balance theory proposed by researchers from time to time but Heider [9] who first proposed the balance theory is still used and is one with well-established theory in the field of psychology.The formalization of Heider's theory by Cartwright and Harary resulting structural balance theory for a graph represents interpersonal network.Here in this paper

to check to evolutionary expect of amino acids, we will apply the theory of balance of a signed graph due to Harary [4]which shows how change of the dyadic sign cause change of the balance for the whole graph. Following is the theorem on balance of a signed graph due to Harary which in particular we use in this context.

Theorem 1. A signed graph is balanced if and only if its vertex set can be partitioned into two classes (one of the two classes may be empty) so that every edge joining vertices within a class is positive and every edge joining vertices between classes is negative [5].

Now we take an example to discuss Harary theorem. Here in Fig.1, edges with positive sign indicate agree or friendly relationship and edges with negative sign indicate disagree or enemy. The circles (A,B,C)and (F,G,H)are in balance state due to partitioning concept but in case of (E, J,F) there is no such partition possible.

3. Properties of Amino Acids

Amino acids are building blocks of proteins. In the evolution of process, amino acid substitutions occur more frequently between structurally more similar amino acids than between dissimilar ones[14],[19]. Again, two amino acids are structurally similar if there have maximum number of common property or physicochemical distance is minimum but at the same time it is very important to study which property influence more in terms of evolution, substitution, proteins folding etc. Different amino acids can differ in many physicochemical properties; indeed, 134 properties were listed by Sneath [21].When an amino acid mutated by another amino acid which is structurally similar (i.e., an exchange between two amino acids separated by a small physicochemical distance) is called a conservative replacement [18],[12].

Amino Acids (A.A) can be classified based on their structure which are directly related to different physico-chemical properties and the structure of their side chains as the i.e. the R chains which are different for 20 amino acids [6]. Here in our study we have considered only four properties viz.,{hydrophilic, Hydrophobic}, {polar, non-polar}, {Aliphatic, Aromatic} and {positive, negative} shown in the Venn diagram Fig.2[11], [1] given below:

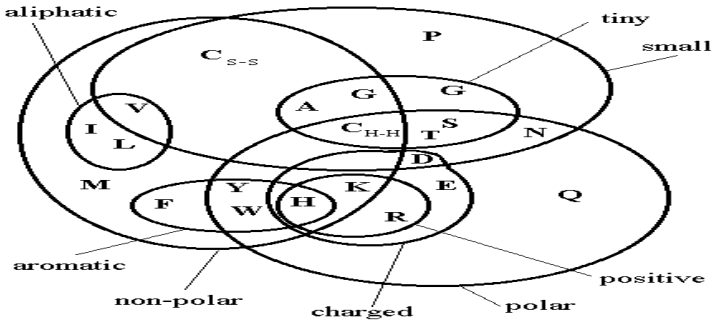

Fig. 2: Venn diagram of physico-chemical properties

Here in Table 1[11] [1], we have assigned neutral value 1 if property exist other wise 0 or neutral 1 if no property exists. Adil et al. [1] studied amino acids networks and based on different centralities values gave conclusion. Here, we have imported same data set as mentioned in [1], to study structural balancing of amino acids networks. From this data we have analyzed amino acids mutational graph in three cases. In first case when there is one property similar or less structurally similar, in case two when there are two properties similar and in case three when there are three properties

A. A.	Hydro phobic	Hydro philic	Aliph atic	Arom atic	Neutral (x)	Polar	Non polar	+ve	-ve	Neutral (y)
G	1	0	0	0	1	0	1	0	0	1
A	1	0	0	0	1	0	1	0	0	1
V	1	0	1	0	0	0	1	0	0	1
M	1	0	0	0	1	0	1	0	0	1
W	1	0	0	1	0	0	1	0	0	1
L	1	0	1	0	0	0	1	0	0	1
I	1	0	1	0	0	0	1	0	0	1
F	1	0	0	1	0	0	1	0	0	1
P	1	0	0	0	1	0	1	0	0	1
Y	1	0	0	1	0	1	0	0	0	1
S	0	1	0	0	1	1	0	0	0	1
T	0	1	0	0	1	1	0	0	0	1
E	0	1	0	0	0	1	0	0	1	0
C	0	1	0	0	1	1	0	0	0	1
N	0	1	0	0	1	1	0	0	0	1
Q	0	1	0	0	1	1	0	0	0	1
D	0	1	0	0	0	1	0	0	1	0
K	0	1	0	0	0	1	0	1	0	0
H	0	1	0	1	0	1	0	1	0	0
R	0	1	0	0	0	1	0	1	0	0

Table -1: Physico-Chemical Property of Amino Acids

similari.e. Maximum structurally similar between two amino acids. Again, as all the properties are not equally important as well as equally influential for mutation or protein synthesis, so it is also important to analyze which property influence more in evolutionary perspectives. As there are more than hundreds of physico-chemical properties so for illustration we have tried to give an observation on substitution which is based on structural similarity of these mentioned properties only.

