

RESEARCH ARTICLE | DECEMBER 22 2023

Some properties of n -cut splicing on labelled semigraphs

FREE

Muhammad Nur Syiham Abdul Razak ✉; Wan Heng Fong; Nor Haniza Sarmin



AIP Conf. Proc. 2975, 020010 (2023)

<https://doi.org/10.1063/5.0183227>



CrossMark

AIP Advances

Why Publish With Us?

- 25 DAYS**
average time to 1st decision
- 740+ DOWNLOADS**
average per article
- INCLUSIVE**
scope

[Learn More](#)

Some Properties of n -Cut Splicing on Labelled Semigraphs

Muhammad Nur Syiham Abdul Razak^{1, a)}, Wan Heng Fong^{1, b)} and
Nor Haniza Sarmin^{1, c)}

¹*Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 UTM
Johor Bahru, Johor, Malaysia.*

^{a)} Corresponding author: mnsyiham2@graduate.utm.my

^{b)}fwh@utm.my

^{c)}nhs@utm.my

Abstract. A semigraph is one of the generalization of graphs. Different from the hypergraph, a semigraph generalizes a graph in a straightforward manner where the edges of the graph are depicted as usual lines as in a graph, instead of sets as in a hypergraph. The vertices of the semigraph can be labelled by assigning symbols to them and the semigraph is called a labelled semigraph. In graph theory, various concepts can be applied on graphs, for instance the splicing system. Splicing system comprises of the field of formal language theory as well as discrete mathematics which focuses on the recombinants behavior of DNA molecules. Labelled semigraph can depict a problem where the labels hold certain values or properties. In DNA computing, a DNA molecule can be illustrated as a labelled semigraph where each vertex is assigned one label representing the nucleotide or base of the DNA molecule; and its cleavage pattern is described as a notion called n -cut splicing. In this research, the idea of n -cut splicing is applied on labelled semigraphs with one symbol. The number of the vertices in a component of n -cut spliced semigraph is firstly determined. Then, the form and order of the new labelled semigraphs generated by the recombination of the labelled n -cut spliced semigraphs are obtained. Besides that, this research also shows that the 180-degree structure of the newly obtained labelled semigraphs is also one of the results of the recombination.

INTRODUCTION

Splicing system is a study that emerged by the relation of computer science, mathematics, biology as well as engineering. The study of this multidisciplinary field has resulted many researches on the splicing system which led to various models of splicing involving a micro-bio molecule known as deoxyribonucleic acid (DNA), called as DNA splicing system. DNA splicing focuses on the recombination behavior of the DNA molecules. However, Freund [1] claimed that it is quite inadequate to describe a process involving three-dimensional molecules as one-dimensional strings model. Hence, Freund described the idea of splicing system by using graph and he introduced graph splicing system to overcome such inadequacy. A graph splicing system emphasizes the cuts on various edges of graphs and the various possible ways to reconnect them to form new graphs.

Graph splicing scheme with at least one defined graph splicing rule is an important component of a graph splicing system. A graph splicing scheme represents and describes a graph splicing system as a whole process. The importance to define at least one graph splicing rule is to restrict the edges that are to be cut. The splicing will never occur if there is no graph splicing rule defined in the graph splicing scheme. In graph splicing systems, multiple types of graph splicing have been introduced by the previous researchers, for instance, Freund [1] in his original publication on graph splicing system extended some ideas on the graph splicing involving linear and circular graphs. Freund introduced two types of graph splicing which are regular splicing and self-splicing. Also, a graph splicing scheme is regular if all graph splicing rules defined are regular.

Besides, a type of graph splicing known as n -cut splicing has been introduced by Jeyabharathi et. al [2] to illustrate the cleavage pattern of DNA strands. Jeyabharathi et. al used the concept of semigraph introduced by Sampthkumar [3] and applied the n -cut splicing which will generate two components of n -cut spliced semigraphs. In n -cut splicing,

the n -cut spliced semigraphs represents the sticky ends of the DNA molecules. Various researches have been done on the n -cut spliced semigraphs such as the characterization of the self-assembled semigraphs and the bipartite structure of the n -cut spliced semigraphs in [4] and [5], respectively. Besides, other concepts and techniques have been applied to n -cut spliced semigraphs. For instance, Thiagarajan et. al stated that new languages can be generated by applying the folding technique, cut vertex and cut edge on the n -cut spliced semigraphs [6-8]. Furthermore, Jeyabharathi et. al. [9] determined the norm of Parikh matrices on n -cut spliced semigraphs and Aisah et. al [10] specified such study on the 2-cut spliced semigraphs and 4-cut spliced semigraphs.

