Discrete Mathematics Letters
www.dmlett.com

Discrete Math. Lett. **9** (2022) 72–79
DOI: 10.47443/dml.2021.s212

Research Article

Entropy measures of distance matrix*

Bünyamin Şahin[1,†], Abdulgani Şahin[2]

[1]*Department of Mathematics, Faculty of Science, Selçuk University, Konya, Turkey*
[2]*Department of Mathematics, Faculty of Science and Letters, Agri Ibrahim Cecen University, Ağrı, Turkey*

(Received: 28 February 2022. Accepted: 19 March 2022. Published online: 25 March 2022.)

Abstract

In this paper, two distance-based entropy measures are studied that were introduced by Bonchev and Trinajstić [*J. Chem. Phys.* **38** (1977) 4517–4533] for interpreting the molecular branching of molecular graphs. One of these entropy measures is based on the distribution of distances in the distance matrix and the other one is based on the distribution of distances in the upper triangular submatrix. The mentioned measures are calculated for paths, stars, complete graphs, cycles, and complete bipartite graphs. The minimal trees with respect to the entropy measures under consideration with fixed diameter are also determined.

Keywords: distance; Wiener index; distance matrix; entropy measure.

2020 Mathematics Subject Classification: 05C09, 05C92.

1. Introduction

The entropy concept was introduced by Shannon in 1948 [17]. The Shannon entropy can be applied to different networks with the possibility of constructing a finite probability scheme for each network. The graph entropy concept was first defined by Rashevsky [15] in 1955. His entropy measure is based on the partition of the vertices with respect to equivalent classes of vertex degrees. Trucco [18] extended this definition by using automorphism groups of graphs. Mowshowitz [13] applied the information theory to different chemical structures and mathematical structures in 1968. Some properties of graph entropies were reported by Das and Shi [5]. Detail about the distance-based entropy measures can be found in [4,14]. The entropy measure based on dominating sets of graphs was introduced recently in [16]. More information about graph entropies can be found in the survey [7] and in the book [6].

Many molecular properties of materials are obtained by molecular topologies [11]. These measures are called topological indices or molecular descriptors in chemical graph theory. Many chemical, physical and biological properties of molecules have good correlations with several well-known topological indices. Therefore, many researchers from a wide range of sciences study on this topic. The first topological index was introduced by Wiener in 1947 [19]. The Wiener index is equal to one half of the sum of distances between every pair of vertices in a graph. The minimal trees of fixed diameter with respect to the Wiener index are characterized by Liu and Pan [12]. Detail about the Wiener index can be found in the excellent survey [8].

So many topological indices have been introduced in the last fifty years. It is understood that they have usually correlated more or less with the relative molecular properties of molecules but the same index can not has high discrimination ability for different molecules [11]. Bonchev and Trinajstić [1] introduced the entropy measures based on distances for interpreting the molecular branching of molecular graphs. Later, they applied the information theory in characterizing chemical structures [2,3]. These molecular descriptors were called information indices and it was shown that the information indices have greater discriminating power for molecules than the respective topological indices [9,10].

2. Preliminaries

Let G be a simple graph with the vertex set $V(G)$ and the edge set $E(G)$. For a vertex $u \in V(G)$, the notation $N_G(u) = \{v \mid uv \in E(G)\}$ denotes the vertices which are adjacent to u and $N_G[u] = \{u\} \cup N_G(u)$. The degree of a vertex u is the cardinality of $N_G(u)$ and it is denoted by $deg_G(u)$ or simply $deg(u)$. A vertex which has degree one is called a leaf. Moreover,

*In memory and honor of Professor Nenad Trinajstić. At the time of finalising this article, the paper [1] was cited more than six hundred fifty times and it is the most cited fifth paper according to his Google Scholar profile. The present paper is concerned with the entropy measures reported in [1].

†Corresponding author (bunyamin.sahin@selcuk.edu.tr).

the distance between the vertices u and v is denoted by $d(u,v)$. The maximum distance between any two vertices of a graph G is called diameter and is denoted by $diam(G)$. The distance matrix $D=[d_{ij}]$, $(i,j=1,2,\ldots,n)$ contains the distances $d_{ij}=d(i,j)$ between any two vertices of a connected graph.

The number of vertices of a graph G is called order and it is denoted by $|V(G)|=n$. The paths, cycles and stars of order n are denoted by P_n, C_n and S_n, respectively. Moreover, complete graphs of order n are denoted by K_n and complete bipartite graphs are denoted by $K_{s,t}$ with bipartite sets $\{u_1,u_2,\ldots,u_s\}$ and $\{v_1,v_2,\ldots,v_t\}$.

