The Uniqueness of D_{MAX}-Matrix Graph Invariants

Matthias Dehmer[1]*, **Yongtang Shi**[2]

1 Institute for Bioinformatics and Translational Research, UMIT, Eduard Wallnoefer Zentrum 1, A-6060, Hall in Tyrol, Austria, 2 Center for Combinatorics and LPMC-TJKLC, Nankai University, Tianjin, P.R. China

Abstract

In this paper, we examine the uniqueness (discrimination power) of a newly proposed graph invariant based on the matrix D_{MAX} defined by Randić et al. In order to do so, we use exhaustively generated graphs instead of special graph classes such as trees only. Using these graph classes allow us to generalize the findings towards complex networks as they usually do not possess any structural constraints. We obtain that the uniqueness of this newly proposed graph invariant is approximately as low as the uniqueness of the Balaban J index on exhaustively generated (general) graphs.

Introduction

Matrix-based descriptors have been developed extensively [1–3]. As a result, the distance matrix, the adjacency matrix and other graph-theoretical matrices [4] have been used to define topological graph measures and to examine their properties [4,5]. A property which has been of considerable interest when designing topological descriptors is referred to as uniqueness [6–8]. Generally, the uniqueness of a structural graph measure relates to the ability to distinguish the structure of non-isomorphic graphs uniquely. From a mathematical point of view, the low uniqueness or high degeneracy of a graph measure under consideration is an undesired aspect as non-isomorphic graphs should be mapped to non-equal values. Such a highly discriminating graph invariant could be then used to distinguish graph structures uniquely and, thus, to perform graph isomorphism testing [9,10]. In the context of graph isomorphism testing, so-called complete graph invariants have been investigated [9,11]. Such a graph invariant has the property that it discriminates all non-isomorphic graphs uniquely (i.e., without any degeneracy) and isomorphic graphs are mapped to equal values [9,11]. For example, Liu and Klein [11] made an attempt to derive complete graph invariants by using eigenvalues. Dehmer et al. [8,9] defined graph entropies which turned out to be the most discriminative measures so far when using exhaustively generated graphs. Clearly, such measures are suitable candidates to test graph isomorphism efficiently [9].

Recently, Randić et al. [2,12] defined so-called D_{MAX} matrices and also topological descriptors thereof. Let $G = (V, E)$ be a finite graph. Then these matrices have been defined by using the ordinary distance matrix D of G such that in each row and column the dominant (largest) distances are used where other elements $\in D$ are set to be zero, see [2,12]. Moreover they defined a new topological index j which has the same definition than the well-known Balaban index J [13] but uses D_{MAX} instead of only using D. Then based on example caclulations, Randić et al. [2] argued that j may be a promising candidate for isomorphism testing, but

they did not examine the problem in depth on wider classes of graphs.

In this paper, we explore the uniqueness of j by employing D_{MAX} on a large scale. For this, we use exhaustively generated graphs with 9 and 10 vertices each [8] and alkane trees $T = (V_T, E_T)$ where $19 \leq |V_t| \leq 22$. Our findings (see section 'Methods and Results') reveal that the uniqueness of j is always worse than the one of J and, thus, the uniqueness of j is insufficient for performing isomorphism testing.

Methods and Results

The Structural Descriptors J and j

Let $G = (V, E)$ be a finite graph. To define the Balaban index J [13,14] of G, let $(D)_{i,j} := (d(v_i, v_j))_{i,j}$ be the distance matrix. $d(v_i, v_j)$ is the topological distance between $v_i \in V$ and $v_i \in V$. For each vertex $v_i \in V$, DS_i denotes the distance sum (row or column sum) by adding the entries in the corresponding row or column of D. Let $\mu := |E| + 1 - |V|$ be the cyclomatic number [14]. Then J has been defined by [13]

$$J(G) := \frac{|E|}{\mu + 1} \sum_{(v_i, v_j) \in E} [DS_i DS_j]^{-\frac{1}{2}}. \quad (1)$$

