

Entropy Measures of Distance Matrix

Bünyamin ŞAHİN^{*1}, Abdulgani ŞAHİN²

¹Selçuk University, Department of Mathematics, Science Faculty, 42130, Konya, TURKEY

bunyamin.sahin@selcuk.edu.tr

² Agri Ibrahim Cecen University, Department of Mathematics, Faculty of Science and Letters, 04000, Ağrı, TURKEY

agani@agri.edu.tr

Abstract. Bonchev and Trinajstić defined two distance based entropy measures to measure the molecular branching of molecular graphs in 1977 [Information theory, distance matrix, and molecular branching, J. Chem. Phys., 38 (1977), 4517–4533]. In this paper we use these entropy measures which are based on distance matrices of graphs. The first one is based on distribution of distances in distance matrix and the second one is based on distribution of distances in upper triangular submatrix. We obtain the two entropy measures of paths, stars, complete graphs, cycles and complete bipartite graphs. Finally we obtain the minimal trees with respect to these entropy measures with fixed diameter.

Keywords. Distance, Wiener Index, Distance Matrix, Entropy Measure

1. Introduction

The entropy concept was introduced by Shannon in 1948 [1]. The Shannon entropy can be applied to different networks via the possibility of constructing a finite probability scheme for each network. The graph entropy concept was defined by Rashevsky [2] in 1955. His entropy measure is based on the partitioning of the vertices with respect to equivalent classes of vertex degrees. Trucco [3] extended this definition by using the automorphism groups of the graphs. Mowshowitz applied the information theory to different chemical structures and mathematical structures in 1968 [4]. Some properties of graph entropies were reported by Das and Shi [5]. Distance based entropy measures studied for example [6,7]. The entropy measure which is based on the dominating sets of the graphs was introduced recently [8]. More information about graph entropies can be found in the survey [9] and in the book [10].

Many molecular properties of materials are obtained by molecular topologies [11]. These measures are called topological indices or molecular descriptors in chemical graph theory. Chemical, physical and biological properties of molecules have good correlations with these 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 [12]. The Wiener index equals to one half of total distances between every pair of vertices in a graph. More results about Wiener index can be found in the excellent survey [13]. Moreover, the minimal trees of fixed diameter with respect to the Wiener index are characterized by Liu and Pan [14].

Too many topological indices have been introduced in the last 50 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 have high discrimination ability for different molecules [11]. Bonchev and Trinajstić introduced entropy measures which are based on distances to interpret the molecular branching of molecular graphs [15]. Later they applied the information theory in characterization of chemical structures [16,17]. 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 [18,19].

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 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, the distance between the vertices u and v is denoted by $d(u, v)$. The maximum distance between two any vertices of a graph G is called diameter and denoted by $\text{diam}(G)$. The distance matrix $D = [d_{ij}]$, $i, j = 1, 2, \dots, n$, contains the distances $d_{ij} = d(i, j)$ between two any 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 $\{v_1, v_2, \dots, v_s\}$ and $\{v_1, v_2, \dots, v_t\}$.

Definition 2.1. For a given probably vector $p = (p_1, p_2, \dots, p_n)$ such that $0 \leq p_i \leq 1$ and $\sum_{i=1}^n p_i = 1$, the Shannon's entropy of p is presented by the following equation [1]

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

Definition 2.2. For a vertex $u \in V(G)$, the total distance of u is introduced by the following equation

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

Definition 2.3. The Wiener index of a graph G is introduced by the following equation [12]

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

We can give the definitions of the information entropies which are introduced by Bonchev and Trinajstić [15]. In a distance matrix of a graph G , a distance i such that $1 \leq i \leq \text{diam}(G)$, appears $2n_i$ times. Then n^2 matrix elements d_{ij} are partitioned into $\text{diam}(G) + 1$ groups with the n zeros which are diagonal matrix elements. Therefore, the probability distribution of the $\text{diam}(G) + 1$ groups for each i -th group is presented in the following table [11]:

Table 1. The distance and probability distributions of distance matrix

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

where $p_i = 2n_i/n^2$ and $p_0 = n/n^2 = 1/n$.

