Efficient Graph Similarity Assessment Method Based on Vectors of Topological Indices

Matthias Dehmer[a,b,c,d], Izudin Redžepović[e], Niko Tratnik[f,g,*], Petra Žigert Pleteršek[h,f]

[a] *Swiss Distance University of Applied Sciences, Brig, Switzerland*

[b] *Department for Biomedical Computer Science and Mechatronics, UMIT—Private University for Health Sciences, Medical Informatics and Technology, Hall in Tyrol, Austria*

[c] *College of Artificial Intelligence, Nankai University, Tianjin, China*

[d] *School of Science, Xian Technological University, Xian, Shaanxi, China*

[e] *State University of Novi Pazar, Serbia*

[f] *University of Maribor, Faculty of Natural Sciences and Mathematics, Slovenia*

[g] *Institute of Mathematics, Physics and Mechanics, Ljubljana, Slovenia*

[h] *University of Maribor, Faculty of Chemistry and Chemical Engineering, Slovenia*

`matthias.dehmer@ffhs.ch, iredzepovic@np.ac.rs, niko.tratnik@um.si, petra.zigert@um.si`

Abstract

Measuring similarity between complex objects is a fundamental task in many scientific fields. When objects are represented as graphs, graph similarity/distance measures offer a powerful framework for quantifying structural resemblance. Those comparative

---

*Corresponding author.

measures play a key role in domains such as network science, chemo-informatics, and social network analysis. While methods like graph edit distance and graph kernels are widely used, they can be computationally intensive or fail to capture fine structural variations, since they require graphs without any structural uncertainty. Another class of methods is based on using topological indices to encode structural information of the graphs, followed by the application of distance or similarity measures for real numbers to obtain corresponding graph-level metrics. In this paper, we introduce a novel class of distance/similarity measures which are based on multiple topological indices. In many cases, they can be computed in polynomial time, which makes our method computationally efficient in practice. We demonstrate its effectiveness through comparisons and show that it captures subtle structural information meaningfully. Additionally, we explore its applicability in two domains: analyzing random graph models in network theory and assessing molecular similarity among isomers in chemoinformatics. These preliminary results suggest that our approach holds promise for graph comparison across disciplines.

1   Introduction

A natural outcome of observing two objects is the recognition of the degree of similarity/distance between them. However, accurately defining or quantifying similarity/distance is challenging due to its multidimensional nature; it's a well-known fact that the terms similarity/distance cannot uniquely be formalized, see [40].

Given that objects from diverse domains are frequently modeled as graphs, studying graph similarity/distance measures or methods emerges as a meaningful and relevant task. Those comparative measures have been crucial across various fields such as network analysis [15], chemoinformatics [4, 44], social sciences [1] offering a quantitative way to assess structural resemblance between graphs [15, 45].

In the following, we briefly survey the most important measures in the scientific literature. The very first contribution when tackling the problem of measuring similarity/distance on chemical graphs has been due to Sussenguth [41]. Strictly speaking, Sussenguth dealt with exact graph matching methods, i.e., measures which are based on isomorphic relations between graphs. Then Zelinka [46] developed the so-called Zelinka-distance;

this is a isomorphism-based distance measure for graphs with the same number of vertices which takes the size of the largest common induced subgraph of the two given graphs into account. Sobik [39] and Kaden [25] generalized this measure for graphs which have different number of vertices. All the mentioned approaches so far belong to the paradigm of exact graph matching.

A well known graph distance metric in the field of inexact graph matching relates to the graph edit distance (GED) due to Bunke [6]. Applications thereof can be found in [18, 36]. Beyond graph edit distance, a variety of approaches have been developed to compare graphs more efficiently or to better capture structural nuances. For instance, graph kernel methods [27] which encode structural properties via kernel functions.

Another prominent category includes measures based on topological indices or feature vectors [7–10], which encode structural information of the graphs by using numerical descriptors; the next step is to use existing similarity/distance measures for real numbers to derive comparative graph measures, see [15]. Previous methods in this field have mostly relied on a single topological index to compute the distance or similarity between graphs. However, a recent approach [38] incorporates 20 different topological indices to calculate graph similarity, but in this method, some information is lost when converting the vector of indices into binary representation, which decreases its sensitivity. Therefore, in the present paper we aim to develop more sensitive approach that would find applications in various fields.

The structure of the paper is as follows. In the next section, we present our methodology. More precisely, distance and similarity measures based on topological indices are defined. Then, in Section 3 we demonstrate that our measure, on the one hand, predicts graph similarity in a manner comparable to an established metric, while on the other hand, it captures subtle differences between distinct graphs more effectively. At the same time, it is significantly less computationally demanding, making it well-suited for use with larger graphs or large-scale datasets.

In the last two sections, we present preliminary results that demonstrate the potential applicability of our measure in network theory and

chemistry. In Section 4, we analyze the behavior of our measure on several well-known random graph models, which are commonly used to simulate specific structural properties of real-world networks, such as the high clustering coefficient in the Watts–Strogatz model [42] or the presence of high-degree nodes in the Barabási-Albert preferential attachment model [2].