In graph theory, a semigraph is called as a labelled semigraph if there is any symbol assigned to the vertices of the graph. A labelled graph is usually used to illustrate problems that hold certain properties or values. For example, a DNA structure is constructed by two strands of oriented chain of nucleotides or bases. There are four types of bases known as Adenine, Thymine, Cytosine and Guanine, represented by symbols A, T, C and G, respectively. In graphs, the structure of the DNA can be illustrated as labelled semigraphs where the vertices represent the bases of the DNA, and the symbols of the bases are assigned to the vertices.

Hence in this research, the n -cut splicing is applied on the labelled semigraphs with one symbol to generate new labelled semigraphs with two symbols. The above result is presented by determining the number of the vertices in the n -cut spliced semigraph. The result is then used to show the form and order of the new labelled semigraphs. Lastly, all cases of the n -cut spliced semigraphs recombine to form the new labelled semigraphs is considered.

In the next section, some concepts and definitions that are needed throughout the study are stated and explained. In the third section, the main results are presented where the n -cut splicing is applied on the labelled semigraphs with one symbol. Finally, the results are concluded in the last section.

PRELIMINARIES

A graph is a mathematical structure constructed by two main objects which are points and lines. A graph is denoted as a two tuple, $G = (V, E)$ where V is a non-empty set of points known as vertices and E is a set of lines known as edges. The set of vertices and edges can also be denoted as $V(G)$ and $E(G)$, respectively. A graph is called as a labelled graph if there is a set of symbols, denoted as A assigned to the vertices of the graph. A labelled graph can be written as $G = (V, E, L)$ where V is the set of vertices, E is the set of edges, and L is a function from V to A . In other words, the function L is a function assigning a symbol to vertices of G , written as $L: x \rightarrow v$ or (v, x) where v is an element of V , meanwhile x is an element of A . In graph theory, a semigraph acts as one of the generalizations of graph, where the vertices are drawn as points and edges are drawn as lines [3]. Besides, a hypergraph is also a type of graphs that is used to generalize graphs. However, a hypergraph is drawn by grouping the vertices of the graph in sets where the sets represent the edges of the graphs. Hence, a semigraph is a more straightforward generalization as compared to hypergraph [3]. A semigraph is denoted as $SG = (V, X)$ where V is the set of vertices and X is the set of edges where connected n -tuples of distinct vertices for $n \geq 2$ and satisfies the following conditions:

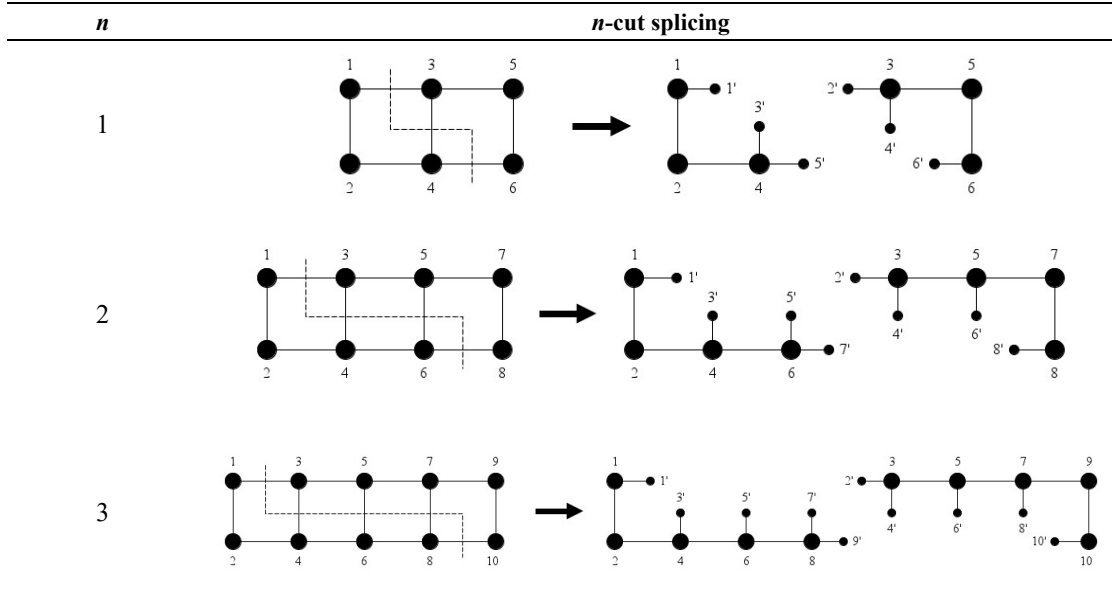
- i. Any two edges have at most one common vertex,
- ii. Two edges (u_1, u_2, \dots, u_n) and (v_1, v_2, \dots, v_m) are equal if
 - a) $m = n$ and
 - b) either $u_i = v_i$ or $u_i = v_{n+1-i}$, where $u_i, v_i \in V$ for $i = 1, 2, \dots, n$ or m .