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

Definition 2.4. For a given distance i in a graph G such that $1 \leq i \leq diam(G)$, the information entropies $I(G)$ and $I^*(G)$ are defined as follows [11,15]

$$I(G) = -\frac{1}{n} \log \frac{1}{n} - \sum_{i=1}^{diam(G)} \frac{2n_i}{n^2} \log \frac{2n_i}{n^2},$$

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

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

In order to make some calculations, we ordered the diagonals of square matrix A . We can order the diagonals of A as follows.

$$\begin{aligned} diag(1) &= \{a_{1n}\}, \\ diag(2) &= \{a_{1n-1}, a_{2n}\}, \dots, \\ diag(n-1) &= \{a_{12}, a_{23}, \dots, a_{n-1n}\}, \\ diag(n) &= \{a_{11}, a_{22}, \dots, a_{nn}\}, \\ diag(n+1) &= \{a_{21}, a_{32}, \dots, a_{nn-1}\}, \dots, \\ diag(2n-2) &= \{a_{n-11}, a_{n2}\}, \\ diag(2n-1) &= \{a_{n1}\}. \end{aligned}$$

It is understood that if A is a symmetric matrix, the following relations are obtained,

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

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

$$|diag(i)| = |diag(2n - i)| = i \text{ for } 1 \leq i \leq n - 1.$$

3. Information Entropies of Some Graphs

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

$$I(P_n) = -\frac{1}{n} \log \frac{1}{n} - \sum_{i=1}^{n-1} \frac{2i}{n^2} \log \frac{2i}{n^2}$$

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

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 distance $n - 1$ appears in $diag(1), diag(2n - 1)$.

Distance $n - 2$ appears in $diag(2), diag(2n - 2)$ and generally distance $n - i$ appears in $diag(i), diag(2n - i)$ with the frequency $2n_i = 2i$ ($1 \leq i \leq n - 1$). The $diag(n)$ contains n zeros.

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

Table 2. Probability distributions for path graph P_n

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

By the definitions of the information entropies I and I^* , we obtain the $I(P_n)$ and $I^*(P_n)$ as given in the following equations

$$I(P_n) = -\frac{1}{n} \log \frac{1}{n} - \sum_{i=1}^{n-1} \frac{2i}{n^2} \log \frac{2i}{n^2},$$

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

Theorem 3.2. The information entropies I and I^* of star graph S_n of order n is 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 the S_n . Then, the distance matrix of S_n is presented as follows.

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

It can be seen that the distance 1 appears in n -th row and n -th column with $2n-2$ times. Out of n zeros and $2n-2$ times 1, the distance 2 appears n^2-3n+2 times.

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

Table 3. Probability distributions for star graph S_n

i	0	1	2
<i>Freq.</i>	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	$\frac{2n-2}{n^2-n}$	$\frac{n^2-3n+2}{n^2-n}$
---------	---	----------------------	--------------------------

By the definitions of the information entropies I and I^* , we obtain the $I(S_n)$ and $I^*(S_n)$ as presented in the following equations

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

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

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

$$I(K_n) = -\frac{1}{n} \log \frac{1}{n} - \frac{n-1}{n} \log \frac{n-1}{n}$$

$$I^*(K_n) = 0.$$

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

Table 4. Probability distributions for complete graph K_n

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

By the definitions of the information entropies I and I^* , we obtain the $I(K_n)$ and $I^*(K_n)$ as given in the following equations. It is clear that the upper triangular matrix contains the distance 1. Then its entropy measure equals to zero.

$$I(K_n) = -\frac{1}{n} \log \frac{1}{n} - \frac{n-1}{n} \log \frac{n-1}{n},$$

$$I^*(K_n) = 0.$$

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

i) If the order of cycle is even, then entropy values are presented by

$$I(C_n) = -\frac{2}{n} \log \frac{1}{n} - \frac{n-2}{n} \log \frac{2}{n},$$

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

ii) If the order of cycle is odd, then entropy values are presented by

$$I(C_n) = -\frac{1}{n} \log \frac{1}{n} - \frac{n-1}{n} \log \frac{2}{n},$$

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

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

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

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

$$\begin{aligned} 2n_1 &= |diag(1)| + |diag(n-1)| + |diag(n+1)| + |diag(2n-1)| \\ &= 1 + n - 1 + n - 1 + 1 = 2n. \end{aligned}$$

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

$$\begin{aligned} 2n_2 &= |diag(2)| + |diag(n-2)| + |diag(n+2)| + |diag(2n-2)| \\ &= 2 + n - 2 + n - 2 + 2 = 2n. \end{aligned}$$

