



Interrelations of Graph Distance Measures Based on Topological Indices

Matthias Dehmer^{1,2*}, Frank Emmert-Streib^{3*}, Yongtang Shi^{4,5*}

1 Department of Computer Science, Universität der Bundeswehr München, Neubiberg, Germany, **2** Division for Bioinformatics and Translational Research, UMIT, Hall in Tyrol, Austria, **3** Computational Biology and Machine Learning Laboratory, Center for Cancer Research and Cell Biology, School of Medicine, Dentistry and Biomedical Sciences, Faculty of Medicine, Health and Life Sciences, Queen's University Belfast, Belfast, United Kingdom, **4** Center for Combinatorics and LPMC-TJKLC, Nankai University, Tianjin, China, **5** College of Computer and Control Engineering, Nankai University, Tianjin, China

Abstract

In this paper, we derive interrelations of graph distance measures by means of inequalities. For this investigation we are using graph distance measures based on topological indices that have not been studied in this context. Specifically, we are using the well-known Wiener index, Randić index, eigenvalue-based quantities and graph entropies. In addition to this analysis, we present results from numerical studies exploring various properties of the measures and aspects of their quality. Our results could find application in chemoinformatics and computational biology where the structural investigation of chemical components and gene networks is currently of great interest.

Citation: Dehmer M, Emmert-Streib F, Shi Y (2014) Interrelations of Graph Distance Measures Based on Topological Indices. PLoS ONE 9(4): e94985. doi:10.1371/journal.pone.0094985

Editor: Matjaz Perc, University of Maribor, Slovenia

Received: February 24, 2014; **Accepted:** March 21, 2014; **Published:** April 23, 2014

Copyright: © 2014 Dehmer et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: Matthias Dehmer and Yongtang Shi thank the Austrian Science Funds for supporting this work (project P26142). Yongtang Shi has also been supported by the National Science Foundation of China. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: matthias.dehmer@umit.at (MD); v@bio-complexity.com (FES); shi@nankai.edu.cn (YS)

Introduction

Methods to determine the structural similarity or distance between graphs have been applied in many areas of sciences. For example, in mathematics [1,2,3], in biology [4,5,6], in chemistry [7,8] and in chemoinformatics [9]. Other application-oriented areas where graph comparison techniques have been employed can be found in [10,11,12]. Note that the terms ‘graph similarity’ or ‘graph distance’ are not unique and strongly depend on the underlying concept. The two main concepts which have been explored extensively are exact and inexact graph matching, see [13,3]. Exact graph matching [2,3] relates to match graphs based on isomorphic relations. An important example is the so-called Zelinka distance [3] which requires computing the maximum common subgraphs of two graphs with the same number of vertices. However, it is evident that this technique is computationally demanding as the subgraph graph isomorphism problem is NP-complete [14]. In contrast to this, inexact or approximative techniques for comparing graphs match graphs in an error-tolerant way, see [13]. A highlight of this development has been the well-known graph edit distance (GED) due to Bunke [15]. String-based techniques also fit into the scheme of approximative graph comparison techniques [1,16]. This approach aims to derive string representations which capture structural information of the underlying networks. By using string alignment techniques, one is able to compute similarity scores of the derived strings instead of matching the graphs by using classical techniques. Concrete examples thereof can be found in [1,16].

As mentioned, numerous graph similarity and distance measures have been explored. But in fact, there is still a lack of a mathematical framework to explore interrelations of these

measures. Suppose let $d_1 : \mathcal{G} \times \mathcal{G} \rightarrow \mathbb{R}_+$ and $d_2 : \mathcal{G} \times \mathcal{G} \rightarrow \mathbb{R}_+$ be two comparative graph measures (i.e., graph similarity or distance measures) which are defined on the graph class \mathcal{G} . Typical questions in this idea group would be to prove interrelations of the measures by means of inequalities such as $d_1 \underset{(>)}{<} d_2$. For instance, inequalities involving graph complexity measures have been inferred by Dehmer et al. [17,18].

The main contribution of this paper is to infer interrelations of graph distance measures. To the best of our knowledge, this problem has not been tackled so far when using graph distance measures. However, interrelations of topological indices interpreted as complexity measures have been studied, see [7,19,20,17,18]. For instance, Bonchev and his co-workers investigated interrelations of branching measures by means of inequalities [7,19,20]. Dehmer [17] examined relations between information-theoretic measures which are based on information functionals and between classical and parametric graph entropies [18]. We here put the emphasis on graph distance measures which are based on so-called topological indices. These measures themselves have not yet been studied. Note that we only consider distance measures (without loss of generality) as they can be easily transformed into graph similarity measures [21]. In order to define these measures concrete, we employ an existing distance measure (see Eq. (6)) and the well-known Randić index [22], the Wiener index [23], eigenvalue-based measures [24], and graph entropies [17,25]. Also, we discuss quality aspects of the measures and state conjectures evidenced by numerical results.

Methods and Results

Topological Indices and Preliminaries

In this section, we introduce the topological indices which are used in the paper. A topological index [23] is a graph invariant, defined by

$$I : \mathcal{G} \rightarrow \mathbb{R}_+. \tag{1}$$

Simple invariants are for instance the number of vertices, the number of edges, vertex degrees, degree sequences, the matching number, the chromatic number and so forth, see [26].

We emphasize that topological indices are graph invariants which characterize its topology. They have been used for examining quantitative structure-activity relationships (QSARs) extensively in which the biological activity or other properties of molecules are correlated with their chemical structures [27]. Topological graph measures have also been applied in ecology [28], biology [29] and in network physics [30,31]. Note that various properties of topological graph measures such as their uniqueness and correlation ability have been examined too [32,33].

Suppose $G=(V,E)$ is a connected graph. The distance between the vertices u and v of G is denoted by $d(u,v)$. The Wiener index of G is denoted by $W(G)$ and defined by

$$W(G) = \sum_{u,v \subseteq V} d(u,v). \tag{2}$$

The name Wiener index or Wiener number for the quantity defined is common in the chemical literature, since Wiener [34] in 1947 seems was the first who considered it. For more results on the Wiener index of trees, we refer to [35].

In 1975, Randić [36] proposed the topological index R (R_{-1} and $R_{-\frac{1}{2}}$) by using the name *branching index* or *connectivity index*, suitable for measuring the extent of branching of the carbon-atom skeleton of saturated hydrocarbons. Nowadays this index is also called the Randić index. In 1998, Bollobás and Erdős [37] generalized this index by replacing $-\frac{1}{2}$ by any real number α , which is called the general Randić index. In fact, the Randić index and the general Randić index became the most popular and most frequently employed structure descriptors used in structural chemistry [38]. For a graph $G=(V,E)$, the Randić index $R(G)$ of G has been defined as the sum of $(d(u)d(v))^{-1/2}$ over all edges uv of G , i.e.,

$$R(G) = \sum_{uv \in E} \frac{1}{\sqrt{d(u)d(v)}}, \tag{3}$$

where $d(u)$ is degree of a vertex u of G . The zeroth-order Randić index due to Kier and Hall [6] is

$${}^0R(G) = \sum_{u \in V(G)} \frac{1}{\sqrt{d(u)}}. \tag{4}$$

For more results on the Randić index and the zeroth-order Randić index, we refer to [39,22,38].

For a given graph G with n vertices, $\lambda_1, \lambda_2, \dots, \lambda_n$ are the eigenvalues of G . The *energy* of a graph G , denoted by $\mathcal{E}(G)$, has been defined by

$$\mathcal{E}(G) = \sum_{i=1}^n |\lambda_i|, \tag{5}$$

due to Gutman in 1977 [40]. For more results on the graph energy, we refer to [41,24,42].

Novel Graph Distance Measures

Now we define the distance measure [21]

$$d(x,y) = 1 - e^{-\left(\frac{x-y}{\sigma}\right)^2}, \tag{6}$$

which is a mapping $d : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}_+$. Obviously it holds $d(x,x)=0$, $d(x,y) \geq 0$, and $d(x,y)=d(y,x)$. In order to translate this concept to graphs, we employ topological indices and obtain

$$d_I(G,H) := d(I(G),I(H)) = 1 - e^{-\left(\frac{I(G)-I(H)}{\sigma}\right)^2} \tag{7}$$

Further we infer a relation between the maximum value of d_I and the extremal values of I .

Observation 1. *Let \mathcal{G} be a class of graphs. Suppose $G,H \in \mathcal{G}$, then G,H are the two graphs attaining the maximum value of d_I if and only if G,H are the graphs attaining the maximum and minimum value of I , respectively.*

Proof. Let $x=|I(G)-I(H)|$, then d_I is a monotone increasing function on x . Therefore, the maximum value of d_I is attained if and only if the maximum value of $|I(G)-I(H)|$ is attained. \square

From Observation 1 and some existing extremal results of topological indices, we obtain some sharp upper bounds of d_I for some classes of graphs. As an example, we list some of those results for trees.

Theorem 1. *Let T and T' be two trees with n vertices. Denote by S_n and P_n the star graph and path graph with n vertices, respectively.*

- (i). The maximum value of $d_W(T,T')$ is attained when T and T' are S_n and P_n , respectively.
- (ii). The maximum value of $d_R(T,T')$ is attained when T and T' are S_n and P_n , respectively.
- (iii). The maximum value of $d_{oR}(T,T')$ is attained when T and T' are S_n and P_n , respectively.
- (iv). The maximum value of $d_{\mathcal{E}}(T,T')$ is attained when T and T' are S_n and P_n , respectively.

Interrelations of Graph Distance Measures

Observe that $0 \leq d_I(G,H) < 1$, which implies that $0 \leq e^{-\left(\frac{I(G)-I(H)}{\sigma}\right)^2} < 1$. Some trivial properties of d_I are as follows. Let \mathcal{G} be a class of graphs and $G,H \in \mathcal{G}$. We get

$$d_I(G,G) = d_I(H,H) = 0; \tag{8}$$

$$d_I(G,H) \geq d_I(G,G) = 0; \tag{9}$$

$$d_I(G,H) = d_I(H,G). \tag{10}$$

However, d_I is not a metric graph distance measure, since the triangle inequality $d_I(G,H) + d_I(H,K) \geq d_I(G,K)$ for $G, H, K \in \mathcal{G}$, does not hold generally. Actually, we obtain a modified version of the triangle inequality.

