"Science Stays True Here"
Journal of Mathematics and Statistical Science (ISSN 2411-2518, USA), Vol.9 (2023), Issue 1, 1-9 | Science Signpost Publishing

Lower Bounds for Zagreb Indices of RNA Graphs Using Graph Algorithms

Elif ERYAŞAR[1], Şerife BÜYÜKKÖSE[2]

eeryasar@sinop.edu.tr , sbuyukkose@gazi.edu.tr

1. Department of Mathematics, Sinop University, 57000, Sinop, Turkey.

2. Department of Mathematics, Gazi University, 06500, Ankara, Turkey (corresponding author).

Abstract

The structure of RNA has been a natural subject for mathematical modeling, including many innovative computational models. The representation of RNA secondary structures as mathematical graphs has been illuminating for RNA structure analysis. In this work, we introduce the first and second Zagreb indices. Then, we obtain tree structures through BFS and DFS algorithms from graph searching algorithms. With the help of these trees, we compare these lower bound values by giving lower bounds for the first and second Zagreb indices of RNA secondary structure graphs.

Keywords: Zagreb indices, Graph search algorithms, RNA graph representation, DFS and BFS trees of RNA

Introduction

In mathematical biology, mathematical methods are applied to biology to solve various modeling and computational problems. An RNA secondary structure can be represented as a graph, and graph theory is used for the analysis and calculation of molecular topology [9]. A graph is shown by a collection of connecting points and lines, who can be separately called vertices and edges [8]. Let be G a simple graph with , δ minimum and Δ maximum degrees, V vertices and E edges. A graph G is represented as $G = (V, E)$. For any $\in V$,d_u denotes the degree of this point [2]. Topological indices are numerical representations of a molecule [18, 20]. They are calculated from the heavy atom graphical characterization of the molecule [14]. One of the first topological indexes was the Wiener index, presented

by the mathematical chemist Wiener in 1947 [17]. It was shown that the Wiener index correlated well with the boiling points of alkanes [18, 25]. The degree-based topological indices are plays an vital role in chemical graph theory. One of the oldest degree-based topological indices is the Zagreb index, which was introduced by Gutman and Trinajstić in 1972 during the analysis of the structure-dependence of the total electron energy [10, 18]. The first Zagreb index $M_1(G)$ and the second Zagreb index $M_2(G)$ are defined as follows [2]:

$$M_1(G) = \sum_{i \in V(G)} {d_i}^2$$

and

$$M_2(G) = \sum_{ij \in E(G)} d_i d_j .$$

Graph models are now used not only to study RNA structure but also to identify modular repetitive RNA units, to sample the conformational space accessible for RNAs, and to predict three-dimensional folds of RNA [15, 20]. It also offers many advantages in new RNA designs. There are many graph representations of RNA secondary structures, including structure-induced representation, tree representation, and binary graph representation [7, 20]. Interest in developing new topological indices based on RNA graphs has been reanimated in recent years because such indices can be used to compare, identify and classify RNAs. The use of topological indices of RNA secondary structure to find quantitative structure-activity relationships (QSAR) can be an alternative to integrate secondary structure or sequence-to sequence similarity [1, 21, 24].

Model and Analysis

Model

Graph Research Algorithms

Many problems in artificial intelligence can be regarded as search problems, including planning, learning, problem solving, and (logical) reasoning [6]. Search problems may frequently be formulated as graph search problems, and can be solved by exploring a space of possible solutions in a more or less systematic order [3, 6]. The commonly used search methods are BFS (Breadth First Search) and DFS

(Deep First Search) algorithms. When applying the BFS algorithm, a starting point is first selected. Then, a tree graph is obtained by going to each neighboring point of this starting point and all the neighbors of the point located there. In the DFS algorithm, on the other hand, after the starting point is selected, it is passed to the point adjacent to this point, and then to a neighboring point of the passed point, and continues until all points of the graph are visited. After this process, a tree graph is obtained again [3, 5]. An important structure closely connected to a graph search is the corresponding search tree. Such a tree contains all the vertices of the graph and for every vertex different from the start vertex correctly one edge to a vertex preceding it in the search ordering. Such trees may be of particular interest as for example the tree obtained by a BFS contains the shortest paths from the root r to all other vertices and the trees generated by DFS can be used for fast planarity testing [11, 12, 19, 22, 23, 26]. In this study, we obtain DFS and BFS trees of RNA graph by applying BFS and DFS algorithms to RNA secondary structure graphs and we plotted RNA graph representations via [27].

