Original article

Study of biological networks using graph theory

Wei Gao [a], Hualong Wu [a,*], Muhammad Kamran Siddiqui [b], Abdul Qudair Baig [c]

[a] School of Information Science and Technology, Yunnan Normal University, Kunming 650500, China
[b] Department of Mathematics, Comsats Institute of Information Technology, Sahiwal 57000, Pakistan
[c] Department of Mathematics, Comsats Institute of Information Technology, Attock Campus, Pakistan

ABSTRACT

As an effective modeling, analysis and computational tool, graph theory is widely used in biological mathematics to deal with various biology problems. In the field of microbiology, graph can express the molecular structure, where cell, gene or protein can be denoted as a vertex, and the connect element can be regarded as an edge. In this way, the biological activity characteristic can be measured via topological index computing in the corresponding graphs. In our article, we mainly study the biology features of biological networks in terms of eccentric topological indices computation. By means of graph structure analysis and distance calculating, the exact expression of several important eccentric related indices of hypertree network and X-tree are determined. The conclusions we get in this paper illustrate that the bioengineering has the promising application prospects.

1. Introduction

In mathematical biology, mathematical methods are applied to biology to deal with various modeling and calculation problems. In the microscopic field of biology, DNA and other protein molecular structure can be represented as a graph, and thus as a mathematical tool, graph theory is introduced to the analysis and calculation of molecular topology.

1.1. Example 1. DNA graphs

The study of DNA sequence is the most important issue in biology science, and there are lots of contributions on DNA analysis and computation from mathematical and algorithmic point of view (Rastgou et al., 2017; Shoaib et al., 2017) pointed out that Hartree-Fock exchange percentage of density functional has a key factor in getting the structure electronic properties. The notable features sequencing platform based on a mathematical framework and it working mechanism has some characteristics, such as: optimizing cost, implement, and sensitivity analysis to different parameters (O'Reilly et al., 2016). The demography and reproductive success and by means of coalescent theory to compute mitochondrial DNA sequences from the Japanese sardine (Niwa et al., 2016; Samina and Khan, 2017; Khan et al., 2017a, 2017b). The DNA storage channel and modeled the read process considering profile vectors. They raised new asymmetric coding tricks to combat the effects of sequencing noise and synthesis, and an asymptotic analysis of the number of profile vectors was also presented. At last, two families of codes for this channel model were constructed. To effectively store FASTQ files raised by big DNA sequencers, Chlopkowski et al. (2016) determined a specialized compressor designing. The ionization potential with single and double excitations was estimated by means of equation of motion coupled cluster trick, and VIEs is estimated in terms of density functional theory with dispersion corrected omega B97x-D (Chakraborty and Ghosh, 2016). A studied on how to build independent spanning trees on hypercubes and how to use them to predict mitochondrial DNA sequence parts through paths on the hypercube (da Silva and Pedrini, 2016). An alignment-free technology for DNA sequence similarity analysis based on graph theory concepts and genetic codes (Jafarzadeh and Iranmanesh, 2016a, 2016b). The new approach to test the DNA sequences using optical joint Fourier transform (Alqallaf and Cherri, 2016). Theoretically studied the transverse electron transport through all four DNA nucleotide bases by electron propagator theory (Kletsov et al., 2015).

Let $k \geqslant 2$ be an integer. The DNA graph is defined by Wang et al. (2008), and it said that a directed graph $D = (V(D), E(D))$ is DNA

---

graph whether it can assign a label $(l_1(x), l_2(x), \ldots, l_k(x))$ of length k to each vertex $x \in V(D)$ as follows:

(1) $l_i(x) \in \{A, C, T, G\}$ where $i \in \{1, \ldots, k\}$;
(2) $(l_1(x), l_2(x), \ldots, l_k(x)) \neq (l_1(y), l_2(y), \ldots, l_k(y))$ if $x \neq y$;
(3) $(x, y) \in E(D)$ if an only if $(l_2(x), \ldots, l_k(x)) = (l_1(y), \ldots, l_{k-1}(y))$.

For a multiset consists of oligonucleotides with length k, a DNA graph can be constructed as follows: set each oligonucleotide with length k from the multiset as a vertex; add an arc between two vertices if the $k - 1$ rightmost nucleotides of first vertex overlap with the $k - 1$ leftmost nucleotides of the second one. Several contributions on DNA graph and DNA mathematical expression (Aram and Iranmanesh, 2012; Pesek and Zerovnik, 2008; Jafarzadeh and Iranmanesh, 2012; Bokhari and Sauer, 2005; Blazewicz et al., 1999; Sa-Ardyen et al., 2003; Pendavingh et al., 2003; Blazewicz et al., 2009; Pevzner et al., 2001; Jafarzadeh and Iranmanesh, 2013; Khan et al., 2017a, 2017b; Shah and Bushnaq, 2017).

1.2. Example 2. Biological networks

The main task of microorganism science is to study the viruses, protozoa, bacteria, euglena, opalinia, fungi, paramecium and amoeba et al. All living organisms consist of cells which are basic structure of life and can be expressed as vertex in the graph model. As an important biology computation model, biological networks are used to deal with biology problems in which its vertex represent cells, genes or proteins, and its edges are expressed as the potential connection between these components. For instance, the mathematical framework of protein interaction, gene expression, carrier transfer information and metabolic networks can be regarded as biological networks. The topological index defined on biological networks can be considered as a numeric function which maps the given structure to a real number, and thus measure its physical, chemical and biological characters. Some contributions on biological networks and other graph applications in biology science can refer to (Jafarzadeh and Iranmanesh, 2016a, 2016b; Nandagopal and Elowitz, 2011; Rajan et al., 2015; Alexander, 2013; Banerjee and Jost, 2007; Ma'ayan, 2008; Mason and Verwoerd, 2006; Lesne, 2006; Eckman and Brown, 2006; Aittokallio and Schwikowski, 2006; El-Ghoul et al., 2006).

In this paper, we focus on the biochemical properties of biological networks via topological index computation. The structure of the rest of this paper is stated as follows: we first give the information about the setting of topological index defined on biological networks; then, the main results and their detailed proofs are presented in Section 3; at last, we discuss the future projects in this filed.

2. Setting

In what follows, let $G = (V(G), E(G))$ be a undirected molecular graph with vertex set $V(G)$ and edge set $E(G)$, where each vertex expresses a cell, gene or protein, and each edge is presented as the connection between two components. A topological index can be a function $f: G \to \mathbb{R}^+$ which can map each molecular graph to a positive real number. Several degree-based or distance-based topological indices like Wiener index, atom-bond connectivity index, harmonic index, sum connectivity index and others are defined to test the chemical, physical, pharmaceutical and biological properties. Furthermore, there are some mention-able work on distance-based and degree-based topological indices of special structures which can be referred to Basavanagoud et al. (2017); Gao et al. (2016a, 2016b, 2016c, 2016d, 2017a, 2017b); Gao and Siddiqui (2017); Gao and Wang (2015, 2016).

