ADV MATH SCI JOURNAL

Advances in Mathematics: Scientific Journal **9** (2020), no.10, 7985–7998 ISSN: 1857-8365 (printed); 1857-8438 (electronic) https://doi.org/10.37418/amsj.9.10.33 Special Issue on AMABDA-2020

ON SOME PROPERTIES OF SPLICED DNA SEMIGRAPH

J. PADMASHREE¹ AND K. THIAGARAJAN

ABSTRACT. Splicing plays a vital role in repairing DNA nucleotides and its sequence due to cancer growth. Semigraph was defined by Sampathkumar as a generalization of a graph. The DNA molecules as a semigraph and splicing schema applied on the semigraph is defined. In this paper, DNA nucleotides in the form of semigraph with end vertices, middle vertices and cut vertices for n bonds are modeled. Splicing of DNA semigraph along cut vertices is shown graphically for generalized semigraph with n bonds. And degrees of vertices, edge degrees of vertices, adjacent edge degrees of vertices and successive adjacent edge degree of vertices are discussed for semigraph with 3 bonds and characterized for n bonds. Relationship between degrees of vertices are carried out using the theorems. Constructed spanning subsemigraphs of the semigraphs in terms of PATH. Also results and observations are carried out for minimum length of covering EV, MV and CV in the Path of G_n when n is odd and even.

1. INTRODUCTION

Cancer disease is the dangerous development of the bundle of afflictions relating strange cell improvement in the body with the probable advancement to sully or increment to various bits of the body through blood and lymph. Normally, the gene in the body balances the regular cell growth when appropriately

¹corresponding author

²⁰²⁰ Mathematics Subject Classification. 97K30, 05C07.

Key words and phrases. Cut Vertices, Edge degrees, End Vertices, Middle vertices, Semigraph and spanning subsemigraphs.

7986

regulating involuntary cell death cells. Also segregation, apoptosis and production of cell system and death cells are inhibited by myriad genes. Cancer cell growth seems mysterious and also develops and affect healthy cells and occupies more space and also it travel to other parts of the body to stem the cancerous cell growth and affect the organ by producing tumour. The challenging part of medical science is to fight with cancer cells in the body and decrease their production.

There are numerous clinical strategies for malignant growth treatment which straightforwardly feeble them and crush them totally concentrating on disease influenced region in the body by medical procedure, chemotherapy, radiation treatment, hormonal treatment, directed treatment and palliative consideration.

During years many researchers started their study and dealt with mathematical modelling of the impact of the cancer cells[1]. A decent overview of numerical models of malignancy development and improvement can be found in the incredible book [7], and a brilliant review of the scope of scientific and computational displaying procedures utilized for organic issues on various scales can be found in the book. We have extended theoretical study in the graphical approach of cancerous DNA structure as Semigraph and characterized the splicing system[3] by applying splicing schema[6] of the spliced semigraph and language generated on folding [10], the spliced semigraph builds the strongest way to precede our research in various aspects and further on the graphene model. Researchers also carried out many discussion on Adjacency of Semigraph[2] and language on Graph splicing[4]. Study on applications of graph theory using algorithmic approach is discussed by Saha Ray[9]. In this paper, we classify all the vertices of semigraph into the cut, middle and end vertices. In particular, four different types of degrees have been defined for each vertex and associations among them have been studied in detail. And the spanning subsemigraphs for semigraph is observed for G_3 in PATH structure and generalized the results to G_n in two cases of n is odd and even.

1.1. Semigraph [8].

Definition 1.1. A semigraph G is a 2 - tuple(V, X) where V is a non-empty set whose components are called vertices of G and X is a set of n - tuples called edges of G of distinct vertices for different $n \ge 2$, fulfilling the accompanying conditions:

(1) The segments of an edge E in X are particular vertices from V.

- (2) Any two edges in X have at most one vertex of V in common.
- (3) Two edges $(x_1, x_2, x_3, ..., x_n)$ and $(y_1, y_2, y_3, ..., y_m)$ are viewed as equivalent if and only if:
 - m = n
 - either $x_i = y_i$ or $x_i = y_{n-i+1}$ for $1 \le i \le n$. Thus the edges $(x_1, x_2, x_3...x_n)$ are the identical as the edge $(x_n, x_{n-1}, ..., x_1)$.