Also, $(v_1, v_2, v_3, \dots, v_n)$ and $(v_n, v_{n-1}, \dots, v_1)$ are said to be the same edge.

The number of vertices in a semigraph is denoted as $|V(SG)|$ and the order of a semigraph denoted as $|SG|$ is the cardinality of its vertex set i.e., $|SG| = |V(SG)|$. Similar to graph, a semigraph can be a labelled semigraph, denoted as $SG(V, X, L)$.

In DNA splicing, one of the interesting parts is the cleavage pattern of the DNA molecules. In graph theory, the cleavage pattern of the DNA molecule can be illustrated and analyzed by the notion of n -cut splicing [2]. Note that the use of n -cut splicing is only specified to the semigraphs illustrating the DNA structure. Firstly, an example of a labelled semigraph $SG_1 = (V_1, X_1, L_1)$ in representing a DNA molecule is given, where $V_1 = \{1, 2, 3, 4, 5, 6, 7, 8\}$, $X_1 = \{(1, 2), (1, 3), (2, 4), (3, 4), (3, 5), (4, 6), (5, 6), (5, 7), (6, 8), (7, 8)\}$, and $L_1 = \{(1, A), (2, T), (3, A), (4, T), (5, G), (6, C), (7, G), (8, C)\}$ is shown. Figure 1 illustrates a DNA molecule in the form of a labelled semigraph.

TABLE 1. The illustration of n -cut splicing to produce two components of n -cut spliced semigraphs.



In this research, all graphs are illustrated as labelled semigraphs of maximum two symbols i.e., $x_1, x_2 \in A$ unless stated otherwise. For instance, a labelled semigraph is illustrated in Figure 4 where the vertices of the labelled semigraph are assigned with a label x_1 .

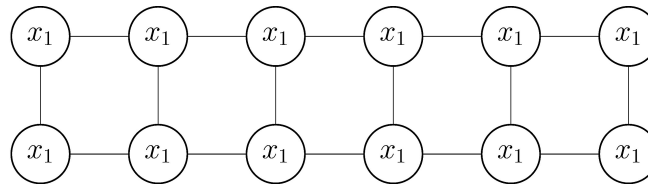


FIGURE 4. A labelled semigraph with the assigned symbol x_1

In the next section, the number of vertices in an n -cut spliced semigraph is determined. Then, the result is used to obtain and determine the form and order of the new labelled semigraphs after the recombination of the n -cut spliced semigraphs.

MAIN RESULTS

In this section, the number of vertices in any component of n -cut spliced semigraph is determined. Also, the n -cut splicing is applied on labelled semigraphs. Then, the form and the order of the new semigraphs after the recombination is presented.

Every graph splicing system focuses on two phases in its whole process in which the first phase is to cut the edges of the graph and the second phase is to recombine them in any possible way. This includes n -cut splicing on semigraphs where two components of n -cut spliced semigraphs are generated from the n -cut splicing and the recombination of the n -cut spliced semigraphs will form new semigraphs. In an n -cut spliced semigraph, the number vertices is one component of concern.

The idea to determine the number of vertices in a component of n -cut spliced semigraph is by observing the vertices that are never cut in an n -cut splicing and any other vertices that are still connected to them. The following proposition on the lower bound and the upper bound of the number of vertices in a component of n -cut spliced semigraphs is presented.