The distance $d(u, v)$ between two vertices u and v in a connected graph is denoted as the length of the shortest path between them. For any vertex $v \in V(G)$, the eccentricity of v is defined as $ec(v) = \max\{d(v, u) | u \in V(G)\}$.

Ghorbani and Khaki (2010) introduced the eccentric version of geometric-arithmetic index as fourth geometric-arithmetic eccentricity index which is stated as

$$GA_4(G) = \sum_{uv \in E(G)} \frac{2\sqrt{ec(u)ec(v)}}{ec(u) + ec(v)}.$$

The fourth Zagreb index is defined by Farahani and Kanna (2015) as

$$Zg_4(G) = \sum_{uv \in E(G)} (ec(u) + ec(v)).$$

Its multiplicative version, named as the fourth multiplicative Zagreb index, is stated by

$$\Pi_4^*(G) = \prod_{uv \in E(G)} (ec(u) + ec(v)).$$

Analogously, the sixth Zagreb index is described as

$$Zg_6(G) = \sum_{uv \in E(G)} ec(u)ec(v).$$

And, the multiplicative version, named as sixth multiplicative Zagreb index is defined by

$$\Pi_6^*(G) = \prod_{uv \in E(G)} ec(u)ec(v).$$

Correspondingly, the fourth Zagreb polynomial and the sixth Zagreb polynomial are defined as

$$Zg_4(G, x) = \sum_{uv \in E(G)} x^{ec(u)+ec(v)}$$

and

$$Zg_6(G, x) = \sum_{uv \in E(G)} x^{ec(u)ec(v)},$$

respectively.

Motivated by Kulli (2016) who introduced the multiplicative version of first atom bond connectivity index, we defined the fifth multiplicative atom bond connectivity index

$$ABC_5 \prod(G) = \prod_{uv \in E(G)} \sqrt{\frac{ec(u) + ec(v) - 2}{ec(u)ec(v)}}.$$

The traditional tree pattern of a tree which is a connected cyclic graph, is usually a binary tree where is composed with vertices, and there are a left reference, a right reference and a data element existing in it. We name the top most vertex root. There are three fields in the vertex of the binary tree. To illustrate, the data are denoted in one field, and the information of left and right sons of the vertex is located in the other two fields. The binary tree can
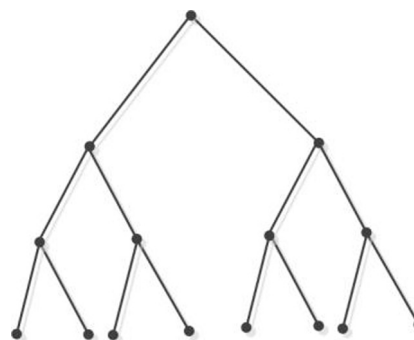

Fig. 1. Binary tree.

meet the requirements to be a complete binary tree (described in Fig. 1), if there are just two descendants in every internal vertex.

The fundamental structure of hypertree k-level is easy to be confirmed as a complete binary tree $HT(k)$. The root vertex of the tree is marked by label 1 and its root is at level 0. If 0 and 1 are added to the labels of the parent vertex separately, we can obtain the labels of left and right children. Then we can denote the children of the vertex x into $2x$ and $2x + 1$. Moreover, other links of hypertree are horizontal and when the label difference of the vertices is 2^{i-2}, the two vertices which share the same level of the tree are joined. As a result, if we add edges (sibling edges) between left and right children of the same parent vertex, we will get the 1-rooted sibling tree ST_k^1 (described in Fig. 2) from the 1-rooted tree T_k^1.

We also obtain the X-tree $XT(k)$ whose structure is described in Fig. 3 from complete binary tree on $2^{k+1} - 1$ vertices of height $2^i - 1$ and adding paths P_i left to right by all the vertices at level i with $i \in \{1, \ldots, k\}$.

3. Main results and proofs

This section mainly aims to calculate a closed result of eccentric related indices. Besides, the result of two kinds of eccentric version Zagreb polynomials for hyper binary trees, and for k-level networks is also aimed to be achieved. Meanwhile, with the purpose to explore the biological properties and activities, this section will borrow the networks to biological networks. There are numerous micro livings such as bacteria, viruses and others in our everyday life, which could be a problem of our life and also a good thing for our health. In terms of the structure of bacteria, they are single celled without true nucleus. The bacteria reproduce quickly and productively by conducting binary fission. During the process, a parental cell is splited into two daughter cells. Sometimes, bacteria and viruses could be dangerous in our life, for they may cause disease or make the disease more serious. With the help of replication, the action of copying, we start the reproduction procedure of viruses. Some diseases like Tuberculosis, cholera, typhoid, influenza, HIV (AIDS), chicken pox might be fatal if they are not considered seriously in time.

In the process, it's the infected person that serves as the media to spread diseases. For example, a person who gets cold and fever will spread the bacteria and viruses whenever they get touch with other people. They can spread by sneezing or by touching/shaking hands with others. Usually, the body fluids play a necessary role in transferring bacteria and viruses among people. We can regard the hypertree network as hypertree biological network. Hence, the parental vertex is considered as the ill person who carries bacteria and viruses, which is shown in Fig. 4. The bacteria reproduce quickly and productively by conducting binary fission, so we can suppose two persons that are infected by an ill person. Their interaction and communications will help to increase the disease level. Maybe one patient is weaker than another one for his immunological deficiency, which arouses the increase in this kind of disease. What we practice in reality to deal with the problem is to take anti-infection medicine. In the below, we obtain a closed result of eccentric related indices and polynomials for hypertree k level $HT(k)$ which is presented in Fig. 5.

Let $\delta = \min\{ec(v)|v \in V(G)\}$ and $\Delta = \max\{ec(v)|v \in V(G)\}$. The edge set $E(G)$ and vertex set $V(G)$ can be classed into the following subsets:

- for any i with $\delta \leq i \leq Delta$, let $D_i = \{u \in V(G)|ec(v) = i\}$ and $d_i = |D_i|$;
- for any i, j meet $\delta \leq i, j \leq \Delta$, let $E_{i,j} = \{e = uv \in E(G)|ec(u) = i$, and $ec(v) = j\}$ and $n_{ij} = |E_{i,j}|$.