Molecular similarity plays a crucial role in chemoinformatics, which has led to the development of various methods for encoding molecular structure [28]. In Section 5, we apply our approach to assess the similarity within a set of decane isomers.

2 Distance and similarity measures based on topological indices

2.1 Distance and similarity measures

We start by stating some definitions regarding distance and similarity measures as they are understood in this work (see also [9]). We do not require these measures to satisfy additional axioms (such as those of a metric). Also, we show how these measures can be converted into each other.

Definition 1. A function $s : X \times X \to \mathbb{R}$ is a *similarity measure* on the set X if

(1) $0 \leq s(x, y) \leq 1$ for all $x, y \in X$,

(2) $s(x, y) = s(y, x)$ for all $x, y \in X$, and

(3) $s(x, x) = 1$ for all $x \in X$.

Definition 2. A function $d : X \times X \to \mathbb{R}$ is a *distance measure* (or *dissimilarity measure*) on the set X if

(1) $d(x, y) \geq 0$ for all $x, y \in X$,

(2) $d(x, y) = d(y, x)$ for all $x, y \in X$, and

(3) $d(x, x) = 0$ for all $x \in X$.

Additionally, if a distance measure d fulfills also the following two properties:

(4) $d(x, y) = 0 \implies x = y$ for all $x, y \in X$, and

(5) $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$,

then we call d a *distance metric*.

Now, let us show that we can easily get a distance measure from a similarity measure and vice versa. Firstly, we start with an arbitrary similarity measure $s : X \times X \to \mathbb{R}$ due to Definition 1. We show that a function $d : X \times X \to \mathbb{R}$, defined by

$$\forall x, y \in X, \ d(x, y) = 1 - s(x, y),$$

is a distance measure. Obviously, for any $x, y \in X$ it holds $s(x, y) \leq 1$ and therefore $d(x, y) = 1 - s(x, y) \geq 0$. Also, we infer

$$d(x, y) = 1 - s(x, y) = 1 - s(y, x) = d(y, x),$$

hence, the symmetry property is fulfilled. Finally, $d(x, x) = 1 - s(x, x) = 1 - 1 = 0$ for any $x \in X$.

This conversion can be very helpful when we start with a given similarity measure s but we need to argue with distance. Note that from $d(x, y) = 1 - s(x, y)$, we also get $0 \leq d(x, y) \leq 1$.

For the other direction, suppose that d is a distance measure on the set X. Firstly, suppose that there exists some constant $M > 0$ such that $d(x, y) \leq M$ for any $x, y \in X$. Then, define the function $s : X \times X \to \mathbb{R}$ as follows:

$$\forall x, y \in X, \ s(x, y) = \frac{M - d(x, y)}{M}. \tag{1}$$

Choose arbitrary $x, y \in X$. Since $d(x, y) \leq M$, we get $0 \leq M - d(x, y) \leq M$ and therefore

$$0 \leq s(x, y) = \frac{M - d(x, y)}{M} \leq 1.$$

On the other hand,

$$s(x, y) = \frac{M - d(x, y)}{M} = \frac{M - d(y, x)}{M} = s(y, x).$$

Finally, we have

$$s(x, x) = \frac{M - d(x, x)}{M} = \frac{M}{M} = 1.$$

Therefore, we have shown that s is a similarity measure.

However, if a distance measure d is not bounded, then the similarity measure s according to Definition 2 can be defined, for example, by

$$\forall x, y \in X, \ s(x, y) = \frac{1}{1 + d(x, y)}.$$

One can easily check that this is a similarity measure.

2.2 Graph-based distance and similarity measures

All graphs considered in this paper are connected. For a graph G, we denote by $V(G)$ the set of vertices of G and by $E(G)$ the set of edges of G. Moreover, $d(u, v)$ is the usual shortest path *distance* between vertices $u, v \in V(G)$ and $\deg(u)$ denotes the *degree* of a vertex $u \in V(G)$. Additionally, the *eigenvalues* of a graph G are the eigenvalues of the adjacency matrix of G.

Now, we present a new method for computing the similarity between two graphs based on topological indices. Let $I_1, \ldots, I_k$ be topological indices, where $k \geq 3$. For a graph G, let $M(G) = \max\{I_1(G), \ldots, I_k(G)\}$ and $m(G) = \min\{I_1(G), \ldots, I_k(G)\}$. Since topological indices can have very different values, we begin by scaling them to enable meaningful comparison. For any $i \in \{1, \ldots, k\}$, define

$$J_i(G) = \frac{I_i(G) - m}{M - m}.$$

Note that $J_i(G) \in [0, 1]$ for any $i \in \{1, \ldots, k\}$. Then, for a given $p \geq 1$ one can calculate the distance between two graphs, G_1 and G_2, as

$$d_p(G_1, G_2) = d_p\Big(\big(J_1(G_1), \ldots, J_k(G_1)\big), \big(J_1(G_2), \ldots, J_k(G_2)\big)\Big),$$

where d_p is a p-distance on $\mathbb{R}^k$, defined as

$$d_p\big((x_1,\ldots,x_k),(y_1,\ldots,y_k)\big) = \left(\sum_{i=1}^{k}|x_i - y_i|^p\right)^{\frac{1}{p}}$$

for any two $(x_1,\ldots,x_k),(y_1,\ldots,y_k) \in \mathbb{R}^k$.