The vertices[5] in a semigraph are recognized into three kinds in particular END vertices(EV), MIDDLE vertices(MV) and CUT vertices(CV), contingent on the splicing applied along the DNA molecules A, C, G and T ie., splicing applied along the cut vertices of an edge in the semigraph. An end vertices in semigraph is represented by thick circle dots whereas thick square shade denotes a middle vertices and small circles denote the cut vertices. Let G_i , $3 \le i \le n$ be the semigraph with '*i*' bonds.

Example 1. DNA particle is a polymer of nucleotides (A, C, G and T) where Adenine consistently bonds with thymine, and cytosine consistently bonds with guanine. The bonding makes the two strands winding around one another in a shape called a double helix. The number of bonds used is the number of linking between vertices (nucleotides). Double stranded DNA molecule sequence with nucleotides in graphical model is depicted as below. Figure 1. represents the theoretical approach on Splicing applied along the cut vertices (small circles) to separate the damaged molecules and using enzymes, the molecules are operated and attached to another sequence of DNA molecules having matching pairs of nucleotides by reading the second pair of DNA sequence along the reverse orientation. For n = 3 bonds, the graphical model as follows:



FIGURE 1. Theoretical Semigraph of DNA molecules on splicing



FIGURE 2. A Semigraph G_3 with cut, middle and end vertices

Figure 2 represents the semigraph structure of the double stranded DNA molecule sequence with cut, middle and end vertices. In the given semigraph G_3 :

$$V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9\}$$

$$X = \{(v_1, v_2), (v_1, v_7, v_3), (v_2, v_4), (v_3, v_8, v_4), (v_3, v_5), (v_4, v_9, v_6), (v_5, v_6)\}.$$

The vertices v_1, v_2, v_5, v_6 are called EV. The vertices v_3, v_4 are MV and vertices v_7, v_8, v_9 are CV.

The vertex set $V = V_1 \cup V_2 \cup V_3$ where V_1, V_2, V_3 represents the set of end vertices, middle vertices and cut vertices respectively. Also |V| denotes the total number of end, middle and cut vertices.

Similarly, the edge set $X = X_1 \cup X_2 \cup X_3$ where X_1, X_2, X_3 represents the set of edges containing end vertices, middle vertices and cut vertices respectively. And |X| represents the total number of edges.

For Figure 2, $|V_1| = 4$, $|V_2| = 2$, $|V_3| = 3$, $|X_1| = 6$, $|X_2| = 5$, $|X_3| = 3$.

Also the vertices v_1, v_7, v_3 are adjacent and share an edge e_2 . Hence semiedges (v_1, v_7) or (v_7, v_1) is denoted as e'_2 .

2. GENERALIZATION OF EV, MV AND CV FOR n BONDS

We derived the generalization of edge vertices, middle vertices and cut vertices for n bonds of Double stranded DNA sequence of nucleotides.

Here, $|V_1| = 4$, $|V_2| = 2n - 4$, $|V_3| = n$, $|X_1| = 6$, $|X_2| = 3n - 4$, $|X_3| = n$ and also |V| = 3n and |X| = 3n - 1.



FIGURE 3. A Semigraph G_n with cut, middle and end vertices3. DEGREES OF SEMIGRAPH VERTICES

Definition 3.1. For any vertex 'v' in a semigraph G = (V, X), the different types of degrees set are defined as below:

- (1) The degree of a vertex 'v', D(v) is the number of edges having 'v' as an end vertex or middle vertex.
- (2) The edge degree of a vertex ED(v), is the number of edges containing the vertex 'v'.
- (3) The adjacent edge degree of a vertex AED(v), is the number of end, middle and cut vertices adjacent to a vertex 'v'.
- (4) The successive adjacent edge degree of a vertex SAED(v), is the number of end, middle and cut vertices direct adjacent to 'v'.

Example 2. The below table depicts the D(v), ED(v), AED(v), SAED(v) of all the vertices for the semigraph given in Figure 2.