Theorem 1. Let $HT(k)$ be the k-level hypertree network. We have

$$GA_4(HT(k)) = 3 + \sum_{i=1}^{k-1}\sum_{p=k}^{2k-2}2^i\left(\frac{4\sqrt{p(p+1)}}{2p+1} + 1\right),$$

$$Zg_4(HT(k)) = 6k + (9k-2)(2^k - 2)(k-1),$$

$$\Pi_4^*(HT(k)) = 6k\prod_{i=1}^{k-1}\prod_{p=k}^{2k-2}2^{2i}(4p+2)(2p+2),$$

$$Zg_6(HT(k)) = 3k^2 + \frac{7k(2^k - 2)(2k^2 - 3k + 1)}{2},$$

$$\Pi_6^*(HT(k)) = 3k^2\prod_{i=1}^{k-1}\prod_{p=k}^{2k-2}2^{2i+1}p(p+1)^3,$$

$$Zg_4(HT(k), x) = 3x^{2k} + \sum_{i=1}^{k-1}\sum_{p=k}^{2k-2}(2^{i+1}x^{2p+1} + 2^i x^{2p+2}),$$

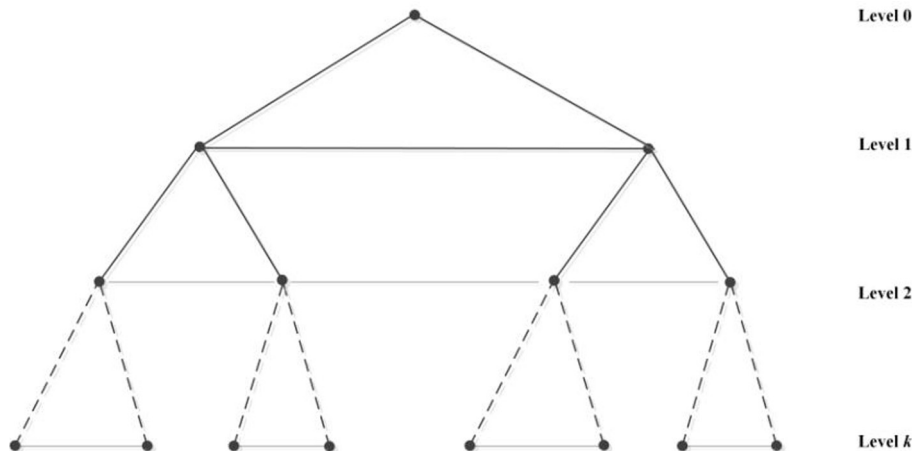$$Zg_6(HT(k), x) = 3x^{k^2} + \sum_{i=1}^{k-1}\sum_{p=k}^{2k-2}(2^{i+1}x^{p(p+1)} + 2^i x^{(p+1)^2}),$$

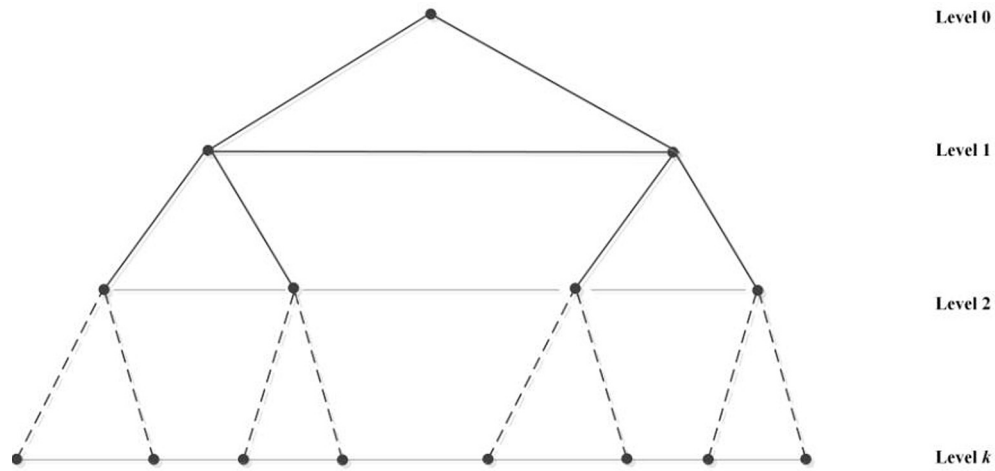**Fig. 2.** 1-rooted sibling tree network ST_k^1.

Fig. 3. k-level X-tree network $XT(k)$.

Fig. 4. Hypertree biological network.

$$ABC_5 \prod(HT(k)) = \frac{3\sqrt{2k-2}}{k} \prod_{i=1}^{k-1} \prod_{p=k}^{2k-2} \frac{2^{2i+1}}{p+1} \sqrt{\frac{4p-2}{p+1}}.$$

Proof of Theorem 1. We prove the result based on the structure analysis and edge dividing technology.

By analyzing the structure of $HT(k)$ and compute the distances from vertices, it's edge set $E(HT(k))$ can be divided into three partitions according to the eccentricities of associated vertices:

- $E_{p,p} = \{e = uv \in E(G) | ec(u) = ec(v) = p\}$ and $n_{pp} = 3 \cdot 2^i$, where $i = 0$ and $p = k$;
- $E_{p,p+1} = \{e = uv \in E(G) | ec(u) = p \quad$ and $\quad ec(v) = p+1\}$ and $n_{p(p+1)} = 2 \cdot 2^i$, where $i \in \{1, \dots, k-1\}$ and $p \in \{k, \dots, 2k-2\}$;
- $E_{p+1,p+1} = \{e = uv \in E(G) | ec(u) = ec(v) = p+1\}$ and $n_{(p+1)(p+1)} = 2^i$,

where $i \in \{1, \dots, k-1\}$ and $p \in \{k, \dots, 2k-2\}$.