Note that $0 \le d_p(G_1, G_2) \le k^{\frac{1}{p}}$ since $J_i(G_\ell) \in [0,1]$ for $i \in \{1,\ldots,k\}$ and $\ell \in \{1,2\}$. If $p=1$, one obtains the *Manhattan distance*:

$$d_1(G_1, G_2) = \sum_{i=1}^{k}\big|J_i(G_1) - J_i(G_2)\big|.$$

Moreover, if $p = 2$, we get the *Euclidean distance*:

$$d_2(G_1, G_2) = \sqrt{\sum_{i=1}^{k}(J_i(G_1) - J_i(G_2))^2}.$$

We chose the d_p distance because it is a standard, well-understood measure for comparing vectors, offering both computational efficiency and flexibility through the choice of p. Note that according to Definition 2, d_p is obviously a distance measure for any chosen set of graphs.

Finally, we define the similarity between two graphs G_1 and G_2 by employing formula (1) as follows:

$$s_p(G_1, G_2) = \frac{k^{\frac{1}{p}} - d_p(G_1, G_2)}{k^{\frac{1}{p}}}.$$

Based on the discussion in Section 2.1, it is obvious that s_p is a similarity measure.

We observe that for a fixed k, the time complexity of the proposed similarity measure is asymptotically equal to the time complexity of the most computationally expensive topological index used in its computation.

Of course, the obtained similarity measure depends on the chosen topological indices $I_1,\ldots,I_k$. Therefore, in the selection of topological indices we decided to include one distance-based index, one degree-based index, two distance-degree-based indices, and two eigenvalue-based indices. Let

G be a connected graph on n vertices with eigenvalues $\lambda_1, \ldots, \lambda_n$. The following indices were used:

1. the *Harary index* [24, 31], defined as

$$H(G) = \sum_{\substack{\{u,v\} \subseteq V(G) \\ u \neq v}} \frac{1}{d(u,v)},$$

2. the *Sombor index* [19], defined as

$$SO(G) = \sum_{uv \in E(G)} \sqrt{(\deg(u))^2 + (\deg(v))^2},$$

3. the *degree distance* [13] (also known as the *Schultz index* [37]), defined as

$$DD(G) = \sum_{\{u,v\} \subseteq V(G)} (\deg(u) + \deg(v))d(u,v),$$

4. the *Gutman index* [20], defined as

$$Gut(G) = \sum_{\{u,v\} \subseteq V(G)} \deg(u)\deg(v)d(u,v),$$

5. the *graph energy* [21], defined as

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

6. the *Estrada index* [16], defined as

$$EE(G) = \sum_{i=1}^{n} e^{\lambda_i}.$$

Indices were chosen based on their prevalence in the literature, with degeneracy [26] taken into consideration. Note that degeneracy of a topological index measures its power to distinguish between graphs. For some

results on this topic see [5, 11, 33, 34].

One of the main characteristics of any computational technique is time complexity, so we analyze this property of our method. Let G_1 and G_2 be connected graphs with n_1 and n_2 vertices, respectively. Denote $n = \max\{n_1, n_2\}$. The distances between all pairs of vertices in a connected graph on n vertices can be computed in $\mathcal{O}\left(n^3\right)$ time, for example by using the Floyd–Warshall algorithm. Additionally, the degrees of all vertices can be computed in $\mathcal{O}\left(n^2\right)$ time when the graph is represented by the adjacency matrix. Consequently, the Harary index, the degree-distance, and the Gutman index can be calculated in $\mathcal{O}\left(n^3\right)$ time, while the Sombor index takes $\mathcal{O}\left(n^2\right)$ time.

On the other hand, approximations of the eigenvalues of a graph can be computed, for example, by using the QR algorithm [17], where each iteration requires $\mathcal{O}\left(n^3\right)$ time (however, since the adjacency matrix is symmetric, the time complexity of an individual step could be reduced after transforming the matrix into another form [12, 30]). Therefore, approximating the eigenvalues using the standard QR algorithm takes at most $\mathcal{O}\left(sn^3\right)$ time, where s denotes the number of iterations needed to achieve the desired numerical precision. Consequently, the same upper bound on time complexity applies to the computation of the graph energy, the Estrada index, and the similarity measure $s_p(G_1, G_2)$, where $p \geq 1$ and $I_1, \ldots, I_6$ are the topological indices mentioned above.