Definition 2.1 (see [17]). *For a given probably vector $p=\{p_1,p_2,\ldots,p_n\}$ such that $0\le p_i\le 1$ and $\sum_{i=1}^n p_i=1$, the ShannonŠs entropy of p is presented by the following equation*

$$I(p)=-\sum_{i=1}^{n} p_i \log p_i.$$

Definition 2.2. *For a vertex $u\in V(G)$, the total distance of u is defined as*

$$D(u)=\sum_{v\in V(G)} d(u,v).$$

Definition 2.3 (see [19]). *The Wiener index of a graph G is defined as*

$$W(G)=\frac{1}{2}\sum_{u\in V(G)} D(u).$$

We now can present the definitions of the information entropies introduced by Bonchev and Trinajstić [1]. In the distance matrix of a graph G, the distance i appears $2n_i$ times, where $1\le i\le diam(G)$. Thus, n^2 elements of the distance matrix of G are partitioned into $diam(G)+1$ groups with n zeros which are diagonal elements of the matrix. Therefore, the probability distribution of the $diam(G)+1$ groups for each i-th group is presented in Table 1 (see [11]), where $p_i=2n_i/n^2$ for $1\le i\le diam(G)$ and $p_0=n/n^2=1/n$.

Table 1: The distance and probability distributions of distance matrix.

Distance	0	1	2	$\cdots$	$diam(G)$
Frequency	n	$2n_1$	$2n_2$	$\cdots$	$2n_{diam(G)}$
Probability	$1/n$	p_1	p_2	$\cdots$	$p_{diam(G)}$

Since the distance matrix is symmetric, in order to simplify the calculations, only the upper triangular submatrix are be used. Thus, $n(n-1)/2$ upper off-diagonal elements are used for the calculations.

Definition 2.4 (see [1,11]). *For a given distance i in a graph G such that $1\le i\le diam(G)$, the information entropies $I(G)$ and $I^*(G)$ are defined as follows*

$$I(G)=-\frac{1}{n}\log\frac{1}{n}-\sum_{i=1}^{diam(G)}\frac{2n_i}{n^2}\log\frac{2n_i}{n^2} \quad \textit{and} \quad I^*(G)=-\sum_{i=1}^{diam(G)}\frac{2n_i}{n(n-1)}\log\frac{2n_i}{n(n-1)}.$$

In order to make some calculations, we order the diagonals of a square matrix

$$A=\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n}\\ a_{21} & a_{22} & \ldots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{n1} & a_{n2} & \ldots & a_{nn}\end{bmatrix}.$$

We order the diagonals of A as follows:

$$\begin{aligned} diag(1) &= \{a_{1n}\},\\ diag(2) &= \{a_{1n-1},a_{2n}\},\ldots,\\ diag(n-1) &= \{a_{12},a_{23},\ldots,a_{n-1n}\},\\ diag(n) &= \{a_{11},a_{22},\ldots,a_{nn}\},\\ diag(n+1) &= \{a_{21},a_{32},\ldots,a_{nn-1}\},\ldots,\\ diag(2n-2) &= \{a_{n-11},a_{n2}\},\end{aligned}$$

$$diag(2n-1) = \{a_{n1}\}.$$

It is noted that if A is a symmetric matrix, then the following relations are hold:

$$\begin{aligned} diag(1) &= diag(2n-1), \\ diag(2) &= diag(2n-2) =, \dots, \\ diag(n-1) &= diag(n+1). \end{aligned}$$

We note that $diag(n)$ is the main diagonal of the matrix A and $|diag(n)| = n$. Moreover, the number of the elements which appear in i-th diagonal is $|diag(i)| = |diag(2n-i)| = i$ for $1 \le i \le n-1$.