Theorem 2. *Let I be a topological index. Let \mathcal{G} be a class of graphs and $G, H, K \in \mathcal{G}$. If*

$$|I(G) - I(K)| \leq |I(G) - I(H)| \quad \text{or} \quad |I(G) - I(K)| \leq |I(H) - I(K)|, \tag{11}$$

then we have $d_I(G,H) + d_I(H,K) \geq d_I(G,K)$.

Proof. We now suppose $|I(G) - I(K)| \leq |I(G) - I(H)|$, since the proof of the other case is similar.

From the inequality $|I(G) - I(K)| \leq |I(G) - I(H)|$, we get

$$e^{-\frac{(I(G)-I(K))^2}{\sigma^2}} \geq e^{-\frac{(I(G)-I(H))^2}{\sigma^2}}. \tag{12}$$

Since $e^{-\frac{(I(H)-I(K))^2}{\sigma^2}} \leq 1$, together with Eq. (12), we have

$$1 + e^{-\frac{(I(G)-I(K))^2}{\sigma^2}} \geq e^{-\frac{(I(G)-I(H))^2}{\sigma^2}} + e^{-\frac{(I(H)-I(K))^2}{\sigma^2}}. \tag{13}$$

Therefore, we have the following inequality,

$$1 - e^{-\frac{(I(G)-I(H))^2}{\sigma^2}} + 1 - e^{-\frac{(I(H)-I(K))^2}{\sigma^2}} - (1 - e^{-\frac{(I(G)-I(K))^2}{\sigma^2}}) \geq 0, \tag{14}$$

i.e., $d_I(G,H) + d_I(H,K) \geq d_I(G,K)$. \square

We emphasize if the Inequalities 11 are satisfied, the modified triangle inequality holds. In practice, the triangle inequality may not be absolutely necessary (e.g., for clustering and classification problems) and is often required to prove properties of the measures.

Theorem 3. *Let I_1 and I_2 be two topological indices. Let \mathcal{G} be a class of graphs and $G, H \in \mathcal{G}$. If*

$$|I_1(G) - I_1(H)| \geq \alpha |I_2(G) - I_2(H)|, \tag{15}$$

then

$$1 - d_{I_1}(G,H) \leq (1 - d_{I_2}(G,H))^{\alpha^2}, \tag{16}$$

where $\alpha > 0$ is a constant.

Proof. Since

$$|I_1(G) - I_1(H)| \geq \alpha |I_2(G) - I_2(H)|, \tag{17}$$

we obtain

$$(I_1(G) - I_1(H))^2 \geq \alpha^2 (I_2(G) - I_2(H))^2. \tag{18}$$

Thus

$$1 - e^{-\frac{(I_1(G)-I_1(H))^2}{\sigma^2}} \geq 1 - e^{-\frac{\alpha^2(I_2(G)-I_2(H))^2}{\sigma^2}} = 1 - \left(e^{-\frac{(I_2(G)-I_2(H))^2}{\sigma^2}} \right)^{\alpha^2}, \tag{19}$$

i.e.,

$$1 - e^{-\frac{(I_1(G)-I_1(H))^2}{\sigma^2}} \geq 1 - \left(1 - (1 - e^{-\frac{(I_2(G)-I_2(H))^2}{\sigma^2}}) \right)^{\alpha^2}. \tag{20}$$

Thus,

$$d_{I_1}(G,H) \geq 1 - \left(1 - d_{I_2}(G,H) \right)^{\alpha^2}. \tag{21}$$

The proof is complete. \square

Suppose I_3 is also a topological index. Then if

$$|I_1(G) - I_1(H)| \leq \beta |I_3(G) - I_3(H)|, \tag{22}$$

we derive similarly

$$1 - d_{I_1}(G,H) \geq (1 - d_{I_3}(G,H))^{\beta^2}, \tag{23}$$

where $\beta > 0$ is a constant. Therefore, we obtain the following theorem.

Theorem 4. *Let I_1, I_2 and I_3 be three topological indices. Let \mathcal{G} be a class of graphs and $G, H \in \mathcal{G}$. If*

$$\beta |I_3(G) - I_3(H)| \geq |I_1(G) - I_1(H)| \geq \alpha |I_2(G) - I_2(H)|, \tag{24}$$

then we infer

$$(1 - d_{I_3}(G,H))^{\beta^2} \leq 1 - d_{I_1}(G,H) \leq (1 - d_{I_2}(G,H))^{\alpha^2}, \tag{25}$$

where $\alpha, \beta > 0$ are constants.

Theorem 5. *Let I_1 and I_2 be two topological indices. Let \mathcal{G} be a class of graphs and $G, H \in \mathcal{G}$. If*

$$|I_1(G) - I_1(H)| \geq |I_2(G) - I_2(H)| + \alpha, \tag{26}$$

then we get

$$1 - d_{I_1}(G, H) \leq e^{-\frac{\alpha^2}{\sigma^2}} e^{-\frac{-2\alpha\sqrt{-\ln(1-d_{I_2}(G,H))}}{\sigma}} (1 - d_{I_2}(G, H)), \tag{27}$$

where $\alpha > 0$ is a constant.

Proof. Since

$$|I_1(G) - I_1(H)| \geq |I_2(G) - I_2(H)| + \alpha, \tag{28}$$

we infer

$$(I_1(G) - I_1(H))^2 \geq (I_2(G) - I_2(H))^2 + 2\alpha|I_2(G) - I_2(H)| + \alpha^2. \tag{29}$$

And therefore,

$$1 - e^{-\frac{(I_1(G)-I_1(H))^2}{\sigma^2}} \geq 1 - e^{-\frac{(I_2(G)-I_2(H))^2 + 2\alpha|I_2(G)-I_2(H)| + \alpha^2}{\sigma^2}} \tag{30}$$

$$= 1 - e^{-\frac{\alpha^2}{\sigma^2}} e^{-\frac{2\alpha|I_2(G)-I_2(H)|}{\sigma^2}} e^{-\frac{(I_2(G)-I_2(H))^2}{\sigma^2}} \tag{31}$$

$$= 1 - e^{-\frac{\alpha^2}{\sigma^2}} e^{-\frac{2\alpha|I_2(G)-I_2(H)|}{\sigma^2}} \left(1 - \left(1 - e^{-\frac{(I_2(G)-I_2(H))^2}{\sigma^2}} \right) \right). \tag{32}$$

Hence,

$$1 - d_{I_1}(G, H) \leq e^{-\frac{\alpha^2}{\sigma^2}} e^{-\frac{2\alpha|I_2(G)-I_2(H)|}{\sigma^2}} (1 - d_{I_2}(G, H)). \tag{33}$$

From the definition of d_{I_2} , i.e.,

$$d_{I_2}(G, H) = 1 - e^{-\left(\frac{I_2(G)-I_2(H)}{\sigma}\right)^2}, \tag{34}$$

we obtain that

$$|I_2(G) - I_2(H)| = \sigma\sqrt{-\ln(1 - d_{I_2}(G, H))}. \tag{35}$$

Finally, by substituting (35) into (33), we get the desired result. \square

Suppose I_3 is also a topological index. Then if

$$|I_1(G) - I_1(H)| \leq |I_3(G) - I_3(H)| + \beta, \tag{36}$$

we have

$$1 - d_{I_1}(G, H) \geq e^{-\frac{\beta^2}{\sigma^2}} e^{-\frac{-2\beta\sqrt{-\ln(1-d_{I_3}(G,H))}}{\sigma}} (1 - d_{I_3}(G, H)), \tag{37}$$

where $\beta > 0$ is a constant. Therefore, we obtain the following theorem.

Theorem 6. Let I_1 , I_2 and I_3 be three topological indices. Let \mathcal{G} be a class of graphs and $G, H \in \mathcal{G}$. If

$$|I_3(G) - I_3(H)| + \beta \geq |I_1(G) - I_1(H)| \geq |I_2(G) - I_2(H)| + \alpha, \tag{38}$$

then we have

$$1 - d_{I_1}(G, H) \geq e^{-\frac{\beta^2}{\sigma^2}} e^{-\frac{-2\beta\sqrt{-\ln(1-d_{I_3}(G,H))}}{\sigma}} (1 - d_{I_3}(G, H)) \tag{39}$$

and

$$1 - d_{I_1}(G, H) \leq e^{-\frac{\alpha^2}{\sigma^2}} e^{-\frac{-2\alpha\sqrt{-\ln(1-d_{I_2}(G,H))}}{\sigma}} (1 - d_{I_2}(G, H)), \tag{40}$$

where $\alpha, \beta > 0$ are constants.

Theorem 7. Let I_1 , I_2 and I_3 be three topological indices. Let \mathcal{G} be a class of graphs and $G, H \in \mathcal{G}$. If

$$|I_1(G) - I_1(H)| \geq |I_2(G) - I_2(H)| \cdot |I_3(G) - I_3(H)|, \tag{41}$$

then we infer

$$\ln(1 - d_{I_1}(G, H)) \leq -\sigma^2 \ln(1 - d_{I_2}(G, H)) \cdot \ln(1 - d_{I_3}(G, H)). \tag{42}$$

Proof. Since

$$|I_1(G) - I_1(H)| \geq |I_2(G) - I_2(H)| \cdot |I_3(G) - I_3(H)|, \tag{43}$$

we derive

$$(I_1(G) - I_1(H))^2 \geq |I_2(G) - I_2(H)|^2 \cdot |I_3(G) - I_3(H)|^2. \tag{44}$$

And therefore,

$$\begin{aligned} 1 - e^{-\frac{(I_1(G)-I_1(H))^2}{\sigma^2}} &\geq 1 - e^{-\frac{(I_2(G)-I_2(H))^2 \cdot (I_3(G)-I_3(H))^2}{\sigma^2}} \\ &\geq 1 - \left(e^{-\frac{(I_3(G)-I_3(H))^2}{\sigma^2}} \right)^{(I_2(G)-I_2(H))^2}, \end{aligned} \tag{45}$$

i.e., $d_{I_1}(G,H) \geq 1 - (1 - d_{I_3}(G,H))^{(I_2(G) - I_2(H))^2}$. Hence we obtain

$$1 - d_{I_1}(G,H) \leq (1 - d_{I_3}(G,H))^{(I_2(G) - I_2(H))^2}, \quad (46)$$

which implies that

$$\ln(1 - d_{I_1}(G,H)) \leq (I_2(G) - I_2(H))^2 \cdot \ln(1 - d_{I_3}(G,H)). \quad (47)$$

By substituting (35) into (47), we easily obtain the assertion of the theorem. \square

By performing a similar proof as in Theorem 7, we obtain a more general result.