In order to test the dependence of our approach on the number of indices used, we expanded the selection of the above mentioned indices with the first Zagreb index [23], the Randić index [32], the resolvent energy [22], and the Wiener index [43]. Therefore, the number of components of vectors in this approach equals $k = 10$. Figure 1 shows the correlation between the similarity s_2 calculated by using 6 indices and the similarity s_2 calculated by using 10 indices for the family of all trees with 7 vertices. We can see that the correlation is very strong ($R = 0.9992$), so the larger number of indices does not significantly affects the similarity results. Therefore, we continue the investigation with only 6 indices.
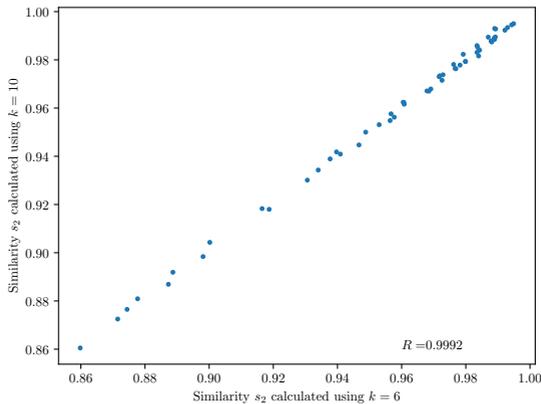
Figure 1. Correlation between similarity s_2 results obtained by $k = 6$ and $k = 10$ for the family of all trees with 7 vertices.

3 Numerical results

In this section, we present several computational results related to graph similarity measures s_1 and s_2 defined in the previous section. We consider the family T_7 of all trees on 7 vertices and the family N_7 of all connected graphs on 7 vertices. Note that T_7 contains 11 trees and N_7 contains 853 graphs. Moreover, if a graph family contains m graphs, then there are $\binom{m}{2}$ pairs to consider. For the family T_7 that means 55 pairs and for the family N_7 that means 363378 pairs of graphs.

We compare similarity measures s_1 and s_2 to a similarity measure obtained by the *graph edit distance* (*GED*). For graphs G_1 and G_2, $GED(G_1, G_2)$ is defined as the minimum number of elementary graph operations that are needed to transform graph G_1 into graph G_2 [36]. This is a widely used measure for quantifying the similarity between two graphs, but computing it is generally NP-hard, making it computationally expensive for large graphs. In [18], a review on the graph edit distance can be found. Note the similarity s_{GED} between graphs G_1 and G_2 can then be defined, for example, as

$$s_{GED} = \frac{1}{GED(G_1, G_2) + 1}.$$

Firstly, we present two figures on which one can see the similarity results for all pairs of graphs, which are ordered according to their similarity, see Figure 2. As already mentioned, the computation of graph edit distance is highly costly, therefore we managed to compute it only for T_7.
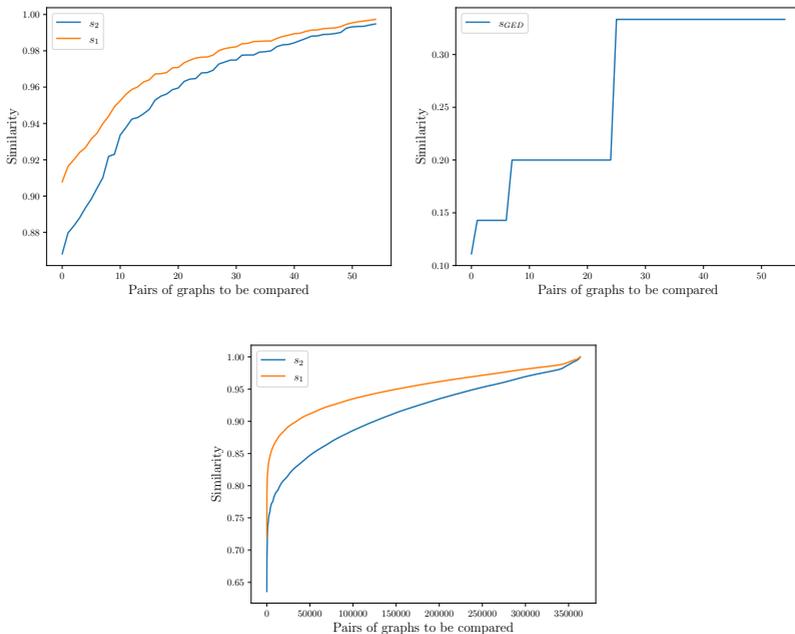
Figure 2. Similarity for all pairs of graphs in families T_7 (first row) and N_7 (second row).

One can observe different behavior of our measures compared to the similarity based on graph edit distance. In particular, s_1 and s_2 distinguish more effectively between pairs of graphs, as shown by the smoothly increasing trends in Figure 2.

