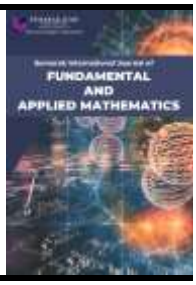
Semarak International Journal of
Fundamental and Applied Mathematics

Journal homepage:
https://semarakilmu.my/index.php/sijfam/index
ISSN: 3030-5527



Spectral Bipartition via Gap Cut on DNA Sequences

Hung Lik Goh[1], Wan Heng Fong[1,*], Sherzod Turaev[2]

[1] Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia, 81310 Skudai, Johor, Malaysia
[2] Department of Computer Science and Software Engineering, College of Information Technology, United Arab Emirates University, P.O. Box 15551, Al Ain, United Arab Emirates

ARTICLE INFO

Article history:
Received 28 June 2024
Received in revised form 3 August 2024
Accepted 27 August 2024
Available online 15 September 2024

Keywords:
DNA; Graph partitioning; Spectral bipartition; Fiedler theory; Algebraic graph theory

ABSTRACT

Deoxyribonucleic Acid (DNA) and graph partitioning are two distinct fields of study which can be linked in the structure of biological networks. Graph partitioning has been extensively studied but not its application in the biological field. This research explored on the application of spectral graph partitioning in DNA splicing, aiming to simulate the cleavage of DNA by performing spectral bipartition on DNA sequences in a DNA splicing system. This research incorporates Fiedler theory and algebraic graph theory, which are commonly utilized in network analysis and the analysis of graph connectivity. Some DNA sequences of even length are selected and expressed in graphical representations. The adjacency matrix, Laplacian matrix, and degree matrix are computed from the graphs, as well as the Fiedler value and Fiedler vector associated with the graphs. Gap cut is used as a method of spectral bipartition which produces two partitions of DNA sequence of unequal lengths. The generalizations of gap cut on DNA sequences of even length are provided as lemmas and theorem.

1. Introduction

Deoxyribonucleic acid (DNA) is a molecule that carries the genetic instructions for the development and function of all living organisms. The DNA molecule is a double helix, composed of two complementary strands of nucleotides that run in opposite directions. Each nucleotide is composed of a sugar (deoxyribose), a phosphate group, and one of the four nitrogenous bases: adenine (A), thymine (T), guanine (G), and cytosine (C). The sequence of these bases along the DNA strand encodes the genetic information [1]. In addition, DNA splicing system has its long history which includes information on the molecular mechanisms of splicing and the role of splicing in DNA sequences. Knowledge on DNA splicing has evolved over the years, as well as the techniques used to analyze DNA splicing such as high-throughput sequencing methods and computational techniques for identifying splice sites and predicting spliced transcripts. In recent studies, Ruslim *et al.,* [2,3] provided deep insight into the application of graph in DNA splicing system. Moreover, Ahmad *et al.,* [4] studied the characteristics of varieties of splicing systems and established their relations with

* *Corresponding author.*
E-mail address: fwh@utm.my

https://doi.org/10.37934/sijfam.3.1.1127 b

second order limit languages. Graph partitioning and DNA splicing are two distinct areas of research, but they can be connected in the context of the structure and organization of biological networks, such as in the molecular network that is formed by the interactions between DNA, Ribonucleic acid (RNA), and proteins in a cell. In the application of graph partitioning in biological field, Gatti *et al.,* [5] introduced graph bisection and partitioning on graph neural networks. Besides that, Kim *et al.,* [6] explored the modularity of RNA structures with tree graph representations by applying graph partitioning to divide an RNA graph into subgraphs. Since there are not many studies on the applications of graph theory and graph partitioning in the field of DNA splicing network, this research has been conducted to explore the application of spectral graph partitioning in the field of DNA splicing systems.

In mathematics, graph theory is a discipline that deals with the study of graphs and their properties. A graph is a mathematical structure that consists of a set of vertices and a set of edges that connect these vertices. The vertices represent objects whereas the edges represent the relationships between the objects [7]. There are many different types of graphs, including simple graphs, directed graphs, weighted graphs and more. Graph theory has been extensively studied and it has applications in a wide range of areas such as material science [8], thermal engineering [9,10] and fluid mechanics [11,12]. On the other hand, graph partitioning is a process of dividing a graph into several smaller subgraphs, or clusters, such that the nodes with each cluster are more strongly connected to each other than they are to nodes in other clusters [13]. The goal of graph partitioning is to find meaningful groupings of nodes in the graph that have some relationship or similarity.

This research aims to simulate the cleavage of DNA by performing spectral graph partitioning on DNA sequences, which is inspired by Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) gene editing technology aiming to edit genes by precisely cutting DNA [14], harness natural DNA repair processes and to modify the gene in the desired manner. We introduce a novel perspective by incorporating Fiedler theory and algebraic graph theory in DNA splicing system, which potentially yields insights and solutions not achievable through traditional methods, thereby fostering advancements in biotechnology. The objectives of this research are to represent DNA sequences in graphical form and to obtain graph partitions of DNA sequences using spectral bipartition.

In the next section, the method of expressing the DNA sequence in graphical representation is introduced, followed by an overview of spectral graph theory which includes adjacency matrix, degree matrix, Laplacian matrix, and Fiedler theory. Moreover, spectral graph partitioning is also presented in the methodology section. Then, the results of spectral bipartition on DNA sequences of length 12, 14 and 16 and the generalizations of spectral bipartition on DNA sequence with even-numbered length are presented in the results and discussions section. The last section of this paper concludes and highlights the findings in this research.

2. Methodology

This methodology section starts with introduction on the methods of expressing a DNA sequence in graphical representations. Additionally, the spectral graph theory is presented and followed by the spectral graph partitioning.

2.1 Graphical Representation of DNA Sequence

Here, random DNA sequences of length 12, 14 and 16 are first expressed in simple undirected graph as graphical representation, respectively. A method of forming vertices of a graph from DNA

sequences is introduced in this phase. Two adjacent elements from the original string form a vertex of the graph. Then, the vertices are connected in the form of a unit distance path graph. The intentions to construct such a graph structure are to represent the sequential order of elements from the original string and to capture the neighbourhood relationships of the elements that form the vertices of the graph. A unit distance path graph P is a simple graph with $|V_P| = |E_P| + 1$, where $|V_P|$ is the number of vertices in P and $|E_P|$ is the number of edges in P, such that all of its vertices and edges can be connected by a single straight line [15], with any edge connecting two vertices having the Euclidean distance of one [16].