Theorem 8. Let $I, I_1, I_2, I_3, \dots, I_k$ be topological indices. Let \mathcal{G} be a class of graphs and $G, H \in \mathcal{G}$. If

$$|I(G) - I(H)| \geq \prod_{j=1}^k |I_j(G) - I_j(H)|, \quad (48)$$

we infer

$$\ln(1 - d_I(G,H)) \leq (-1)^{k-1} \sigma^{2k-2} \prod_{j=1}^k \ln(1 - d_{I_j}(G,H)). \quad (49)$$

Theorem 9. Let I_1, I_2 and I_3 be three topological indices. Let \mathcal{G} be a class of graphs and $G, H \in \mathcal{G}$. If

$$|I_1(G) - I_1(H)| = c_1 |I_2(G) - I_2(H)| + c_2 |I_3(G) - I_3(H)|, \quad (50)$$

where $c_1, c_2 > 0$, then we get

$$d_{I_1}(G,H) = 1 - (1 - d_{I_2}(G,H))^{c_1^2} \cdot (1 - d_{I_3}(G,H))^{c_2^2} \cdot e^{-2c_1 c_2 \sqrt{\ln(1 - d_{I_2}(G,H)) \ln(1 - d_{I_3}(G,H))}}. \quad (51)$$

Proof. Since

$$|I_1(G) - I_1(H)| = c_1 |I_2(G) - I_2(H)| + c_2 |I_3(G) - I_3(H)|, \quad (52)$$

we derive

$$|I_1(G) - I_1(H)|^2 = c_1^2 |I_2(G) - I_2(H)|^2 + c_2^2 |I_3(G) - I_3(H)|^2 + 2c_1 c_2 |I_2(G) - I_2(H)| |I_3(G) - I_3(H)|. \quad (53)$$

Therefore,

$$1 - e^{-\frac{(I_1(G) - I_1(H))^2}{\sigma^2}} = \frac{c_1^2 |I_2(G) - I_2(H)|^2 + c_2^2 |I_3(G) - I_3(H)|^2 + 2c_1 c_2 |I_2(G) - I_2(H)| |I_3(G) - I_3(H)|}{\sigma^2} \quad (54)$$

$$= 1 - e^{-\frac{c_1^2 (I_2(G) - I_2(H))^2}{\sigma^2}} e^{-\frac{c_2^2 (I_3(G) - I_3(H))^2}{\sigma^2}} e^{-\frac{2c_1 c_2 |I_2(G) - I_2(H)| |I_3(G) - I_3(H)|}{\sigma^2}}, \quad (55)$$

which implies

$$d_{I_1}(G,H) = 1 - (1 - d_{I_2}(G,H))^{c_1^2} (1 - d_{I_3}(G,H))^{c_2^2} e^{-\frac{2c_1 c_2 |I_2(G) - I_2(H)| |I_3(G) - I_3(H)|}{\sigma^2}}. \quad (56)$$

By applying the substitutions

$$|I_2(G) - I_2(H)| = \sigma \sqrt{-\ln(1 - d_{I_2}(G,H))} \quad (57)$$

and

$$|I_3(G) - I_3(H)| = \sigma \sqrt{-\ln(1 - d_{I_3}(G,H))} \quad (58)$$

into (56), we obtain the final result. \square

By performing a similar proof as in Theorem 9, we obtain a more general result again.

Theorem 10. Let $I, I_1, I_2, I_3, \dots, I_k$ be topological indices. Let \mathcal{G} be a class of graphs and $G, H \in \mathcal{G}$. If

$$|I(G) - I(H)| = \sum_{j=1}^k c_j |I_j(G) - I_j(H)|, \quad (59)$$

where $c_j > 0$ for $1 \leq j \leq k$, then we infer

$$d_I(G,H) = 1 - \prod_{j=1}^k (1 - d_{I_j}(G,H))^{c_j^2} \cdot \prod_{a,b \in [k], a \neq b} e^{-2c_a c_b \sqrt{\ln(1 - d_{I_a}(G,H)) \ln(1 - d_{I_b}(G,H))}}. \quad (60)$$

Graph Distance Measures Based on Randić Index

In this section, we consider the values of the graph distance measure based on the Randić index and other topological indices for some classes of graphs. Denote by W and R the Wiener index and Randić index, respectively.

Theorem 11. Let \mathcal{G} be a class of regular graphs with n vertices and I is an arbitrary topological index. For two graphs $G, H \in \mathcal{G}$, we infer

$$d_I(G,H) \geq d_R(G,H) = 0. \quad (61)$$

Proof. Let G and H be two regular graphs of order n . By the definition of the Randić index, we obtain that $R(G) = R(H) = \frac{n}{2}$, which implies that $R(G) - R(H) = 0$. Therefore, we infer $d_R(G,H) = 0$. Since $d_I(G,H) \geq 0$ for any topological index, then we obtain the desired inequality.

By using the definition of the zeroth-order Randić index for two graphs with the same degree sequences, we obtain that ${}^0R(G) = {}^0R(H)$. Therefore, we get the following theorem.

Theorem 12. *Let \mathcal{G} be a class of graphs with the same degree sequences and I is an arbitrary topological index. Then for two graphs $G, H \in \mathcal{G}$, we infer*

$$d_I(G, H) \geq d_{0R}(G, H) = 0. \tag{62}$$

For a given graph G of order n , we get $\sqrt{n-1} \leq R(G) \leq \frac{n}{2}$ (see [39]). Thus,

$$|R(G) - R(H)| \leq \frac{n}{2} - \sqrt{n-1}. \tag{63}$$

From (63), we infer an upper bound for $d_R(G, H)$.

Theorem 13. *Let G and H be two connected graphs of order n . Then we get*

$$d_R(G, H) \leq 1 - e^{-\frac{n^2 + 4n - 4 - 4n\sqrt{n-1}}{4\sigma^2}}. \tag{64}$$

The equality holds if and only if G and G' are S_n and a regular graph, respectively.

A path $P = x_0x_1x_2 \dots x_k$ is *pendent* if $d(x_0) \geq 3$, $d(x_k) = 1$ and $d(x_i) = 2$ for all $1 \leq i \leq k$. Especially, a vertex v is *pendent* if $d(v) = 1$. Suppose u and v are two pendent vertices, and u' the unique neighbor of u . We define an operation as follows: deleting the edge uu' and adding the edge uv . We call this operation “transfer u to v ”.

Theorem 14. *Let $G = (V, E)$ be a graph with n vertices. Denote by P_1 and P_2 the two pendent paths attaching to the same vertex such that $|P_1| \geq |P_2| \geq 1$. Denote by H the graph obtained by transferring the pendent vertex of P_2 to the pendent vertex of P_1 . Then we have*

$$d_W(G, H) > d_R(G, H). \tag{65}$$

Proof. Let $G = (V, E)$ be a graph with n vertices. Suppose $P_1 = uu_1u_2 \dots u_a$ and $P_2 = uv_1v_2 \dots v_b$ with $a \geq b \geq 1$. Since P_1 and P_2 are two pendent paths attaching to the same vertex, then we get

$$n - a - b \geq 2. \tag{66}$$

By using the definition of H , we infer $H = G - v_{b-1}v_b + u_av_b$. By using the definition of d_I , we only need to show

$$|W(G) - W(H)| > |R(G) - R(H)|. \tag{67}$$

Observe that $V(G) = V(H) = V$. We will discuss the difference of the distances between two vertices in G and H . Let x and y be two vertices of G . If $x, y \in V \setminus \{v_b\}$, then we have $d_H(x, y) = d_G(x, y)$. Now we suppose $x = v_b$. If $y \notin V(P_1) \cup V(P_2)$, then

$$d_H(x, y) - d_G(x, y) = (a + 1) - b = a - b + 1. \tag{68}$$

Observe that

$$\begin{aligned} d_G(v_b, u) + \sum_{i=1}^{b-1} d_G(v_i) + \sum_{j=1}^a d_G(u_j) = \\ d_H(v_b, u) + \sum_{i=1}^{b-1} d_H(v_i) + \sum_{j=1}^a d_H(u_j). \end{aligned} \tag{69}$$

Therefore, we have

$$\begin{aligned} W(H) - W(G) = \sum_{y \notin V(P_1) \cup V(P_2)} (d_H(x, y) - d_G(x, y)) = \\ (n - a - b - 1)(a - b + 1) > 0, \end{aligned} \tag{70}$$

i.e.,

$$|W(G) - W(H)| = (n - a - b - 1)(a - b + 1). \tag{71}$$

For $b \geq 3$, it is easy to verify $R(G) = R(H)$. Therefore $|W(G) - W(H)| > |R(G) - R(H)|$ holds.

For $b = 2$, from (66), we have $2 \leq a \leq n - 4$ and $|W(G) - W(H)| = (n - a - 3)(a - 1)$. By performing some elementary calculations, we get

$$(n - a - 3)(a - 1) > \frac{n}{2} - \sqrt{n-1}, \tag{72}$$

i.e.,

$$a^2 - (n - 2)a + \frac{3n}{2} - \sqrt{n-1} - 3 < 0 \tag{73}$$

for $2 \leq a \leq n - 4$ and each value of n . Therefore, from (63), we infer $|W(G) - W(H)| > |R(G) - R(H)|$.