Moreover, we examine correlations between similarity measures s_1, s_2, and s_{GED}, see Figure 3. We notice that the results obtained by similarities s_1 and s_2 are comparable. Both of them better discriminate between pairs of graphs in comparison to s_{GED}. However, they show the same trend since with increasing s_{GED} also s_1 and s_2 increase. Based on these findings, we see that it is enough to proceed with s_2.
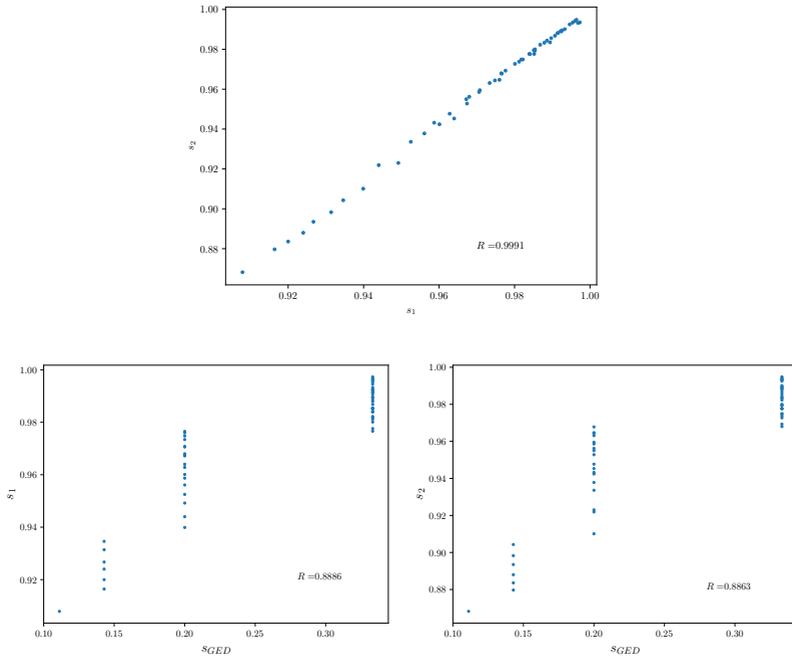
**Figure 3.** Correlations between s_1, s_2 and s_{GED} in case of T_7.

Finally, to evaluate the similarity results, we examine the pairs of graphs with minimum and maximum similarity for both families of graphs. Graphs with the minimum and maximum similarity s_2 in family T_7 are depicted in Figure 4 and Figure 5, respectively. Note that there is only one pair of trees that achieve minimum, while on the other hand there several pairs with the maximum similarity. Such behavior occurs also in the family N_7. Graphs with the minimum and maximum similarity s_2 in this family are depicted in Figure 6 and Figure 7, respectively.

# 4 Application in network theory

In this section, we test our methodology on three different random network models in order to see weather it can discriminate among them. In particular, we use the Erdös-Rényi model, the Barabási-Albert model (generating scale-free networks), and the Watts-Strogatz model (with "small-world"
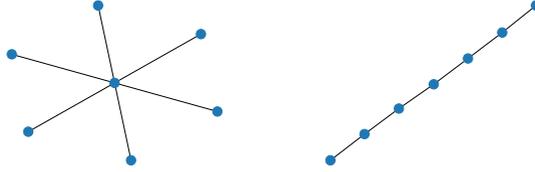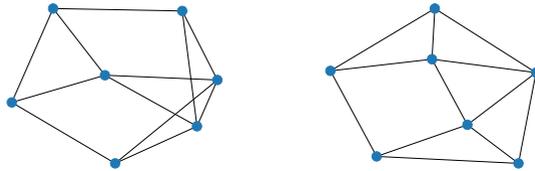
Figure 4. Uniquely determined pair of graphs with the minimum similarity s_2 in family T_7.
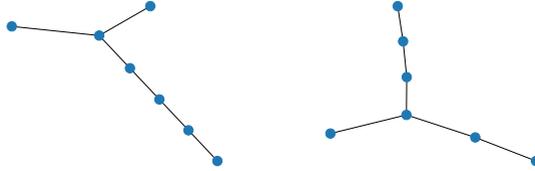
Figure 5. One pair of graphs with the maximum similarity s_2 in family T_7.
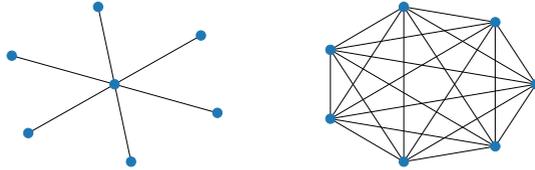
Figure 6. Uniquely determined pair of graphs with the minimum similarity s_2 in family N_7.



Figure 7. One pair of graphs with the maximum similarity s_2 in family N_7.

properties). We selected these three models since they are among the most commonly used models for generating random networks [29].

We generate three graphs, ER_1, ER_2, and ER_3, in the Erdös-Rényi

model using the parameters $n = 100$ (the number of nodes in the graph) and $p = 0.1$ (the probability that an edge exists between any pair of nodes). Next, graphs BA_1, BA_2, and BA_3 are generated by the Barabási-Albert model using the parameters $n = 100$ (the number of nodes in the graph) and $m = 3$ (the number of edges each new node creates when it is added). Finally, graphs WS_1, WS_2, and WS_3 are generated by the Watts-Strogatz model using the parameters $n = 100$ (the number of nodes in the graph), $k = 6$ (each node is adjacent to 6 nearest neighbors in a ring topology), and $p = 0.7$ (the rewiring probability). One example graph for each model is depicted in Figure 8.
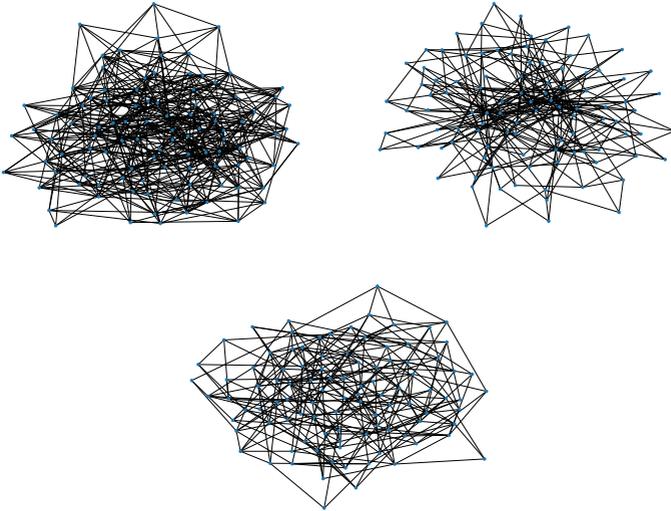
Figure 8. Example networks generated by Erdös-Rényi model (upper left), Barabási-Albert model (upper right), and Watts-Strogatz model (bottom).