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

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

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

Table 5. Probability distributions for path graph C_n for n is even

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

By the definitions of the information entropies I and I^* , we obtain the $I(C_n)$ and $I^*(C_n)$ as given in the following equations for n is even.

$$\begin{aligned}
 I(C_n) &= -2\frac{1}{n}\log\frac{1}{n} - \left(\frac{n}{2} - 1\right)\frac{2}{n}\log\frac{2}{n} \\
 &= -\frac{2}{n}\log\frac{1}{n} - \frac{n-2}{n}\log\frac{2}{n}, \\
 I^*(C_n) &= -\left(\frac{n}{2} - 1\right)\frac{2}{n-1}\log\frac{2}{n-1} - \frac{1}{n-1}\log\frac{1}{n-1} \\
 &= -\frac{n-2}{n-1}\log\frac{2}{n-1} - \frac{1}{n-1}\log\frac{1}{n-1}.
 \end{aligned}$$

ii) If the order of a cycle is odd, then its diameter equals to $(n-1)/2$. Therefore, the distance matrix of C_n is presented as follows for n is odd.

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

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

$$\begin{aligned}
 2n_1 &= |diag(1)| + |diag(n-1)| + |diag(n+1)| + |diag(2n-1)| \\
 &= 1 + n - 1 + n - 1 + 1 = 2n.
 \end{aligned}$$

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

$$\begin{aligned} 2n_2 &= |diag(2)| + |diag(n-2)| + |diag(n+2)| + |diag(2n-2)| \\ &= 2 + n - 2 + n - 2 + 2 = 2n. \end{aligned}$$

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

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

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

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

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

By the definitions of the information entropies I and I^* , we obtain the $I(C_n)$ and $I^*(C_n)$ as in the following equations for n is odd,

$$\begin{aligned} I(C_n) &= -\frac{1}{n} \log \frac{1}{n} - \left(\frac{n-1}{2}\right) \frac{2}{n} \log \frac{2}{n} \\ &= -\frac{1}{n} \log \frac{1}{n} - \frac{n-1}{n} \log \frac{2}{n}, \\ I^*(C_n) &= -\left(\frac{n-1}{2}\right) \frac{2}{n-1} \log \frac{2}{n-1} \\ &= \log \frac{n-1}{2}. \end{aligned}$$

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

$$I\left(K_{\frac{n}{2}, \frac{n}{2}}\right) = -\frac{1}{n} \log \frac{1}{n} - \frac{1}{2} \log \frac{1}{2} - \frac{n-2}{2n} \log \frac{n-2}{2n}$$

$$I^*\left(K_{\frac{n}{2}, \frac{n}{2}}\right) = -\frac{n}{2n-2} \log \frac{n}{2n-2} - \frac{n-2}{2n-2} \log \frac{n-2}{2n-2}$$

Proof. Let $K_{\frac{n}{2}, \frac{n}{2}}$ be a complete bipartite graph with bipartite sets $A = \{v_1, v_3, \dots, v_{n-1}\}$ and $B = \{v_2, v_4, \dots, v_n\}$.

Therefore, the distance matrix of $K_{\frac{n}{2}, \frac{n}{2}}$ is presented as follows.

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

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

$$\begin{aligned} 2n_1 &= 2(|diag(1)| + |diag(3)| + \dots + |diag(n-1)|) \\ &= 2(1 + 3 + \dots + n-1) = \frac{n^2}{2} \end{aligned}$$

It is obtained that distance 2 appears in $diag(2), diag(4), \dots, diag(n-2)$ in upper triangular submatrix and in $diag(n+2), diag(n+4), \dots, diag(2n-2)$ in lower triangular submatrix. Then

$$\begin{aligned} 2n_2 &= 2(|diag(2)| + |diag(4)| + \dots + |diag(n-2)|) \\ &= 2(2 + 4 + \dots + n-2) = \frac{n^2 - 2n}{2} \end{aligned}$$