RNA Graph Representation

Ribonucleic Acids (RNAs) are among the most important molecules in life [13, 15]. Graph theory has been very useful in modeling and comparing RNA structures and generating new RNA structures [13, 15, 16, 20]. There are different representations of RNA graphs, from a very simplified representation where the number of vertices is minimal, where each nucleotide is a vertex, depending on the type of vertices and edges assigned [7]. In this study, we will use the graph representation resulting from the structure. Here, each nucleotide corresponds to a corner point and each hydrogen bond or phosphodiester bond to an edge. (Figure 1)

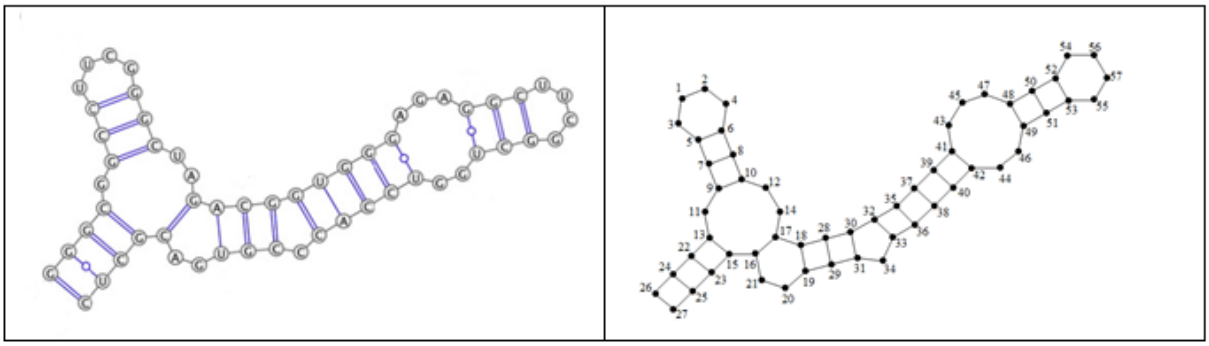

Figure 1: Ribosomal RNA and graph represented

First and Second Zagreb İndices and BFS-DFS trees of RNA Graphs

We show graph representations of RNA secondary structures as G_{RNA}. First and second Zagreb indices of RNA structure grapH G_{RNA} given in Figure 1:

$$M_1(G_{RNA}) = 21.2^2 + 36.3^2$$
$$= 84 + 324$$
$$= 408$$

and

$$M_2(G_{RNA}) = 12(2.2) + 18(2.3) + 45(3.3)$$
$$= 48 + 108 + 405$$
$$= 561.$$

When applying the algorithm to the RNA graph, we specifically select one of the vertices with the highest degree. Here, point 10 is chosen as the starting point. In addition, the BFS and DFS trees of the RNA Graph given in figure 1 are as given in figure 2. Here we denote the BFS tree as G_{T-BFS} and the DFS tree as G_{T-DFS}.

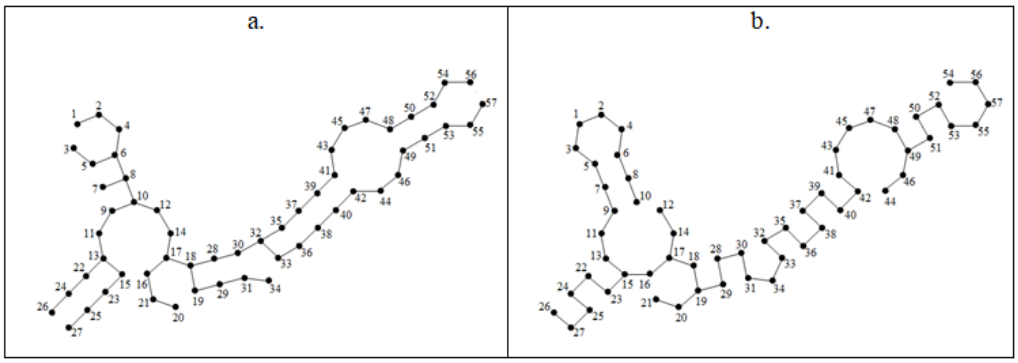

Figure 2: $\boldsymbol{G_{T-BFS}}$ tree, b. $\boldsymbol{G_{T-DFS}}$ tree

Analysis and Discussion

Lower bounds of Zagreb indices of RNA graphs with the help of BFS tree

Theorem 1. For an $n-$point G_{RNA} secondary structure graph

$$4n - r(\Delta^2 - 4) - 6k \leq M_1(G_{RNA})$$

there is inequality. Here k is the sum of the number of loops and ends in the RNA structure.