Then, we calculated the similarity s_2 for each pair of the considered random networks. Let $\mathcal{RN}$ be the set of all 9 considered networks. Moreover, if

$$m = \min\{s_2(G_1, G_2) \mid G_1, G_2 \in \mathcal{RN}, \ G_1 \neq G_2\} \text{ and}$$
$$M = \max\{s_2(G_1, G_2) \mid G_1, G_2 \in \mathcal{RN}, \ G_1 \neq G_2\},$$

then for any pair of networks G_1, G_2 from $\mathcal{RN}$ we calculate the scaled similarity

$$s_2'(G_1, G_2) = \frac{s_2(G_1, G_2) - m}{M - m}$$

to highlight the differences in the similarities. The results are gathered in Table 1.

	ER_1	ER_2	ER_3	BA_1	BA_2	BA_3	WS_1	WS_2
ER_2	0.80							
ER_3	0.95	0.75						
BA_1	0.11	0.00	0.13					
BA_2	0.11	0.00	0.14	0.96				
BA_3	0.11	0.00	0.13	0.97	1.00			
WS_1	0.27	0.13	0.30	0.76	0.79	0.78		
WS_2	0.26	0.12	0.29	0.77	0.80	0.79	0.99	
WS_3	0.26	0.12	0.29	0.76	0.80	0.79	1.00	1.00

Table 1. Scaled similarity s_2' between random networks obtained by three different models.

As one may see, our approach is able to distinguish among networks obtained by different models, namely the highest similarities are achieved for networks within the same model. On the other hand, significantly lower similarities are obtained for pairs of networks belonging to different models. For example, scaled similarities between Erdös-Rényi networks and Barabási-Albert networks are at most 0.14. Slightly higher scaled similarities are obtained between Erdös-Rényi networks and Watts-Strogatz networks. On the other hand, scaled similarities between Barabási-Albert networks and Watts-Strogatz networks lie between 0.76 and 0.80. These findings show that our methodology may be applicable in network science for classification and comparison of networks. Namely, with further tuning and refinement, the approach can be adapted to capture the structural nuances of more complex networks, making it applicable to a broader range of real-world networks.

5 Application in chemistry

In this section, we present results demonstrating the potential application of our method in chemistry, specifically in the assessment of molecular similarity. To evaluate its performance, we carried out a comparative analysis on a set of 75 decane isomers, comparing our approach with two established descriptors commonly used for molecular similarity evaluation. Specifically, the molecular structures of decane isomers were encoded using two types of molecular descriptors: Morgan circular fingerprints (with a radius of 2 and 2048 bits) [35] and MACCS keys [14]. Then, the similarity between two molecules was calculated using the Tanimoto index [3], a widely accepted standard in molecular similarity computations. The correlations between the obtained similarities and the s_2 similarity for all pairs of decane isomers are depicted in Figure 9.