To study the substitution relation with structural similarity of amino acids [14], [19], we have considered here point mutation of amino acids. Amino

acids are possible combination of three bases called codons and as there are four bases, (Adenine (A), Cytosine (C), Guanine (G), or Thymine (T/U)), so 64 total possible codons out of which three are stop codons. These codons are formed due to change of one base position or called one-point mutation and create a network as describe in [2]. From the network point of view and importance of individual node i.e., amino acids have studied by different researchers [9], [2] which ultimately give the evolutionary aspects of amino acids and proteins as well. We have categorized our study in two main parts, in the very first approach, we have considered point mutation with respect to the above-mentioned properties and in second approach for a random selection of property we have tried to showcase which properties influence more in evolutionary process.

4. Signed Graph in Amino Acids Network

We have considered first point mutation graph [2] of 20 amino acids. Adil et al. [2] created point mutation graph from the standard genetic codon table based on first base, second base and third base mutation and applied different centralities measures. Here, we have adopted same point mutation graph [2] for studding the evolutionary relationship over physico-chemical properties of amino acids. We have discussed below three cases for one property matches, two property matches and three properties from the table 1.

Case1: We have assigned a positive(+)sign in point mutation graph between two amino acids if there is at least one property common to both and else assigned a negative(-)sign.

Case2: We have assigned a positive (+) sign in point mutation graph between two amino acids if there are at least two properties common to both and else assigned a negative (-) sign.

Case3: We have assigned a positive (+) sign in point mutation graph between two amino acids if there are at least three properties common to both and else assigned a negative (-) sign.

So, Case 1 gives weakest compatibility signed graph and Case3 gives strongest compatibility signed graph to check the evolutionary aspects in terms of structural similarity.

Case1: Here, in the Fig.3representing signed graph of amino acids, we have assigned a positive (+) sign between two amino acids if there is a point mutation as well as one property similar between 20 amino acids, otherwise assigned negative (-) sign to edges.

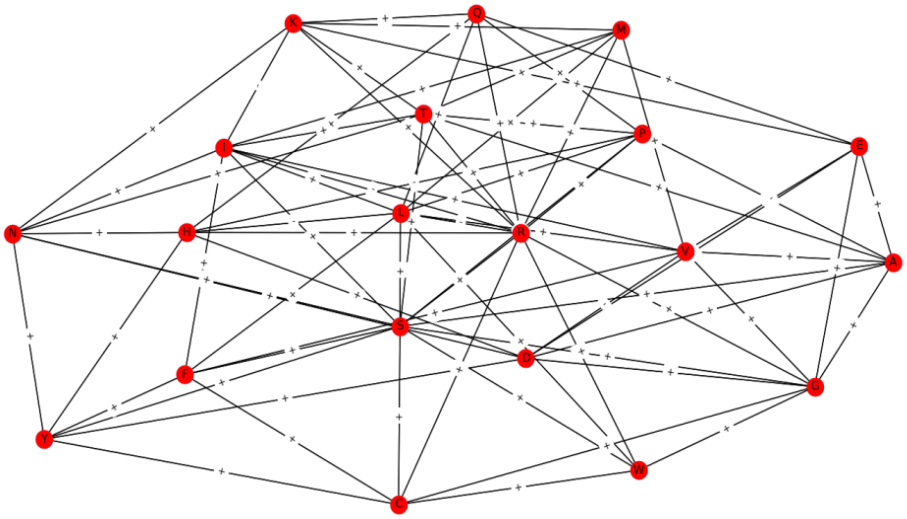

Fig.3: One Property Signed Graph

From the Fig. 3 we have obtained the pairs of positive edges as -(Y,S), (Y,H), (Y,N), (Y,F), (Y, D), (H, D), (H,R), (H,Q), (H, N), (G,S), (G,D), (G,A), (G,V), (G,R), (G, W), (G,E), (G, C), (P, S), (P,L), (P, A), (P,R), (P, T), (P,Q), (S, L), (S, A), (S,R), (S, W), (S,T), (S, Q), (K, M), (K, R), (K,T), (K, E), (K,Q), (K,N), (D,A), (D,E), (D, N), (M, L), (M,V), (M, R), (M,T), (M, I), (L, V), (L,W), (L,Q), (L, F), (L,I), (A, V), (A,T), (A,E), (V,F), (V,I), (R, W), (R,T), (R, Q), (R, C), (W,C), (T,N), (T, I), (N, I), (F,C), (F, I).