Proposition 1. Let SG be any semigraph with its order the same with the number of vertices. If an n -cut splicing is applied on SG and generates two components of n -cut spliced semigraphs denoted as SSG , then $n + 2$ and $|SG| - (n + 2)$, respectively, where SSG denotes any of the components of n -cut spliced semigraphs with $|V(SSG)|$ denotes the number of vertices.

Proof.

The proposition is proved by contradiction. Supposed that if $|V(SSG)| < n + 2$ or $|V(SSG)| > |SG| - (n + 2)$, then no n -cut spliced semigraph exist after n -cut splicing is applied on SG . Assume that $|V(SSG)| < n + 2$ or $|V(SSG)| > |SG| - (n + 2)$ is true. Let an n -cut splicing be applied on SG that generates two components of n -cut spliced semigraphs. Since the edge connecting the two vertices from the left most and the edge connecting the two vertices from the right most of the graph are never cut, then there will always be at least two vertices for each component of n -cut spliced semigraph. Then, for every edge cut by the n -cut splicing, there exist at least n vertices connected to the two vertices from left most (or the right most) which implies that $|V(SSG)| \geq n + 2$ hence contradicting that $|V(SSG)| < n + 2$. Next, since $|V(SSG)| \geq n + 2$, if a component of n -cut spliced semigraph has at least $|V(SSG)| = n + 2$ vertices, then another component of n -cut spliced semigraph has at least $|V(SSG)| \leq |SG| - (n + 2)$ which also contradicts $|V(SSG)| > |SG| - (n + 2)$. Therefore, if an n -cut splicing is applied on a semigraph SG and generates two components of n -cut spliced semigraphs with SSG denoting any of the components, then $n + 2 \leq |V(SSG)| \leq |SG| - (n + 2)$. ■

The above proposition shows that any component of n -cut spliced semigraph consists of at least $n + 2$ vertices and at most $|SG| - (n + 2)$ vertices. For an example, if a 1-cut splicing is applied on a semigraph SG with order $|SG| = 8$, the semigraphs will be cut into two components, in which one is the left overhang and another one is the right overhang. Hence, if one of the overhangs consists of three vertices since $n + 2 = 1 + 2$, then another overhang consists of five vertices obtained by $|SG| - (n + 2) = 8 - 3$.

Next, Propositions 2 and 3 are presented to determine the form and order of the new semigraphs after the recombination of two labelled n -cut spliced semigraphs. Propositions 2 and 3 are proved by analysing the number of vertices in the components of n -cut spliced semigraphs and by considering both left and right overhangs for the splicing on each labelled semigraphs, respectively.

Proposition 2. Let two labelled semigraphs be denoted as SG_i and SG_j for $i \neq j$ such that $|SG_i| = |V(SG_i)| = k$ and $|SG_j| = |V(SG_j)| = l$ where all vertices in SG_i and SG_j labelled with x_1, x_1 and x_2, x_2 , respectively. Suppose an n -cut splicing is applied on SG_i, SG_i and SG_j , two components of labelled n -cut spliced semigraphs are generated from splicing of each labelled semigraphs denoted as SSG_i and SSG_j where $|V(SSG_i)| = r$ and $|V(SSG_j)| = s$, respectively. Then the resulted graphs after the recombination of the labelled n -cut spliced semigraphs are labelled semigraphs SG_α, SG_α with two symbols, x_1 and x_2 with order $|SG_\alpha| = r + s$ such that $2(n + 2) \leq r + s \leq (k + l) - 2(n + 2)$.

Proof.

Let SG_i and SG_j be two labelled semigraphs where $|SG_i| = k$ and $|SG_j| = l$, respectively. Suppose that an n -cut splicing is applied on both SG_i and SG_j , then four components of labelled n -cut spliced semigraphs are generated. Let the number of vertices of any component of labelled n -cut spliced semigraphs denoted as SSG_i generated from SG_i be equal to r such that $n + 2 \leq r \leq k - (n + 2)$; meanwhile let the number of vertices of any component of labelled n -cut spliced semigraphs denoted as SSG_j generated from SG_j be equal to s such that $n + 2 \leq s \leq l - (n + 2)$. After the recombination of SSG_i and SSG_j , the resulted labelled semigraph SG_α consists $r + s$ number of vertices where r number of vertices are labelled with x_1 and s number of vertices are labelled with x_2 . Then $(n + 2) + (n + 2) \leq r + s \leq k - (n + 2) + (l - (n + 2))$ which implies that $2(n + 2) \leq r + s \leq (k + l) - 2(n + 2)$. Hence, $|SG_\alpha| = r + s$. ■