For example, given a string ABCDEF, the vertices are formed by adjacent elements from the string, which are [AB], [BC], [CD], [DE], [EF]. In order to identify the vertex and the order of the elements in the original string, the square bracket is used for each pair of adjacent elements. The unit distance path graph is expressed in Figure 1.

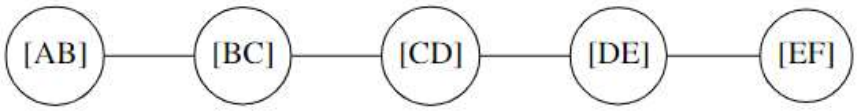


Fig. 1. Unit distance path graph of the string ABCDEF

Moreover, if there is any repeated pair of elements in the string, then the vertices are labelled with subscripts to distinguish the order of the elements in the string. For example, for the string ABCDAB, the vertices are formulated in the form of $[AB]_1$, $[BC]_1$, $[CD]_1$, $[DA]_1$, $[AB]_2$. Since the vertex [AB] occurred twice, then the first vertex [AB] is labelled as $[AB]_1$ whereas the second vertex $[AB]_2$ is labelled as $[AB]_2$. The unit distance path graph of the string ABCDAB is presented in Figure 2.

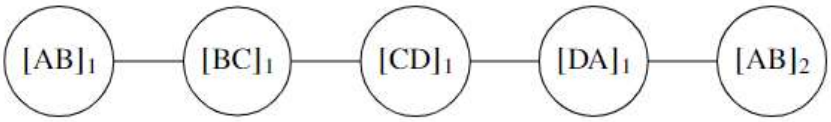


Fig. 2. Unit distance path graph of the string ABCDAB

2.2 Spectral Graph Theory

Spectral graph theory is the study of spectrum of various matrix representations of a given graph, such as the adjacency matrix, Laplacian matrix, and other related matrices [17]. The spectrum of a matrix is the set of all its eigenvalues. An eigenvalue is a scalar that satisfies a certain equation, involving the matrix and a corresponding eigenvector [18]. Besides that, adjacency matrix, degree matrix and Laplacian matrix are the key components for spectral graph theory.

2.2.1 Adjacency matrix

According to Spielman and Teng [19], consider an undirected graph G of n nodes with vertex set, $V = \{V_1, V_2, \dots, V_n\}$. For all $i, j = 1, 2, 3, \dots, n$, the adjacency matrix of the graph is denoted by $A(G)$, an $n \times n$ matrix with n real eigenvalues, where the entries of $A(G)$, A_{ij} are given by Eq. (1):

$$A_{ij} = \begin{cases} 1; \text{ if } V_i \text{ is adjacent to } V_j, \\ 0; \text{ otherwise.} \end{cases} \tag{1}$$

2.2.2 Degree matrix

Let d_i be the degree of the vertex V_i, the degree matrix of the graph is denoted by $D(G)$, where the entries of $D(G)$, D_{ij} are given by Eq. (2):

$$D_{ij} = \begin{cases} d_i; \text{ if } i = j, \\ 0; \text{ otherwise.} \end{cases} \tag{2}$$

2.2.3 Laplacian matrix

By denoting $L(G)$ to be the Laplacian matrix of the graph, the entries of $L(G)$, L_{ij} are given by Eq. (3):

$$L_{ij} = \begin{cases} d_i; & \text{if } i = j, \\ -1; & \text{if } V_i \text{ is adjacent to } V_j, \\ 0; & \text{otherwise.} \end{cases} \tag{3}$$

Furthermore, $L(G)$ is an $n \times n$ matrix with n real eigenvalues, which can be expressed in the form:

$$L(G) = D(G) - A(G). \tag{4}$$

2.2.4 Fiedler theory

Fiedler theory, which includes Fiedler value and Fiedler vector plays a crucial role in this research. Named after Miroslav Fiedler, Fiedler value is defined as the second smallest eigenvalue of the Laplacian matrix of a graph, whereas the eigenvector associated to the Fiedler value is known as Fiedler vector [20]. The entries of Fiedler vector are sorted in ascending order, and the graph can be partitioned by choosing a splitting value that separates the two parts. In addition, if two connected subgraphs with same size can be produced by spectral graph partitioning at the median of the Fiedler vector, then the entries of Fiedler vector must have balanced sign patterns [21].

Let $\bar{u} = \{u_1, u_2, u_3, \dots, u_n)^T$ be the Fiedler vector of $L(G)$, where u_i denotes the entries of Fiedler vector for $i = 1, 2, \dots, n$. The choices of splitting value, s, plays an important role in spectral partitioning such that two partitions are produced, one partition with $u_i > s$ and another with $u_i \leq s$, where $i = 1, 2, 3, \dots, n$. Such spitting is called a Fiedler cut. According to Spielman and Teng [19],

there are some popular choices for the Fiedler cut, such as bisection cut, ratio cut, gap cut, and sign cut.

2.3 Spectral Graph Partitioning

According to Shewchuk [22], spectral graph partitioning is a technique used in graph theory to divide a graph into several smaller subgraphs, or clusters. Spectral graph partitioning is based on the eigenvectors and eigenvalues of the Laplacian matrix of the graph and the idea behind it is to map the graph to a lower-dimensional space, such as a space of eigenvectors in order to find clusters in that space. In spectral graph partitioning, the eigenvectors of the Laplacian matrix of the graph are used to construct a mapping from the nodes of the graph to a lower-dimensional space. This research only includes spectral graph partitioning producing two partitions of DNA sequence, which is also known as spectral bipartition.

The procedures of spectral bipartition begin with the formulation of adjacency matrix and Laplacian matrix of each graph, followed by computing the eigenvalues and eigenvectors of the Laplacian matrix in order to identify the Fiedler value and Fiedler vector. Then, the partitions of the graph are obtained by choosing the splitting value corresponding to the gap cut to bipartition the Fiedler vector. Lastly, the partitioned vectors are mapped back to the vertices of the graph of DNA sequence, then the vertices split into their respective elements producing two partitions of DNA sequence.

In this research, gap cut is used in spectral bipartition, where the splitting value s is the largest gap of the sorted entries of Fiedler vector. Here, the gaps between each adjacent entry are computed by the absolute difference between the adjacent entries:

$$G_i = |u_{i+1} - u_i|, \tag{5}$$

where u_i are the entries of the Fiedler vector and G_i is the gap value, for $i = 1, 2, \ldots, n-1$.