We now check for balancing of this signed graph as shown in Fig. 3using Theorem 1of Harary mentioned above. On applying the conditions of the theorem, we found that it is not possible to partition the set of vertices into two classes so that every edge joining vertices within a class is positive and every edge joining vertices between classes is negative. Clearly therefore the structure is not a balanced or stable one. In this context, it may explain in this way that the percentage of structural similarities among the amino acids is less here or we can say that more dissimilar the amino acids and hence there will be chance of further conformation of the secondary proteins structure which may be unstable. Also, here we may interpret like this, frequency of substitution or point mutation is more when there is a percentage of structural similarity between two random amino acids in terms of properties is less due to which there are some portion of proteins where may be conformation occur later.

Case2: In the below Fig. 4, we have modeled a signed graph of amino acids where edges are assigned a positive (+) sign between two amino acids if there is a point mutation with two properties similar between them, otherwise assigned negative (-) sign to edges.

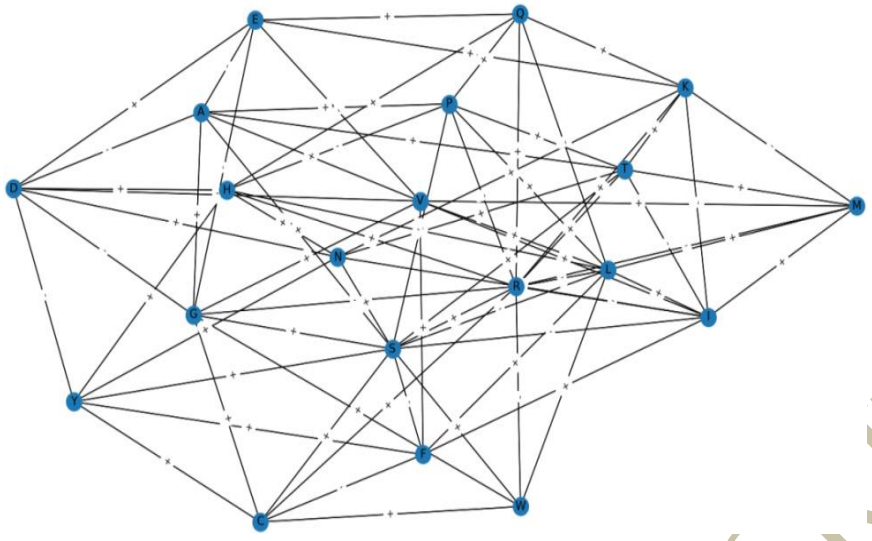

Fig.4: Two Properties Signed Graph

From the fig. 4 we have the pair have positive edges as (Y,S), (Y,H), (Y,N), (Y, F), (H,R), (H,Q), (H, N), (H,D), (G,S), (G, A), (G, V), (G,W), (G, C), (P, L), (P, A), (P,T), (P,Q), (S,A), (S, R), (S,W), (S, T), (S, N), (S, C), (K,R), (K,T), (K,E), (K, Q), (K, N), (D,E), (D,N), (M, L), (M, V), (M, T), (M, I), (L,V), (L,W), (L,F), (L,I), (A, V), (A, T), (V,F), (V, I), (R, T), (R,Q), (R, C), (W, C), (T,N), (E,Q), (F, I).

It is noted that here also we have not been able to partition the set of vertices into two classes of a balanced signed graph given by Theorem 1, although in this case percentage of structural similarity is more compared to Case1 still hence a smaller number of mutations is occurred. Here, it is not possible to partition the signed graph into two sets as shown in Fig.4, therefore graph is not balanced one and this may result in conformation of secondary structure occurrence in the evolution process.

Case3: Here, each edge assigned a positive (+) sign in the mutation graph if there is maximum of three properties similarity between two amino acids.