3. Information entropies of some graphs

Theorem 3.1. *The information entropies I and I^* of the path graph P_n of order n are given by the following formulas:*

$$I(P_n) = -\frac{1}{n}\log\frac{1}{n} - \sum_{i=1}^{n-1}\frac{2i}{n^2}\log\frac{2i}{n^2} \quad \textit{and} \quad I^*(P_n) = -\sum_{i=1}^{n-1}\frac{2i}{n^2-n}\log\frac{2i}{n^2-n}.$$

Proof. The distance matrix of P_n is presented as follows:

$$D(P_n) = \begin{bmatrix} 0 & 1 & 2 & 3 & \dots & n-2 & n-1 \\ 1 & 0 & 1 & 2 & \dots & n-3 & n-2 \\ 2 & 1 & 0 & 1 & \dots & n-4 & n-3 \\ 3 & 2 & 1 & 0 & \dots & n-5 & n-4 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ n-2 & n-1 & \dots & \dots & \dots & 0 & 1 \\ n-1 & n-2 & \dots & \dots & \dots & 1 & 0 \end{bmatrix}.$$

It can be seen that the distance $n-1$ appears in $diag(1)$ and $diag(2n-1)$. The distance $n-2$ appears in $diag(2)$ and $diag(2n-2)$, and generally the distance $n-i$ appears in $diag(i)$ and $diag(2n-i)$ with the frequency $2n_i = 2i$ $(1 \le i \le n-1)$. The $diag(n)$ contains n zeros.

From the matrix $D(P_n)$, we obtain the frequency of distances ($Freq.$), probability distributions of distances in the distance matrix (p_i) and probability distributions of distances in the upper triangular distance matrix (p_i^*) as shown in Table 2.

Table 2: Probability distributions for the path graph P_n.

i	0	1	2	3	$\cdots$	$n-2$	$n-1$
$Freq.$	n	$2n-2$	$2n-4$	$2n-6$	$\cdots$	4	2
p_i	$\frac{1}{n}$	$\frac{2n-2}{n^2}$	$\frac{2n-4}{n^2}$	$\frac{2n-6}{n^2}$	$\cdots$	$\frac{4}{n^2}$	$\frac{2}{n^2}$
p_i^*	0	$\frac{2n-2}{n^2-n}$	$\frac{2n-4}{n^2-n}$	$\frac{2n-6}{n^2-n}$	$\cdots$	$\frac{4}{n^2-n}$	$\frac{2}{n^2-n}$

Now, by using the definitions of the information entropies I and I^*, we obtain the required formulas. □

Theorem 3.2. *The information entropies I and I^* of the star graph S_n of order n are given by the following formulas,*

$$I(S_n) = -\frac{1}{n}\log\frac{1}{n} - \frac{2n-2}{n^2}\log\frac{2n-2}{n^2} - \frac{n^2-3n+2}{n^2}\log\frac{n^2-3n+2}{n^2},$$
$$I^*(S_n) = -\frac{2n-2}{n^2-n}\log\frac{2n-2}{n^2-n} - \frac{n^2-3n+2}{n^2-n}\log\frac{n^2-3n+2}{n^2-n}.$$

Proof. Let S_n be a star of order n with the vertex set $\{v_1, v_2, \dots, v_n\}$ such that v_n is the central vertex of S_n. Then, the distance matrix of S_n is presented as follows:

$$D(S_n) = \begin{bmatrix} 0 & 2 & 2 & \dots & 2 & 1 \\ 2 & 0 & 2 & \dots & 2 & 1 \\ 2 & 2 & \ddots & \ddots & \vdots & 1 \\ \vdots & \vdots & \ddots & \ddots & 2 & \vdots \\ 2 & 2 & \dots & 2 & \ddots & 1 \\ 1 & 1 & \dots & \dots & 1 & 0 \end{bmatrix}.$$

Note that the distance 1 appears only in the n-th row and in the n-th column with $2n-2$ times. Therefore, the distance matrix $D(S_n)$ consists of n zeros, $2n-2$ times 1, and n^2-3n+2 times 2.

From the matrix $D(S_n)$, we obtain the frequency of distances ($Freq.$), probability distributions of distances in the distance matrix (p_i), and probability distributions of distances in the upper triangular submatrix (p_i^*) as shown in Table 3.

Table 3: Probability distributions for the star graph S_n.

i	0	1	2
$Freq.$	n	$2n-2$	n^2-3n+2
p_i	$\frac{1}{n}$	$\frac{2n-2}{n^2}$	$\frac{n^2-3n+2}{n^2}$
p_i^*	0	$\frac{2n-2}{n^2-n}$	$\frac{n^2-3n+2}{n^2-n}$

Now, by using the definitions of the information entropies I and I^*, we obtain the $I(S_n)$ and $I^*(S_n)$. □

Theorem 3.3. *The information entropies I and I^* of the complete graph K_n of order n are given by the following formulas:*

$$I(K_n) = -\frac{1}{n}\log\frac{1}{n} - \frac{n-1}{n}\log\frac{n-1}{n} \quad \textit{and} \quad I^*(K_n) = 0.$$

Proof. It is known that the distance matrix of the complete graph K_n is consisted of n times 0 which are diagonal elements and n^2-n times 1 which are the off-diagonal elements of the distance matrix $D(K_n)$. Therefore, we obtain the probability distributions of K_n as shown in Table 4.