For $b = 1$, from (66), we have $1 \leq a \leq n - 3$ and $|W(G) - W(H)| = a(n - a - 2)$. By performing some elementary calculations, we obtain

$$a(n - a - 2) > \frac{n}{2} - \sqrt{n-1}, \tag{74}$$

i.e.,

$$a^2 - (n - 2)a + \frac{n}{2} - \sqrt{n-1} < 0 \tag{75}$$

for $1 \leq a \leq n - 3$ and each value of n . Therefore, from (63), we infer $|W(G) - W(H)| > |R(G) - R(H)|$. The proof is complete. \square

This theorem can be used to compare the values of the distance measure by using trees. Let \mathcal{T}_n be the set of trees with n vertices and

$$T_n^0 = \{T^0 : T^0 \in \mathcal{T}_n \text{ and each pendent path of } T^0 \text{ has length one}\}. \tag{76}$$

Observe that for every $T \in \mathcal{T}_n \setminus T_n^0$, there must be a tree $T^0 \in T_n^0$ such that T can be obtained from T^0 by repeatedly transferring pendent vertices. Therefore, we obtain the following corollary.

Corollary 1. *Let $T \in \mathcal{T}_n \setminus T_n^0$, there exists a tree $T^0 \in T_n^0$ such that $d_W(T, T^0) > d_R(T, T^0)$.*

Actually, numerical experiments show that for any two trees $T, T' \in \mathcal{T}_n$, the inequality $d_W(T, T') > d_R(T, T')$ holds. We state the result as a conjecture.

Conjecture 1. *Let T and T' be any two trees with n vertices. Then*

$$d_W(T, T') \geq d_R(T, T') \tag{77}$$

holds.

As an example, we consider (all) 23 trees with 8 vertices and calculate all possible values of $d_W(T, T')$ (blue) and $d_R(T, T')$ (red) as shown in Figure 1. From Figure 1, we observe that $d_W(T, T') \geq d_R(T, T')$ holds for each pair of trees T and T' .

Graph Distance Measures Based on Graph Entropy

In this section, we consider graph distance measures which are based on graph entropy and other topological indices for some classes of graphs.

In order to start, we reproduce the definition of Shannon's entropy [43]. Let $p = (p_1, p_2, \dots, p_n)$ be a probability vector, namely, $0 \leq p_i \leq 1$ and $\sum_{i=1}^n p_i = 1$. The Shannon's entropy of p has been defined by

$$I(p) = - \sum_{i=1}^n p_i \log p_i. \tag{78}$$

We denote by d_{Ip} the graph distance measure based on $I(p)$.

In the following, we infer an upper bound for $d_{Ip}(G, H)$.

Theorem 15. *Let G and H be two graphs with the same vertex set. Denote by $p = (p_1, p_2, \dots, p_n)$ and $p' = (p'_1, p'_2, \dots, p'_n)$ be the probability vectors of G and H , respectively. If $p_i \leq p'_i$ for each i , then we infer*

$$d_{Ip}(G, H) < 1 - e^{-\frac{A^2}{\sigma^2}}, \tag{79}$$

where $A = \sum_{i=1}^n \left(p'_i \log \left(1 + \frac{1}{p_i} \right) + \log(p'_i + 1) \right)$.

Proof. Since $p_i \leq p'_i$ for each i , then we obtain $p_i < p'_i + 1$ and $\log p_i < \log(p'_i + 1)$. Then we have

$$p_i \log p_i < (p'_i + 1) \log(p'_i + 1) \tag{80}$$

$$= p'_i \log(p'_i + 1) + \log(p'_i + 1) \tag{81}$$

$$= p'_i \log \left(p'_i \left(1 + \frac{1}{p_i} \right) \right) + \log(p'_i + 1) \tag{82}$$

$$= p'_i \log p'_i + p'_i \log \left(1 + \frac{1}{p_i} \right) + \log(p'_i + 1). \tag{83}$$

Therefore, we get the inequality,

$$I(p) = - \sum_{i=1}^n p_i \log p_i > I(p') - \sum_{i=1}^n \left(p'_i \log \left(1 + \frac{1}{p_i} \right) + \log(p'_i + 1) \right) = I(p') - A, \tag{84}$$

i.e., $I(p') - I(p) < A$. Hence,

$$d_{Ip}(G, H) = 1 - e^{-\frac{(I(p) - I(p'))^2}{\sigma^2}} < 1 - e^{-\frac{A^2}{\sigma^2}}. \tag{85}$$

The desired inequality holds. \square

In [25], Dehmer and Mowshowitz generalized the definition of graph entropy by using information functionals. Let $G = (V, E)$ be a connected graph. For a vertex $v_i \in V$, we define

$$p(v_i) := \frac{f(v_i)}{\sum_{j=1}^{|V|} f(v_j)}, \tag{86}$$

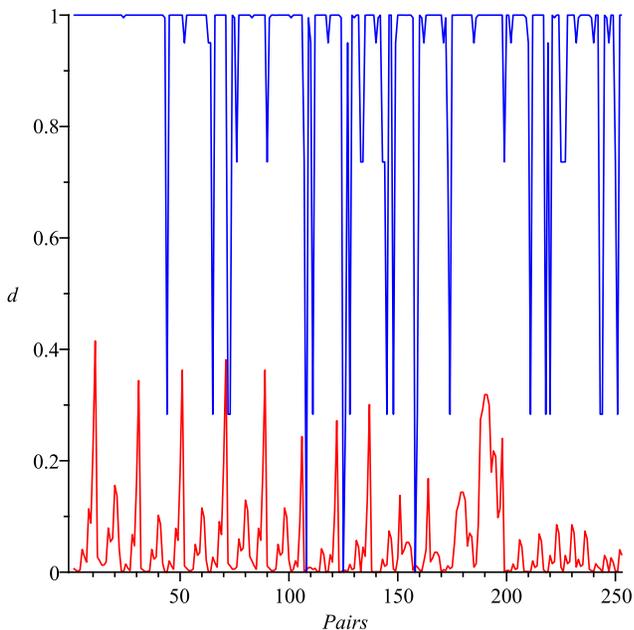


Figure 1. All the values of $d_W(T, T')$ (blue) and $d_R(T, T')$ (red). The Y-axis denotes the values of the distance measure and the X-axis denotes the graph pairs. doi:10.1371/journal.pone.0094985.g001

where f represents an arbitrary information functional. By substituting $p(v_i)$ to (78), we have

$$If(G) = \log\left(\sum_{i=1}^{|V|} f(v_i)\right) - \sum_{i=1}^{|V|} \frac{f(v_i)}{\sum_{j=1}^{|V|} f(v_j)} \log f(v_i). \quad (87)$$

We denote by d_{If} the graph distance measure based on If .

Relations between $d_{\mathcal{E}}(G,H)$ and $d_{I_g}(G,H)$

Denote by $\lambda_1, \lambda_2, \dots, \lambda_n$ the eigenvalues of a graph G . By setting $f(v_i) = |\lambda_i|$ in (87), we obtain a new expression of the graph entropy namely

$$I_g(G) = \log\left(\sum_{i=1}^n |\lambda_i|\right) - \frac{1}{\sum_{i=1}^n |\lambda_i|} \sum_{i=1}^n |\lambda_i| \log |\lambda_i|. \quad (88)$$

Recall that the energy of G is defined as $\mathcal{E}(G) = \sum_{i=1}^n |\lambda_i|$. Then we infer

$$I_g(G) = \log(\mathcal{E}(G)) - \frac{1}{\mathcal{E}(G)} \sum_{i=1}^n |\lambda_i| \log |\lambda_i|. \quad (89)$$

From the definition of $I_g(G)$, it is interesting to investigate the relation between the graph distance measures $d_{\mathcal{E}}$ and d_{I_g} .

Theorem 16. *Let G and H be two graphs of order n with $\mathcal{E}(G) > \mathcal{E}(H)$. Denote by $\lambda_1, \lambda_2, \dots, \lambda_n$ and $\lambda'_1, \lambda'_2, \dots, \lambda'_n$ the eigenvalues of G and H , respectively. Let $\lambda' = \max_{1 \leq i \leq n} \{|\lambda'_i|\}$ and $\lambda = \min_{1 \leq i \leq n} \{|\lambda_i|\}$. Then we get*

$$\ln(1 - d_{I_g}(G,H)) \geq \frac{\ln(1 - d_{\mathcal{E}}(G,H))}{(\xi \ln 2)^2} - \frac{2(\log \frac{\lambda'}{\lambda}) \sqrt{-\ln(1 - d_{\mathcal{E}}(G,H))}}{\sigma \xi \ln 2} - \frac{(\log \frac{\lambda'}{\lambda})^2}{\sigma^2}, \quad (90)$$

where $\xi \in (\mathcal{E}(H), \mathcal{E}(G))$ is a constant.