3. Results

In this research, random DNA sequences of length 12, 14 and 16 are selected, namely GTACCGCGTACA (length 12), AGTCGTACCGTACG (length 14) and CTAGGTACATGACCGT (length 16). The results based on these three DNA sequences are sufficient to observe some patterns in terms of Fielder vector and the length of the partitions. The results of graphical representations and graph partitions of spectral bipartition via gap cut are presented in this section.

3.1 Results of DNA Sequence GTACCGCGTACA

The DNA sequence of length 12, GTACCGCGTACA, is first expressed in a unit distance path graph, where the graphical representation of GTACCGCGTACA is shown in Figure 3.

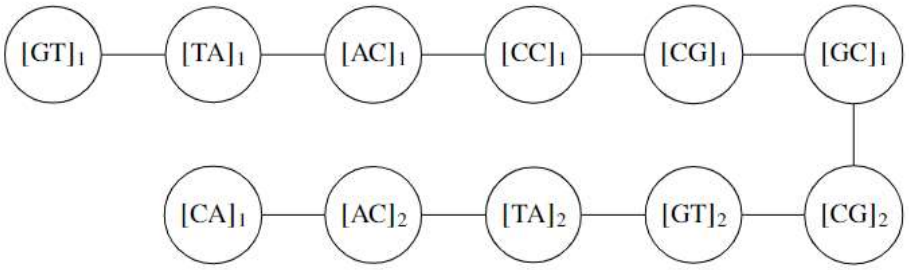


Fig. 3. Unit distance path graph of DNA sequence GTACCGCGTACA

The graph in Figure 3 is denoted as G_1 and its adjacency matrix, $A(G_1)$, degree matrix, $D(G_1)$, and Laplacian matrix, $L(G_1)$ are computed and presented in the following:

$$A(G_1) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \tag{6}$$

$$D(G_1) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{7}$$

$$L(G_1) = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}. \tag{8}$$

Then, the eigenvalues and eigenvectors of $L(G_1)$ are computed and tabulated in Table 1.

Table 1
Sets of eigenvalues and eigenvectors of $L(G_1)$

i	Eigenvalue, λ_i	Eigenvector, v_i
1	0	$(1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1)^T$
2	0.0810	$(-1 \quad -0.9190 \quad -0.7635 \quad -0.5462 \quad -0.2846 \quad 0 \quad 0.2846 \quad 0.5462 \quad 0.7635 \quad 0.9190 \quad 1)^T$
3	0.6903	$(-1 \quad 0.3091 \quad 0.5944 \quad 1.0822 \quad 0.8308 \quad 0 \quad -0.8308 \quad -1.0882 \quad -0.5944 \quad 0.3097 \quad 1)^T$
4	1.7154	$(-1 \quad 0.7154 \quad 1.2036 \quad -0.3728 \quad -1.3097 \quad 0 \quad 1.3097 \quad 0.3728 \quad -1.2036 \quad -0.7154 \quad 1)^T$
5	2.8308	$(-1 \quad 1.8308 \quad -0.5211 \quad -1.3979 \quad 1.6825 \quad 0 \quad -1.6825 \quad 1.3979 \quad 0.5211 \quad -1.8308 \quad 1)^T$
6	0.3175	$(1 \quad 0.6825 \quad 0.1483 \quad -0.4330 \quad -0.8768 \quad -1.0422 \quad -0.8768 \quad -0.4330 \quad 0.1483 \quad 0.6825 \quad 1)^T$
7	3.6825	$(-1 \quad 2.6825 \quad -3.5133 \quad 3.2287 \quad -1.9190 \quad 0 \quad 1.9190 \quad -3.2287 \quad 3.5133 \quad -2.6825 \quad 1)^T$
8	1.1692	$(1 \quad -0.1682 \quad -1.1406 \quad -0.7784 \quad 0.4938 \quad 1.1887 \quad 0.4938 \quad -0.7784 \quad -1.1406 \quad -0.1692 \quad 1)^T$
9	2.2846	$(1 \quad -1.2846 \quad -0.6344 \quad 1.4652 \quad 0.2173 \quad -1.5270 \quad 0.2173 \quad 1.4652 \quad -0.6344 \quad -1.2846 \quad 1)^T$
10	3.3097	$(1 \quad -2.3097 \quad 2.0251 \quad -0.3426 \quad -1.5764 \quad 2.4072 \quad -1.5764 \quad -0.3426 \quad 2.0251 \quad -2.3091 \quad 1)^T$
11	3.9190	$(1 \quad -2.9190 \quad 4.6015 \quad -5.9112 \quad 6.7420 \quad -7.0267 \quad 6.7420 \quad -5.9112 \quad 4.6015 \quad -2.9190 \quad 1)^T$

The second smallest eigenvalue of $L(G_1)$ is identified as the Fiedler value, which is 0.0810. Furthermore, the eigenvector corresponding to the Fiedler value is identified as the Fiedler vector, which $(-1 \quad -0.9190 \quad -0.7635 \quad -0.5462 \quad -0.2846 \quad 0 \quad 0.2846 \quad 0.5462 \quad 0.7635 \quad 0.9190 \quad 1)^T$. Then, the vertices of G_1 can be labelled according to the entries of the Fiedler vector, as shown in Figure 4.

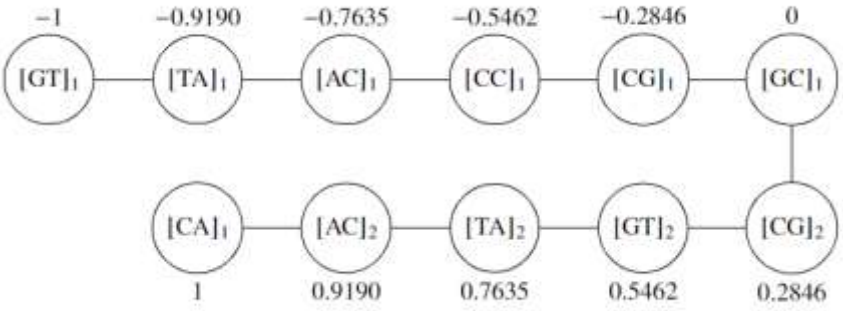


Fig. 4. Unit distance path graph of DNA sequence GTACCGCGTACA with labels

For gap cut, the gaps between each adjacent entry of Fiedler vector are computed, where the computations corrected up to four decimal places, as shown in Table 2, with u_{i+1} and u_i are adjacent entries of Fiedler vector whereas G_i is the gap value for $i = 1, 2, 3, \ldots, 10$.

