Applied Mathematics and Computation 266 (2015) 623–633

Contents lists available at ScienceDirect

Applied Mathematics and Computation

journal homepage: www.elsevier.com/locate/amc

Graph distance measures based on topological indices revisited

Matthias Dehmer [a,b], Frank Emmert-Streib [c,d], Yongtang Shi [e,*]

[a] Department of Computer Science, Universität der Bundeswehr München, Werner-Heisenberg-Weg 39, Neubiberg 85577, Germany
[b] Department of Biomedical Computer Science and Mechatronics, UMIT, Eduard Wallnoefer Zentrum 1, Hall in Tyrol, A-6060, Austria
[c] Computational Medicine and Statistical Learning Laboratory, Department of Signal Processing, Tampere University of Technology, Finland
[d] Institute of Biosciences and Medical Technology Tampere 33520, Finland
[e] Center for Combinatorics and LPMC-TJKLC, Nankai University, Tianjin 300071, China

ARTICLE INFO

Keywords:
Distance measure
Edit distance
Topological index
Extremal graph

ABSTRACT

Graph distance measures based on topological indices have been already explored by Dehmer et al. Also, inequalities for those graph distance measures have been proved. In this paper, we continue studying such comparative graph measures based on the well-known Wiener index, graph energy and Randić index, respectively. We prove extremal properties of the graph distance measures for some special classes of graphs. To demonstrate useful properties of the measures, we also discuss numerical results. To conclude the paper we state some open problems.

1. Introduction

Determining the structural similarity or distance of graphs has been a striking research problem. Methods to do so have been applied in many disciplines, such as mathematics [1–4], biology [5,6], chemistry [7,8] and chemoinformatics [9,10]. Other application-oriented areas where graph comparison techniques have been employed can be found in [11]. Note that the problem of matching graphs quantitatively is a classical problem and has been already tackled in the sixties; Sussenguth [12] explored the graph matching problem based on graph isomorphism before the seminal work done by Vizing [13] and Zelinka [14]. To study classical and recent results on graph matching in more detail, we refer to our recent survey [15].

In order to explain the contribution of this paper, we first explain the underlying concepts in more detail. Given two graphs with n vertices, G_1 and G_2, the *edit distance* between G_1 and G_2, denoted by $GED(G_1, G_2)$ [16], is the minimum cost caused by the number of edge additions and/or deletions that are needed transform G_1 into G_2. *GED* has been introduced by Bunke [16]. For more results on the graph edit distance, we refer to [3,17–19].

A topological index [20] is a graph invariant, defined by

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

Topological indices are graph invariants which characterize the graph structure quantitatively. 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 structure, see [21]. Topological graph measures have also been applied in ecology [22], biology [23] and network physics [24,25]. Note that various properties of topological graph measures such as their uniqueness and correlation ability have been examined too [26,27].

* Corresponding author. Tel.: +862223503682; fax: 862223509272
E-mail addresses: matthias.dehmer@unibw.de (M. Dehmer), v@bio-complexity.com (F. Emmert-Streib), shi@nankai.edu.cn (Y. Shi).

http://dx.doi.org/10.1016/j.amc.2015.05.072

As a consequence, thousands of topological indices have been developed describing structural, physical and chemical properties of molecular graphs. Topological indices can be divided into several categories such as degree-based indices, distance-based indices and spectrum-based indices. Obviously, such categorization is not unique. For instance, prominent degree-based indices [28] comprise the Randić index [29], the general Randić index [30], the zeroth order general Randić index [31–36], the eccentricity index [37–39]. Known distance-based indices [40] include the Wiener index [41–45], the Szeged index [46,47], the ABC index [48,49], the Kirchhoff index [50–52], the Harary index [53] etc., see [54]. By using eigenvalues of graphs, various of graph energies have been developed [55,56], e.g., graph energy [55,56], the Randić energy [57], the incidence energy [58], the Laplacian energy [59], the matching energy [60,61], the energy of a matrix [62], the HOMO-LUMO index [63,64], the inertia [39,65], the Estrada index [66] and so forth [67,68]. This list shows that numerous indices have been developed; however the feasibility of many of these measures have not yet been demonstrated.

In this paper we continue exploring the graph distance measure introduced in [69]

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

Note that

$$d(x,y) = 1 - e^{-(\frac{x-y}{\sigma})^2}, \tag{3}$$

is a distance measure for real numbers, see [70]. We emphasize that we have already obtained related results in [69]. In particular, we investigated some measures $d_I(G, G')$, where G, G' are general graphs.

The main contribution of the paper is to focus on graphs whose distance can be obtained by performing only one graph edit operation, i.e., $GED(G, G') = 1$. This enables us to demonstrate which measure is the most powerful and useful one. That means we continue investigating $d_I(G, G')$ and their extremal properties for some special graph classes. As indices I, we use the Wiener index, the graph energy and the Randić index, respectively. Our numerical results show that the distance measure d_R is more powerful on the used data sets; R is the well-known Randić index.

2. Preliminaries

Denote by K_n, P_n and S_n the complete graph, the path graph and the star graph with n vertices, respectively. A *chemical graph* is a graph with maximum degree of 4. We call a chemical graph *chemical tree* if it is a tree, see [71]. The graph S_n^+ is obtained from S_n by joining two leaves of S_n. Denote by P_n^k the unicyclic graph obtained by connecting a vertex of C_k with a leaf of P_{n-k}.

Observe that for two connected graphs with n vertices,

$$0 \le GED(G_1, G_2) \le \frac{(n-1)(n-2)}{2}.$$

It is obvious that $GED(G_1, G_2) = 0$ if and only if $G_1 \cong G_2$, and $GED(G_1, G_2) = \frac{(n-1)(n-2)}{2}$ if and only if one of them is the complete graph; the second graphs represents a tree.

Before we start, we state an important observation

Observation 1. For $G, H \in \mathcal{G}$, G, H are the two graphs attaining the maximum value of d_I if and only if G, H are the graph attaining the maximum and minimum value of I, respectively.

Proof. Let $x = |I(G) - I(H)|$, then d_I is a monotone increasing function depending 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. The result follows. □

3. Graph distance measures based on Wiener index

3.1. Wiener index

Suppose $G = (V, E)$ is connected. 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

$$W(G) = \sum_{u,v \subseteq V} d(u,v). \tag{4}$$

In [72] and in many subsequent contributions, it has been demonstrated that if T is any n-vertex tree different from P_n and S_n, then $W(S_n) < W(T) < W(P_n)$.

Lemma 1 ([72]). *Among all trees of order n, the star S_n minimizes the Wiener index and the path P_n maximizes the Wiener index. In addition, $W(P_n) = \binom{n+1}{3}$ and $W(S_n) = (n-1)^2$.*

In [73], the authors determined the extremal values of the Wiener index among all unicyclic graphs.

Lemma 2 ([73]). *Let G be a unicyclic graph of order n, then*

$$n(n-2) \le W(G) \le \frac{n^3 - 7n + 12}{6}.$$

The left equality holds if and only if $G \cong S_n^+$; the right equality holds if and only if $G \cong P_n^3$.