Proposition 3. Let $L(S)$ be the set of all semigraphs resulted by the recombination of the labelled n -cut spliced semigraphs. If SG_α is a labelled semigraph in $L(S)$ such that the vertices are labelled by two symbols, x_1 and x_2 , then there exists a labelled semigraph SG_β in $L(S)$ where the labelled semigraph is the 180-degree structure of SG_α .

Proof.

Let SG_i and SG_j be two labelled semigraphs with all the vertices assigned with x_1 and x_2 , respectively. Suppose SSG_i and SSG_j denote the vertices of the n -cut spliced semigraphs after n -cut splicings are applied on SG_i and SG_j , respectively such that $|V(SSG_i)| = r$ and $|V(SSG_j)| = s$. In this proof, we simplify the notation of vertices of the labelled n -cut spliced semigraphs by their labels which are x_1^r and x_2^s . Since x_1^r can be either a left overhang or a right overhang from splicing SG_i ; as well as x_2^s can be either a left overhang and a right overhang by splicing SG_j , hence if x_1^r is a left overhang, then it will recombine with x_2^s which is a right overhang and produces a new labelled semigraph SG_α where the vertices are denoted as $x_1^r x_2^s$. Also, if x_1^r is a right overhang, then it will recombine with x_2^s which is a left overhang and produces another new labelled semigraph SG_β where the vertices are denoted as $x_2^s x_1^r$. Therefore, $SSG_\alpha, SSG_\beta \in L(S)$. ■

The above two propositions have shown that for any recombination of two labelled n -cut spliced semigraphs of one symbol, the new semigraphs obtained must be in the form of labelled semigraphs with two symbols. Also, for every semigraph obtained by the recombination of the labelled n -cut spliced semigraph, there exists its 180-degree structure as one of the resulted new semigraphs.

The following example illustrates the n -cut splicing on two labelled semigraphs of one symbol in obtaining new semigraphs.

Example 1.

Let $A = \{x_1, x_2\}$ be a set of two symbols and two labelled semigraphs be denoted as SG_1 and SG_2 where all vertices are assigned with symbols x_1 and x_2 , respectively as shown in the following figure.

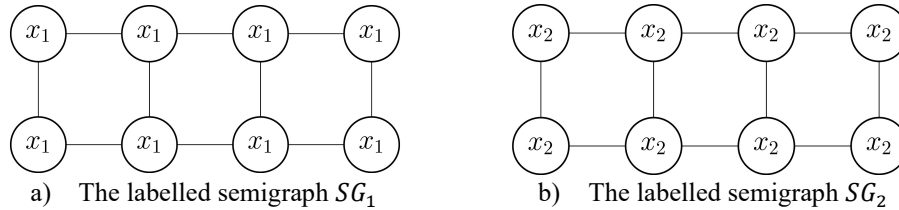


FIGURE 5. Two labelled semigraphs SG_1 and SG_2

If a 1-cut splicing is applied on both SG_1 and SG_2 , hence there will be two possible sites for the cut to occur on each of the labelled semigraph. Figure 6 illustrates all possible sites where PC_{1a} and PC_{1b} denote the possible sites of 1-cut splicing to occur on SG_1 ; meanwhile PC_{2a} and PC_{2b} denote the possible sites of 1-cut splicing to occur on SG_2 that are illustrated in Figure 6.