Proof. Let G and H be two graphs of order n . Let $\mathcal{E} = \mathcal{E}(G)$ and $\mathcal{E}' = \mathcal{E}(H)$ with $\mathcal{E} > \mathcal{E}'$. Then we get

$$I_g(G) - I_g(H) = \left(\log(\mathcal{E}) - \frac{1}{\mathcal{E}} \sum_{i=1}^n |\lambda_i| \log |\lambda_i|\right) - \left(\log(\mathcal{E}') - \frac{1}{\mathcal{E}'} \sum_{i=1}^n |\lambda'_i| \log |\lambda'_i|\right) \quad (91)$$

$$= (\log(\mathcal{E}) - \log(\mathcal{E}')) + \left(\frac{1}{\mathcal{E}'} \sum_{i=1}^n |\lambda'_i| \log |\lambda'_i| - \frac{1}{\mathcal{E}} \sum_{i=1}^n |\lambda_i| \log |\lambda_i|\right) \quad (92)$$

$$\leq (\log(\mathcal{E}) - \log(\mathcal{E}')) + \left(\frac{\log \lambda'}{\mathcal{E}'} \sum_{i=1}^n |\lambda'_i| - \frac{\log \lambda}{\mathcal{E}} \sum_{i=1}^n |\lambda_i|\right) \quad (93)$$

$$= (\mathcal{E} - \mathcal{E}') \cdot \frac{1}{\xi \ln 2} + \log\left(\frac{\lambda'}{\lambda}\right), \quad (94)$$

where $\xi \in (\mathcal{E}', \mathcal{E})$. Thus,

$$d_{I_g}(G,H) = 1 - e^{-\frac{(I_g(G) - I_g(H))^2}{\sigma^2}} \quad (95)$$

$$\leq 1 - e^{-\frac{1}{\sigma^2} \left((\mathcal{E} - \mathcal{E}') \cdot \frac{1}{\xi \ln 2} + \log\left(\frac{\lambda'}{\lambda}\right) \right)^2} \quad (96)$$

$$= 1 - e^{-\frac{(\mathcal{E} - \mathcal{E}')^2}{\sigma^2} \frac{1}{(\xi \ln 2)^2}} \cdot e^{-\frac{\mathcal{E} - \mathcal{E}'}{\sigma^2} \frac{2 \log(\frac{\lambda'}{\lambda})}{\xi \ln 2}} \cdot e^{-\frac{1}{\sigma^2} \left(\log\left(\frac{\lambda'}{\lambda}\right)\right)^2} \quad (97)$$

$$= 1 - (1 - d_{\mathcal{E}}(G,H))^{\frac{1}{(\xi \ln 2)^2}} \cdot e^{-\frac{2(\frac{\lambda'}{\lambda})}{\sigma \xi \ln 2} \sqrt{-\ln(1 - d_{\mathcal{E}}(G,H))}} \cdot e^{-\frac{(\frac{\lambda'}{\lambda})^2}{\sigma^2}}, \quad (98)$$

i.e.,

$$1 - d_{I_g}(G,H) \geq (1 - d_{\mathcal{E}}(G,H))^{\frac{1}{(\xi \ln 2)^2}} \cdot e^{-\frac{2(\frac{\lambda'}{\lambda})}{\sigma \xi \ln 2} \sqrt{-\ln(1 - d_{\mathcal{E}}(G,H))}} \cdot e^{-\frac{(\frac{\lambda'}{\lambda})^2}{\sigma^2}}. \quad (99)$$

Taking logarithm for the two sides of the above inequality, we have

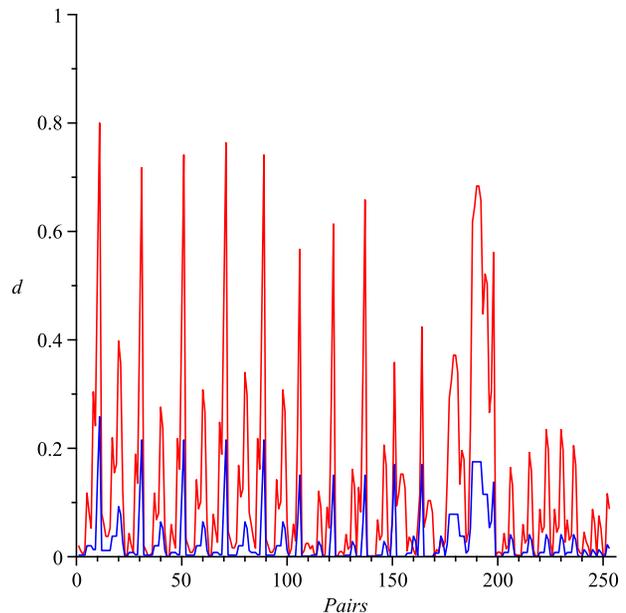


Figure 2. Values of $d_{\mathcal{E}}(T, T')$ (red) and $d_{I_g}(T, T')$ (blue). The Y-axis denotes the values of the distance measure and the X-axis denotes the graph pairs. doi:10.1371/journal.pone.0094985.g002

$$\ln(1 - d_{I_g}(G, H)) \geq \frac{\ln(1 - d_{\mathcal{E}}(G, H))}{(\xi \ln 2)^2} - \frac{2(\log \frac{\lambda'}{\lambda})\sqrt{-\ln(1 - d_{\mathcal{E}}(G, H))}}{\sigma \xi \ln 2} - \frac{(\log \frac{\lambda'}{\lambda})^2}{\sigma^2} \quad (100)$$

The required inequality holds. \square
 Actually, numerical experiments show that for any two distinct trees $T, T' \in \mathcal{T}_n$, $d_{\mathcal{E}}(T, T') \geq d_{I_g}(T, T')$ holds. See Figure 2 as an example, in which we consider (all) 23 trees with 8 vertices and calculate all possible values of $d_{\mathcal{E}}(T, T')$ (red) and $d_{I_g}(T, T')$ (blue). We state this observation as a conjecture.

Conjecture 2. *Let T and T' be any two distinct trees with n vertices. Then*

$$d_{\mathcal{E}}(T, T') \geq d_{I_g}(T, T') \quad (101)$$

holds.

Using a similar proof method of Theorem 16, we can obtain a generalization for the distance measure based on I_f (see Eq. (87)). Let f be an arbitrary information functional and $f(G) = \sum_{i=1}^n f(v_i)$ be a topological index.

Theorem 17. *Let G and H be two graphs of order n with $f(G) > f(H)$. Let $\lambda' = \max_{1 \leq i \leq n} \{f(v_i) : v_i \in G\}$ and $\lambda = \min_{1 \leq j \leq n} \{f(v_j) : v_j \in H\}$. Then we have*

$$\ln(1 - d_{I_f}(G, H)) \geq \frac{\ln(1 - d_f(G, H))}{(\eta \ln 2)^2} - \frac{2(\log \frac{\lambda'}{\lambda})\sqrt{-\ln(1 - d_f(G, H))}}{\sigma \eta \ln 2} - \frac{(\log \frac{\lambda'}{\lambda})^2}{\sigma^2} \quad (102)$$

where $\eta \in (f(H), f(G))$ is a constant.

Dehmer and Mowshowitz [44] introduced a new class of measures (called here generalized measures) that derive from functions such as those defined by Rényi's entropy and Daróczy's entropy. Let G be a graph of order n . Then

$$I_1(G) := \sum_{i=1}^n \frac{f(v_i)}{\sum_{j=1}^n f(v_j)} \left[1 - \frac{f(v_i)}{\sum_{j=1}^n f(v_j)} \right]. \quad (103)$$

If we let $f(v_i) = |\lambda_i|$, then we can obtain the new generalized entropy based on eigenvalues. We denote the entropy by

$$I_{g_1}(G) := \sum_{i=1}^n \frac{|\lambda_i|}{\sum_{j=1}^n |\lambda_j|} \left[1 - \frac{|\lambda_i|}{\sum_{j=1}^n |\lambda_j|} \right]. \quad (104)$$

For a given graph $G = (V, E)$ with n vertices, denote by $\lambda_1, \lambda_2, \dots, \lambda_n$ the eigenvalues of G . By substituting $\mathcal{E} = \sum_{i=1}^n |\lambda_i|$ into equality (104), we have

$$I_{g_1}(G) = \sum_{i=1}^n \frac{|\lambda_i|}{\sum_{j=1}^n |\lambda_j|} \left[1 - \frac{|\lambda_i|}{\sum_{j=1}^n |\lambda_j|} \right] \quad (105)$$

$$= \sum_{i=1}^n \frac{|\lambda_i|}{\mathcal{E}} \left(1 - \frac{|\lambda_i|}{\mathcal{E}} \right) = \frac{1}{\mathcal{E}^2} \sum_{i=1}^n |\lambda_i|(\mathcal{E} - |\lambda_i|) \quad (106)$$

$$= \frac{1}{\mathcal{E}^2} \left(\mathcal{E}^2 - \sum_{i=1}^n |\lambda_i|^2 \right) = 1 - \frac{2m}{\mathcal{E}^2}. \quad (107)$$

The last equality holds since $\sum_{i=1}^n \lambda_i^2 = 2m$. By the following theorem, we study the relation between $d_{\mathcal{E}}$ and $d_{I_{g_1}}$.

Theorem 18. *Let \mathcal{G} be a class of graphs with n vertices and m edges. For two graphs $G, H \in \mathcal{G}$, let $\mathcal{E} = \mathcal{E}(G)$ and $\mathcal{E}' = \mathcal{E}(H)$. Then we get*

$$d_{\mathcal{E}}(G, H) > d_{I_{g_1}}(G, H) \quad (108)$$

and

$$d_{\mathcal{E}}(G, H) \leq d_{I_{g_1}}(G, H) + \left(1 - \frac{16m^2}{\mathcal{E}^3 \mathcal{E}'^3} \right) \frac{(\mathcal{E} - \mathcal{E}')^2}{\sigma^2} e^{\alpha}, \quad (109)$$

where $\alpha \in (1 - \frac{16m^2}{\mathcal{E}^3 \mathcal{E}'^3}, 1)$ is a constant.

Proof. Let G and H be two graphs with n vertices and m edges. Without loss of generality, we suppose $\mathcal{E} \geq \mathcal{E}'$.

To show the first inequality, it suffices to prove

$$|\mathcal{E} - \mathcal{E}'| > |I_{g_1}(G) - I_{g_1}(H)|. \quad (110)$$

Then from (107), we derive

$$I_{g_1}(G) - I_{g_1}(H) = -\frac{2m}{\mathcal{E}^2} + \frac{2m}{\mathcal{E}'^2} = \frac{2m(\mathcal{E}^2 - \mathcal{E}'^2)}{\mathcal{E}^2 \mathcal{E}'^2}. \quad (111)$$

If we want to prove

$$\mathcal{E} - \mathcal{E}' \geq \frac{2m(\mathcal{E}^2 - \mathcal{E}'^2)}{\mathcal{E}^2 \mathcal{E}'^2}, \quad (112)$$

we only need to show

$$2m(\mathcal{E} + \mathcal{E}') \leq \mathcal{E}^2 \mathcal{E}'^2. \quad (113)$$

From a well-known bound of energy $\mathcal{E}' \geq 2\sqrt{m}$, we have $\mathcal{E}'^2 > 2m$ and $\mathcal{E}^2 > (\mathcal{E} + \mathcal{E}')$. Therefore, $d_{\mathcal{E}}(G, H) > d_{I_{g_1}}(G, H)$ holds.