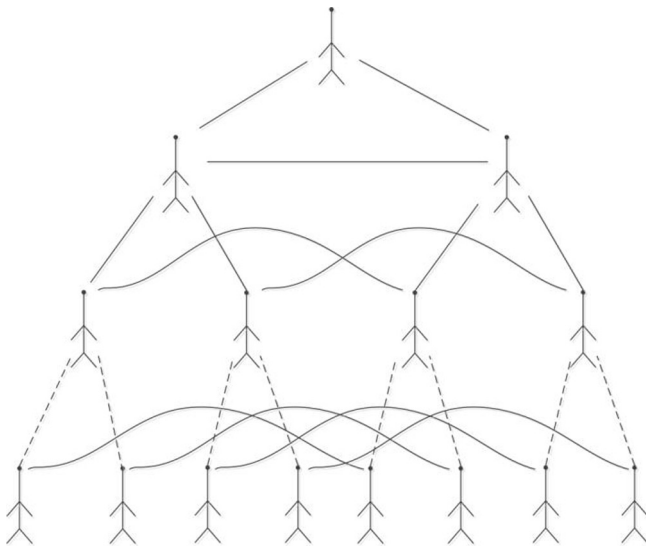Form the definitions of eccentric version indices, we get
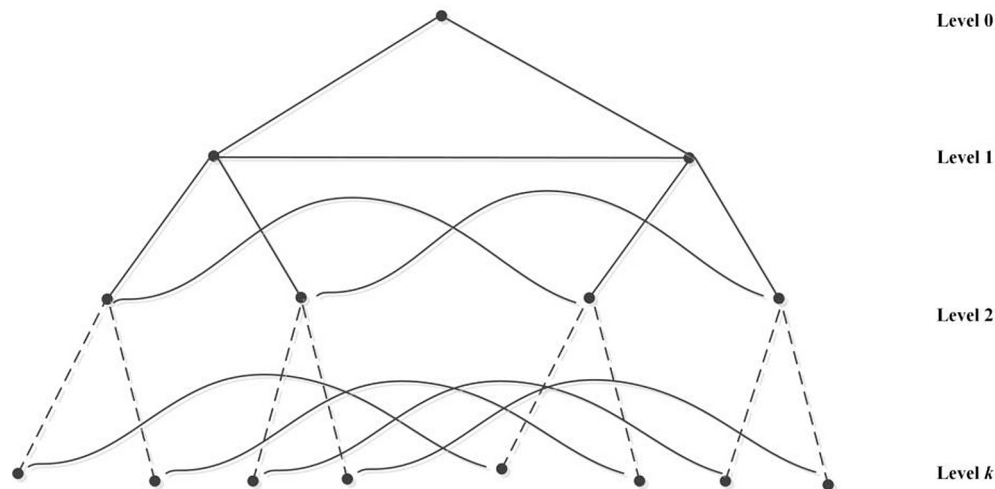

Fig. 5. k-level hypertree network $HT(k)$.

$$GA_4(HT(k)) = \sum_{i=0} 3 \cdot 2^i \sum_{p=k} \frac{2\sqrt{p^2}}{p+p} + \sum_{i=1}^{k-1} 2 \cdot 2^i \sum_{p=k}^{2k-2} \frac{2\sqrt{p(p+1)}}{p+p+1}$$

$$+ \sum_{i=1}^{k-1} 2^i \sum_{p=k}^{2k-2} \frac{2\sqrt{(p+1)(p+1)}}{p+1+p+1}$$

$$= 3 + \sum_{i=1}^{k-1} \sum_{p=k}^{2k-2} 2^i \left(\frac{4\sqrt{p(p+1)}}{2p+1} + 1 \right),$$

$$Zg_4(HT(k)) = \sum_{i=0} 3 \cdot 2^i \sum_{p=k} (p+p) + \sum_{i=1}^{k-1} 2 \cdot 2^i \sum_{p=k}^{2k-2} (p+p+1)$$

$$+ \sum_{i=1}^{k-1} 2^i \sum_{p=k}^{2k-2} (p+1+p+1)$$

$$= 6k + \sum_{i=1}^{k-1} \sum_{p=k}^{2k-2} 2^i (6p+4)$$

$$= 6k + (9k-2)(2^k - 2)(k-1),$$

$$\Pi_4^*(HT(k)) = \prod_{i=0} 3 \cdot 2^i \prod_{p=k} (p+p) \times \prod_{i=1}^{k-1} 2 \cdot 2^i \prod_{p=k}^{2k-2} (p+p+1)$$

$$\times \prod_{i=1}^{k-1} 2^i \prod_{p=k}^{2k-2} (p+1+p+1)$$

$$= 6k \prod_{i=1}^{k-1} \prod_{p=k}^{2k-2} 2^{2i} (4p+2)(2p+2),$$

$$Zg_6(HT(k)) = \sum_{i=0} 3 \cdot 2^i \sum_{p=k} (p \cdot p) + \sum_{i=1}^{k-1} 2 \cdot 2^i \sum_{p=k}^{2k-2} (p \cdot (p+1))$$

$$+ \sum_{i=1}^{k-1} 2^i \sum_{p=k}^{2k-2} (p+1)^2 = 3k^2 + \sum_{i=1}^{k-1} \sum_{p=k}^{2k-2} 2^i (3p^2 + 4p + 1)$$

$$= 3k^2 + \frac{7k(2^k-2)(2k^2 - 3k + 1)}{2},$$

$$\Pi_6^*(HT(k)) = \prod_{i=0} 3 \cdot 2^i \prod_{p=k} p^2 \times \prod_{i=1}^{k-1} 2 \cdot 2^i \prod_{p=k}^{2k-2} (p(p+1))$$

$$\times \prod_{i=1}^{k-1} 2^i \prod_{p=k}^{2k-2} (p+1)^2$$

$$= 3k^2 \prod_{i=1}^{k-1} \prod_{p=k}^{2k-2} 2^{2i+1} p(p+1)^3,$$

$$Zg_4(HT(k), x) = \sum_{i=0} 3 \cdot 2^i \sum_{p=k} x^{p+p} + \sum_{i=1}^{k-1} 2 \cdot 2^i \sum_{p=k}^{2k-2} x^{p+p+1}$$

$$+ \sum_{i=1}^{k-1} 2^i \sum_{p=k}^{2k-2} x^{p+1+p+1}$$

$$= 3x^{2k} + \sum_{i=1}^{k-1} \sum_{p=k}^{2k-2} (2^{i+1} x^{2p+1} + 2^i x^{2p+2}),$$

$$Zg_6(HT(k), x) = \sum_{i=0} 3 \cdot 2^i \sum_{p=k} x^{p^2} + \sum_{i=1}^{k-1} 2 \cdot 2^i \sum_{p=k}^{2k-2} x^{p \cdot (p+1)} + \sum_{i=1}^{k-1} 2^i \sum_{p=k}^{2k-2} x^{(p+1)^2}$$

$$= 3x^{k^2} + \sum_{i=1}^{k-1} \sum_{p=k}^{2k-2} (2^{i+1} x^{p(p+1)} + 2^i x^{(p+1)^2}),$$

$$ABC_5 \prod(HT(k)) = \prod_{i=0} 3 \cdot 2^i \prod_{p=k} \sqrt{\frac{p+p-2}{p \cdot p}} \times \prod_{i=1}^{k-1} 2$$

$$\cdot 2^i \prod_{p=k}^{2k-2} \sqrt{\frac{p+1+p-2}{p \cdot (p+1)}}$$

$$\times \prod_{i=1}^{k-1} 2^i \prod_{p=k}^{2k-2} \sqrt{\frac{p+1+p+1-2}{(p+1) \cdot (p+1)}}$$

$$= \frac{3\sqrt{2k-2}}{k} \prod_{i=1}^{k-1} \prod_{p=k}^{2k-2} \frac{2^{2i+1}}{p+1} \sqrt{\frac{4p-2}{p+1}}.$$

Fig. 6. X-tree biological network.