A critical analysis to examine the uniqueness of J and other quantities has recently been carried out by Dehmer et al. [8] based on using exhaustively generated (general) graphs. In this sense Dehmer et al. examined the limitations of the Balaban J index and found that this index is quite unstable [9]; here that means there is a strong dependency between the sample size of the graph set and the uniqueness [8]. To study the technical details and the precise definitions, we refer to [8,9]. Moreover, the findings of Dehmer et al. [8] revealed that the uniqueness of J by using exhaustively generated graphs is poor. For example by using the class N_{10} (all non-isomorphic graphs with 10 vertices), $|N_{10}| = 11.716571$, the

Balaban index J could only discriminate 20% of N_{10} uniquely. Nevertheless, J has high uniqueness for alkane trees and isomers [8,9,13].

To define j, we require the definition of D_{MAX} [12]:

$$(D_{MAX})_{i,j} := \begin{cases} (D)_{i,j} & : (D)_{i,j} \geq \min\{R_i, C_j\} \\ 0 & : (D)_{i,j} < \min\{R_i, C_j\} \end{cases}$$

Following Randić et al., the topological index j is just the D_{MAX} analog to Balaban's J index, see [2]. Based on the fact that j can discriminate the remaining isomers of n-dodecane and D_{MAX} has often a different structure compared to D, Randić et al. concluded that D_{MAX} and, hence, j may be a promising tool for graph isomorphism testing, see [2]. In the next section, we see that this statement has been too premature when evaluating j on general and exhaustively generated graphs. By evaluating characteristic properties (e.g., the uniqueness) of topological graph measures on such (general) graphs, one can conclude how would the index behave in the context of using complex networks.

Results

Before interpreting Table 1, we explain its notation. We here used the graph classes N_i, $i = 9,10$ [8] and C_i, $19 \leq i \leq 22$ [8]. Again N_i is the class of all exhaustively generated non-isomorphic and connected graphs with i vertices [8]. C_i is the class of exhaustively generated non-isomorphic and connected alkane trees [8]. ndv stands for the number of non-distinguishable values [8] and $S := \dfrac{|\mathcal{G}| - ndv}{|\mathcal{G}|}$ where $\mathcal{G}$ is a class of graphs, see [7].

Table 1 shows numerical results when comparing j and J on the just explained graph classes. We observe that the uniqueness of j is quite poor for all graph classes. In case of using N_9, the uniqueness of j is approximately as low as the uniqueness of J. That means both topological indices can only discriminate about 39% out of 261080 graphs. By considering the results for J, we see that J possesses high uniqueness when using C_i. Note that this has already been found by Balaban [13] and Dehmer et al. [8]. But it is surprising that the uniqueness of j is, without exception, much

Table 1. The uniqueness of j and J measured by ndv and S.

Uniqueness of j and J				
	Graph class	cardinality	ndv	S
j based on D_{Max}	N_9	261080	165109	0.36759
	N_{10}	11716571	9476268	0.1912081
	C_{19}	148284	144041	0.028614
	C_{20}	366319	359177	0.019496
	C_{21}	910726	898838	0.013053
	C_{22}	2278658	2258804	0.00871302
J	N_9	261080	156674	0.399900
	N_{10}	11716571	9307263	0.205633
	C_{19}	148284	5967	0.95975
	C_{20}	366319	44800	0.877702
	C_{21}	910726	45703	0.949816
	C_{22}	2278658	306911	0.865310

doi:10.1371/journal.pone.0083868.t001

Table 2. The values of J and j by using the isomers of n-dodecane.

Graph ID	J	j
1	4.252509	13.15875
2	4.252509	16.04085
3	3.752273	15.86058
4	3.752273	11.99851
5	4.135003	15.04953
6	4.135003	11.07033
7	3.575256	13.16886
8	3.575256	11.81169
9	3.773441	21.73837
10	3.773441	7.43385
11	3.954123	12.76649
12	3.954123	10.84469

doi:10.1371/journal.pone.0083868.t002

worse than the one of J. Table 2 shows that j can discriminate the isomers of n-dodecane for which the Balaban J index is pairwisely degenerated.