End Vertices	v_1	v_2	v_5	v_6
D(v)	2	2	2	2
ED(v)	2	2	2	2
AED(v)	3	2	2	3
SAED(v)	2	2	2	2

Table 1: Degrees of End vertices

Tab	le 2	: Degree	s of	Middl	e vertices
-----	------	----------	------	-------	------------

Middle Vertices	v_3	v_4
D(v)	3	3
ED(v)	3	3
AED(v)	5	5
SAED(v)	3	3

Table 3: Degrees of Cut vertices

Cut Vertices	v_7	v_8	v_9
D(v)	0	0	0
ED(v)	1	1	1
AED(v)	2	2	2
SAED(v)	2	2	2

4. DISCUSSION OF RESULTS ON DEGREES OF VERTICES

From the obtained degrees, edge degrees, adjacent edge degrees and successive adjacent edge degree of EV, MV and CV for n bonds, there is an gradual adjacency of vertices on each level of splicing of semigraph with respective to increasing number of bonds. The figure 4 represents the D(v), ED(v), AED(v) and SAED(v) for 3, 4 and 5 bonds.



FIGURE 4. Graph of degrees of vertices for 3, 4 and 5 bonds

The subsequent theorems present the properties of different degrees of vertices and association between one another.

Theorem 4.1. Let $V = v_1, v_2, ..., v_m$, $X = X_1, X_2, X_3$ and $X_i = E_1, E_2, ..., E_q$, i = 1, 2, 3 and $1 \le q \le n$ in a semigraph G = (V, X).

(1)
$$\sum_{i=1}^{m} D(v_i) = 2 \sum_{i=1}^{3} |X_i|$$

(2)
$$\sum_{i=1}^{m} ED(v_i) = 2 \sum_{i=1}^{3} |X_i| + |X_3| = \sum_{i=1}^{m} D(v_i) + |X_3|$$

Proof.

- (1) Here V_1, V_2 and V_3 are set of end vertices, middle vertices and cut vertices respectively. It is obvious observation that if a cut vertex $v \in V_3$, then D(v) = 0. If $v \in V_1$ and D(v) = p, then there are 'p' edges adjacent to an end vertex in X_1 . i.e., v is an end vertex for all p edges. Each edge containing v is adjacent to another end vertex or middle vertex. Hence each edge in X_1 is counted twice for each end vertex and its adjacent middle vertices that belong to X_2 . Similarly for any vertex $v \in V_2$, then each edge in X_2 is counted twice for each middle vertex and its adjacent end vertices that belong to X_1 . Hence proves (1).
- (2) For any $u \in V_1$, $v \in V_2$ and u and v are adjacent in X_1 , the edge degree is counted twice for each u and v. Also each edge is counted twice for each degree of end vertex and middle vertex which is already depicted in proof (1). In particular, the cut vertex $w \in V_3$ have each edge degree as one which is directly lies in the adjacent edges having end vertices in X_1 and middle vertices in X_2 . Thus edge degree of cut vertices, V_3 is $|X_3|$. Hence this proves (2).

Theorem 4.2. For any $v \in V_1$ or V_2 in a semigraph G = (V, X), D(v) = ED(v) = SAED(v).

Proof. If an end vertex $u \in V_1$, then the number of edges containing an end vertex u is same as the number of edges having u as end vertex. Similarly, if a middle vertex $v \in V_2$, then the number of edges containing a middle vertex v is same as the number of edges having v as middle vertex. Hence D(v) = ED(v). Suppose an end vertex $u \in V_1$, then there is exactly one end vertex or middle vertex in vertex set V_1 or V_2 is successive adjacent to u in X_1 and thus the number of edges successively adjacent to u is equal to number of edges containing u as an end vertex. Also for any middle vertex $v \in V_2$, then there is exactly one end vertex or end vertex or middle vertex is successive adjacent to v in X_2 and thus the number of edges successively adjacent to v is equal to number of edges containing v as an middle vertex. Hence D(v) = ED(v) = SAED(v).