Thus, the desired results are obtained.

In one word, due to people's social activity, the infected diseases could spread among people directly or indirectly. As a biological network, the X-tree is another form of binary hypertree network. We can regard the parental vertex of X-tree network as the victim, and it will spread the bacteria and viruses among people through the contact with others in in hypertree biological network. We can see that there is a strong relationship among the whole infected persons at each level X-tree in biological network, which is represented in Fig. 6. In contrast, the relevance between two infected persons and the reasons of abundant infections are dealt. As a result, the infections will increase at much higher level than hypertree biological network.

Theorem 2. Let $XT(k)$ be the k-level X-tree with $k \geqslant 3$. We get

$$GA_4(XT(k)) = (3k-5)2^{k-1} - 3k^2 + 6k + 4 + \sum_{i=1}^{k-2}\sum_{p=k}^{2k-3}(6 \cdot 2^i + 12)$$
$$\times \frac{\sqrt{p(p+1)}}{2p+1} + 4(2^{k-2} + 1)\frac{\sqrt{(2k-2)(2k-1)}}{4k-3},$$

$$Zg_4(XT(k)) = 174k - 37k \cdot 2^{k-1} + 13 \cdot 2^{k-1} - 81k^2 + 9k^3 + 9k^2 2^k - 74,$$

$$\Pi_4^*(XT(k)) = 6k(4k-3)(2^{2k}-16)(2k-1) \times \prod_{i=1}^{k-2}\prod_{p=k+1}^{2k-2} 6p(2^i-1)$$
$$\times \prod_{i=1}^{k-2}\prod_{p=k}^{2k-3}(3 \cdot 2^i + 6)(2p+1),$$

$$Zg_6(XT(k)) = 51k2^{k-2} - 151k - 3 \cdot 2^k + \frac{367}{2}k^2 - \frac{141k^3}{2} + 7k^4$$
$$- 71k^2 2^{k-2} + 7k^3 2^k + 38,$$

$$\Pi_6^*(XT(k)) = 6k^2(2k-2)(2k-1)^3(2^{2k-2}-4)\prod_{i=1}^{k-2}\prod_{p=k+1}^{2k-2} 3p^2(2^i-1)$$
$$\times \prod_{i=1}^{k-2}\prod_{p=k}^{2k-3} 3p(p+1)(2^i+2),$$

$$Zg_4(XT(k),x) = 6x^{2k} + \sum_{i=1}^{k-2}\sum_{p=k+1}^{2k-2}(3 \cdot 2^i - 3)x^{2p} + \sum_{i=1}^{k-2}\sum_{p=k}^{2k-3}(3 \cdot 2^i$$
$$+ 6)x^{2p+1} + (2^{k-1}+2)x^{4k-3} + (2^{k-1}-2)x^{4k-2},$$

$$Zg_6(XT(k),x) = 6x^{k^2} + \sum_{i=1}^{k-2}\sum_{p=k+1}^{2k-2}(3 \cdot 2^i - 3)x^{p^2} + \sum_{i=1}^{k-2}\sum_{p=k}^{2k-3}(3 \cdot 2^i + 6)x^{p(p+1)}$$
$$+ (2^{k-1}+2)x^{(2k-2)(2k-1)} + (2^{k-1}-2)x^{(2k-1)^2},$$

$$ABC_5\prod(XT(k)) = \frac{3(2^{2k}-16)}{k(2k-1)}\sqrt{\frac{(4k-5)(k-1)}{2k-1}} \times \prod_{i=1}^{k-2}\prod_{p=k+1}^{2k-2}$$
$$\times \frac{(3 \cdot 2^i - 3)\sqrt{2p-2}}{p}$$
$$\times \prod_{i=1}^{k-2}\prod_{p=k}^{2k-3}(3 \cdot 2^i + 6)\sqrt{\frac{2p-1}{p(p+1)}}.$$

Proof of Theorem 2. Again, we prove the conclusions in light of the structure analysis and edge dividing technology.

By analyzing the structure of $XT(k)$ and compute the distances from vertices, it's edge set $E(XT(k))$ can be cut into four subsets by using the eccentricities of associated vertices:

- $E_{p,p}$ = {e = uv ∈ E(G)|ec(u) = ec(v) = p}: $n_{pp}^1 = 6 \cdot 2^i$, where $i = 0$ and $p = k$; $n_{pp}^2 = 3 \cdot 2^i - 3$, where $i \in \{1, \ldots, k-2\}$ and $p \in \{k+1, \ldots, 2k-2\}$. Thus, $n_{pp} = n_{pp}^1 + n_{pp}^2$.
- $E_{p,p+1}$ = {e = uv ∈ E(G)|ec(u) = p and ec(v) = p+1} and $n_{p(p+1)} = 3 \cdot 2^i + 6$, where $i \in \{1, \ldots, k-2\}$ and $p \in \{k, \ldots, 2k-3\}$;
- $E_{2p-2,2p-1}$ = {e = uv ∈ E(G)|ec(u) = 2p - 2 and ec(v) = 2p - 1} and $n_{(2p-2)(2p-1)} = 2 \cdot 2^i + 2$, where $i = k - 2$ and $p = k$;
- $E_{2p-1,2p-1}$ = {e = uv ∈ E(G)|ec(u) = ec(v) = 2p - 1} and $n_{(2p-1)(2p-1)} = 2 \cdot 2^i - 2$, where $i = k - 2$ and $p = k$.