Now we show the second inequality. From (111), we have

$$|\mathcal{E} - \mathcal{E}'| - |I_{g_1}(G) - I_{g_1}(H)| = \mathcal{E} - \mathcal{E}' - \frac{2m(\mathcal{E}^2 - \mathcal{E}'^2)}{\mathcal{E}^2 \mathcal{E}'^2} \quad (114)$$

$$= (\mathcal{E} - \mathcal{E}') \left(1 - \frac{2m(\mathcal{E} + \mathcal{E}')}{\mathcal{E}^2 \mathcal{E}'^2} \right) \tag{115}$$

$$\leq (\mathcal{E} - \mathcal{E}') \left(1 - \frac{2m \cdot 2\sqrt{\mathcal{E}\mathcal{E}'}}{\mathcal{E}^2 \mathcal{E}'^2} \right) \tag{116}$$

$$= (\mathcal{E} - \mathcal{E}') \left(1 - \frac{4m}{\mathcal{E}^{3/2} \mathcal{E}'^{3/2}} \right). \tag{117}$$

Therefore, we have

$$|I_{g_1}(G) - I_{g_1}(H)| \geq (\mathcal{E} - \mathcal{E}') \frac{4m}{(\mathcal{E}\mathcal{E}')^{3/2}}.$$

From the definition of the distance measure, by some elementary calculations, we finally infer

$$d_{\mathcal{E}}(G, H) - d_{I_{g_1}}(G, H) = \left(1 - e^{-\frac{(\mathcal{E} - \mathcal{E}')^2}{\sigma^2}} \right) - \left(1 - e^{-\frac{(I_{g_1}(G) - I_{g_1}(H))^2}{\sigma^2}} \right) \tag{118}$$

$$= e^{-\frac{(I_{g_1}(G) - I_{g_1}(H))^2}{\sigma^2}} - e^{-\frac{(\mathcal{E} - \mathcal{E}')^2}{\sigma^2}} \tag{119}$$

$$\leq e^{-\frac{(\mathcal{E} - \mathcal{E}')^2}{\sigma^2} \frac{16m^2}{(\mathcal{E}\mathcal{E}')^3}} - e^{-\frac{(\mathcal{E} - \mathcal{E}')^2}{\sigma^2}} \tag{120}$$

$$= \left(1 - \frac{16m^2}{\mathcal{E}^3 \mathcal{E}'^3} \right) \frac{(\mathcal{E} - \mathcal{E}')^2}{\sigma^2} e^{\alpha}, \tag{121}$$

where $\alpha \in (1 - \frac{16m^2}{\mathcal{E}^3 \mathcal{E}'^3}, 1)$ is a constant.

The proof is complete. \square

Relations between $d_I(G, H)$ and $d_{I_f}(G, H)$

Let $G = (V, E)$ be a connected graph with n vertices, m edges and degree sequence (d_1, d_2, \dots, d_n) , where $d_i = d(v_i)$ for $1 \leq i \leq n$. By setting $f(v_i) = d_i^k$ in (87), we can obtain the new entropy based on degree powers, denoted by $I_f^k(G)$

$$I_f^k(G) = \log \left(\sum_{i=1}^n d_i^k \right) - \frac{1}{\sum_{i=1}^n d_i^k} \sum_{i=1}^n d_i^k \log d_i^k. \tag{122}$$

For $k = -1/2$, the expression $\sum_{i=1}^n f(v_i) = \sum_{i=1}^n \frac{1}{\sqrt{d_i}}$ is just the zeroth-order Randić index ${}^0R(G)$. Then by using Theorem 17, we obtain the following result.

Theorem 19. Let G and H be two graphs of order n with ${}^0R(G) > {}^0R(H)$. Let

$$\lambda' = \max_{1 \leq i \leq n} \left\{ \frac{1}{\sqrt{d(v_i)}} : v_i \in G \right\} \text{ and } \lambda = \min_{1 \leq j \leq n} \left\{ \frac{1}{\sqrt{d(v_j)}} : v_j \in H \right\}. \tag{123}$$

Then we have

$$\ln(1 - d_{I_f^{-1/2}}(G, H)) \geq \frac{\ln(1 - d_{0_R}(G, H))}{(\eta \ln 2)^2} - \frac{2(\log \frac{\lambda'}{\lambda}) \sqrt{-\ln(1 - d_{0_R}(G, H))}}{\sigma \eta \ln 2} - \frac{(\log \frac{\lambda'}{\lambda})^2}{\sigma^2}, \tag{124}$$

where $\eta \in (f(H), f(G))$ is a constant.

For $k = 1$, we get

$$I_f^1(G) = \log(2m) - \frac{1}{2m} \sum_{i=1}^n d_i \log d_i. \tag{125}$$

Furthermore, by the definition of $I_f^k(G)$, for two graphs with the same degree sequences, we obtain that $I_f^k(G) = I_f^k(H)$. Therefore, we get the following result.

Theorem 20. Let \mathcal{G} be a class of graphs with the same degree sequences and I is an arbitrary topological index. Then for two graphs $G, H \in \mathcal{G}$, we infer

$$d_I(G, H) \geq d_{0_R}(G, H) = d_{I_f^k}(G, H) = 0. \tag{126}$$

By using the similar proof method applied in Theorem 14, we obtain a weaker result.

Theorem 21. Let $T = (V, E)$ be a tree with n vertices. Denote by P_1 and P_2 two pendent paths attaching to the same vertex such that $|P_1| \geq |P_2| \geq 1$. Denote by T' the tree obtained by transferring the pendent vertex of P_2 to the pendent vertex of P_1 . Then we have

$$d_W(T, T') > d_{I_f^1}(T, T') \text{ and } d_R(T, T') > d_{I_f^1}(T, T'). \tag{127}$$

Proof. Let $T = (V, E)$ be a tree with n vertices. Suppose $P_1 = uu_1u_2 \dots u_a$ and $P_2 = uv_1v_2 \dots v_b$ with $a \geq b \geq 1$. Denote by x the degree of u , i.e., $x = d(u)$. Since P_1 and P_2 are two pendent paths attaching to the same vertex, then we have $n - a - b \geq 2$. By using the definition of T' , we have $T' = T - v_{b-1}v_b + u_a v_b$. By using the definition of d_I , we only need to show

$$\begin{aligned} |W(T) - W(T')| &> |I_f^1(T) - I_f^1(T')| \text{ and} \\ |R(T) - R(T')| &> |I_f^1(T) - I_f^1(T')|. \end{aligned} \tag{128}$$

For a tree T with n vertices, we get $I_f^1(S_n) \leq I_f^1(T) \leq I_f^1(P_n)$. By performing elementary calculations, we get

$$|I_f^1(T) - I_f^1(T')| \leq I_f^1(P_n) - I_f^1(S_n) = \frac{\log(n-1)}{2} - \frac{n-3}{n-2}. \tag{129}$$

Observe that $V(T) = V(T') = V$. We first discuss the difference of the distances between two vertices in T and T' . Let x and y be

two vertices of T . If $x, y \in V \setminus \{v_b\}$, then we have $d_{T'}(x, y) = d_T(x, y)$. Now we suppose $x = v_b$. If $y \notin V(P_1) \cup V(P_2)$, then $d_{T'}(x, y) - d_T(x, y) = (a + 1) - b = a - b + 1$. Observe that

$$d_T(v_b, u) + \sum_{i=1}^{b-1} d_T(v_i) + \sum_{j=1}^a d_T(u_j) = d_{T'}(v_b, u) + \sum_{i=1}^{b-1} d_{T'}(v_i) + \sum_{j=1}^a d_{T'}(u_j). \tag{130}$$

Therefore, we get $|W(T) - W(T')| = (n - a - b - 1)(a - b + 1)$.

For $b \geq 2$, it is easy to verify that $If_1(T) = If_1(T')$, i.e., $|If_1(T) - If_1(T')| = 0$. Then,

$$\begin{aligned} |W(T) - W(T')| &> |If_1(T) - If_1(T')| \quad \text{and} \\ |R(T) - R(T')| &> |If_1(T) - If_1(T')|. \end{aligned} \tag{131}$$

In the following, we suppose $b = 1$.

We obtain $1 \leq a \leq n - 3$ and $|W(T) - W(T')| = a(n - a - 2)$. By performing elementary calculations, we get

$$a(n - a - 2) > \frac{\log(n - 1)}{2} - \frac{n - 3}{n - 2}, \tag{132}$$

for $1 \leq a \leq n - 3$ and each value of n . Therefore, $|W(T) - W(T')| > |If_1(T) - If_1(T')|$.