From the $D(K_{\frac{n}{2}, \frac{n}{2}})$ matrix, the probability distributions of distances are 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 the definitions of the information entropies I and I^* , we obtain the $I\left(K_{\frac{n}{2}, \frac{n}{2}}\right)$ and $I^*\left(K_{\frac{n}{2}, \frac{n}{2}}\right)$ as in the following equations,

$$I\left(K_{\frac{n}{2}, \frac{n}{2}}\right) = -\frac{1}{n} \log \frac{1}{n} - \frac{1}{2} \log \frac{1}{2} - \frac{n-2}{2n} \log \frac{n-2}{2n},$$

$$I^*\left(K_{\frac{n}{2}, \frac{n}{2}}\right) = -\frac{n}{2n-2} \log \frac{n}{2n-2} - \frac{n-2}{2n-2} \log \frac{n-2}{2n-2}.$$

4. Some Relations with Respect to 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 R^n such that for a vector $x = (x_1, x_2, \dots, x_n) \in R^n$. Thus, we use $x_1 \geq x_2 \geq \dots \geq x_n$.

Definition 4.1. For $x, y \in R^n$, $x < y$ if ([5])

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

When, $x < y$, x is said to be majorized by y (y majorizes x).

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

Theorem 4.2. 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 $p(H) < p(G)$, then we obtain that $I(G) \leq I(H)$.

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

Lemma 4.3. Let G be a graph and $u, v \in V(G)$. Let $G_{s,t}$ be a graph which is obtained from G by attaching s, t pendant vertices to the vertices v, u (respectively) such that s and t are almost

equal. Therefore, 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

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

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 $x = (\dots, p_2(G_{s,t}), \dots)$ and $x' = (\dots, p_2(G_{s-1,t+1}), \dots)$ 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 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 distance 2 is increased $2t$ times because of the leaf y is attached to u . Then the difference of probabilities of 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 $x < x'$ 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 upper triangular submatrix and the result is obtained.

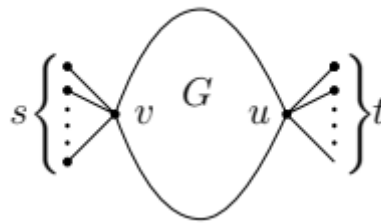


Figure 1. The graph $G_{s,t}$

Let $P_{d+1}: v_0 v_1 \dots v_d$ 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).

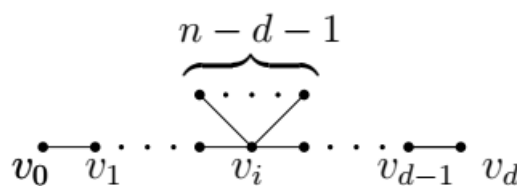


Figure 2. The graph $P_{d+1,i,k}$

Lemma 4.4. For any $1 \leq j < i \leq \left\lfloor \frac{d}{2} \right\rfloor$, it is obtained that

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

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

Proof. *i)* Assume that a leaf y which is attached to i -th vertex of P_{d+1} is move on $i + 1$ -th vertex of P_{d+1} such that $i < i + 1 \leq \left\lfloor \frac{d}{2} \right\rfloor$. Now we ordered the distances from y to other vertices v_0, v_1, \dots, v_d such that $i + 1, i, \dots, 1, 2, \dots, d - i + 1$ in $P_{d+1,i,1}$, respectively. Similarly, the distances from y to other vertices v_0, v_1, \dots, v_d are ordered $i + 2, i + 1, \dots, 1, 2, \dots, d - i$ 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 $\left\lfloor \frac{d}{2} \right\rfloor$ -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 for k leaves as $I(P_{d+1,i,k}) \leq I(P_{d+1,j,k})$.

ii) It is obtained by a similar way to item (*i*).

By Lemma 4.4 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 4.5. For $1 \leq i \leq \left\lfloor \frac{d}{2} \right\rfloor$ the information indices are ordered as follows

$$i) I\left(P_{n-2,\left\lfloor \frac{d}{2} \right\rfloor,1}\right) \leq I\left(P_{n-2,\left\lfloor \frac{d}{2} \right\rfloor-1,1}\right) \leq \dots \leq I(P_{n-2,1,1}),$$

$$ii) I^*\left(P_{n-2,\left\lfloor \frac{d}{2} \right\rfloor,1}\right) \leq I^*\left(P_{n-2,\left\lfloor \frac{d}{2} \right\rfloor-1,1}\right) \leq \dots \leq I^*(P_{n-2,1,1}).$$

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 4.3 and Lemma 4.4 as in the following corollary.