Form the definitions of eccentric version indices, we get

$$GA_4(XT(k)) = \sum_{i=0}6 \cdot 2^i\sum_{p=k}\frac{2\sqrt{p^2}}{p+p} + \sum_{i=1}^{k-2}(3 \cdot 2^i - 3)\sum_{p=k+1}^{2k-2}\frac{2\sqrt{p^2}}{p+p}$$

$$+ \sum_{i=1}^{k-2}(3 \cdot 2^i + 6)\sum_{p=k}^{2k-3}\frac{2\sqrt{p(p+1)}}{p+p+1}$$

$$+ \sum_{i=k-2}(2 \cdot 2^i + 2)\sum_{p=k}\frac{2\sqrt{(2p-2)(2p-1)}}{2p-2+2p-1}$$

$$+ \sum_{i=k-2}(2 \cdot 2^i - 2)\sum_{p=k}\frac{2\sqrt{(2p-1)(2p-1)}}{2p-1+2p-1}$$

$$= 6 + (k-2)\sum_{i=1}^{k-2}(3 \cdot 2^i - 3) + \sum_{i=1}^{k-2}\sum_{p=k}^{2k-3}(6 \cdot 2^i + 12)\frac{\sqrt{p(p+1)}}{2p+1}$$

$$+ (2 \cdot 2^{k-2} + 2)\frac{2\sqrt{(2k-2)(2k-1)}}{4k-3} + (2 \cdot 2^{k-2} - 2)$$

$$= (3k-5)2^{k-1} - 3k^2 + 6k + 4 + \sum_{i=1}^{k-2}\sum_{p=k}^{2k-3}(6 \cdot 2^i + 12)\frac{\sqrt{p(p+1)}}{2p+1}$$

$$+ (2 \cdot 2^{k-2} + 2)\frac{2\sqrt{(2k-2)(2k-1)}}{4k-3},$$

$$Zg_4(XT(k)) = \sum_{i=0}6 \cdot 2^i\sum_{p=k}(p+p) + \sum_{i=1}^{k-2}(3 \cdot 2^i - 3)\sum_{p=k+1}^{2k-2}(p+p)$$

$$+ \sum_{i=1}^{k-2}(3 \cdot 2^i + 6)\sum_{p=k}^{2k-3}(p+p+1)$$

$$+ \sum_{i=k-2}(2 \cdot 2^i + 2)\sum_{p=k}(2p-2+2p-1)$$

$$+ \sum_{i=k-2}(2 \cdot 2^i - 2)\sum_{p=k}(2p-1+2p-1)$$

$$= 12k + \sum_{i=1}^{k-2}\sum_{p=k+1}^{2k-2} 6p(2^i-1) + \sum_{i=1}^{k-2}\sum_{p=k}^{2k-3}(3 \cdot 2^i + 6)(2p+1)$$

$$+ (4k-3)(2^{k-1}+2) + (2k-1)(2^k-4)$$

$$= 174k - 37k \cdot 2^{k-1} + 13 \cdot 2^{k-1} - 81k^2 + 9k^3$$

$$+ 9k^2 2^k - 74,$$

$$\Pi_4^*(XT(k)) = \prod_{i=0} 6 \cdot 2^i \prod_{p=k}(p+p) \times \prod_{i=1}^{k-2}(3 \cdot 2^i - 3)\prod_{p=k+1}^{2k-2}(p+p)$$

$$\times \prod_{i=1}^{k-2}(3 \cdot 2^i + 6)\prod_{p=k}^{2k-3}(p+p+1)$$

$$\times \prod_{i=k-2}(2 \cdot 2^i + 2)\prod_{p=k}(2p-2+2p-1)$$

$$\times \prod_{i=k-2}(2 \cdot 2^i - 2)\prod_{p=k}(2p-1+2p-1)$$

$$= 6k(2^k+4)(4k-3)(2^k-4)(2k-1) \times \prod_{i=1}^{k-2}\prod_{p=k+1}^{2k-2}6p(2^i-1)$$

$$\times \prod_{i=1}^{k-2}\prod_{p=k}^{2k-3}(3 \cdot 2^i + 6)(2p+1),$$

$$Zg_6(XT(k)) = \sum_{i=0} 6 \cdot 2^i \sum_{p=k}(p \cdot p) + \sum_{i=1}^{k-2}(3 \cdot 2^i - 3)\sum_{p=k+1}^{2k-2}(p \cdot p)$$

$$+ \sum_{i=1}^{k-2}(3 \cdot 2^i + 6)\sum_{p=k}^{2k-3}(p \cdot (p+1))$$

$$+ \sum_{i=k-2}(2 \cdot 2^i + 2)\sum_{p=k}(2p-2)\cdot(2p-1)$$

$$+ \sum_{i=k-2}(2 \cdot 2^i - 2)\sum_{p=k}(2p-1)\cdot(2p-1)$$

$$= 6k^2 + \sum_{i=1}^{k-2}\sum_{p=k+1}^{2k-2}3p^2(2^i-1) + \sum_{i=1}^{k-2}\sum_{p=k}^{2k-3}p(p+1)(3 \cdot 2^i + 6)$$

$$+ (2^{k-1}+2)(2k-2)(2k-1) + (2^{k-1}-2)(2k-1)^2$$

$$= 51k2^{k-2} - 151k - 3 \cdot 2^k + \frac{367}{2}k^2 - \frac{141k^3}{2}$$

$$+ 7k^4 - 71k^2 2^{k-2} + 7k^3 2^k + 38,$$

$$\Pi_6^*(XT(k)) = \prod_{i=0} 6 \cdot 2^i \prod_{p=k}(p \cdot p) \times \prod_{i=1}^{k-2}(3 \cdot 2^i - 3)\prod_{p=k+1}^{2k-2}(p \cdot p)$$

$$\times \prod_{i=1}^{k-2}(3 \cdot 2^i + 6)\prod_{p=k}^{2k-3}(p \cdot (p+1))$$

$$\times \prod_{i=k-2}(2 \cdot 2^i + 2)\prod_{p=k}((2p-2)\cdot(2p-1))$$

$$\times \prod_{i=k-2}(2 \cdot 2^i - 2)\prod_{p=k}((2p-1)\cdot(2p-1))$$

$$= 6k^2(2k-2)(2k-1)^3(2^{2k-2}-4)\prod_{i=1}^{k-2}\prod_{p=k+1}^{2k-2}3p^2(2^i-1)$$

$$\times \prod_{i=1}^{k-2}\prod_{p=k}^{2k-3}3p(p+1)(2^i+2),$$

$$Zg_4(XT(k),x) = \sum_{i=0} 6 \cdot 2^i \sum_{p=k}x^{p+p} + \sum_{i=1}^{k-2}(3 \cdot 2^i - 3)\sum_{p=k+1}^{2k-2}x^{p+p}$$

$$+ \sum_{i=1}^{k-2}(3 \cdot 2^i + 6)\sum_{p=k}^{2k-3}x^{p+p+1} + \sum_{i=k-2}(2 \cdot 2^i + 2)\sum_{p=k}x^{2p-2+2p-1}$$