Proof: Let the graph obtained as a result of applying the BFS algorithm to a G_{RNA} structure be represented by G_{T-BFS}. In this case, if there are k loops and ends in the G_{RNA} tree structure, there are at most $2k$ 1 $-$order points in the G_{T-BFS} tree. Also, if we assume that there are at most r points of

maximum degree in the G_{T-BFS} tree, the number of points with a degree of 2 is at most $n-2k-r$. In that case

$$2k.1^2 + r.\Delta^2 + (n-2k-r).2^2 \leq M_1(G_{RNA})$$
$$2k + r\Delta^2 + 4n - 8k - 4r \leq M_1(G_{RNA})$$
$$4n - r(\Delta^2 - 4) - 6k \leq M_1(G_{RNA}).$$

Where, A being $4n - r(\Delta^2 - 4) - 6k = A$; The first Zagreb index of the G_{RNA} construct is a lower bound for M_1.

Theorem 2. For an $n-$point G_{RNA} secondary structure graph

$$2(\Delta r + 2n - 2k - 2r) \leq M_2(G_{RNA})$$

there is inequality. Here k is the sum of the number of loops and ends in the G_{RNA} structure.

Proof: If an $n-$point G_{RNA} structure has k loops and endpoints, then there are $2k$ points with a degree of 1 in the G_{T-BFS} tree obtained by applying the BFS algorithm. Let's assume that the degrees of the vertices connected to these points with a degree of 1 are maximum 2. On the other hand, if we assume that there are r points with Δ maximum degree in the G_{T-BFS} tree, let the other endpoints of the edges connected to this Δ maximum degree r point have a maximum degree of 2. Thus, $n-2k-r$ points are left in the tree graph with an end (or related) and Δ maximum degree and not connected to these points. Assuming that every vertex is of 2 degrees, except for all points of Δ degree and 1 degree

$$(1.2)2k + (\Delta.2)r + (2.2)(n-2k-z) \leq M_2(G)$$
$$= 4k + 2\Delta r + 4n - 8k - 4z \leq M_2(G)$$
$$= 2(\Delta r + 2n - 2k - 2r) \leq M_2(G).$$

Where $2(\Delta r + 2n - 2k - 2r) = C$, where C; The second Zagreb index of the G_{RNA} construct is a lower bound for M_2.

Lower bounds of Zagreb indices of RNA graphs with the help of DFS tree

Theorem 3. For an $n-$point G_{RNA} secondary structure graph

$$\Delta^2 - 6k + 4(n-1) \leq M_1(G_{RNA})$$

there is inequality. Here k is the sum of the number of loops and ends in the RNA structure.

Proof: Let the graph obtained as a result of applying the DFS algorithm to a G_{RNA} structure be represented by G_{T-DFS}. If there are k loops and ends in the n-point G_{RNA} tree structure, there are at most $2k$ $1-$order points. In addition, there are $n-2k-1$ 2 graded points, 1 of which is Δ maximum graded point. In that case

$$2k.1 + 1.\Delta^2 + 2^2(n - 2k - 1) \leq M_1(G_{RNA})$$
$$\Delta^2 + 2k + 4(n - 2k - 1) \leq M_1(G_{RNA})$$
$$\Delta^2 + 2k + 4n - 8k - 4 \leq M_1(G_{RNA})$$
$$\Delta^2 - 6k + 4(n - 1) \leq M_1(G_{RNA})$$

Where $\Delta^2 - 6k + 4(n-1) = B$ being B; The first Zagreb index of the G_{RNA} construct is a lower bound for M_1.

Theorem 4. For an n −point G_{RNA} secondary structure graph

$$2(\Delta t + 2n - 2k - 2t - 2) \leq M_2(G_{RNA})$$

Here is inequality. Here k is the sum of the number of loops and ends in the G_{RNA} structure.