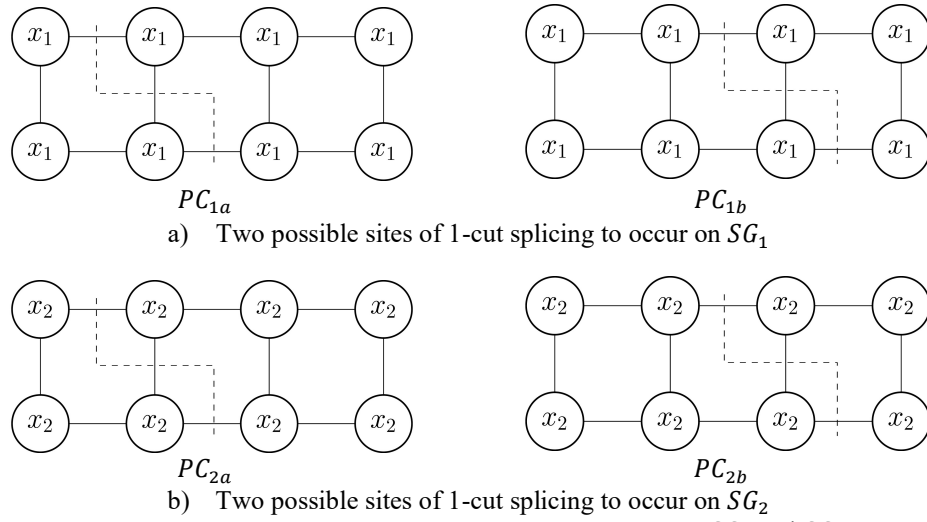


FIGURE 6. All possible sites for the 1-cut splicing to occur on SG_1 and SG_2

Hence, there are eight components of labelled n -cut spliced semigraphs generated by considering all possible sites of applying 1-cut splicing on SG_1 and SG_2 , depicted in Figure 7.

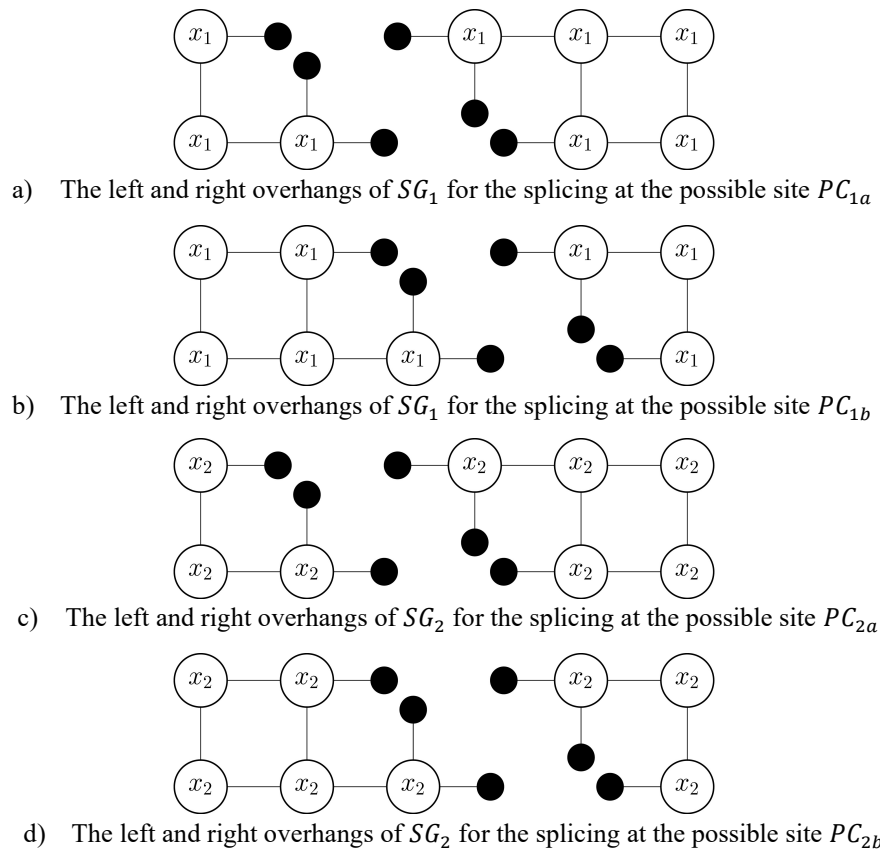


FIGURE 7. The left and right overhangs for each case of splicing SG_1SG_1 and SG_2 at the possible sites $PC_{1a}, PC_{1b}, PC_{2a}$ and PC_{2b}

After the recombination of the labelled n -cut spliced semigraphs in Figure 7, it shown that the resulted new labelled semigraphs are labelled with two symbols such that every new labelled semigraph generated has its 180-degree

structure i.e. SG_4, SG_6, SG_8 and SG_{10} are the 180-degree structure of SG_3, SG_5, SG_7 and SG_9 , respectively and vice versa as illustrated in Figure 8.