$$+ \sum_{i=k-2}(2 \cdot 2^i - 2)\sum_{p=k}x^{2p-1+2p-1}$$

$$= 6x^{2k} + \sum_{i=1}^{k-2}\sum_{p=k+1}^{2k-2}(3 \cdot 2^i - 3)x^{2p} + \sum_{i=1}^{k-2}\sum_{p=k}^{2k-3}(3 \cdot 2^i + 6)x^{2p+1}$$

$$+ (2 \cdot 2^{k-2}+2)x^{4k-3} + (2 \cdot 2^{k-2}-2)x^{4k-2},$$

$$Zg_6(XT(k),x) = \sum_{i=0} 6 \cdot 2^i \sum_{p=k}x^{p \cdot p} + \sum_{i=1}^{k-2}(3 \cdot 2^i - 3)\sum_{p=k+1}^{2k-2}x^{p \cdot p}$$

$$+ \sum_{i=1}^{k-2}(3 \cdot 2^i + 6)\sum_{p=k}x^{p \cdot (p+1)} + \sum_{i=k-2}(2 \cdot 2^i + 2)\sum_{p=k}x^{(2p-2)\cdot(2p-1)}$$

$$+ \sum_{i=k-2}(2 \cdot 2^i - 2)\sum_{p=k}x^{(2p-1)\cdot(2p-1)}$$

$$= 6x^{k^2} + \sum_{i=1}^{k-2}\sum_{p=k+1}^{2k-2}(3 \cdot 2^i - 3)x^{p^2} + \sum_{i=1}^{k-2}\sum_{p=k}^{2k-3}(3 \cdot 2^i + 6)x^{p(p+1)}$$

$$+ (2 \cdot 2^{k-2}+2)x^{(2k-2)(2k-1)} + (2 \cdot 2^{k-2}-2)x^{(2k-1)^2},$$

$$ABC_5\prod(XT(k)) = \prod_{i=0} 6 \cdot 2^i \prod_{p=k}\sqrt{\frac{p+p-2}{p \cdot p}}$$

$$\times \prod_{i=1}^{k-2}(3 \cdot 2^i - 3)\prod_{p=k+1}^{2k-2}\sqrt{\frac{p+p-2}{p \cdot p}}$$

$$\times \prod_{i=1}^{k-2}(3 \cdot 2^i + 6)\prod_{p=k}^{2k-3}\sqrt{\frac{p+p+1-2}{p(p+1)}}$$

$$\times \prod_{i=k-2}(2 \cdot 2^i + 2)\prod_{p=k}\sqrt{\frac{2p-2+2p-1-2}{(2p-2)(2p-1)}}$$

$$\times \prod_{i=k-2}(2 \cdot 2^i - 2)\prod_{p=k}\sqrt{\frac{2p-1+2p-1-2}{(2p-1)(2p-1)}}$$

$$= \frac{6\sqrt{2k-2}}{k}\frac{2^{2k-2}-4}{2k-1}\sqrt{\frac{8k-10}{2k-1}}$$

$$\times \prod_{i=1}^{k-2}\prod_{p=k+1}^{2k-2}\frac{(3 \cdot 2^i - 3)\sqrt{2p-2}}{p}$$

$$\times \prod_{i=1}^{k-2}\prod_{p=k}^{2k-3}(3 \cdot 2^i + 6)\sqrt{\frac{2p-1}{p(p+1)}}.$$

Hence, we yield the expected conclusions.

4. Conclusions and future problems

In this paper, we discuss the theoretical topics in some biology problems, and finally determine the eccentric topological indices of biological networks in view of structure analysis, distance calculating and mathematical derivation. Since the biological networks can help the researchers better understand how infectious diseases fast increase through infected persons, the theoretical results have the wide and promising application prospects in biological, medical and pharmacy engineering.

The following topics can be considered as the future work:

- How to structure a kind of graph to express the microstructure under more biology problems.
- Design an algorithm with low computational complexity to express the complex DNA gene structure and apply it in gene mutation.
- How to put the known theoretical results into biological reverse engineering.

Conflict of interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgements

References

Aittokallio, T., Schwikowski, B., 2006. Graph-based methods for analysing networks in cell biology. Brief. Bioinform. 7 (3), 243.

Alexander, S.A., 2013. Infinite graphs in systematic biology, with an application to the species problem. Acta. Biotheor. 61 (2), 181–201.

Alqallaf, A.K., Cherri, A.K., 2016. Dna sequencing using optical joint fourier transform. Optik – Int. J. Light Electron Opt. 127 (4), 1929–1936.

Aram, V., Iranmanesh, A., 2012. 3D-dynamic representation of DNA sequences. MATCH-Commun. Math. Comput. Chemist. 67, 809–816.

Banerjee, A., Jost, J., 2007. Graph spectra as a systematic tool in computational biology. Discrete Appl. Math. 157 (10), 2425–2431.

Basavanagoud, B., Desai, V.R., Patil, S., 2017. (β, α) -Connectivity index of graphs. Appl. Math. Nonlinear Sci. 2 (1), 369–374.

Blazewicz, J., Bryja, M., Figlerowicz, M., Gawron, P., Kasprzak, M., Kirton, E., et al., 2009. Brief communication: whole genome assembly from 454 sequencing output via modified dna graph concept. Comput. Biol. Chem. 33 (3), 224–230.

Blazewicz, J., Hertz, A., Kobler, D., De Werra, D., 1999. On some properties of dna graphs. Discrete Appl. Math. 98 (1–2), 1–19.

Bokhari, S.H., Sauer, J.R., 2005. A parallel graph decomposition algorithm for dna sequencing with nanopores. Bioinformatics 21 (7), 889–896.

Chakraborty, R., Ghosh, D., 2016. The effect of sequence on the ionization of guanine in dna. Phys. Chem. Chem. Phys. Pccp 18 (9), 6526–6533.

Chlopkowski, M., Antczak, M., Slusarczyk, M., Wdowinski, A., Zajaczkowski, M., Kasprzak, M., 2016. High-order statistical compressor for long-term storage of dna sequencing data. RAIRO - Operations Research.

Da Silva, E., Pedrini, H., 2016. Inferring patterns in mitochondrial dna sequences through hypercube independent spanning trees. Comput. Biol. Med. 70, 51–57.

Eckman, B.A., Brown, P.G., 2006. Graph data management for molecular and cell biology. Ibm J. Res. Dev. 50 (6), 545–560.

El-Ghoul, M., El-Ahmady, A.E., Homoda, T., 2006. On chaotic graphs and applications in physics and biology. Chaos, Solitons Fractals 27 (1), 159–173.

Farahani, M.R., Kanna, R.M.R., 2015. Fourth Zagreb index of circumcoronene series of benzenoid. Leonardo Electron. J. Pract. Technol. 27, 155–161.