Theorem 4.3. Let $v \in V_3$ be a cut vertex in a semigraph G = (V, X). Then

- a) D(v) = 0
- *b)* ED(v) = 1

c) AED(v) = 2
d) SAED(v) = AED(v)
e) D(v) < ED(v) < AED(v) and D(v) < ED(v) < SAED(v)

Proof. By definition, D(v) is the number of edges having v as end vertex. For $v \in V_3$, there is no such edge exists. Hence D(v) = 0. This proves (a). It is easy from the observation that the cut vertices in a semigraph 'G' having exactly one edge in X. Hence for any cut vertex 'v', ED(v) = 1. This proves (b). Let $v \in V_3$ is a cut vertex. Then there is an edge E contains exactly two end vertex or middle vertex or both end vertex and middle vertex and is adjacent to v in E i.e., $(v_1, v, v_2) \in E$ is an edge in the X_3 . Hence AED(v) = 2. This proves (c). If $v \in V_3$ and E contains v, there is exactly one edge containing v in E. It is also evident that there are two vertices successive adjacent vertices to v in E. Hence SAED(v) = 2. By proof of (c), SAED(v) = AED(v). This proves (d). By the above proof of (a), (b), (c) and (d), D(v) < ED(v) < AED(v) and D(v) < ED(v) < SAED(v). This proves (e).

Theorem 4.4. Let $v \in V_3$ be a cut vertex in a semigraph G = (V, X). Then $\sum SAED(v) = \sum AED(v) = 2|X_3|$.

Proof. Let $v \in V_3$. Then $v = (v_1, v, v_2) \in E$ for some E in X_3 . Note that the two middle or end vertices v_1 and v_2 are successively adjacent to v. Thus for any v, there is an existence of two vertices must be adjacent to v. Thus SAED(v) = 2. If there are $|X_3| = n$ cut vertices in G, then the sum of successive adjacent degree of all n cut vertices is 2n which is $2|X_3|$. Hence by (d) of Theorem. 3, $\sum SAED(v) = \sum AED(v) = 2|X_3|$.

4.1. Definitions.

4.1.1. Subsemigraph. Let G = (V, X) be a semigraph. Then a semigraph $H = (V_1, X_1)$ is said to be Subsemigraph of G, denoted by $H \subseteq G$, if:

- (1) $V_1 \subseteq V$
- (2) $X_1 \subseteq X$
- (3) An edge (x₁, x₂, x₃, ..., x_n) in X₁ implies (x₁, x₂, x₃, ..., x_n) in X. In other words, the end vertices, middle vertices and cut vertices in V₁. Say, x₁, x₂ in V₁ and x₁ is adjacent to x₂ in H. Then x₁, x₂ in V and x₁ is adjacent to x₂ in G.

4.1.2. Spanning Subsemigraph. Let G = (V, X) be a semigraph and subsemigraph is $H = (V_1, X_1)$. Then the spanning subsemigraph S of G is subsemigraph of G and denoted by $S(G) \subseteq G$, if the vertex set containing EV, MV and CV of S are equal to vertices of G, i.e., V(S) = V(G).

4.1.3. Walk. A walk in Semigraph G is an alternating sequence of vertices (EV, MV and CV) and edges $v_1e_1v_2e_2....e_nv_n$. In case, the starting and ending vertices are Cut vertices in a walk of G, then the walk in semigraph traverse in semiedge associated with cut vertex i.e., $v_1e'_1v_2e_2....e'_nv_n$.

4.1.4. *Trail.* A trail in Semigraph G is an open walk with no edges or semiedges repeated twice in the traverse sequence of EV, MV, CV.

4.1.5. *Path.* A path in Semigraph G is an open trail with no end vertices, middle vertices and cut vertices repeated twice in the traverse sequence. In other words, A path is ordered sequence of vertices and edges which occur only once travelling from starting vertex to ending vertex. The number of occurrence of the edges and semiedges in the sequence of a path (P) from end, middle or cut vertex 'u' to another 'v' is called the length of the walk which is denoted by $\pi(P)$ or $\pi(u, v)$. A semigraph G_n is said to be connected if any two end vertices or middle vertices or cut vertices of G_n are linked by a path.

Observation 1. A path on n vertices with (n-1) edges is denoted by P_n and $\pi(P)$ is (n-1).

Observation 2. The length of edge containing 2 end vertices or 2 middle vertices is always 1. And the length of semiedge containing end vertex and cut vertex or cut vertex and middle vertex is always $\frac{1}{2}$.