Table 2
Computation of gap values between adjacent entries of Fiedler vector of G_1

i	u_{i+1}	u_i	$G_i = \lvert u_{i+1} - u_i \rvert$
1	-0.9190	-1	0.0810
2	-0.7635	-0.9190	0.1555
3	-0.5462	-0.7635	0.2173
4	-0.2846	-0.5462	0.2616
5	0	-0.2846	0.2846
6	0.2846	0	0.2846
7	0.5462	0.2846	0.2616
8	0.7635	0.5462	0.2173
9	0.9190	0.7635	0.1555
10	1	0.9190	0.0810

From Table 1, the largest gap is 0.2846, therefore the splitting value for gap cut is 0.2846. Hence, the vertex $[CG]_2$ is the splitting vertex such that the DNA sequence GTACCGCGTACA is bipartitioned, producing two partitions namely GTACCGC and GTACA, which are of length seven and length five respectively.

3.2 Results of DNA Sequence AGTCGTACCGTACG

The DNA sequence of length 14, AGTCGTACCGTACG, is first expressed in a unit distance path graph, where the graphical representation of AGTCGTACCGTACG is shown in Figure 5.

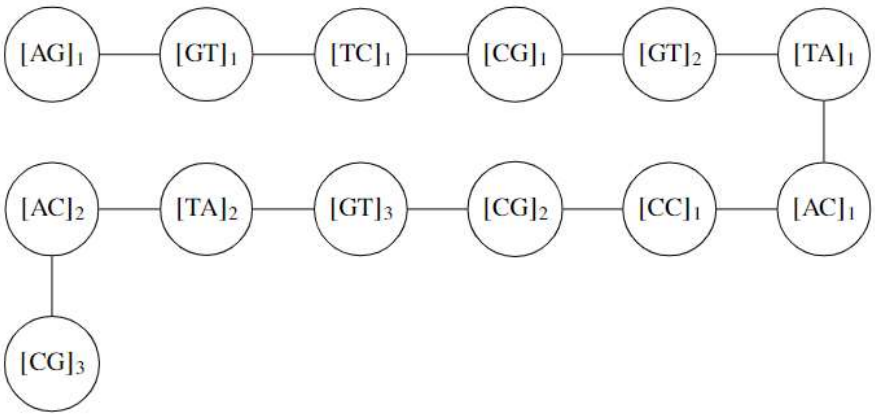


Fig. 5. Unit distance path graph of DNA sequence AGTCGTACCGTACG

By denoting the graph in Figure 5 as G_2, its adjacency matrix, $A(G_2)$, degree matrix, $D(G_2)$, and Laplacian matrix, $L(G_2)$ are computed and presented in the following:

$$A(G_2) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \tag{9}$$

$$D(G_2) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{10}$$

$$L(G_2) = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}. \tag{11}$$

Additionally, the eigenvalues and eigenvectors of $L(G_2)$ are computed and tabulated in Table 3.

Table 3
Sets of eigenvalues and eigenvectors of $L(G_2)$

i	Eigenvalue, λ_i	Eigenvector, v_i
1	0	$(1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1\ 1)^T$
2	0.2291	$(1\ 0.7709\ 0.3652\ -0.1241\ -0.5851\ -0.9120\ -1.0300\ -0.9120\ -0.5851\ -0.1241\ 0.3652\ 0.7709\ 1)^T$
3	0.8639	$(1\ 0.1361\ -0.8453\ -1.0965\ -0.4005\ 0.6416\ 1.1294\ 0.6416\ -0.4005\ -1.0965\ 0.8453\ 0.1361\ 1)^T$
4	1.7589	$(1\ -0.7589\ -1.1830\ 0.4737\ 1.2972\ -0.1610\ -1.3360\ 0.1610\ 1.2972\ 0.4737\ -1.1830\ -0.7589\ 1)^T$
5	2.7092	$(1\ -1.7092\ 0.2122\ 1.5587\ -1.3177\ -0.6242\ 1.7604\ -0.6242\ -1.3177\ 1.5587\ 0.2122\ -1.7092\ 1)^T$
6	3.4970	$(1\ -2.4970\ 2.7381\ -1.6020\ -0.3399\ 2.1108\ -2.8200\ 2.1108\ -0.3399\ -1.6020\ 2.7381\ -2.4970\ 1)^T$
7	3.9419	$(1\ -2.9419\ 4.7128\ -6.2098\ 7.3459\ -8.0552\ 8.2962\ -8.0552\ 7.3459\ -6.2098\ 4.7128\ -2.9419\ 1)^T$
8	0.0581	$(-1\ 0.9419\ -0.8290\ -0.6680\ -0.4681\ -0.2411\ 0\ 0.2411\ 0.4681\ 0.6680\ 0.8290\ 0.9419\ 1)^T$
9	0.5030	$(-1\ -0.4970\ 0.2559\ 0.8802\ 1.0617\ 0.7092\ 0\ -0.7092\ -1.0617\ -0.8802\ -0.2559\ 0.4970\ 1)^T$
10	1.2908	$(-1\ 0.2908\ 1.2062\ 0.5647\ -0.8058\ -1.1361\ 0\ 1.1361\ 0.8058\ -0.5647\ -1.2062\ -0.2908\ 1)^T$
11	2.2411	$(-1\ -1.2411\ 0.7008\ -1.4100\ -0.3609\ 1.4970\ 0\ -1.4970\ 0.3609\ 1.4100\ -0.7008\ -1.2411\ 1)^T$
12	3.1361	$(-1\ 2.1361\ -1.4269\ -0.5150\ 2.0120\ -1.7710\ 0\ 1.7710\ -2.0120\ 0.5150\ 1.4269\ -2.1361\ 1)^T$
13	3.7709	$(-1\ 2.7709\ -3.9070\ 4.1481\ -3.4389\ 1.9419\ 0\ -1.9419\ 3.4389\ -4.1481\ 3.9070\ -2.7709\ 1)^T$

The second smallest eigenvalue of $L(G_2)$ is identified as the Fiedler value, which is 0.0581. Moreover, the eigenvector corresponding to the Fiedler value is identified as Fiedler vector which is, $(-1\ -0.9419\ -0.8290\ -0.6680\ -0.4681\ -0.2411\ 0\ 0.2411\ 0.4681\ 0.6680\ 0.8290\ 0.9419\ 1)^T$. Then, the vertices of G_2 can be labelled according to the entries of the Fiedler vector, as shown in Figure 6.