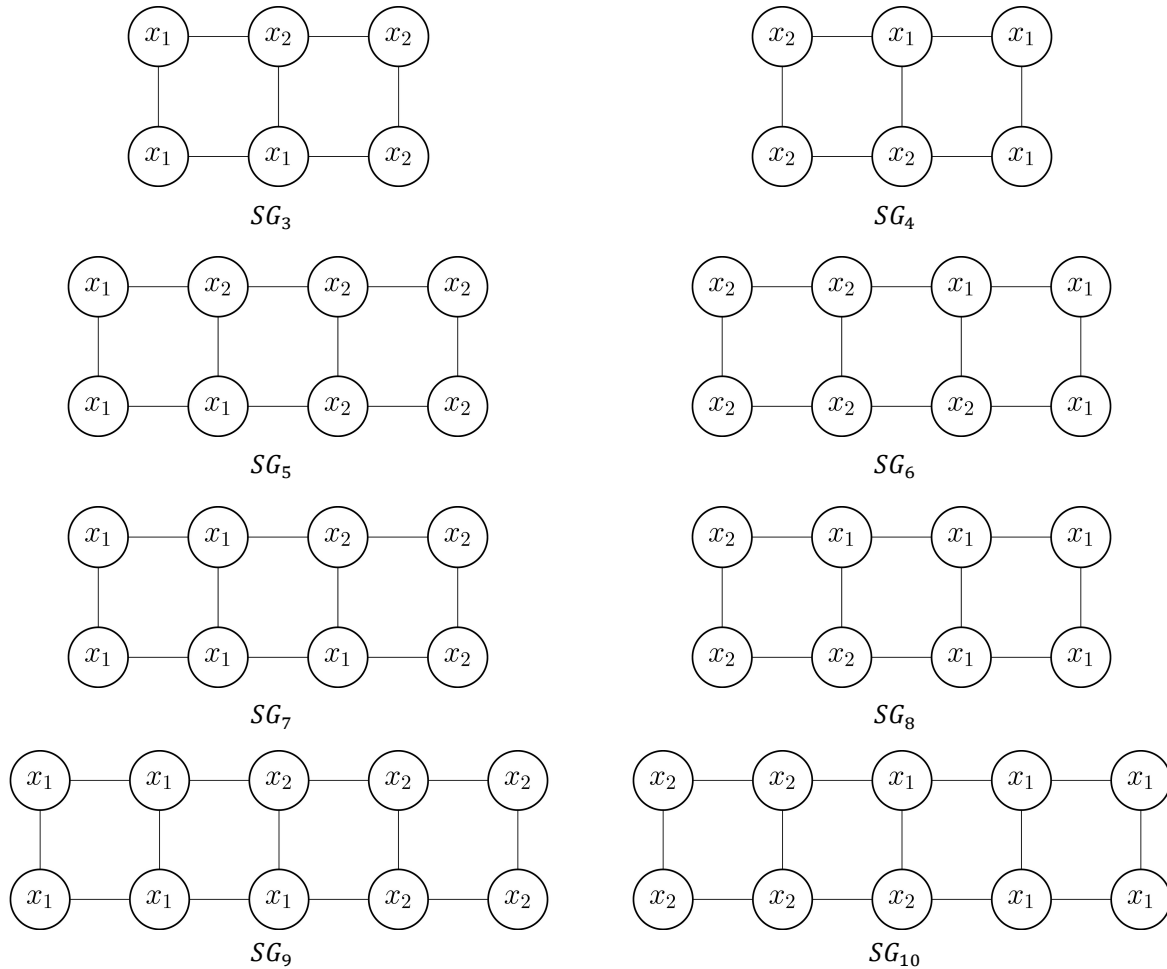


FIGURE 8. The possible resulted new labelled semigraphs after the recombination of the labelled n -cut spliced semigraphs

Note that the recombination stated is only considered for the case where the labelled n -cut spliced semigraph from SG_1 recombines with the labelled n -cut spliced semigraph from SG_2 . In other words, a labelled n -cut spliced semigraph does not recombine with the other labelled n -cut spliced semigraph from the same initial labelled semigraph.