As observed from the results, the Tanimoto similarity based on Morgan fingerprints generally yields lower similarity scores for decane isomers compared to both MACCS key-based Tanimoto similarity and the s_2 structural similarity measure. This discrepancy likely arises from the nature of Morgan fingerprints, which emphasize local atomic environments and specific topological features. In contrast, MACCS keys capture broader, predefined substructural patterns that are more consistent across isomeric forms. Consequently, MACCS-based and s_2 similarities tend to align more closely. It is also noticeable that both Tanimoto similarity measures, based on Morgan fingerprints and MACCS keys, tend to produce a relatively high number of molecule pairs with identical similarity scores. In contrast, such behavior is less apparent with the s_2 similarity values. This suggests that our method may offer increased sensitivity to subtle structural differences among isomers. These findings indicate that our methodology may be well-suited for molecular similarity assessment. Furthermore, with the application of an appropriate weighting scheme to molecular graphs, the approach could potentially be extended to datasets containing more structurally diverse molecules.
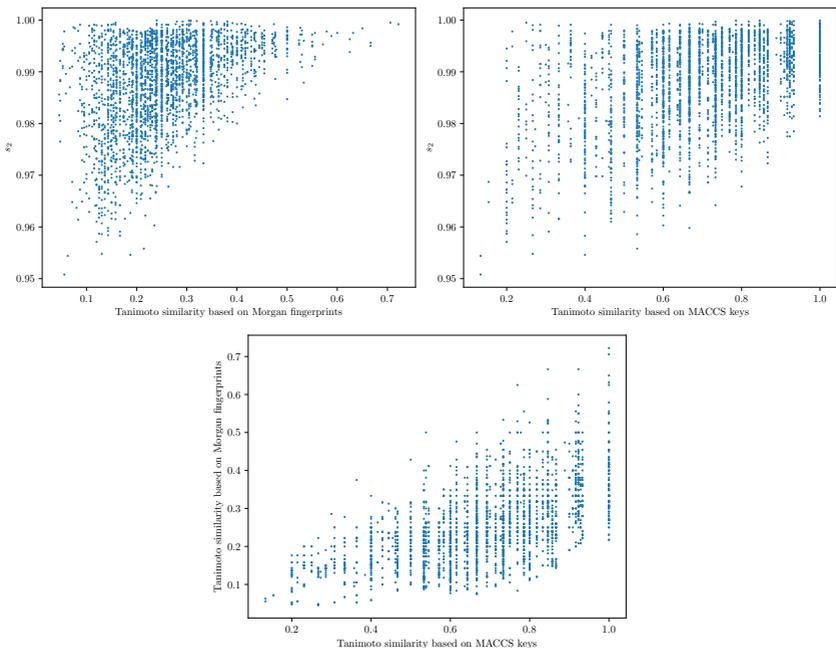
Figure 9. Correlation between Tanimoto similarity based on Morgan fingerprints and s_2 (upper left), Tanimoto similarity based on MACCS keys and s_2 (upper right), and Tanimoto similarity based on MACCS keys and Morgan fingerprints (bottom), all for decane isomers.

6   Summary and conclusion

In this paper, we introduced new graph similarity and distance measures. While traditional methods such as graph edit distance and single-index topological measures provide valuable insights, they often face limitations in sensitivity or computational efficiency. The approach presented in this paper addresses these challenges by leveraging multiple topological indices to capture subtle structural differences while maintaining low computational cost.

Our results indicate that this method not only aligns well with established metrics but also demonstrates strong potential for large-scale applications and domain-specific analyses, including network modeling and molecular similarity studies. Specifically, the developed approach ef-

fectively distinguishes between real-world network models and measures molecular similarity with performance that exceeds that of established molecular similarity metric. Furthermore, applying suitable weighting schemes to both molecular graphs and network structures could be a promising way to extend the approach to datasets with more structurally diverse molecules and complex real-world networks.

References

[1] C. G. Akcora, B. Carminati, E. Ferrari, User similarities on social networks, *Soc. Netw. Anal. Min.* **3** (2013) 475–495.

[2] R. Albert, A.-L. Barabási, Statistical mechanics of complex networks, *Rev. Mod. Phys.* **74** (2002) 47–97.

[3] D. Bajusz, A. Rácz, K. Héberger, Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.* **7** (2015) #20.

[4] V. Batagelj, Similarity measures between structured objects, in: A. Graovac (Ed.), *Proceedings of an International Course and Conference on the Interfaces between Mathematics, Chemistry and Computer Sciences, Studies in Physical and Theoretical Chemistry, Vol. 63*, Elsevier, Amsterdam, 1989.

[5] S. Brezovnik, M. Dehmer, N. Tratnik, P. Žigert Pleteršek, On the Wiener-like root-indices of graphs, *Comp. Appl. Math.* **44** (2025) #333.

[6] H. Bunke, What is the distance between graphs?, *Bull. EATCS* **20** (1983) 35–39.

[7] M. Dehmer, Z. Chen, F. Emmert-Streib, Y. Shi, S. Tripathi, A. Musa, A. Mowshowitz, Properties of graph distance measures by means of discrete inequalities, *Appl. Math. Model.* **59** (2018) 739–749.

[8] M. Dehmer, Z. Chen, Y. Shi, Y. Zhang, S. Tripathi, M. Ghorbani, A. Mowshowitz, F. Emmert-Streib, On efficient network similarity measures, *Appl. Math. Comput.* **362** (2019) #124521.

[9] M. Dehmer, F. Emmert-Streib, Comparing large graphs efficiently by margins of feature vectors, *Appl. Math. Comput.* **188** (2007) 1699–1710.

[10] M. Dehmer, F. Emmert-Streib, Y. Shi, Interrelations of graph distance measures based on topological indices, *PLoS ONE* **9** (2014) #e94985.

[11] M. Dehmer, M. Grabner, B. Furtula, Structural discrimination of networks by using distance, degree and eigenvalue-based measures, *PLoS ONE* **7** (2012) #e38564.

[12] J. W. Demmel, *Applied Numerical Linear Algebra*, SIAM, Philadelphia, 1997.

[13] A. A. Dobrynin, A. A. Kochetova, Degree distance of a graph: A degree analogue of the Wiener index, *J. Chem. Inf. Comput. Sci.* **34** (1994) 1082–1086.

[14] J. L. Durant, B. A. Leland, D. R. Henry, J. G. Nourse, Reoptimization of MDL keys for use in drug discovery, *J. Chem. Inf. Model.* **42** (2002) 1273–1280.