5. CONSTRUCTING SPANNING SUBSEMIGRAPHS OF SEMIGRAPH IN FORM OF PATH

In this section, few spanning subsemigraphs of Semigraph G_3 in the PATH structure including all end vertices, middle vertices and cut vertices are constructed as follows:

From the Figure 5 spanning subsemigraph $S_1, S_2, S_3, S_4, S_5, S_6$ of G_3 , the different traversing along all end vertices, middle vertices and cut vertices is stated as below:



FIGURE 5. Spanning subsemigraphs of Semigraph $'G'_3$

The path of S_1 , $P(S_1)$ is $v_7e'_2v_1e_1v_2e_3v_4e'_4v_8e'_4v_3e_5v_5e_7v_6e'_6v_9$ The path of S_2 , $P(S_2)$ is $v_2e_1v_1e'_2v_7e'_2v_3e'_4v_8e'_4v_4e'_6v_9e'_6v_6e_7v_5$ The path of S_3 , $P(S_3)$ is $v_2e_1v_1e'_2v_7e'_2v_3e_5v_5e_7v_6e'_6v_9e'_6v_4e'_4v_8$ The path of S_4 , $P(S_4)$ is $v_8e'_4v_4e_3v_2e_1v_1e'_2v_7e'_2v_3e_5v_5e_7v_6e'_6v_9e'_6v_6e_7v_5$ The path of S_5 , $P(S_5)$ is $v_8e'_4v_3e'_2v_7e'_2v_1e_1v_2e_3v_4e'_6v_9e'_6v_6e_7v_5$ The path of S_6 , $P(S_6)$ is $v_7e'_2v_1e_1v_2e_3v_4e'_6v_9e'_6v_6e_7v_5e_5v_3e'_4v_8$

Similarly, spanning subsemigraphs can be found in any semigraph with n number of bonds.

5.1. Minimum length of the path $\pi(P(S(G_n)))$. Consider the spanning subsemigraphs $\{S_i\}$, for some *i* of semigraph G_n . Then the minimum length of the path in Spanning subsemigraph of $S(G_n)$ is obtained by traversing along all EV, MV and CV covering in the short distance which is denoted by $\pi(P(S(G_n)))$.

Observation 3. From the figure 5. the spanning subsemigraphs $S(G_3)$ are Hamiltonian path[5] and minimum length of the path covering all end, middle and cut vertices among $S(G_3)$ is $P(S_2)$.

Observation 4. (From Figure 6) For Spanning subsemigraphs $S(G_3)$, the minimum length of the path covering all vertices $\pi(P(S(G_2)))$ is 5. whereas for spanning subsemigraphs $S(G_4)$, the minimum length of the path covering all vertices $\pi(P(S(G_4)))$ is $(7 + \frac{1}{2})$.



FIGURE 6. Minimum length of the Spanning Subsemigraph of G_3 and G_4

Corollary 5.1. Let $S(G_n)$ be the set of spanning subsemigraphs of semigraph G_n . Then the length of the walk in the $S(G_n)$, $\pi(P(S(G_n)))$ is always integer and

$$\pi \left(P\left(S\left(G_{n}\right)\right) \right) = \begin{cases} (2n-1)/2 & \text{if } n \text{ is odd and } n = 3, 5, \dots \\ (2n-1) & \text{if } n \text{ is even and } n = 4, 6, \dots \end{cases}$$

Result 1. Let $S(G_n)$ be the set of spanning subsemigraphs of semigraph G_n , And $\pi(P(S(G_n)))$ is the minimum length of the path in $S(G_n)$. Then:

- (1) Starting and ending vertices are end vertices in the path, $P(S(G_n))$ when n is odd.
- (2) Starting and ending vertices are end vertex and cut vertex in the path, $P(S(G_n))$ when n is even.

Result 2. Let $S(G_n)$ be the set of spanning subsemigraphs of semigraph G_n , $\pi(P(S(G_n)))$ is the minimum length of the path in $S(G_n)$. Then $P(S(G_n)))$ contains atleast 4 semiedges to cover all EV, MV and CV in $S(G_n)$.

Theorem 5.1. The minimum length of walk 'W' from one EV, MV or CV to another is a path in semigraph G_n .

Proof. Suppose there is a walk 'W' from one end vertex, middle vertex or cut vertex u to another end vertex, middle vertex or cut vertex v such that $v \neq u$. Then there exists a minimum length of the walk W from u to v in semigraph G_n . Now we claim that W is a path. We proof this by contrary that walk W in G_n is not a path. For some EV, MV or CV say z in W occurs twice on walk W. i.e., W = uzyzv. On deleting the vertex y from W, there is a walk W = uzv from u to v which is contradiction to the minimality of W.