Table 4: Probability distributions for the complete graph K_n.

i	0	1
$Freq.$	n	n^2-n
p_i	$\frac{1}{n}$	$\frac{n^2-n}{n^2}$
p_i^*	0	1

By using the definitions of the information entropies I and I^*, we obtain the required formulas. □

Theorem 3.4. *The information entropies I and I^* of the cycle graph C_n of order n are given by the following formulas.*

i) If the order of C_n is even, then entropy values are computed as

$$I(C_n) = -\frac{2}{n}\log\frac{1}{n} - \frac{n-2}{n}\log\frac{2}{n} \quad \textit{and} \quad I^*(C_n) = -\frac{n-2}{n-1}\log\frac{2}{n-1} - \frac{1}{n-1}\log\frac{1}{n-1}.$$

ii) If the order of C_n is odd, then entropy values are computed as

$$I(C_n) = -\frac{1}{n}\log\frac{1}{n} - \frac{n-1}{n}\log\frac{2}{n} \quad \textit{and} \quad I^*(C_n) = \log\frac{n-1}{2}.$$

Proof. i) If the order of C_n is even then, its diameter is equal to $n/2$. Therefore, the distance matrix of C_n is presented as follows:

$$D(C_n) = \begin{bmatrix} 0 & 1 & 2 & \cdots & \cdots & \frac{n}{2} & \cdots & \cdots & 2 & 1 \\ 1 & 0 & 1 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 2 \\ 2 & 1 & \ddots & 1 & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \frac{n}{2} \\ \frac{n}{2} & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 2 \\ 2 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 1 \\ 1 & 2 & \cdots & \cdots & \frac{n}{2} & \cdots & \cdots & 2 & 1 & 0 \end{bmatrix}.$$

It can be seen that the distance 1 appears in $diag(1)$, $diag(n-1)$ in the upper triangular submatrix and in $diag(n+1)$, $diag(2n-1)$ in the lower triangular submatrix. Thus,

$$2n_1 = |diag(1)| + |diag(n-1)| + |diag(n+1)| + |diag(2n-1)| = 1 + n - 1 + n - 1 + 1 = 2n.$$

Also, the distance 2 appears in $diag(2)$, $diag(n-2)$ in the upper triangular submatrix and in $diag(n+2)$, $diag(2n-2)$ in the lower triangular submatrix. Thus,

$$2n_2 = |diag(2)| + |diag(n-2)| + |diag(n+2)| + |diag(2n-2)| = 2 + n - 2 + n - 2 + 2 = 2n.$$

Generally, we obtain that $2n_i = 2n$ for $1 \le i \le (n/2) - 1$. We can investigate the frequency of the $diam(C_n) = n/2$. The distance $n/2$ appears in $diag(n/2)$ in the upper triangular submatrix and in $diag(3n/2)$ in the lower triangular submatrix. Therefore, we obtain

$$2n_{(\frac{n}{2})} = \left|diag\left(\frac{n}{2}\right)\right| + \left|diag\left(\frac{3n}{2}\right)\right| = \frac{n}{2} + \frac{n}{2} = n.$$

From the matrix $D(C_n)$, the probability distributions of distances are obtained as shown in Table 5.

Table 5: Probability distributions for the cycle graph C_n when n is even.

i	0	1	2	3	$\cdots$	$\frac{n}{2}-1$	$\frac{n}{2}$
$Freq.$	n	$2n$	$2n$	$2n$	$\cdots$	$2n$	n
p_i	$\frac{1}{n}$	$\frac{2}{n}$	$\frac{2}{n}$	$\frac{2}{n}$	$\cdots$	$\frac{2}{n}$	$\frac{1}{n}$
p_i^*	0	$\frac{2}{n-1}$	$\frac{2}{n-1}$	$\frac{2}{n-1}$	$\cdots$	$\frac{2}{n-1}$	$\frac{1}{n-1}$

By using the definitions of the information entropies I and I^*, we obtain $I(C_n)$ and $I^*(C_n)$.