To prove the other inequality, we need more detailed discussion. By using the definition of graph entropy, we get

$$|If_1(T) - If_1(T')| = \frac{1}{2(n - 1)} [x \log x - (x - 1) \log(x - 1) - 2]. \tag{133}$$

Let S be the set of the neighbors of vertex u , which does not contain u_1 and v_1 . Denote by d_j the degree of a vertex in S , where $j = 1, 2, \dots, x - 3$. If $a = 1$, then

$$|R(T) - R(T')| = \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2(x - 1)}} - \frac{2}{\sqrt{x}} + \left(\frac{1}{\sqrt{x - 1}} - \frac{1}{\sqrt{x}} \right) \sum_{j=1}^{x-3} \frac{1}{\sqrt{d_j}}. \tag{134}$$

By performing some calculations, we can show that for $x \geq 3$ and $n \geq 9$,

$$\begin{aligned} \frac{1}{\sqrt{2}} + \frac{1}{\sqrt{2(x - 1)}} - \frac{2}{\sqrt{x}} &> \\ \frac{1}{2(n - 1)} [x \log x - (x - 1) \log(x - 1) - 2], \end{aligned} \tag{135}$$

i.e., $|R(T) - R(T')| > |If_1(T) - If_1(T')|$ for $n \geq 9$. For smaller n , we verify this inequality directly. If $a \geq 2$, then we have

$$|R(T) - R(T')| = \frac{1}{2} + \frac{1}{\sqrt{2(x - 1)}} - \frac{1}{\sqrt{x}} - \frac{1}{\sqrt{2x}} + \left(\frac{1}{\sqrt{x - 1}} - \frac{1}{\sqrt{x}} \right) \sum_{j=1}^{x-3} \frac{1}{\sqrt{d_j}}. \tag{136}$$

We can show that for $x \geq 4$ and $n \geq 13$,

$$\begin{aligned} \frac{1}{2} + \frac{1}{\sqrt{2(x - 1)}} - \frac{1}{\sqrt{x}} - \frac{1}{\sqrt{2x}} &> \\ \frac{1}{2(n - 1)} [x \log x - (x - 1) \log(x - 1) - 2], \end{aligned} \tag{137}$$

i.e., $|R(T) - R(T')| > |If_1(T) - If_1(T')|$ for $n \geq 13$. For smaller n , we verify this inequality directly. Now suppose $x = 3$, then there is only one vertex in S whose degree is at most $n - 4$. Therefore by using (133) and (136), we get

$$|If_1(T) - If_1(T')| = \frac{1}{2(n - 1)} (3 \log 3 - 4) \tag{138}$$

and

$$|R(T) - R(T')| > 1 - \frac{1}{\sqrt{3}} - \frac{1}{\sqrt{6}} + \left(\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{3}} \right) \frac{1}{\sqrt{n - 4}}. \tag{139}$$

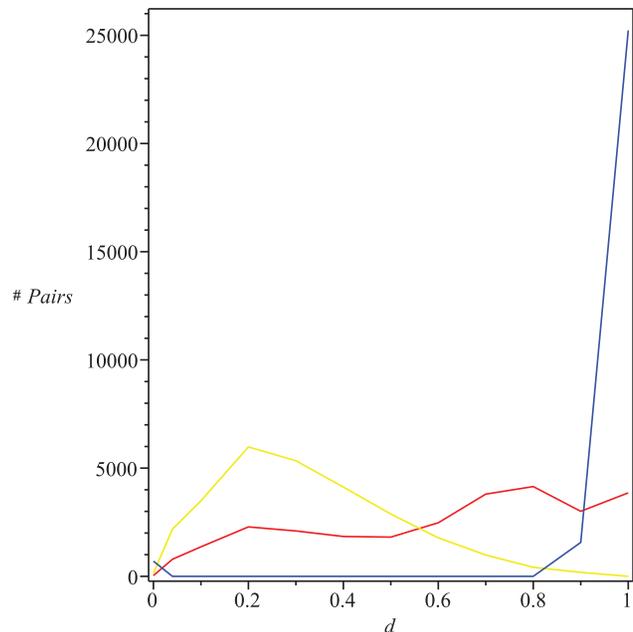


Figure 3. Values of $d_{If_1}(T, T')$ (blue) and $d_R(T, T')$ (red). The Y-axis denotes the values of the distance measure and the X-axis denotes the graph pairs. doi:10.1371/journal.pone.0094985.g003

We can verify

$$1 - \frac{1}{\sqrt{3}} - \frac{1}{\sqrt{6}} + \left(\frac{1}{\sqrt{2}} - \frac{1}{\sqrt{3}}\right) \frac{1}{\sqrt{n-4}} > \frac{1}{2(n-1)}(3 \log 3 - 4) \quad (140)$$

for each n , i.e., $|R(T) - R(T')| > |If_1(T) - If_1(T')|$. □

From Theorem 14 and 21, we obtain the following corollary.

Corollary 2. Let $T = (V, E)$ be a tree with n vertices. Denote by P_1 and P_2 the two pendent paths attaching to the same vertex such that $|P_1| \geq |P_2| \geq 1$. Denote by T' the tree obtained by transferring the pendent vertex of P_2 to the pendent vertex of P_1 . Then we have

$$d_W(T, T') > d_R(T, T') > d_{If_1}(T, T'). \quad (141)$$

Therefore, we obtain a similar result to comparing the values of distance measures of trees.

Corollary 3. Let $T \in \mathcal{T}_n \setminus \mathcal{T}_n^0$, there exists a tree $T^0 \in \mathcal{T}_n^0$ such that $d_W(T, T^0) > d_R(T, T^0) > d_{If_1}(T, T^0)$.

Actually, our numerical results (see section ‘Numerical Results’) show that for any two trees $T, T' \in \mathcal{T}_n$, the following inequality may hold.

Conjecture 3. Let T and T' be any two trees with n vertices. Then

$$d_W(T, T') \geq d_R(T, T') \geq d_{If_1}(T, T') \quad (142)$$

holds.

By way of example, we consider all 23 trees of 8 vertices and calculate all possible values of $d_{If_1}(T, T')$ (blue) and $d_R(T, T')$ (red), respectively, as shown in Figure 3. From Figure 3, we observe that

$$d_{If_1}(T, T') \leq d_R(T, T') \quad (143)$$

holds for each pair of trees T and T' .

Numerical Results

In this section, we interpret the numerical results. First, we consider all trees with 8 vertices. The number of trees is 23 and the number of pairs is 253 (see [45]). From the curves shown by Figure 1, we see that both measures $d_W(T, T')$ (blue) and $d_R(T, T')$ (red) satisfy the inequality Eq. (77). From the curves shown by Figure 2, we observe that both measures $d_E(T, T')$ (red) and $d_{I_E}(T, T')$ (blue) satisfy the inequality Eq. (101). From the curves shown by Figure 3, we also learn that both measures $d_{If_1}(T, T')$ (blue) and $d_R(T, T')$ (red) fulfill the inequality Eq. (143). By using this method, several other inequalities could be generated and verified graphically.

Figures 4 and 5 show the numerical results by using the graph distance measures based on graph energy \mathcal{E} , the Wiener index W and the Randić index R , respectively. We consider all trees with 11 vertices. The number of trees is 235 and the number of pairs is 27495 (see [45]). By Figure 4, we depict the distributions of the ranked distance values, that is, d_E (red), d_W (blue), and d_R (yellow). First and foremost, we see that the measured values of all three measures cover the entire interval $[0, 1]$. This indicates that the measures are generally useful as they are well defined. By

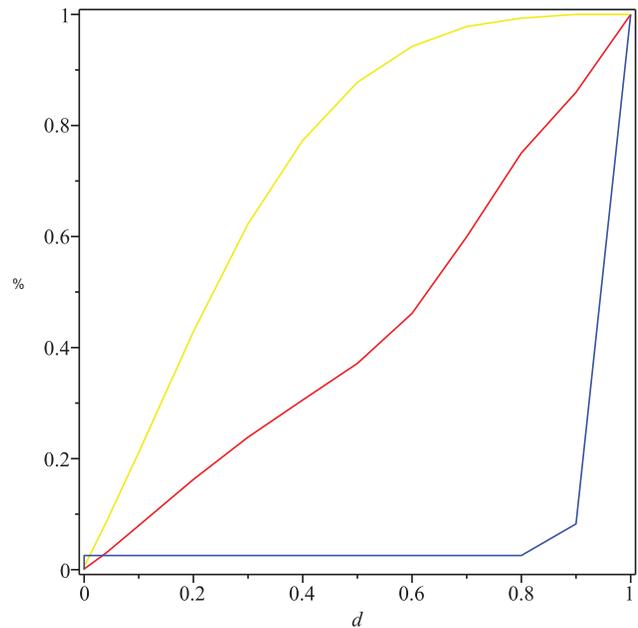


Figure 4. Distributions of the ranked values of the distance measure d_E (red), d_W (blue), d_R (yellow). The X-axis denotes the values of the distance measure. The Y-axis denotes the number of graph pairs. doi:10.1371/journal.pone.0094985.g004

considering d_W , we observe that only a relatively little number of pairs have a measured value ≤ 0.8 . But a large number of pairs possess distance values ≥ 0.8 . When considering d_R , the situation is reverse. The distance values of d_E seem to slightly increase with some up- and downturns. However, Figure 4 does not comment on the ability of the graph distance measures to classify graphs

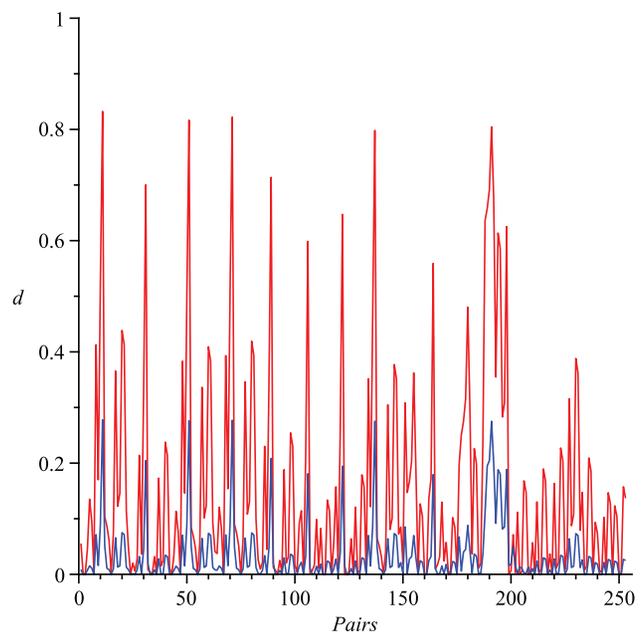


Figure 5. The X-axis denotes the values of the distance measures d_E (red), d_W (blue), d_R (yellow). The Y-axis represents the percentage rate of all graphs studied. doi:10.1371/journal.pone.0094985.g005

efficiently. This needs to be examined in the future and would far beyond the scope of this paper.