Proof: If an n −point G_{RNA} structure has k loops and endpoints, then there are $2k$ points with a degree of 1 in the G_{T-DFS} tree obtained by applying the DFS algorithm. Let's assume that the degrees of the vertices connected to these points with a degree of 1 are maximum 2. On the other hand, if we assume that there is 1 point with Δ maximum degree in the G_{T-DFS} tree, let the vertex t and other endpoints connected to this Δ maximum degree point have a maximum degree of 2. Thus, there are $n - 2k - t - 1$ points in the tree graph that are end (or related) and Δ maximum degree and not connected to these points. Assuming that every vertex is of 2 degrees, except for all points of Δ degree and 1 degree

$$(1.2)2k + (\Delta.2)t + (2.2)(n - 2k - t - 1) \leq M_2(G)$$
$$= 4k + 2\Delta t + 4n - 8k - 4t - 4 \leq M_2(G)$$
$$= 2(\Delta t + 2n - 2k - 2t - 2) \leq M_2(G).$$

Where $2(\Delta t + 2n - 2k - 2t - 2) = D,\ D$; The second Zagreb index of the G_{RNA} construct is a lower bound for M_2.

Comparisons

Theorem 5. For an n −point G_{RNA} secondary structure graph

$$A = 4n - r(\Delta^2 - 4) - 6k \leq M_1(G_{RNA})$$

and

$$B = \Delta^2 - 6k + 4(n - 1) \leq M_1(G_{RNA})$$

and $B \geq A$. Therefore, the lower bound for the tree structure obtained as a result of the BFS algorithm for $M_1(G_{RNA})$ gave better results compared to the DFS algorithm.

Proof:

$$A = 4n - r(\Delta^2 - 4) - 6k$$

and

$$B = \Delta^2 - 6k + 4(n-1)$$

and

$$B - A = \Delta^2 - 4 + r(\Delta^2 - 4)$$
$$= (\Delta^2 - 4)(1 + r)$$

$B - A \geq 0$ is obtained with $\Delta \geq 2$ and $r > 1$. Therefore, $B \geq A$ is

$$A \leq B \leq M_1(G_{RNA})$$

inequality is achieved.

Theorem 6. For an n −point G_{RNA} secondary structure graph

$$C = 2(\Delta r + 2n - 2k - 2r) \leq M_2(G_{RNA})$$

and

$$D = 2(\Delta t + 2n - 2k - 2t - 2) \leq M_2(G_{RNA})$$

including $C \geq D$. Therefore, the lower bound for the tree structure obtained as a result of the DFS algorithm for $M_2(G_{RNA})$ gave better results compared to the BFS algorithm.

Proof:

$$C = 2(\Delta r + 2n - 2k - 2r)$$

and

$$D = 2(\Delta t + 2n - 2k - 2t - 2)$$

and

$$C - D = 2(\Delta r + 2n - 2k - 2r) - 2(\Delta t + 2n - 2k - 2t - 2)$$
$$= 2\Delta(r + t) - 4(r + t) + 4$$
$$= (2\Delta - 4)(r + t) + 4$$

$C - D \geq 0$ is obtained with $\Delta \geq 2$ and $r > 1, t \geq 1$. Therefore, $C \geq D$ is

$$D \leq C \leq M_2(G_{RNA})$$

inequality is achieved.

Conclusions

Many studies have been conducted on the Zagreb indices and many results have been obtained on this subject. However, there was no lower bound finding study of Zagreb indices of any RNA graph with the help of BFS-DFS algorithms. Here, we gave theorems for the lower bounds of Zagreb indices of RNA graphs with the help of BFS and DFS trees and completed our proofs by making various comparisons.

Acknowledgment

The authors thank the Scientific and Technological Research Council of Turkey (TÜBİTAK) for their support within the scope of the project numbered 2021/1.2210-A.

References

[1]. Aguero-Chapin, G., Antunes, A., Ubeira, F. M., Chou, K. C., and Gonzalez-Diaz, H. (2008). Comparative study of topological indices of macro/supramolecular RNA complex networks. *Journal of Chemical Information and Modeling*, *48*(11), 2265-2277.

[2]. Büyükköse, Ş. (2021). Graf Teori (Third Edition). Ankara: Nobel Academic Publishing, 1-170.

[3]. Çölkesen, R. (2015). Bilişim Matematiği (First Edition). İstanbul: Papatya Publishing, 349-382.