ii) If the order of C_n is odd, then its diameter is equal to $(n-1)/2$. Thus, the distance matrix of C_n is presented as:

$$D(C_n) = \begin{bmatrix} 0 & 1 & 2 & \cdots & \cdots & \frac{n-1}{2} & \frac{n-1}{2} & \cdots & \cdots & 2 & 1 \\ 1 & 0 & 1 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 2 \\ 2 & 1 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \frac{n-1}{2} \\ \frac{n-1}{2} & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \frac{n-1}{2} \\ \frac{n-1}{2} & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 2 \\ 2 & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & 1 \\ 1 & 2 & \cdots & \cdots & \frac{n-1}{2} & \frac{n-1}{2} & \cdots & \cdots & 2 & 1 & 0 \end{bmatrix}.$$

It is noted that the distance 1 appears in $diag(1)$, $diag(n-1)$ in the upper triangular submatrix and in $diag(n+1)$, $diag(2n-1)$ in the lower triangular submatrix. Then,

$$2n_1 = |diag(1)| + |diag(n-1)| + |diag(n+1)| + |diag(2n-1)| = 1 + n - 1 + n - 1 + 1 = 2n.$$

Also, the distance 2 appears in $diag(2)$, $diag(n-2)$ in the upper triangular submatrix and in $diag(n+2)$, $diag(2n-2)$ in the lower triangular submatrix. Then,

$$2n_2 = |diag(2)| + |diag(n-2)| + |diag(n+2)| + |diag(2n-2)| = 2 + n - 2 + n - 2 + 2 = 2n.$$

This continues and finally, we compute the frequency of $diam(C_n) = (n-1)/2$. The distance $(n-1)/2$ appears in $diag((n-1)/2)$, $diag((n+1)/2)$ in the upper triangular submatrix and in $diag((3n-1)/2)$, $diag((3n+1)/2)$ in the lower triangular submatrix. Therefore,

$$2n_{(\frac{n-1}{2})} = \left|diag\left(\frac{n-1}{2}\right)\right| + \left|diag\left(\frac{n+1}{2}\right)\right| + \left|diag\left(\frac{3n-1}{2}\right)\right| + \left|diag\left(\frac{3n+1}{2}\right)\right| = 2n.$$

From the $D(C_n)$ matrix, the probability distributions of distances is shown in Table 6.

By the definitions of the information entropies I and I^*, we obtain $I(C_n)$ and $I^*(C_n)$.

Table 6: Probability distributions for the path graph C_n when n is odd.

i	0	1	2	3	$\cdots$	$\frac{n-1}{2}$
$Freq.$	n	$2n$	$2n$	$2n$	$\cdots$	$2n$
p_i	$\frac{1}{n}$	$\frac{2}{n}$	$\frac{2}{n}$	$\frac{2}{n}$	$\cdots$	$\frac{2}{n}$
p_i^*	0	$\frac{2}{n-1}$	$\frac{2}{n-1}$	$\frac{2}{n-1}$	$\cdots$	$\frac{2}{n-1}$

□

Theorem 3.5. *The information entropies I and I^* of the complete bipartite graph $K_{\frac{n}{2},\frac{n}{2}}$ of order n are given as*

$$I(K_{\frac{n}{2},\frac{n}{2}}) = -\frac{1}{n}\log\frac{1}{n} - \frac{1}{2}\log\frac{1}{2} - \frac{n-2}{2n}\log\frac{n-2}{2n} \quad \textit{and} \quad I^*(K_{\frac{n}{2},\frac{n}{2}}) = -\frac{n}{2n-2}\log\frac{n}{2n-2} - \frac{n-2}{2n-2}\log\frac{n-2}{2n-2}.$$

Proof. Let $A = \{v_1, v_3, \dots, v_{n-1}\}$ and $B = \{v_2, v_4, \dots, v_n\}$ be the bipartite sets of $K_{\frac{n}{2},\frac{n}{2}}$. The distance matrix of $K_{\frac{n}{2},\frac{n}{2}}$ is

$$D(K_{\frac{n}{2},\frac{n}{2}}) = \begin{bmatrix} 0 & 1 & 2 & \dots & 1 & 2 & 1 \\ 1 & 0 & 1 & 2 & \ddots & 1 & 2 \\ 2 & 1 & \ddots & 1 & \ddots & \ddots & 1 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 1 & \ddots & \ddots & \ddots & \ddots & \ddots & 2 \\ 2 & 1 & \ddots & \ddots & \ddots & \ddots & 1 \\ 1 & 2 & 1 & \dots & 2 & 1 & 0 \end{bmatrix}.$$