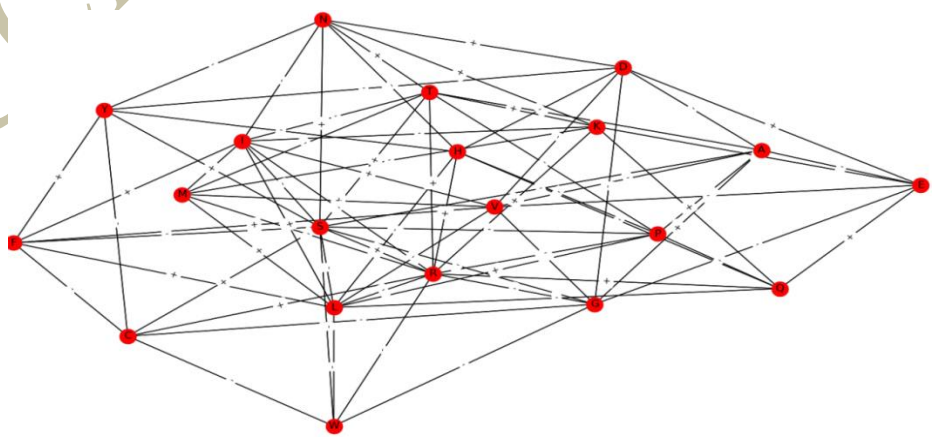

Fig.5: Three Properties Signed Graph

Herein Fig. 5 the vertices which represent the positive edges are- (Y,F), (G,A), (G,V), (P,L), (P,A), (M,L), (M,V), (M,I), (V,F), (V,I), (F,L), (F,I), (L,I), (H,R), (S,R), (S,T), (S,N), (S,C), (K,R), (K,T), (K,E), (K,Q), (K,N), (D,E), (D,N), (R,T), (R,C), (T,N), (E,Q).

Now, we separate all these vertices into two disjoint subsets, v_1 and v_2 where v_1={(Y,F, G,A, P,M, I, V,F,L} and v_2={H,R,S,T,N,C,K,E,Q, D}. From these vertex sets v_1 and v_2 it is observed that within the set there is a positive edge but between the set there is a negative (-) edge. From this we can say that it fulfills the Harary Balance Theory (Theorem 1). Here, it is observed that the frequency of substitution is less and structural similarities percentage is maximum than previous two cases. From this we may conclude like there is an interrelationship between structural similarities with substitution. In the evolution process the amino acids with maximum structural similarity play a crucial role than dissimilar ones. Also, for a stable or balance conformation of proteins we may say that substitutions preferentially occur between pairs of amino acids heaving maximum similar properties. This formulation may not be applicable to all the variable sites of a protein and may be confined to a particular targeted region in the whole sequence and thus the understanding of normal and abnormal behavior of protein sequence is the only key to prevent and cure the diseases.

As there are more than hundred numbers of physico-chemical and physical properties exists, it is therefore desirable to study which properties of amino acids are more important than others in determining the substitution rate and pattern of protein evolution. Here, in this section we have tried to emphasize individual property influence from Table 1, which has a significant contribution on protein synthesis. For that, we have considered randomly three properties form the Table 1 and divide it into three groups and later apply the same approach as in Case1, Case2 and Case3 respectively. In Group1, we have considered {Aliphatic, Aromatic}, {Polar, Non polar} and {+ve and –ve}, in Group2 we have considered {Hydrophobic, hydrophilic} and {+ve and –ve}, {aliphatic and aromatic} and in Group 3 we have considered {Hydrophobic, hydrophilic}, {polar, non-polar}, {aliphatic and aromatic} accordingly. Now, considering earlier proposed approach here also inGroup1, Group2 and Group3 i.e., we have assigned a positive (+) sign to an edge between two amino acids if there are at least one and two properties matches and evaluate each of them. Now, we have observed that in both the group i.e. in Group1 and Group 2 there is no maximum structural similarities among the amino acids i.e., no two properties matches. So, it does not fulfill the signed graph structural balance property and we may say the individual influential of properties in both Group1 and Group2 is less for protein synthesis. Again, in case of Group3, as there is a maximum structural similarity among amino acids as shown in table1 i.e., maximum two properties match out of three, the

mutational graph gives a balanced network and this it may be due to influence of hydrophobic, hydrophilic, polar, non-polar property in evolution process. From all this discussion we may conclude that out of four considered properties hydrophobic, hydrophilic, polar, and non-polar these two properties play an important role in evolution of amino acids in the process of protein synthesis. Here, in this manuscript we have tried to study the theoretical analysis of structural relationship with respect to substitution of amino acids with the help of signed graph approach and hence might not bear any homology to the actual biological evolutionary process.