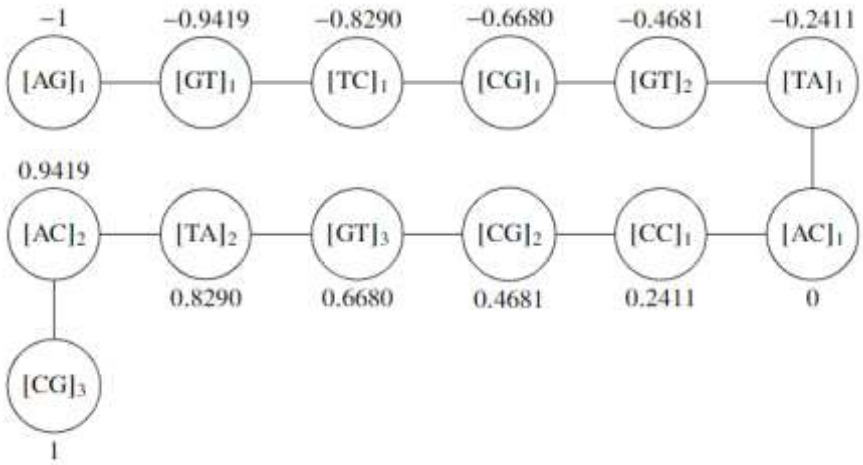


Fig. 6. Unit distance path graph of DNA sequence AGTCGTACCGTACG with labels

For gap cut, the gaps between each adjacent entry of Fiedler vector are computed, where the computations corrected up to four decimal places and tabulated in Table 4, with u_{i+1} and u_i are adjacent entries of Fiedler vector whereas G_i is the gap value for $i = 1, 2, 3, \dots, 12$.

Table 4
Computation of gap values between adjacent entries of Fiedler vector of G_2

i	u_{i+1}	u_i	$G_i = \lvert u_{i+1} - u_i \rvert$
1	-0.9419	-1	0.0581
2	-0.8290	-0.9419	0.1129
3	-0.6680	-0.8290	0.1610
4	-0.4681	-0.6680	0.1999
5	-0.2411	-0.4681	0.2270
6	0	-0.2411	0.2411
7	0.2411	0	0.2411
8	0.4681	0.2411	0.2270
9	0.6680	0.4681	0.1999
10	0.8290	0.6680	0.1610
11	0.9419	0.8290	0.1129
12	1	0.9419	0.0581

From Table 2, it is clear that the largest gap is 0.2411, therefore the splitting value for gap cut is 0.2411. Consequently, the vertex $[CC]_2$ is the splitting vertex such that the DNA sequence AGTCGTACCGTACG is bipartitioned, producing two partitions namely AGTCGTAC and CGTACG, which are of length eight and length six respectively.

3.3 Results of DNA Sequence CTAGGTACATCACCGT

The DNA sequence of length 16, CTAGGTACATCACCGT, is expressed in a unit distance path graph, where the graphical representation of CTAGGTACATCACCGT is shown in Figure 7.

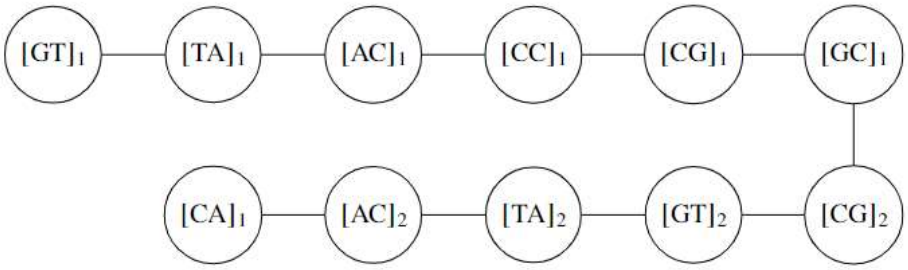


Fig. 7. Unit distance path graph of DNA sequence CTAGGTACATCACCGT

The graph in Figure 7 is denoted as G_3 and its adjacency matrix, $A(G_3)$, degree matrix, $D(G_3)$, and Laplacian matrix, $L(G_3)$ are computed and presented in the following:

$$A(G_3) = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}, \tag{12}$$

$$D(G_3) = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{13}$$

$$L(G_3) = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1 \end{pmatrix}. \tag{14}$$

Then, the eigenvalues and eigenvectors of $L(G_3)$ are computed and tabulated in Table 5.

Table 5
Sets of eigenvalues and eigenvectors of $L(G_3)$

i	Eigenvalue, λ_i	Eigenvector, v_i
1	0	$(1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1\ \ 1)^T$
2	1	$(-1\ \ 0\ \ 1\ \ 1\ \ 0\ \ -1\ \ -1\ \ 0\ \ 1\ \ 1\ \ 0\ \ -1\ \ -1\ \ 0\ \ 1)^T$
3	3	$(1\ \ -2\ \ 1\ \ 1\ \ -2\ \ 1\ \ 1\ \ -2\ \ 1\ \ 1\ \ -2\ \ 1\ \ 1\ \ -2\ \ 1)^T$
4	2.6180	$(-1\ \ 1.6180\ \ 0\ \ -1.6180\ \ 1\ \ 1\ \ -1.6180\ \ 0\ \ 1.6180\ \ -1\ \ -1\ \ 1.6180\ \ 0\ \ -1.6180\ \ 1)^T$
5	1.3820	$(1\ \ -0.3820\ \ -1.2361\ \ -0.3820\ \ 1\ \ 1\ \ -0.3820\ \ -1.2361\ \ -0.3820\ \ 1\ \ 1\ \ -0.3820\ \ -1.2361\ \ -0.3820\ \ 1)^T$
6	3.6180	$(1\ \ -2.6180\ \ 3.2361\ \ -2.6180\ \ 1\ \ 1\ \ -2.6180\ \ 3.2361\ \ -2.6180\ \ 1\ \ 1\ \ -2.6180\ \ 3.2361\ \ -2.6180\ \ 1)^T$
7	0.3820	$(-1\ \ -0.6180\ \ 0\ \ 0.6180\ \ 1\ \ -1\ \ -0.6180\ \ 0\ \ 0.6180\ \ 1\ \ -1\ \ -0.6180\ \ 0\ \ 0.6180\ \ 1)^T$
8	0.1729	$(1\ 0.8271\ 0.5112\ 0.1069\ -0.3159\ -0.6841\ -0.9340\ -1.0223\ -0.9340\ -0.6841\ -0.3159\ 0.1069\ 0.5112\ 0.8271\ 1)^T$
9	0.6617	$(1\ 0.3383\ -0.5473\ -1.0707\ -0.8856\ -0.1144\ 0.7325\ 1.0946\ 0.7325\ -0.1144\ -0.8856\ -1.0707\ -0.5473\ 0.3383\ 1)^T$
10	2.2091	$(1\ -1.2091\ -0.7472\ 1.3653\ 0.4618\ -1.4618\ -0.1562\ 1.4945\ -0.1562\ -1.4618\ 0.4618\ 1.3653\ -0.7472\ -1.2091\ 1)^T$
11	3.9563	$(1\ -2.9563\ 4.7833\ -6.4014\ 7.7397\ -8.7397\ 9.3577\ -9.5668\ 9.3577\ -8.7397\ 7.7397\ -6.4014\ 4.7833\ -2.9563\ 1)^T$
12	0.0437	$(-1\ -0.9563\ -0.8708\ -0.7472\ -0.5910\ -0.4090\ -0.2091\ 0\ 0.2091\ 0.4090\ 0.5910\ 0.7472\ 0.8708\ 0.9563\ 1)^T$
13	1.7909	$(-1\ 0.7909\ 1.1654\ -0.5473\ -1.2798\ 0.2798\ 1.3383\ 0\ -1.3383\ -0.2798\ 1.2798\ 0.5473\ -1.1654\ -0.7909\ 1)^T$
14	3.3383	$(-1\ -2.3383\ -2.1292\ 0.5112\ 1.4451\ -2.4451\ 1.8271\ 0\ -1.8271\ 2.4451\ -1.4451\ -0.5112\ 2.1292\ -2.3383\ 1)^T$
15	3.8271	$(-1\ 2.8271\ -4.1654\ 4.7834\ -4.5743\ 3.5743\ -1.9563\ 0\ 1.9563\ -3.5743\ 4.5743\ -4.7834\ 4.1654\ -2.8271\ 1)^T$