It is noted that the distance 1 appears in $diag(1), diag(3), \dots, diag(n-1)$ in the upper triangular submatrix and in $diag(n+1)$, $diag(n+3), \dots, diag(2n-1)$ in the lower triangular submatrix. Then,

$$2n_1 = 2(\,|diag(1)| + |diag(3)| + \dots + |diag(n-1)|\,) = 2(1+3+\dots+n-1) = \frac{n^2}{2}.$$

Also, the distance 2 appears in $diag(2)$, $diag(4)$, $\dots$, $diag(n-2)$ in the upper triangular submatrix and in $diag(n+2)$, $diag(n+4)$, $\dots$, $diag(2n-2)$ in the lower triangular submatrix. Then,

$$2n_2 = 2(\,|diag(2)| + |diag(4)| + \dots + |diag(n-2)|) = 2(2+4+\dots+n-2) = \frac{n^2-2n}{2}.$$

Thus, from the $D(K_{\frac{n}{2},\frac{n}{2}})$ matrix, the probability distributions of distances are obtained as shown in Table 7.

Table 7: Probability distributions for the complete bipartite graph $K_{\frac{n}{2},\frac{n}{2}}$.

i	0	1	2
$Freq.$	n	$\frac{n^2}{2}$	$\frac{n^2-2n}{2}$
p_i	$\frac{1}{n}$	$\frac{1}{2}$	$\frac{n-2}{2n}$
p_i^*	0	$\frac{n}{2n-2}$	$\frac{n-2}{2n-2}$

By using the definitions of the information entropies I and I^*, we obtain $I(K_{\frac{n}{2},\frac{n}{2}})$ and $I^*(K_{\frac{n}{2},\frac{n}{2}})$. □

4. Some relations with respect to the information entropies I and I^*

In order to make some comparisons, we use majorization method [5]. We consider non-increasing arrangement of each vector in $\mathbb{R}^n$. Consider a vector $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ such that $x_1 \geq x_2 \geq \dots \geq x_n$.

Definition 4.1 (see [5]). *For $x, y \in \mathbb{R}^n$, $x \prec y$ if*

$$G = \begin{cases} \sum_{i=1}^{k} x_i \leq \sum_{i=1}^{k} y_i, & i = 1, 2, \dots, n-1 \\ \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \end{cases}.$$

When $x \prec y$, x is said to be majorized by y (or y majorizes x).

Let $p(G) = (p(v_1), p(v_2), \ldots, p(v_n))$ be a probably vector of a graph G for the vertex set $\{v_1, v_2, \ldots, v_n\}$ such that $p(v_1) \geq p(v_2) \geq \ldots \geq p(v_n)$ and $\sum_{i=1}^{n} p(v_i) = 1$. Since the fuction $h(x) = -x \log x$ is a concave function for $x > 0$, we give an essential theorem as used in [5] for the information entropy $I(G) = -\sum_{i=1}^{n} p(v_i) \log p(v_i)$.

Theorem 4.1. *Let H and G be two non-isomorphic graphs of order n and $p(H)$, $p(G)$ be the probability vectors of H and G, respectively. If $p(H) \prec p(G)$, then the inequality $I(G) \leq I(H)$ holds.*

If the diameter of two graphs are equal, then the distance matrices of these graphs are consisted of the same group distances but their frequencies can be different. It implies that the entropy measures I, I^* can be compared by majorization method.

Lemma 4.1. *Let G be a graph and $u, v \in V(G)$. If $G_{s,t}$ is the graph obtained from G by attaching s, t leaves to the vertices v, u (respectively), the information entropies of the graphs $G_{s,t}$, $G_{s-i,t+i}$ and $G_{s+i,t-i}$ (see Figure 1) are compared by the following inequalities with fixed diameter:*

i)
$$I(G_{s-i,t+i}) \leq I(G_{s,t}) \text{ for } 1 \leq i \leq s \quad \text{or} \quad I(G_{s+i,t-i}) \leq I(G_{s,t}) \text{ for } 1 \leq i \leq t.$$

ii)
$$I^*(G_{s-i,t+i}) \leq I^*(G_{s,t}) \text{ for } 1 \leq i \leq s \quad \text{or} \quad I^*(G_{s+i,t-i}) \leq I^*(G_{s,t}) \text{ for } 1 \leq i \leq t.$$

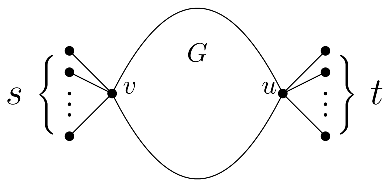

Figure 1: The graph $G_{s,t}$.