A hypothesis is that the sparseness of D_{MAX} leads to this effect described above. So this matrix can not capture the complexity of the used graphs meaningfully and, thus, j is degenerated for most of the graphs. This result shows the complexity of the problem to construct highly unique graph measures on general and exhaustively generated graphs.

Summary and Conclusion

This paper investigated the uniqueness of the recently developed topological index j introduced by Randić et al. [2]. j has been defined quite similarly as it is based on the novel matrix D_{MAX} instead of D. Based on small tests and by only using example graphs, Randić et al. [2] hypothesized j has higher uniqueness than J and, the index j which combined with index J, may suffice to resolve the graph isomorphism issue for most cases of molecular graphs.

In this paper we have evaluated this hypothesis on a large scale by using general graphs. In fact, our study disproved this conjecture and demonstrated that the uniqueness of j is quite poor by using general exhaustively generated graphs and alkane trees. As future work, we plan to determine so-called degeneracy classes analytically for performing a proper mathematical treatment of the problem. In any way, the search for highly discriminating graph invariants should be continued [8,9,15,16]. Following Randić et al. [2], such measures could be used as a prescreening method and would eliminate need for detailed and elaborate tests on large number of cases. Also, this fact has already been raised by Dehmer et al. [9] where they developed information-theoretic network measures with very low degeneracy on exhaustively generated graphs for graph isomorphism testing.

Acknowledgments

Author Contributions

Wrote the paper: MD YS.

References

1. Devillers J, Balaban AT (1999) Topological Indices and Related Descriptors in QSAR and QSPR. Gordon and Breach Science Publishers. Amsterdam, The Netherlands.
2. Randić M, Orel R, Balaban AT (2013) d_{MAX} matrix graph invariants as graph descriptors. graphs having the same balaban j index. MATCH Commun Math Comput Chem 70: 221–238.
3. Todeschini R, Consonni V, Mannhold R (2002) Handbook of Molecular Descriptors. Wiley-VCH. Weinheim, Germany.
4. Janežič D, Miležević A, Nikolić S, Trinajstić N (2007) Graph-Theoretical Matrices in Chemistry. Mathematical Chemistry Monographs. University of Kragujevac and Faculty of Science Kragujevac.
5. Dehmer M, Sivakumar L, Varmuza K (2012) Uniquely discriminating molecular structures using novel eigenvalue-based descriptors. MATCH Communications in Mathematical and in Computer Chemistry 67: 147–172.
6. Bonchev D, Mekenyan O, Trinajstić N (1981) Isomer discrimination by topological information approach. J Comp Chem 2: 127–148.
7. Konstantinova EV (1996) The discrimination ability of some topological and information distance indices for graphs of unbranched hexagonal systems. J Chem Inf Comput Sci 36: 54–57.
8. Dehmer M, Grabner M, Varmuza K (2012) Information indices with high discriminative power for graphs. PLoS ONE 7: e31214.
9. Dehmer M, Grabner M, Mowshowitz A, Emmert-Streib F (2012) An efficient heuristic approach to detecting graph isomorphism based on combinations of highly discriminating invariants. Advances in Computational Mathematics.
10. McKay BD (2010). Nauty. http://cs.anu.edu.au/~bdm/nauty/.
11. Liu X, Klein DJ (1991) The graph isomorphism problem. Journal of Computational Chemistry 12: 1243–1251.
12. Randić M (2013) d_{MAX} matrix of dominant distances in a graph. MATCH Commun Math Comput Chem 70: 239–258.
13. Balaban AT (1982) Highly discriminating distance-based topological index. Chem Phys Lett 89: 399–404.
14. Balaban AT, Balaban TS (1991) New vertex invariants and topological indices of chemical graphs based on information on distances. J Math Chem 8: 383–397.
15. Diudea MV, Ilić A, Varmuza K, Dehmer M (2011) Network analysis using a novel highly discriminating topological index. Complexity 16: 32–39.
16. Xu CYHL (1996) On highly discriminating molecular topological index. J Chem Inf Comput Sci 36: 82–90.