The second smallest eigenvalue of $L(G_3)$ is identified as the Fiedler value, which is 0.0437. The eigenvector corresponding to the Fiedler value is identified as the Fiedler vector, which is $(-1\ -0.9563\ -0.8708\ -0.7472\ -0.5910\ -0.4090\ -0.2091\ 0\ 0.2091\ 0.4090\ 0.5910\ 0.7472\ 0.8708\ 0.9563\ \ 1)^T$. Then, the vertices of G_3 can be labelled according to the entries of the Fiedler vector, as shown in Figure 8.

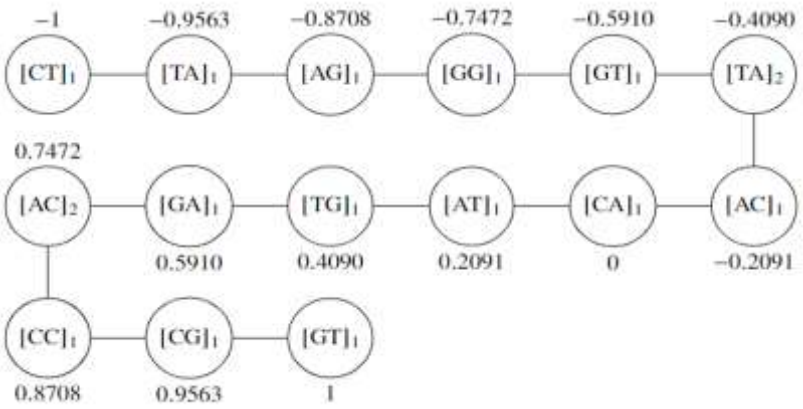


Fig. 8. Unit distance path graph of DNA sequence CTAGGTACATCACCGT with labels

For gap cut, the gaps between each adjacent entries of Fiedler vector are computed with the computations corrected up to four decimal places, as shown in Table 6, with u_{i+1} and u_i are adjacent entries of Fiedler vector whereas G_i is the gap value for $i = 1, 2, 3, \dots, 14$.

Table 6
Computation of gap values between adjacent entries of Fiedler vector of G_3

i	u_{i+1}	u_i	$G_i = \lvert u_{i+1} - u_i \rvert$
1	-0.9663	-1	0.0437
2	-0.8708	-0.9663	0.0855
3	-0.7472	-0.8708	0.1236
4	-0.5910	-0.7472	0.1562
5	-0.4090	-0.5910	0.1820
6	-0.2091	-0.4090	0.1999
7	0	-0.2091	0.2091
8	0.2091	0	0.2091
9	0.4090	0.2091	0.1999
10	0.5910	0.4090	0.1820
11	0.7472	0.5910	0.1562
12	0.8708	0.7472	0.1236
13	0.9563	0.8708	0.0855
14	1	0.9563	0.0437

Based on Table 6, the largest gap is 0.2091, which implies that the splitting value for gap cut is 0.2091. Thus, the vertex $[AT]_1$ is the splitting vertex such that the DNA sequence CTAGGTACATCACCGT is bipartitioned, producing two partitions namely CTAGGTACA and TGACCGT, which are of length nine and length seven respectively.

3.4 Generalization of Spectral Bipartition via Gap Cut on DNA Sequences with Even Length

After performing spectral bipartition via gap cut on DNA sequences of length 12, 14 and 16, some observations can be made based on the patterns of the results in order to generalize spectral bipartition via gap cut on DNA sequences with even-numbered length.

The Fiedler vector based on the graphs G_1, G_2 and G_3 have the form such that the median of the entries of Fiedler vector is zero and the entries on the left hand side of the median are of opposite signs with the entries on the right hand side of the median. This leads to Lemma 1 on the properties of the entries of Fiedler vector.

Lemma 1. The Fiedler vector v_f of a unit distance path graph G with $2n-1$ vertices with $n \in \mathbb{N}$, has the form of $(-u_{2n-1} \quad -u_{2n-2} \quad \dots \quad -u_{n+1} \quad 0 \quad u_{n+1} \quad \dots \quad u_{2n-2} \quad u_{2n-1})^T$, such that $u_{2n-1} > u_{2n-2} > \cdots > u_{n+2} > u_{n+1} > 0$, where u_i are the entries of the v_f for $i = n+1, n+2, \dots, 2n-1$.