Proof. i) Assume that a leaf y is removed from v and it is attached to u. Therefore, the tree $G_{s-1,t+1}$ is obtained. Let $x = (\ldots, p_2(G_{s,t}), \ldots)$ and $x' = (\ldots, p_2(G_{s-1,t+1}), \ldots)$ be nonincreasing probably vectors of $G_{s,t}$ and $G_{s-1,t+1}$. The total distance from y to other leaves which are incident to v is $2(s-1)$. Moreover, the total distance from $s-1$ leaves to y is also $2(s-1)$. It means that if y is removed from v, the frequency of the distance 2 is decreased $2(s-1)$ times in the distance matrix of $G_{s,t}$. In the graph $G_{s-1,t+1}$, the frequency of the distance 2 is increased $2t$ times because of the leaf y is attached to u. Then, the difference of probabilities of the distance 2 in the distance matrices of $G_{s-1,t+1}$ and $G_{s,t}$ is

$$p_2(G_{s-1,t+1}) - p_2(G_{s,t}) = \frac{2t}{n^2} - \frac{2(s-1)}{n^2} = \frac{2t - 2s + 2}{n^2} \geq 0.$$

It implies that $x \prec x'$ and $I(G_{s-i,t+i}) \leq I(G_{s,t})$ or $I(G_{s+i,t-i}) \leq I(G_{s,t})$.

ii) The same argument is used for the upper triangular submatrix and so the result is obtained.

□

Let $P_{d+1} : v_0 v_1 \ldots v_d$ be a path of order $d+1$. The graph which is obtained from P_{d+1} by attaching $k = n - d - 1$ leaves to i-th vertex of P_{d+1} is denoted by $P_{d+1,i,k}$ (see Figure 2).

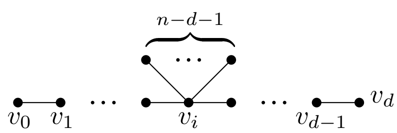

Figure 2: The graph $P_{d+1,i,k}$.

Lemma 4.2. *For $1 \leq j < i \leq \lfloor \frac{d}{2} \rfloor$, the following inequalities hold*

i) $I(P_{d+1,i,k}) \leq I(P_{d+1,j,k})$,

ii) $I^*(P_{d+1,i,k})) \leq I^*(P_{d+1,j,k})$.

Proof. i) Assume that a leaf y which is attached to the i-th vertex of P_{d+1} is moved on the $(i+1)$-th vertex of P_{d+1} such that $1 < i+1 \leq \lfloor \frac{d}{2} \rfloor$. Now, we order the distances from y to other vertices $v_0, v_1, \ldots, v_d$ such that $i+1, i, \ldots, 1, 2, \ldots, d-i+1$ in $P_{d+1,i,1}$, respectively. Similarly, distances from y to other vertices $v_0, v_1, \ldots, v_d$ are ordered $i+2, i+1, \ldots, 1, 2, \ldots, d-i$ in $P_{d+1,i+1,1}$, respectively. It means that the frequencies of greater distances are decreased in the distance matrix and the frequencies of small distances are increased. This trend continues until the leaf y arrives to the $\lfloor \frac{d}{2} \rfloor$-th vertex of the graph. Then, the probably vector of $P_{d+1,i,1}$ is majorized by the probably vector of $P_{d+1,i+1,1}$ and $I(P_{d+1,i+1,1}) \leq I(P_{d+1,i,1})$. It can be generalized as $I(P_{d+1,i,k}) \leq I(P_{d+1,j,k})$ for k leaves.

ii) It is obtained by the same way as used in the proof of item (i).

□

By Lemma 4.2, the information indices I, I^* of trees in the family of $P_{n-2,i,1}$ for $d = n-2$ and $k = 1$ are ordered as in the following corollary.

Corollary 4.1. *For $1 \leq i \leq \lfloor \frac{d}{2} \rfloor$, the information indices are ordered as follows $I(P_{n-2,\lfloor \frac{d}{2} \rfloor,1}) \leq I(P_{n-2,\lfloor \frac{d}{2} \rfloor-1,1}) \leq \cdots \leq I(P_{n-2,1,1})$ and $I^*(P_{n-2,\lfloor \frac{d}{2} \rfloor,1}) \leq I^*(P_{n-2,\lfloor \frac{d}{2} \rfloor-1,1}) \leq \cdots \leq I^*(P_{n-2,1,1})$.*

The results which are obtained in the previous corollary are generalized to graphs with diameter $3 \leq d \leq n-3$ and $k = n-d-1$ by Lemma 4.1 and Lemma 4.2 as in the following corollary.