[4]. Eryaşar, E. (2022). Graf Algoritmaları ile RNA Graflarının Zagreb İndeksi için Alt Sınırlar. Master Thesis. Graduate School of Natural and Applied Sciences, Ankara. 44.

[5]. Eryaşar, E., Büyükköse, Ş. (2022). Graf Algoritmaları ve Yapay Zekâya Katkıları. Yapay Zekâ ve Büyük Veri Kitap Serisi 4. (First Edition). Ankara: Nobel Publishing, 81-106.

[6]. Everitt, T., and Hutter, M. (2015). A topological approach to meta-heuristics: analytical results on the BFS vs. DFS algorithm selection problem. *arXiv:1509.02709*.

[7]. Galindo, J. F., Rubiano, G., and Daza, E. E. (2014). Pretopological spaces as a classification tool for RNAs Represented as a succession. *MATCH Commun. Math. Comput. Chem*, *72*, 453-474.

[8]. Gao, W., Farahani, M., Husin, M., and Wang, S. (2017). On the edge-version atom-bond connectivity and geometric arithmetic indices of certain graph operations, *Appl. Math. Comput*, 308 , 11-17.

[9]. Gao, W., Wu, H., Siddiqui, M. K., and Baig, A. Q. (2018). Study of biological networks using graph theory. *Saudi Journal of Biological Sciences*, *25*(6), 1212-1219.

[10]. Gutman, I. (2013). Degree-based topological indices. *Croatica Chemica Acta*, 86(4), 351-361.

[11]. Hopcroft, J., and Tarjan, R. (1973). Algorithm 447: Efficient algorithms for graph manipulation. *Communications of the ACM*, 16(6), 372-378.

[12]. Hopcroft, J., and Tarjan, R. (1974). Efficient planarity testing. *Journal of the ACM (JACM)*, 21(4), 549-568.

[13]. Jain, S., & Schlick, T. (2017). F-RAG: generating atomic coordinates from RNA graphs by fragment assembly. *Journal of Molecular Biology*, 429(23), 3587-3605.

[14]. Jude, T. P., Panchadcharam, E., & Masilamani, K. (2019). Computation of Zagreb and Randić indices of two biodegradable dendrimers used in cancer therapy. *Ceylon Journal of Science*, *48*(4), 359-366.

[15]. Kim, N., Fuhr, K. N., and Schlick, T. (2013). Graph applications to RNA structure and function. *Biophysics of RNA Folding*, 23-51, Springer, New York, NY.

[16]. Koessler, D. R., Knisley, D. J., Knisley, J., AND Haynes, T. (2010, October). A predictive model for secondary RNA structure using graph theory and a neural network. *BMC Bioinformatics.* 11(6), 1-10.

[17]. Klavˇzar, S., Gutman, I. (1996). The Szeged and the Wiener index of graphs, *Appl. Math. Lett*. 9, 45-49.

[18]. Liu, J. B., Javed, S., Javaid, M., and Shabbir, K. (2019). Computing first general Zagreb index of operations on graphs. *IEEE Access*, 7, 47494-47502.

[19]. Scheffler, R. (2022). On the recognition of search trees generated by BFS and DFS. *Theoretical Computer Science*.

[20]. Schlick, T. (2018). Adventures with RNA graphs. *Methods*, 143, 16-33.

[21]. Shu, W., Bo, X., Zheng, Z., and Wang, S. (2008). A novel representation of RNA secondary structure based on element-contact graphs. *BMC Bioinformatics*, 9(1), 1-18.

[22]. Tarjan, R. (1972). Depth-first search and linear graph algorithms. *SIAM Journal on Computing*, 1(2), 146-160.

[23]. Tarjan, R. E. (1976). Edge-disjoint spanning trees and depth-first search. *Acta Informatica*, 6(2), 171-185.

[24]. Wang, C., Liu, J. B., and Wang, S. (2017). Sharp upper bounds for multiplicative Zagreb indices of bipartite graphs with given diameter. *Discrete Applied Mathematics*, 227, 156-165.

[25]. Wiener, H. (1947). Structural determination of paraffin boiling points. *Journal of the American Chemical Society*, 69(1), 17-20.

[26]. Zhai, Y., Liu, J.B. S. and Wang, S. (2017) Structure properties of Koch networks based on networks dynamical systems, *Complexity*, 2017, 7.

[27]. https://graphonline.ru/en