Theorem 5.2. $\pi(u, v) \leq \pi(u, z) + \pi(z, v)$ for all end, middle and cut vertices u, v, z in $V(G_n)$ of semigraph G_n if z lies in the minimum length of path from u to v.

Proof. Suppose W is the minimum length of the path from u to z and W' is the minimum length of the path from z to v. Then WzW' is a walk from u to v which is of length $\pi(u, z) + \pi(z, v)$. This summation is an upper bound on the minimum length of the path from u to v (by Theorem 5.1).

The below figure represents the minimum length of covering end vertices, middle vertices and cut vertices in the path formed by spanning subsemigraphs of G_n for both the odd and even number of bonds (n = 3, 4, 5, 6, 7, 8).



FIGURE 7. Minimum length of the path $\pi(P(S(G_n))), n = 3, 4, 5, 6, 7, 8$

6. Application

Graph theory is widely utilized in numerous applications. Significantly, the development of the spanning subsemigraphs of graphical model of DNA atom nucleotides as semigraph structure and finding the minimum length of path between end vertices, middle vertices and cut vertices in this paper can be utilized for building transportations framework, in which the association between the two roads viewed as a vertex and the road between two places(vertices) is viewed as an edge. Also, their route framework happens on the base of the most limited way between two vertices. In the event that, because of traffic and pinnacle long stretches of movement, the intermediate place which is cut vertex is built to reduce the transportation time and deferral in expectations of products.

CONCLUSION

The concept of adjacency between end vertices, middle vertices and cut vertices in a semigraph plays a vital role in the DNA nucleotide sequence. The adjacency relation between degrees of vertices is discussed. And constructed the spanning subsemigraphs of the semigraphs in the form of PATH structure and discussed the optimal path in which the minimum length of the path travelled along all EV, MV and CV. Hence, it may be interesting to study the language on the DNA molecule sequence in form of semigraph to pass the security information using decrypted message.

REFERENCES

- [1] L.M. ADLEMAN: *Molecular computation of solutions to combinatorial problems*, Science, New Series 266.5187, (1994), 1021-1024.
- [2] C.M. DESHPANDE, Y.S. GAIDHANI: *About adjacency matrix of semigraphs*, International Journal of Applied Physics and Mathematics, **2**(4) (2012), 250-252.
- [3] R. FREUND: Splicing Systems on Graphs Proc. Intelligence in Neural and Biological Systems, IEEE Press, 189-194. (1995).
- [4] D.K. GNANAMALAR, K.G. SUBRAMANIAN, D.T. GNANARAJ: A Note on Graph Splicing Languages, Lecture Notes in Computer Science, 2340 (2001), 381-390.
- [5] F. HARARY: Graph Theory, Addition Wesley, Reading Mass, (1972).
- [6] T. HEAD: *Splicing schemes and DNA*, in "Lindenmayer Systems: Impacts on Theoretical Computer Science and Developmental Biology, (1992), 371-383.
- [7] PREZIOSI (ED.): Cancer Modelling and Simulation, Chapman & Hall/CRC (2003).
- [8] E. SAMPATHKUMAR: Semigraphs and their applications, Technical Report [DST/MS/022/94], Department of Science & Technology, Govt. of India, August, (1999).
- [9] S. SAHA RAY: *Graph Theory with Algorithms and its Applications*, Springer India (2013). DOI:10.1007/978-81-322-0750-4_2
- [10] K. THIAGARAJAN, J. PADMASHREE, N. PONNAMMAL: Semigraph Folding Approach for Generalization of Planar Triangulation, Indian Journal of Science and Technology, 9(48) (2016), 1-9.

DEPARTMENT OF MATHEMATICS, RESEARCH AND DEVELOPMENT CENTRE, BHARATHIAR UNIVERSITY, COIMBATORE, TAMILNADU, INDIA. *Email address*: padmasathiru@gmail.com

PENTAGRAM RESEARCH CENTRE PRIVATE LIMITED, HYDERABAD, TELANGANA, INDIA. *Email address*: vidhyamannan@yahoo.com