Gao, W., Siddiqui, M.K., 2017. Molecular descriptors of nanotube, oxide, silicate, and triangulene networks. J. Chem., 2017, Article ID 6540754 (10p) doi:10.1155/2017/6540754.

Gao, W., Siddiqui, M.K., Imran, M., Jamil, M.K., Farahani, M.R., 2016a. Forgotten topological index of chemical structure in drugs. Saudi Pharm. J. 24 (3), 258–264.

Gao, W., Farahani, M.R., Shi, L., 2016b. The forgotten topological index of some drug structures. Acta Med. Med. 32, 579–585.

Gao, W., Wang, W.F., Farahani, M.R., 2016c. Topological indices study of molecular structure in anticancer drugs. J. Chem., 2016, Article ID 3216327 (8p). doi:10.1155/2016/3216327.

Gao, W., Wang, W.F., Jamil, M.K., Farahani, M.R., 2016d. Electron energy studying of molecular structures via forgotten topological index computation. J. Chem., 2016, Article ID 1053183 (7p). doi:10.1155/2016/1053183.

Gao, W., Baig, A.Q., Ali, H., Sajjad, W., Farahani, M.R., 2017a. Margin based ontology sparse vector learning algorithm and applied in biology science. Saudi J. Biol. Sci. 24 (1), 132–138.

Gao, W., Yan, L., Shi, L., 2017b. Generalized Zagreb index of polyomino chains and nanotubes. Optoelectron. Adv. Mater. - Rapid Commun. 11 (1–2), 119–124.

Gao, W., Wang, W.F., 2015. The vertex version of weighted wiener number for bicyclic molecular structures. Comput. Math. Methods Med., 2015, Article ID 418106 (10p) doi:10.1155/2015/418106.

Gao, W., Wang, W.F., 2016. The eccentric connectivity polynomial of two classes of nanotubes. Chaos, Solitons Fractals 89, 290–294.

Ghorbani, M., Khaki, A., 2010. A note on the fourth version of geometric-arithmetic index. Optoelectron. Adv. Mater. - Rapid Commun. 4 (12), 2212–2215.

Jafarzadeh, N., Iranmanesh, A., 2013. C-curve: a novel 3d graphical representation of dna sequence based on codons. Math. Biosci. 241 (2), 217.

Jafarzadeh, N., Iranmanesh, A., 2012. A novel graphical and numerical representation for analyzing DNA sequences based on codons. MATCH-Commun. Math. Comput. Chem. 68, 611–620.

Jafarzadeh, N., Iranmanesh, A., 2016a. A new graph theoretical method for analyzing DNA sequences based on genetic codes. MATCH-Commun. Math. Comput. Chem. 75 (3), 731–742.

Jafarzadeh, N., Iranmanesh, A., 2016. Application of graph theory to biological problems, Studia Ubb Chemia, LXI, 9-16.

Khan, F.J., Sarmin, N.H., Khan, A., Khan, H.U., 2017. New types of fuzzy interior ideals of ordered semigroups based on fuzzy points. Matriks Sains Matematik 1 (1), 25–33.

Khan, H.U., Khan, A., Khan, F.M., Li, Y., 2017. Generalized Bi-ideal of Ordered Semigroup Related to Intuitionistic Fuzzy Point. Matriks Sains Matematik 1 (1), 09–15.

Kletsov, A.A., Glukhovskoy, E.G., Chumakov, A.S., Ortiz, J.V., 2015. Ab initio electron propagator calculations of transverse conduction through dna nucleotide bases in 1-nm nanopore corroborate third generation sequencing. Biochim. Biophys. Acta (BBA) – General Subjects, 1860(1), 140–145.

Kulli, V.R., 2016. Multiplicative connectivity indices of certain nanotubes. Ann. Pure Appl. Math. 12 (2), 169–176.

Lesne, A., 2006. Complex networks: from graph theory to biology. Lett. Math. Phys. 78 (3), 235–262.

Ma'ayan, A., 2008. Network integration and graph analysis in mammalian molecular systems biology. Iet Syst. Biol. 2 (5), 206–221.

Mason, O., Verwoerd, M., 2006. Graph theory and networks in biology. Iet Syst. Biol. 1 (2), 89–119.

Nandagopal, N., Elowitz, M.B., 2011. Synthetic biology: integrated gene circuits. Science 333 (6047), 1244–1248.

Niwa, H.S., Nashida, K., Yanagimoto, T., 2016. Reproductive skew in Japanese sardine inferred from DNA sequences. ICES J. Mar. Sci. 73 (9), 2181–2189.

O'Reilly, E., Baccelli, F., De Veciana, G., Vikalo, H., 2016. End-to-end optimization of high-throughput DNA sequencing. J. Comput. Biol. 23 (10), 789–800.

Pendavingh, R., Schuurman, P., Woeginger, G.J., 2003. Recognizing dna graphs is difficult. Discrete Appl. Math. 127 (1), 85–94.

Pesek, J., Zerovnik, A., 2008. Numerical characterization of modified Hamori curve representation of DNA sequences. MATCH-Commun. Math. Comput. Chem. 60, 301–312.

Pevzner, P.A., Tang, H., Waterman, M.S., 2001. An Eulerian path approach to dna fragment assembly. Proc. Nat. Acad. Sci. USA 98 (17), 9748.

Rajan, R.S., Anitha, J., Rajasingh, I., 2015. 2-power domination in certain interconnection networks. Proc. Comput. Sci. 57, 738–744.

Rastgou, A., Soleymanabadi, H., Bodaghi, A., 2017. DNA sequencing by borophene nanosheet via an electronic response: a theoretical study. Microelectron. Eng. 169, 9–15.

Sa-Ardyen, P., Jonoska, N., Seeman, N.C., 2003. Self-assembling dna graphs. Nat. Comput. 2 (4), 427–438.

Samina, Shah K., Khan, R.A., 2017. Study of nonlocal boundary value problems of non-integer order hybrid differential equations. Matriks Sains Matematik 1 (1), 21–24.

Shah, K., Bushnaq, S., 2017. Investigating at least one solution to a systems of multi-points boundary value problems. Matriks Sains Matematik 1 (1), 16–20.

Shoaib, M., Sarwar, M., Hussain, M., Ali, G., 2017. Existence and uniqueness of common fixed point for mappings satisfying integral type contractive conditions in G-metric spaces. Matriks Sains Matematik 1 (1), 01–08.

Wang, S.Y., Yuan, J., Lin, S.W., 2008. Dna labelled graphs with dna computing. Sci. China Math. 51 (3), 437–452.