Corollary 4.2. *Assume that the diameter d satisfies $3 \leq d \leq n-3$ and $k = n-d-1$. The information indices are ordered as follows $I(P_{d,\lfloor \frac{d}{2} \rfloor,k}) \leq I(P_{d,\lfloor \frac{d}{2} \rfloor-1,k}) \leq \cdots \leq I(P_{d,1,k})$ and $I^*(P_{d,\lfloor \frac{d}{2} \rfloor,k}) \leq I^*(P_{d,\lfloor \frac{d}{2} \rfloor-1,k}) \leq \cdots \leq I^*(P_{d,1,k})$.*

5. Conclusion

There are many open problems concerning the entropy measures considered in this paper. The extremal trees as well as the extremal unicyclic and bicyclic graphs can be obtained with respect to these entropy measures with different fixed parameters. It is observed that the distribution of vertex degrees of graphs is well studied but the distribution of distances is not studied well in the literature. The entropy measures considered in this paper are useful tools for the distribution of distances in a graph. Also, the average distance $\mu(G)$ of a graph G is defined (see [8]) as

$$\mu(G) = \frac{W(G)}{n(n-1)}.$$

The average distance μ related to the entropy measure I^*. It seems to be an interesting problem to find relations between μ and I^*.

References

[1] D. Bonchev, N. Trinajstić, Information theory, distance matrix, and molecular branching, *J. Chem. Phys.* **67** (1977) 4517–4533.
[2] D. Bonchev, N. Trinajstić, On topological characterization of molecular branching, *Int. J. Quantum Chem.* **14** (1978) 293–303.
[3] D. Bonchev, N. Trinajstić, Chemical information theory: structural aspects, *Int. J. Quantum Chem.* **22** (1982) 463–480.
[4] Z. Chen, M. Dehmer, Y. Shi, A note on distance-based graph entropies, *Entropy* **16** (2014) 5416–5427.
[5] K. C. Das, Y. Shi, Some properties on entropies of graphs, *MATCH Commun. Math. Comput. Chem.* **78** (2017) 259–272.
[6] M. Dehmer, F. Emmert-Streib, Z. Chen, X. Li, Y. Shi (Eds.), *Mathematical Foundations and Applications of Graph Entropy*, Wiley, Weinheim, 2016.
[7] M. Dehmer, A. Mowshowitz, A history of graph entropy measures, *Inform. Sci.* **181** (2011) 57–78.
[8] R. C. Entringer, I. Gutman, A. A. Dobrynin, Wiener index of trees: theory and applications, *Acta Appl. Math.* **66** (2001) 211–249.
[9] E. V. Konstantinova, The discrimination ability of some topological and information distance indices for graphs of unbranched hexagonal systems, *J. Chem. Inf. Comput. Sci.* **36** (1996) 54–57.
[10] E. V. Konstantinova, M. V. Diudea, The Wiener polynomial derivatives and other topological indices in chemical research, *Croat. Chem. Acta* **73** (2000) 383–403.
[11] E. V. Konstantinova, On some applications of information indices in chemical graph theory, In: R. Ahlswede, L. Bäumer, N. Cai, H. Aydinian, V. Blinovsky, C. Deppe, H. Mashurian, *General Theory of Information Transfer and Combinatorics*, Springer, Berlin, 2006, pp. 831–852.
[12] H. Liu, X. F. Pan, On the Wiener index of trees with fixed diameter, *MATCH Commun. Math. Comput. Chem.* **60** (2008) 85–94.
[13] A. Mowshovitz, Entropy and the complexity of graphs: II. The information content of digraphs and infinite graphs, *Bull. Math. Biophys.* **30** (1968) 225–240.
[14] A. Mowshowitz, M. Dehmer, The Hosoya entropy of a graph, *Entropy* **17** (2015) 1054–1062.
[15] N. Rashevsky, Life, information theory and topology, *Bull. Math. Biophys.* **17** (1955) 229–235.
[16] B. Sahin, New network entropy: the domination entropy of graphs, *Inform. Process. Lett.* **174** (2022) #106195.
[17] C. Shannon, W. Weaver, *The Mathematical Theory of Communications*, University of Illinois, Urbana, 1949.
[18] E. Trucco, A note of the information content of graphs, *Bull. Math. Biophys.* **18** (1956) 129–135.
[19] A. H. Wiener, Structural determination of paraffin boiling points, *J. Amer. Chem. Soc.* **69** (1947) 17–20.