[15] F. Emmert-Streib, M. Dehmer, Y. Shi, Fifty years of graph matching, network alignment and network comparison, *Inf. Sci.* **346−347** (2016) 180–197.

[16] E. Estrada, Characterization of 3D molecular structure, *Chem. Phys. Lett.* **319** (2000) 713–718.

[17] J. G. F. Francis, The QR transformation. A unitary analogue to the LR transformation - part 1, *Comput. J.* **4** (1961) 265–271.

[18] X. Gao, B. Xiao, D. Tao, L. Xuelong, A survey of graph edit distance, *Pattern Anal. Appl.* **13** (2010) 113–129.

[19] I. Gutman, Geometric approach to degree–based topological indices: Sombor indices, *MATCH Commun. Math. Comput. Chem.* **86** (2021) 11–16.

[20] I. Gutman, Selected properties of the Schultz molecular topological index, *J. Chem. Inf. Comput. Sci.* **34** (1994) 1087–1089.

[21] I. Gutman, The energy of a graph, *Ber. Math. Stat. Sekt. Forschungsz. Graz* **103** (1978) 1–22.

[22] I. Gutman, B. Furtula, E. Zogić, E. Glogić, Resolvent energy of graphs, *MATCH Commun. Math. Comput. Chem.* **75** (2016) 279–290.

[23] I. Gutman, N. Trinajstić, Graph theory and molecular orbitals. Total π-electron energy of alternant hydrocarbons, *Chem. Phys. Lett.* **17** (1972) 535–538.

[24] O. Ivanciuc, T.S. Balaban, A. T. Balaban, Reciprocal distance matrix, related local vertex invariants and topological indices, *J. Math. Chem.* **12** (1993) 309–318.

[25] F. Kaden, Graphmetriken und Distanzgraphen, *ZKI-Informationen, Akad. Wiss. DDR* **2** (1982) 1–63.

[26] 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.

[27] N. M. Kriege, F. D. Johansson, C. Morris, A survey on graph kernels, *Appl. Netw. Sci.* **5** (2020) #6.

[28] A. G. Maldonado, J. P. Doucet, M. Petitjean, B. T. Fan, Molecular similarity and diversity in chemoinformatics: From theory to applications, *Mol. Divers.* **10** (2006) 39–79.

[29] M. Newman, *Networks: An Introduction*, Oxford University Press, 2010.

[30] J. M. Ortega, H. F. Kaiser, The LL^T and QR methods for symmetric tridiagonal matrices, *Comput. J.* **6** (1963) 99–101.

[31] D. Plavšić, S. Nikolić, N. Trinajstić, Z. Mihalić, On the Harary index for the characterization of chemical graphs, *J. Math. Chem.* **12** (1993) 235–250.

[32] M. Randić, Characterization of molecular branching, *J. Am. Chem. Soc.* **97** (1975) 6609–6615.

[33] I. Redžepović, Chemical applicability of Sombor indices, *J. Serb. Chem. Soc.* **86** (2021) 445–457.

[34] I. Redžepović, B. Furtula, On degeneracy of $\mathcal{A}$-eigenvalue–based molecular descriptors and r-equienergetic chemical trees, *MATCH Commun. Math. Comput. Chem.* **84** (2020) 385–397.

[35] D. Rogers, M. Hahn, Extended-connectivity fingerprints, *J. Chem. Inf. Model.* **50** (2010) 742–754.

[36] A. Sanfeliu, K.-S. Fu, A distance measure between attributed relational graphs for pattern recognition, *IEEE Trans. Syst. Man Cybern.* **13** (1983) 353–362.

[37] H. P. Schultz, Topological organic chemistry. 1. Graph theory and topological indices of alkanes, *J. Chem. Inf. Comput. Sci.* **29** (1989) 227–228.

[38] M. S. Öz, A method for measuring similarity or distance of molecular and arbitrary graphs based on a collection of topological indices, *J. Chemom.* **39** (2025) e70047.

[39] F. Sobik, Graphmetriken und Klassifikation strukturierter Objekte, *ZKI-Informationen, Akad. Wiss. DDR* **2**(82) (1982) 63–122.

[40] F. Sobik, Modellierung von Vergleichsprozessen auf der Grundlage von Ähnlichkeitsmaßen für Graphen, *ZKI-Informationen, Akad. Wiss. DDR* **4** (1986) 104–144.

[41] E. H. Sussenguth, *Structure Matching in Information Processing*, Ph.D. thesis, Harvard University, Cambridge, 1964.

[42] D. J. Watts, S. H. Strogatz, Collective dynamics of "small-world" networks, *Nature* **393** (1998) 440–442.

[43] H. Wiener, Structural determination of paraffin boiling points, *J. Am. Chem. Soc.* **69** (1947) 17–20.

[44] P. Willet, *Similarity and Clustering in Chemical Information Systems*, Res. Stud. Press, Letchworth, 1987.

[45] P. Wills, F. G. Meyer, Metrics for graph comparison: A practitioner's guide, *PLoS ONE* **15** (2020) #e0228728.

[46] B. Zelinka, On a certain distance between isomorphism classes of graphs, *Časopis pro pěst. Mathematiky* **100** (1975) 371–373.