5. Conclusion

In this manuscript, we have attempted a general model to study the intensity of structural similarity concerning some properties of amino acids for point mutation. Here we have used a graph theory tool called signed graph to study the balance property for the different case studies. We have observed from our analysis that amino acids network with maximum structural similarity is a balanced network and hence the protein evolution is stable. We can say that there may be no further conformation of the protein. Correspondingly, we have attempted to find the individual significance or impact of properties in amino acids evolution. Among the four physico-chemical properties hydrophobic, hydrophilic with polar and non-polar effects play a crucial role in the evolution process. This is only our theoretical perception for structural equalization property and our findings may differ for all the regions of proteins.

Acknowledgement: We are thankful to the anonymous reviewer for their constructive as well as useful suggestions.

References

[1]Akhtar, A., (2015), Graph theoretic approach to analyze amino acid network, Int. J. Adv. Appl. Math. And Mech2, 31-37.

[2]Akhtar, Adil., and Ali, Tazid., (2014), Analysis of Unweighted Amino Acids Network, International Scholarly Research Notices (Hindawi).

[3]Aref, S., and Wilson, M., (2019), Balance and Frustration in Signed Networks. 7. 163–189.

[4]Davis, J. A., (1963), Structural balance, mechanical solidarity, and interpersonal relations. Am. J. Sociol. 68, 444–462.

[5]F. Harary.,(1954), On the Notion of Balance of a Signed Graph, Michigan Math. J. (2) 143-146.

[6]Grantham, R., (1974), Amino acid difference formula to help explain protein evolution, Science (New York, N.Y.), 185(4154):862-864.

[7]Graur, D., (1985), Amino acid composition and the evolutionary rates of protein-coding genes. J Mol Evol 22, 53–62.

[8]Guha, R., Kumar, R., Raghavan, P., and Tomkins, A. (2004). Propagation of trust and distrust, Proceedings of the 13th International Conference on World Wide Web, WWW '04 (New York, NY: ACM), 403–412.

[9]Heider F., (1946), Attitudes and cognitive organization. Journal of Psychology; 21(1): 107-112.

[10]Kim, J., Park, H., Lee, J.-E., and Kang, U. (2018). Side: representation learning in signed directed networks, in Proceedings of the 2018 World Wide Web Conference on World Wide Web (International World Wide Web Conferences Steering Committee) (Lyon), 509–518.

[11]Lamy, J., Berthelot, H., & Favre, M. (2016). Rainbow Boxes: A Technique for Visualizing Overlapping Sets and an Application to the Comparison of Drugs Properties. 2016 20th International Conference Information Visualization (IV), 253-260.

[12]Miyata, T., Miyazawa, S., & Yasunaga, T, (1979), Two types of amino acid substitutions in protein evolution. Journal of molecular evolution, 12(3), 219–236.

[13]Primetis, E., and Chavlis, S., and Pavlidis, P., (2018), Evolutionary Models of Amino Acid Substitutions Based on the Tertiary Structure of their Neighborhoods.

[14] S, Sodhani., Q, Meng., and T, Jian., (2019), Attending Over Triads for Learning Signed Network Embedding, Frontiers Big Data 2., page6.

[15] Wang, S., Aggarwal, C., Tang, J., and Liu, H., (2017), Attributed signed network embedding, in Proceedings of the 2017 ACM on Conference on Information and Knowledge Management (Singapore: ACM), 137–146.

[16] Wang, S., Tang, J., Aggarwal, C., Chang, Y., and Liu, H. (2017b). Signed network embedding in social media, in Proceedings of the 2017 SIAM International Conference on Data Mining (Houston, TX: SIAM), 327–335.

[17] Weber, C., and Whelan, S., (2019), Physicochemical Amino Acid Properties Better Describe Substitution Rates in Large Populations, Molecular biology and evolution, 36. 679-690.

[18] Xia, X., Li., W, (1998), (2016), What Amino Acid Properties Affect Protein Evolution?, J Mol Evol 47, 557–564.

[19] Zuckerkandl, E., Pauling, L., (1965), Evolutionary divergence and convergence in proteins. In: Bryson V, Vogel J (eds) Evolving genes and proteins. Academic Press, New York, pp 97–166.

[20] Wu B.Y., Chen JF. (2013), Balancing a Complete Signed Graph by Editing Edges and Deleting Nodes. In: Chang RS., Jain L., Peng SL. (eds), Advances in Intelligent Systems and Applications - Volume 1. Smart Innovation, Systems and Technologies, Springer, Berlin, Heidelberg, Volume 1, pp 79-88.

[21] Sneath, P. (1995). Thirty Years of Numerical Taxonomy. Systematic Biology, 44(3), 281-298. doi:10.2307/2413593