Proof. Firstly, the Fiedler vector is the eigenvector corresponding with the Fiedler value, where the entries of the Fiedler vector, u_i with $i = 1,2, \dots, 2n-1$, are associated with the vertices of the graph. Consider that $v_f = (u_1 \quad u_2 \quad \dots \quad u_{n-1} \quad u_n \quad u_{n+1} \quad \dots \quad u_{2n-1})^T$, such that $u_1 < u_2 < \cdots < u_{2n-1}$, is the Fiedler vector of the unit distance path graph G with $2n-1$ vertices, where the vertices are given as $V = \{V_1, V_2, \dots, V_{2n-1}\}$. The unit distance path graph G is shown in Figure 9.

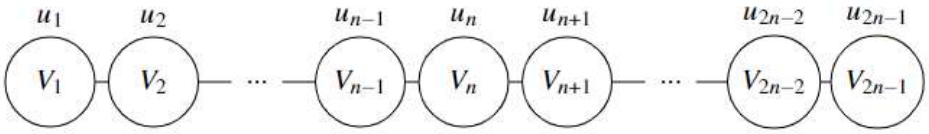


Fig. 9. Unit distance path graph of G labelled with the entries of v_f

The path graph of G is symmetrical at the vertex V_n since the graphical structures on the left-hand side and the right-hand side of V_n are identical. Thus, the median of the entries of v_f, which is u_n must be equal to zero. Since two connected subgraphs with the same size can be produced by spectral graph partitioning at V_n, the entries of v_f must have balanced sign patterns. It follows that the entries from u_1 to u_{n-1} are negative and the entries from u_{n+1} to u_{2n-1} are positive since the entries of v_f are sorted in ascending order. Moreover, with identical graphical structure of both sides of V_n, it can be deduced that $u_{n-1} = -u_{n+2}$ since V_{n-1} corresponds to V_{n+1}. Similarly, $u_{n-2} = -u_{n+2}$ since V_{n-2} corresponds toV_{n+2}. With the same reasoning, it follows that $u_{n-3} = -u_{n+3}, u_{n-4} = -u_{n+4}, \dots, u_2 = -u_{2n-2}, u_1 = -u_{2n-1}$. Hence, v_f has the form:

$$v_f = (-u_{2n-1} \quad -u_{2n-2} \quad \dots \quad -u_{n+1} \quad 0 \quad u_{n+1} \quad \dots \quad u_{2n-2} \quad u_{2n-1})^T,$$

and the proof is completed.

One of the observations made from the computation of gap values in DNA sequence of length 12, 14 and 16 is that the median of the entries and its neighboring entries have the largest gap value. Consequently, Lemma 2 on the largest gap of the adjacent entries of v_f is presented.

Lemma 2. For $v_f = (-u_{2n-1} \quad -u_{2n-2} \quad \dots \quad -u_{n+1} \quad 0 \quad u_{n+1} \quad \dots \quad u_{2n-2} \quad u_{2n-1})^T$ such that $u_{2n-1} > u_{2n-2} > \cdots > u_{n+2} > u_{n+1} > 0$, where $n \in \mathbb{N}$ where u_i are the entries of the Fiedler vector for $i = n+1, n+2, \dots, 2n-1$, the largest gap of the adjacent entries is u_{n+1}.

Proof. The gap value between adjacent entries of v_f is computed by using the absolute difference equation, $G_i = |u_{i+1} - u_i|$,where u_i are the entries of v_f and G_i is the gap value, for $i = 1,2,3, \dots, 2n-1$. Computing the gap value of u_{2n-2} and u_{2n-1} yields the following inequalities:

$$\begin{aligned} |u_{2n-1} - u_{2n-2}| &= |u_{2n-2} - u_{2n-1}| \\ &< |u_{2n-2} - u_{2n-3}| \\ &< |u_{2n-3} - u_{2n-4}| \\ &\vdots \\ &< |u_{n+2} - u_{n+1}| \\ &< |u_{n+1} - 0|. \end{aligned}$$

Hence,$|u_{2n-1} - u_{2n-2}| < |u_{2n-2} - u_{2n-3}| < \cdots < |u_{n+2} - u_{n+1}| < |u_{n+1} - 0| = u_{n+1}$, and the largest gap is u_{n+1}. The proof is then completed.

It is important to note that the gap value of the entries $-u_{n+1}$ and 0 is the same with the gap value of the entries 0 and u_{n+1} since $|u_{n+1} - 0| = |0 - (-u_{n+1})| = u_{n+1}$. Next, the largest gap of the adjacent entries which equals to u_{n+1}, is selected as the splitting value of gap cut, which leads to Theorem 1.

Theorem 1. For DNA sequence $b_1b_2b_3 \dots b_{2n-1}b_{2n}$ with even length, where $n \in \mathbb{N}$, gap cut on $b_1b_2b_3 \dots b_{2n-1}b_{2n}$ produces two partitions of length $n+1$ and length $n-1$ respectively.

Proof. With arbitrary DNA sequence $b_1b_2b_3 \dots b_{2n-1}b_{2n}$ of even length, the vertices of $b_1b_2b_3 \dots b_{2n-1}b_{2n}$ are in the form of $[b_1b_2], [b_2b_3], [b_3b_4], \dots, [b_{2n-2}b_{2n}]$. Clearly, graph G has $2n-1$ vertices. Next, denote the vertex set of G, $V = \{V_1, V_2, \dots, V_{2n-1}\}$ in the form $V_k = [b_kb_{k+1}]$ for $k = 1, 2, 3, \dots, 2n-1$. From Lemma 1, the Fiedler vector v_f has the form of $(-u_{2n-1} \quad -u_{2n-2} \quad \dots \quad -u_{n+1} \quad 0 \quad u_{n+1} \quad \dots \quad u_{2n-2} \quad u_{2n-1})^T$, with $u_{2n-1} > u_{2n-2} > \dots > u_{n+2} > u_{n+1} > 0$. Furthermore, with the aid of Lemma 2, the largest gap of the adjacent entries which equals to u_{n+1} is selected as the splitting value of gap cut. Referring to Figure 9, vertex $V_{n+1} = [b_{n+1}b_{n+2}]$ is the splitting point since vertex V_{n+1} is assigned with the splitting value u_{n+1}. Consequently, the DNA sequence $b_1b_2b_3 \dots b_{2n-1}b_{2n}$ is bipartitioed such that two partitions are produced, namely $b_1b_2b_3 \dots b_nb_{n+1}$ and $b_{n+2}b_{n+3}b_{n+4} \dots b_{2n-1}b_{2n}$, with length $n+1$ and length $n-1$ respectively, therefore the proof is completed.