From Example 1, two labelled semigraphs with order eight are presented. By observing Figure 7, it is shown that the lower and upper bound of the number of vertices in an n -cut spliced semigraph after 1-cut splicing is applied are $n + 2 = 1 + 2 = 3$ and $|SG| - (n + 2) = 8 - (1 + 2) = 8 - 3 = 5$, respectively. Also, the vertices resulted in new labelled semigraphs labelled with two symbols such that the minimum order of the labelled semigraph is $2(n + 2) = 2(1 + 2) = 6$; meanwhile the maximum order of the labelled semigraph is $(|SG_1| + |SG_2|) - 2(n + 2) = (8 + 8) - 6 = 10$.

Hence, it shown that applying an n -cut splicing on a semigraph will generate two components of n -cut spliced semigraphs where each component has at least $n + 2$ or at most $|SG| - (n + 2)$ number of vertices. It is also found that the vertices in a new labelled semigraph obtained after the recombination of two n -cut spliced semigraphs from splicings of two initial labelled semigraphs. Besides, the total number of vertices in the new labelled semigraph is at least $2(n + 2)$ and at most the sum of the order of the two initial labelled semigraphs minus $2(n + 2)$. Also, the new labelled semigraph obtained has its 180-degree structure by considering all cases of recombination.

CONCLUSION

In graph theory, the structure of a DNA molecule can be illustrated as a labelled semigraph and its cleavage pattern can be described as n -cut splicing. In this research, the idea of n -cut splicing is applied on the labelled semigraph with one symbol to generate new labelled semigraphs with two symbols. The above idea is achieved by firstly determining the number of vertices in a component of n -cut spliced semigraph. It is found that a component of n -cut spliced semigraph consists of at least $n + 2$ and at most $|SG| - (n + 2)$ number of vertices. Next, the form of the new labelled semigraph and its order is obtained, where the new labelled semigraph must consist of two symbols generated by the recombination of two labelled semigraphs of one symbol and the order is at least $2(n + 2)$ and at most the sum of order of the two initial labelled semigraphs minus $2(n + 2)$. Also, it is shown that for every generated new labelled semigraph, there exists another new labelled semigraph generated which is the 180-degree structure of itself.

ACKNOWLEDGMENTS

The authors would like to acknowledge Universiti Teknologi Malaysia (UTM) for the financial support through UTMShine Grant Vote number 09G89. The authors would also like to thank UTM for the funding of this research through UTM Fundamental Research Grant (UTMFR) Vote Number 20H70.

REFERENCES

1. R. Freund, "Splicing Systems on Graphs". in *1st Int. Symp. on Intelligence in Neural and Biological Systems (INBS'95)* (IEEE, Washington DC, 1995) pp. 189-194.
2. S. Jeyabharathi, J. Padmashree, S. S. Selvi and K. Thiagarajan, "Semigraph Structure on DNA Splicing System," in *6th Int. Conf. on Bio-Inspired Computing: Theories and Applications BIC-TA* (IEEE Computer Science, Washington DC, 2011) pp. 182-187.
3. E. Sampathkumar, C. M. Deshpande, B. Y. Bam, L. Pushpalatha and V. Swaminathan, *Semigraphs and Its Application*. (Academy of Discrete Mathematics and Applications, India, 2019).
4. S. Jeyabharathi, P. R. Indhu and S. S. Selvi, *Int. J. Comput. Techn. & Appl.* **5**, 774-780 (2014).
5. S. Jeyabharathi, M. Angayarkanni, S. S. Selvi and R. Anusha, *Int. J. Comput. Appl.* **6**, 56-62 (2012).
6. K. Thiagarajan, J. Padmashree and S. Jeyabharathi, "Evolutively Finite Words on Semigraph Folding," in *3rd Int. Conf. on Applied Mathematics and Pharmaceutical Sciences*, (Singapore, 2013) pp. 1-8.
7. K. Thiagarajan, J. Padmashree and P. Natarajan, *Indian J. Sci. Technol.* **9**, 1-9(2016).
8. K. Thiagarajan and J. Padmashree, *Journal of Physics: Conf. Series* **1000**, 1-11 (2018).
9. S. Jeyabharathi, K. Thiagarajan and S. S. Selvi, *Int. J. Adv. Eng. Techn.* **VII**, 81-84 (2016).
10. I. Aisah, P. R. E. Jayanti and A. K. Supriatna, in *IOP Conf. Series: Materials Science and Engineering*, 567 (IOP Publishing, 2019) pp. 1-8.