Furthermore, we have computed the cumulative distributions by using the measures d_E (red), d_W (blue), d_R (yellow), respectively, as shown in Figure 5. In general, the computation of the cumulative distribution may serve as a preprocessing step when analyzing graphs structurally. In fact, we see how many percent of the 235 graphs have a distance value which is less or equal d . Also, Figure 5 shows that the value distributions are quite different. From Figure 5, we see that the curve for d_W strongly differs from d_E and d_R . When considering d_R , we also observe that about 80% of the 235 trees have a distance value approximately ≤ 0.5 . That means most of the trees are quite dissimilar according to d_R . For d_W , the situation is absolutely reverse. Here 80% of the trees have a distance value approximately ≤ 0.98 . Finally evaluating the graph distance measure d_E on these trees reveals that about 80% of the trees possess a distance value approximately ≤ 0.85 . In summary, we conclude from Figure 5 that all three measures capture the distance between the graphs quite differently. But nevertheless, this does not imply that the quality of one measure may be worse than another. Again, an important issue of quality is fulfilled as the measures turned out to be well defined, see Figure 4. Another crucial issue would be evaluating the classification ability which is future work.

Summary and Conclusion

In this paper, we have studied interrelations of graph distance measures which are based on distinct topological indices. In order to do so, we employed the Wiener index, the Randić index, the zeroth-order Randić index, the graph energy, and certain graph entropies [25]. In particular, we have obtained inequalities involving the novel graph distance measures. Evidenced by a numerical analysis we also found three conjectures dealing with relations between the distance measures on trees.

From Theorem 1, we see that the star graph and the path graph maximize d_I among all trees with a given number of vertices, for any topological index we considered here. Actually, this also holds for some other topological indices, such as the Hosoya index [46,47], the Merrifield-Simmons index [48,49,47], the Estrada index [50,51,52], and the Szeged index [53,54]. All other

References

- Dehmer M, Mehler A (2007) A new method of measuring similarity for a special class of directed graphs. *Tatra Mountains Mathematical Publications* 36: 39–59.
- Sobik F (1982) Graphmetriken und klassifikation strukturierter objekte, ZKI-informationen. *Akad Wiss DDR* 2: 63–122.
- Zelinka B (1975) On a certain distance between isomorphism classes of graphs. *Časopis propest Matematiky* 100: 371–373.
- Emmert-Streib F (2007) The chronic fatigue syndrome: A comparative pathway analysis. *J Comput Biology* 14: 961–972.
- Junker B, Schreiber B (2008) *Analysis of Biological Networks*. Wiley-Interscience, Berlin.
- Kier L, Hall L (2002) The meaning of molecular connectivity: A bimolecular accessibility model. *Croat Chem Acta* 75: 371–382.
- Bonchev D, Trinajstić N (1977) Information theory, distance matrix and molecular branching. *J Chem Phys* 67: 4517–4533.
- Skvortsova M, Baskin I, Stankevich I, Palyulin V, Zefirov N (1998) Molecular similarity. 1. analytical description of the set of graph similarity measures. *J Chem Inf Comput Sci* 38: 785–790.
- Varmuza K, Scsibrany H (2000) Substructure isomorphism matrix. *J Chem Inf Comput Sci* 40: 308–313.
- Mehler A, Weiß P, Lücking A (2010) A network model of interpersonal alignment. *Entropy* 12: 1440–1483.
- Hsieh S, Hsu C (2008) Graph-based representation for similarity retrieval of symbolic images. *Data Knowl Eng* 65: 401–418.
- Dehmer M, Mehler A (2007) A new method of measuring similarity for a special class of directed graphs. *Tatra Mountains Mathematical Publications* 36: 39–59.
- Bunke H (2000) Recent developments in graph matching. In: 15-th International Conference on Pattern Recognition. pp. 117–124.

theorems we have proved in this paper shed light on the problem of proving interrelations of the measures. We believe that such statements help to understand the measures more thoroughly and, finally, they are useful to establish new applications employing quantitative graph theory [55]. We emphasize that the star graph and the path graph are apparently the two most dissimilar trees among all trees. Similar observations can also be obtained for unicyclic graphs or bicyclic graphs. Therefore, in the future, we would like to explore which classes of graphs have this property, i.e., identifying graphs (such as the path graph and the star graph) which maximize or minimize d_I .

Another direction for future work is to compare the values of $d_I(G, G')$ where G, G' are general graphs. For example, we could assume that G and G' are obtained by only one graph edit operation, i.e., $\text{GED}(G, G') = 1$, see [15]. Then, all the graph which fulfill this equation are (by definition) similar. This construction could help to study the sensitivity of the measures thoroughly. Note that similar properties of topological indices have already been investigated, see [56]. As a conclusive remark, we mention that dynamics models on spatial graphs have been studied by Perc and Wang and other researchers, see [57,58]. It would be interesting to study the distance measures in this mathematical framework as well.

Supporting Information

Supporting Information S1 CSV file containing descriptor values of 235 trees by using the Randić index.
(CSV)

Supporting Information S2 CSV file containing descriptor values of 235 trees by using graph energy.
(CSV)

Supporting Information S3 CSV file containing descriptor values of 235 trees by using the Wiener index.
(CSV)

Author Contributions

Wrote the paper: MD YS FES.

24. Gutman I, Li X, Zhang J (2009) Graph energy. In: Dehmer M, Emmert-Streib F, editors, *Analysis of Complex Networks. From Biology to Linguistics*, Wiley-VCH, pp. 145–174. Weinheim.
25. Dehmer M, Mowshowitz A (2011) A history of graph entropy measures. *Inform Sciences* 181: 57–78.
26. Harary F (1969) *Graph Theory*. Addison Wesley Publishing Company, Reading, MA, USA.
27. Dehmer M, Varmuza K, Bonchev D, editors (2012) *Statistical Modelling of Molecular Descriptors in QSAR/QSPR*. Quantitative and Network Biology. Wiley-Blackwell.
28. Ulanowicz RE (2001) Information theory in ecology. *Computers and Chemistry* 25: 393–399.
29. Emmert-Streib F, Dehmer M (2011) Networks for systems biology: Conceptual connection of data and function. *IET Systems Biology* 5: 185–207.
30. Wilhelm T, Hollunder J (2007) Information theoretic description of networks. *Physica A* 388: 385–396.
31. Solé RV, Valverde S (2004) Information theory of complex networks: On evolution and architectural constraints. In: *Lecture Notes in Physics*, volume 650, pp. 189–207.
32. Bonchev D, Mekenyan O, Trinajstić N (1981) Isomer discrimination by topological information approach. *J Comp Chem* 2: 127–148.
33. Dehmer M, Emmert-Streib F, Tripathi S (2013) Large-scale evaluation of molecular descriptors by means of clustering. *PLoS ONE* 8: e83956.
34. Wiener H (1947) Structural determination of paran boiling points. *J Amer Chem Soc* 69: 17–20.
35. Dobrynin A, Entringer R, Gutman I (2001) Wiener index of trees: theory and application. *Acta Appl Math* 66: 211–249.
36. Randić M (1975) On characterization of molecular branching. *J Amer Chem Soc* 97: 6609–6615.
37. Bollobás B, Erdős P (1998) Graphs of extremal weights. *Ars Combin* 50: 225–233.
38. Li X, Gutman I (2006) *Mathematical Aspects of Randić-Type Molecular Structure Descriptors*. University of Kragujevac and Faculty of Science Kragujevac.
39. Li X, Shi Y (2008) A survey on the randić index. *MATCH Commun Math Comput Chem* 59: 127–156.
40. Gutman I (1977) Acyclic systems with extremal hückel π -electron energy. *Theor Chim Acta* 45: 79–87.
41. Gutman I (2001) The energy of a graph: old and new results. In: Betten A, Kohnert A, Laue R, Wassermann A, editors, *Algebraic Combinatorics and Applications*, Springer-Verlag, pp. 196–211.
42. Li X, Shi Y, Gutman I (2012) *Graph Energy*. Springer, New York.
43. Shannon C, Weaver W (1949) *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, USA.
44. Dehmer M, Mowshowitz A (2011) Generalized graph entropies. *Complexity* 17: 45–50.
45. Read R, Wilson R (1998) *An Atlas of Graphs*. Clarendon Press, Oxford.
46. Hosoya H (1971) Topological index. a newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. *Bull Chem Soc Jpn* 4: 2332–2339.
47. Wagner S, Gutman I (2010) Maxima and minima of the hosoya index and the merrifield-simmons index: A survey of results and techniques. *Acta Appl Math* 112: 323–346.
48. Merrifield R, Simmons H (1980) The structure of molecular topological spaces. *Theor Chim Acta* 55: 55–75.
49. Merrifield R, Simmons H (1989) *Topological Methods in Chemistry*. Wiley, New York.
50. Estrada E (2000) Characterization of 3d molecular structure. *Chem Phys Lett* 319: 713–718.
51. Deng H, Radenković S, Gutman I (2009) The estrada index. In: Cvetković D, Gutman I, editors, *Applications of Graph Spectra*, Math. Inst. pp. 123–140. Belgrade.
52. Gutman I, Deng H, Radenković S (2011) The estrada index: An updated survey. In: Cvetković D, Gutman I, editors (2011) *Selected Topics on Applications of Graph Spectra*, Math. Inst. pp. 155–174.
53. Gutman I (1994) A formula for the wiener number of trees and its extension to graphs containing cycles. *Graph Theory Notes of New York* 27: 9–15.
54. Simić S, Gutman I, Baltić V (2000) Some graphs with extremal sezed index. *Math Slovaca* 50: 1–15.
55. Dehmer M, Emmert-Streib F (2014) *Quantitative Graph Theory. Theory and Applications*. CRC Press. In press.
56. Furtula B, Gutman I, Dehmer M (2013) On structure-sensitivity of degree-based topological indices. *Applied Mathematics and Computation* 219: 8973–8978.
57. Perc M, Wang Z (2010) Heterogeneous aspirations promote cooperation in the prisoners dilemma game. *PLoS ONE* 5: e15117.
58. Jin Q, Wang L, Xia C, Wang Z (2014) Spontaneous symmetry breaking in interdependent networked game. *Scientific Reports* 4: 4095.