4. Conclusions

In this paper, spectral bipartition via gap cut is applied to random selected DNA sequences of length 12, 14 and 16, which are expressed in unit distance path graphs. The results show that spectral bipartition via gap cut on DNA sequences of 12, 14 and 16 produces two partitions of unequal lengths. Additionally, a few observations based on the spectral bipartition via gap cut on DNA sequences of length 12, 14 and 16 are discussed, leading to generalization of spectral bipartition via gap cut on DNA sequence with even length which includes the pattern of the entries of the Fiedler vector and the largest gap of adjacent entries of the Fiedler vector in the form of lemmas. Consequently, spectral bipartition via gap cut on DNA sequence with even length is presented which results in two partitions with length $n+1$ and $n-1$ respectively.

Acknowledgement

The authors would like to acknowledge Universiti Teknologi Malaysia (UTM) and Research Management Centre for the financial support through Universiti Teknologi Malaysia Fundamental Research (UTMFR) Vote Number QJ130000.3854.22H45.

References

[1] Rettner, Rachael. "What is DNA?." Live Science (2021).
[2] Ruslim, Nooradelena Mohd, Marta Elizabeth, Yuhani Yusof, Mohd Sham Mohamad, and Noraziah Adzhar. "Deoxyribonucleic acid (DNA) splicing system from graph theoretic perspective." In *Journal of Physics: Conference Series*, vol. 1988, no. 1, p. 012081. IOP Publishing, 2021. https://doi.org/10.1088/1742-6596/1988/1/012081
[3] Ruslim, Nooradelena Mohd, Yuhani Yusof, Mohd Sham Mohamad, and Mohammad Hassan Mudaber. "A Bibliometric Review on Deoxyribonucleic Acid (DNA) Splicing System." *Journal of Advanced Research in Micro and Nano Engineering* 18, no. 1 (2024): 123-137. https://doi.org/10.37934/armne.18.1.123137
[4] Ahmad, Muhammad Azrin, Nor Haniza Sarmin, Wan Heng Fong, Yuhani Yusof, and Noraziah Adzhar. "On the new relation of second order limit language and other different types of splicing system." In *AIP Conference Proceedings*, vol. 2266, no. 1. AIP Publishing, 2020. https://doi.org/10.1063/5.0018075
[5] Gatti, Alice, Zhixiong Hu, Tess Smidt, Esmond G. Ng, and Pieter Ghysels. "Deep learning and spectral embedding for graph partitioning." In *Proceedings of the 2022 SIAM Conference on Parallel Processing for Scientific Computing*, pp. 25-36. Society for Industrial and Applied Mathematics, 2022.. https://doi.org/10.1137/1.9781611977141.3
[6] Kim, Namhee, Zhe Zheng, Shereef Elmetwaly, and Tamar Schlick. "RNA graph partitioning for the discovery of RNA modularity: a novel application of graph partition algorithm to biology." *PloS one* 9, no. 9 (2014): e106074. https://doi.org/10.1371/journal.pone.0106074

[7] Wilson, Robin J. Introduction to graph theory. Pearson Education India, 1979.
[8] Said, Nurlaela Muhammad, Mohd Basri Ali, and Kamarul Ariffin Zakaria. "Correlation of Absorb Energy with PSD Energy and Area under Strain-Time Grap." *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences* 49, no. 2 (2018): 126-137.
[9] Benslimane, Farida, Fatah Bounaama, and Belkacem Draoui. "Bond Graph Modeling and Control of A Single-Zone Building in a Semi-Arid Region for Thermal Comfort." *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences* 61, no. 1 (2019): 94-105.
[10] Zainudin, Amira Syuhada, and Abdul Rahim Othman. "Thermal Stability of PALF-PP and PALF-PLA for Natural Fiber Honeycomb Core Materials." *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences* 87, no. 1 (2021): 20-29. https://doi.org/10.37934/arfmts.87.1.2029
[11] Afikuzzaman, Mohammad, Mohammad Ferdows, Raushan Ara Quadir, and Md Mahmud Alam. "MHD Viscous incompressible Casson fluid flow with hall current." *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences* 60, no. 2 (2019): 270-282.
[12] Zulkiflee, Fasihah, Sharidan Shafie, and Ahmad Qushairi Mohamad. "Unsteady free convection flow of nanofluids between vertical oscillating plates with mass diffusion." *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences* 76, no. 2 (2020): 118-131. https://doi.org/10.37934/arfmts.76.2.118131
[13] Bichot, Charles-Edmond, and Patrick Siarry, eds. Graph partitioning. John Wiley & Sons, 2013. https://doi.org/10.1002/9781118601181
[14] Christie, Kathleen A., Jimmy A. Guo, Rachel A. Silverstein, Roman M. Doll, Megumu Mabuchi, Hannah E. Stutzman, Jiecong Lin et al. "Precise DNA cleavage using CRISPR-SpRYgests." *Nature biotechnology* 41, no. 3 (2023): 409-416. https://doi.org/10.1038/s41587-022-01492-y
[15] Gross, Jonathan L., Jay Yellen, and Mark Anderson. Graph theory and its applications. Chapman and Hall/CRC, 2018. https://doi.org/10.1201/9780429425134
[16] Horvat, Boris, and Tomaž Pisanski. "Products of unit distance graphs." *Discrete mathematics* 310, no. 12 (2010): 1783-1792. https://doi.org/10.1016/j.disc.2009.11.035
[17] Hogben, Leslie. "Spectral graph theory and the inverse eigenvalue problem of a graph." *The Electronic Journal of Linear Algebra* 14 (2005): 12-31. https://doi.org/10.13001/1081-3810.1174
[18] Horn, Roger A., and Charles R. Johnson. Matrix analysis. Cambridge university press, 2012.
[19] Spielman, Daniel A., and Shang-Hua Teng. "Spectral partitioning works: Planar graphs and finite element meshes." In *Proceedings of 37th conference on foundations of computer science*, pp. 96-105. IEEE, 1996. https://doi.org/10.1109/SFCS.1996.548468
[20] Fiedler, Miroslav. "A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory." *Czechoslovak mathematical journal* 25, no. 4 (1975): 619-633. https://doi.org/10.21136/CMJ.1975.101357
[21] Kim, Sooyeong, and Steve Kirkland. "Fiedler vectors with unbalanced sign patterns." *Czechoslovak Mathematical Journal* 71, no. 4 (2021): 1071-1098. https://doi.org/10.21136/CMJ.2021.0198-20
[22] Shewchuk, Jonathan Richard. Allow me to introduce spectral and isoperimetric graph partitioning. Technical Report, 2016.