Corollary 4.6. Assume that the diameter is $3 \leq d \leq n - 3$ and the $k = n - d - 1$. Then the information indices are ordered as follows.

$$i) I\left(P_{d,\left\lfloor \frac{d}{2} \right\rfloor,k}\right) \leq I\left(P_{d,\left\lfloor \frac{d}{2} \right\rfloor-1,k}\right) \leq \dots \leq I(P_{d,1,k}),$$

$$ii) I^*\left(P_{d,\left\lfloor \frac{d}{2} \right\rfloor,k}\right) \leq I^*\left(P_{d,\left\lfloor \frac{d}{2} \right\rfloor-1,k}\right) \leq \dots \leq I^*(P_{d,1,k}).$$

Conclusion

Prof. Nenad Trinajstić passed away on 27 August 2021. The paper [15] was cited more than six hundred times and it is the most cited fifth paper according to his scholar page. Therefore we are very happy to improve the results about the entropy measures reported in [15].

There are many open problems in this topic. The entropy measures of different graphs are obtained. Moreover, extremal trees, extremal unicyclic and bicyclic graphs can be obtained with respect to the entropy measures with different parameters.

It is observed that the distribution of vertex degrees of graphs is well studied but distribution of distances does not studied as long as degree distributions in the literature. Then, the entropy measures which are reported in this paper are useful tools for distribution of distances in a graph.

Average distance of a graph G ($\mu(G)$) is defined as [13]

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

It is more related to the entropy measure I^* . It is an interesting problem to find the relations between μ and I^* .

References

- [1] C. Shannon and W. Weaver, Mathematical Theory of Communications, University of Illinois, Urbana, 1949.
- [2] N. Rashevsky, Life, information theory and topology, Bull. Math. Biophys., 17 (1955), 229–235.
- [3] E. Trucco, A note of the information content of graphs, Bull. Math. Biophys., 17 (1956), 129–135.
- [4] A. Mowshovitz, The information content of digraphs and infinite graphs, Bull. Math. Biophys., 30 (1968), 225–240.
- [5] K.C. Das, Y. Shi, Some Properties on Entropies of Graphs, MATCH Commun. Math. Comput. Chem., 78 (2017), 259-272.
- [6] A. Mowshowitz, M. Dehmer, The Hosoya entropy of a graph, Entropy 17 (2015) 1054–1062.
- [7] Z. Chen, M. Dehmer, Y. Shi, A Note on Distance-based Graph Entropies, Entropy 16 (2014), 546-5417.
- [8] B. Şahin, New Network Entropy: The Domination Entropy of Graphs, Information Processing Letters, 174 (2022), 106195.
- [9] M. Dehmer, A. Mowshowitz, A history of graph entropy measures, Information Sciences, 181 (2011) 57–78.
- [10] M. Dehmer, Z. Chen, X. Li, Y. Shi, F. Emmert-Streib, Mathematical Foundations and Applications of Graph Entropy, Wiley-Blackwell, 2016.
- [11] E.V. Konstantinova, On some applications of information indices in chemical graph theory, In: Ahlswede R. et al. (Eds.): General Theory of Information Transfer and Combinatorics, Lecture Notes in Computer Science, 4123 (2006), 831-852.

- [12] A. H. Wiener, Structural determination of paraffin boiling points, *J. Am. Chem. Soc.*, 69 (1947), 17-20.
- [13] R.C. Entringer, I. Gutman, A.A. Dobrynin, Wiener index of trees: theory and applications, *Acta Appl. Math.*, 66 (2001), 211-249.
- [14] H. Liu, X.F. Pan, On the Wiener Index of Trees with Fixed Diameter, *MATCH Commun. Math. Comput. Chem.*, 60 (2008), 85-94.
- [15] D. Bonchev, N. Trinajstić, Information theory, distance matrix, and molecular branching, *J. Chem. Phys.*, 38 (1977), 4517–4533.
- [16] D. Bonchev, N. Trinajstić, On topological characterization of molecular branching, *Int. J. Quantum Chem.*, S12 (1978), 293–303.
- [17] D. Bonchev, N. Trinajstić, Chemical information theory, structural aspects, *Intern. J. Quantum Chem. Symp.*, 16 (1982), 463-480.
- [18] 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.
- [19] E.V. Konstantinova, M.V. Diudea, The Wiener Polynomial Derivatives and Other Topological Indices in Chemical Research, *Croatica Chemica Acta*, 73 (2) (2000), 383-403.